

UNIVERSITÉ LOUIS PASTEUR STRASBOURG I  
FACULTÉ DES SCIENCES ÉCONOMIQUES ET DE GESTION

---

Thèse  
de Doctorat ès Sciences Économiques

---

**MODÈLES DE DURÉE MULTIVARIÉS  
AVEC HÉTÉROGÉNÉITÉ MULTIPLE :  
APPLICATIONS AU MARCHÉ DU  
TRAVAIL**

*Présentée et soutenue publiquement par*

**GUILLAUME HORNY**

**Jury**

DIRECTEUR DE THÈSE	<b>François Laisney</b> Université Louis Pasteur - Strasbourg I
RAPPORTEURS EXTERNES	<b>Jean Pierre Florens</b> Université des Sciences Sociales - Toulouse I <b>Denis Fougère</b> CREST-INSEE, Paris
RAPPORTEUR INTERNE	<b>Bertrand Koebel</b> Université Louis Pasteur - Strasbourg I
SUFFRAGANTS	<b>Bernhard Boockmann</b> ZEW, Mannheim <b>Gérard Van den Berg</b> Free University Amsterdam



*La Faculté n'entend donner aucune approbation ou improbation aux opinions émises dans les thèses. Ces opinions doivent être considérées comme propres à leurs auteurs.*



*À mes parents*



# Remerciements

La présence de remerciements au début d'une thèse est d'une régularité empirique remarquable. L'aspect systématique de l'exercice, souvent riche en hyperboles, rend d'autant plus compliquée l'écriture de remerciements qui soient sincères sans être convenus. Éviter le recours aux phrases toutes faites est difficile, et constitue pourtant le seul moyen d'évoquer fidèlement l'attachement aux différentes personnes qui ont participé directement et indirectement à ce travail.

Mes premiers remerciements s'adressent à François Laisney, mon directeur de thèse. J'ai d'abord pu bénéficier de son enseignement, puis de sa disponibilité. Outre ses compétences et sa rigueur, il a su gérer avec maestria l'équilibre délicat de l'évolution de la thèse, à savoir cadrer un projet tout en encourageant l'initiative. A l'heure où j'écris ces lignes, nous nous connaissons depuis près de 7 ans et je continue à découvrir ses nombreuses qualités humaines. En témoigne l'intérêt permanent porté à ce travail, qui va au delà de la curiosité scientifique.

J'exprime toute ma gratitude à Gérard Van den Berg. Il est un acteur incontournable de la scène internationale de l'économétrie des modèles de durée, et l'importance de ses travaux ainsi que le rôle central qu'ils jouent dans cette thèse font que je suis particulièrement touché qu'il ait accepté d'évaluer ce travail. Je le remercie également de m'avoir permis de séjourner quelques temps au Tinbergen Institute, à Amsterdam. J'ai pu bénéficier de ses précieux conseils, si naturels pour lui et si édifiants pour moi.

Je tiens à remercier Jean Pierre Florens d'avoir accepté d'être rapporteur externe. Les remarques qu'il a émises lors du Workshop "European Unemployment" du ZEW (2005) ont influencé la suite de mes travaux.

Je remercie Denis Fougère de me faire l'honneur d'être rapporteur externe. Ses nombreux travaux contribuent pour beaucoup à la vitalité de l'économétrie des modèles de durée, en répondant simultanément aux besoins d'études novatrices, de méthodes adaptées et de recommandations politiques.

Je suis reconnaissant à Bernhard Boockmann d'avoir accepté d'être membre de mon jury. Il a su m'intéresser au fonctionnement de l'OIT et a contribué à l'élaboration de cette thèse. Sa grande culture est l'une des raisons qui m'incite à retourner régulièrement au ZEW, à Mannheim.

Je remercie également Bertrand Koebel d'avoir accepté la lourde tâche de rapporteur interne. Les discussions que nous avons eues au sujet de cette thèse ont toujours été fructueuses, et j'espère que sa patience n'aura pas été déçue.

Toute ma gratitude revient également à Arthur van Soest pour cet été 2000 au CentER, à Tilburg. Quelques unes des conditions initiales prévalant

à la trajectoire de cette thèse ont été fixées à ce moment.

Je remercie également Bernhard Boockmann, Dragana Djurdjevic, Rute Mendes et Gérard Van den Berg pour leurs collaborations sur certains chapitres. J'exprime ma reconnaissance à Théophile Azomahou, Mathieu Brezovski et Viktor Winschel pour leurs remarques sur différentes parties de ces travaux, ainsi qu'aux participants des séminaires d'économétrie de Strasbourg (les deux ensembles n'étant pas exclusifs). Je remercie aussi Narayan Sastry, pour les explications fournies quant au fonctionnement de son algorithme.

Cette thèse doit beaucoup aux commentaires obtenus lors des congrès, et je remercie les organisateurs et participants de l'ESEM (2004, 2006), SMYE (2006), SESAC (2006) ainsi que les séminaires du Tinbergen Institute (2005), du ZEW (2001, 2004 et 2006) et les journées du BETA (2004, 2006).

Alors que cette période de ma vie touche à sa fin et que le probabilité de transiter de l'état de "doctorant" vers un état "autre" n'a jamais été aussi proche de 1, j'adresse toute ma reconnaissance à ceux qui m'ont entouré pendant le doctorat. Je pense en particulier aux collègues passés et présents du bureau 126, à Yves, Thomas, Ghazi, Blandine, Mohammed, Li, Irem, Carinne et Morad. Parmi les compagnons de mes nombreuses journées au PEGE, qui m'ont tous apporté quelque chose dans un domaine particulier, figurent dans le désordre Thierry, Rachel, Paul, Christophe, Stéphane, Charlotte, Benoît, Ami et René. Je remercie plus généralement mes collègues de la Faculté de Sciences Économiques (Strasbourg I), de l'Institut du Travail, de l'Institut de Préparation à l'Administration Générale, de l'Institut d'Études Politiques et de l'École Doctorale.

Cette thèse ne serait pas ce qu'elle est sans le temps passé à faire d'autres choses, et une pensée toute particulière va vers Claude, Julien et Marie pour l'organisation des ACDD, les Gazettes Cournot, la représentation de l'École Doctorale mais aussi les sorties raquettes, randonnées, vélo et autres moments que je n'oublie pas.

Enfin, j'exprime ma reconnaissance à mes parents pour leurs nombreux sacrifices, leur soutien permanent et leur affection.





# Table des matières

<b>Introduction</b>	<b>1</b>
1 Une petite histoire des modèles de durée . . . . .	4
2 Hétérogénéité observée et non observée . . . . .	7
3 Outils actuels et limites . . . . .	14
4 Méthodologie et Résultats . . . . .	16
<b>1 Revue de littérature : propriétés, identification et inférence dans le modèle MPH</b>	<b>19</b>
Résumé . . . . .	20
1.1 Éléments de base des modèles de durée . . . . .	21
1.2 Le modèle de Cox . . . . .	34
1.3 Le modèle de Mélange de Hasards Proportionnels . . . . .	50
1.4 Identification . . . . .	66
1.5 Inférence basée sur la vraisemblance . . . . .	78
1.6 Les approches basées sur la vraisemblance partielle . . . . .	84
1.7 L'approche Bayésienne . . . . .	88
1.8 Conclusion . . . . .	94
<b>2 Estimation par vraisemblance partielle de modèles MPH à effets aléatoires : un algorithme EM basé sur la vraisemblance pénalisée</b>	<b>95</b>
Résumé . . . . .	96
Introduction . . . . .	97
2.1 Le modèle MPH à effets aléatoires . . . . .	98
2.2 Inférence avec l'algorithme EM . . . . .	100
2.3 Estimation de modèles MPH à un effet aléatoire . . . . .	103
2.4 Simulations par méthodes de Monte Carlo . . . . .	104
2.5 Ratification des conventions du BIT . . . . .	107
2.6 Conclusion . . . . .	108
<b>3 Une approche Bayésienne du comportement de ratification des conventions de l' OIT</b>	<b>109</b>
Résumé . . . . .	110

Introduction . . . . .	111
3.1 Les données . . . . .	113
3.2 Les modèles . . . . .	117
3.3 Inférence Bayésienne . . . . .	119
3.4 Résultats . . . . .	121
Conclusion . . . . .	126
<b>4 La mobilité professionnelle au Portugal : une approche Bayésienne avec données d'appariement employeurs-employés</b>	<b>129</b>
Résumé . . . . .	130
Introduction . . . . .	131
4.1 Les données . . . . .	132
4.2 Modèles . . . . .	135
4.3 Inférence Bayésienne . . . . .	139
4.4 Résultats . . . . .	140
Conclusion . . . . .	144
<b>5 Conclusion</b>	<b>147</b>
<b>A Annexes du Chapitre 1</b>	<b>153</b>
A.1 Calcul de la fonction de survie conditionnelle aux variables explicatives observées dans le cadre d'erreurs de mesure sur les régresseurs . . . . .	154
A.2 Calcul de l'effet croisé du temps et des variables explicatives sur la hasard moyen . . . . .	154
A.3 Épisodes uniques et identification . . . . .	156
A.4 Preuve de l'identification sans variable explicative du modèle de Box-Cox . . . . .	158
A.5 Preuve de l'identification en présence de régresseurs endogènes	160
A.6 Identification du modèle de Weibull sans supposer $E(V) < \infty$	163
A.7 Épisodes multiples et identification . . . . .	164
A.8 Identification du modèle MMPH à deux effets aléatoires . . . .	165
A.9 Espérances conditionnelles dans un modèle à deux termes d'hétérogénéité gamma . . . . .	166
A.10 Vraisemblance partielle pénalisée et hétérogénéité gamma . . .	169
<b>B Annexes du Chapitre 2</b>	<b>173</b>
B.1 Calcul des espérances conditionnelles . . . . .	174
B.2 Évaluation de la matrice d'information . . . . .	176
B.3 Calcul de la matrice d'information en présence d'hétérogénéité gamma . . . . .	177

B.4	Programme de l'algorithme EMPL . . . . .	180
B.5	Programme de l'algorithme EM accéléré . . . . .	189
B.6	Résultats concernant la ratification des conventions du BIT . .	202
<b>C</b>	<b>Annexes du Chapitre 3</b>	<b>203</b>
C.1	Identification du MPH à deux effets aléatoires en présence d'épisodes multiples . . . . .	204
C.2	Construction de la Vraisemblance Partielle . . . . .	205
C.3	Programme du modèle à deux effets aléatoires . . . . .	206
<b>D</b>	<b>Annexes du Chapitre 4</b>	<b>209</b>
D.1	Calcul des espérances et variances conditionnelles . . . . .	210
D.2	Statistiques descriptives des durées . . . . .	210
D.3	Statistiques descriptives des variables explicatives . . . . .	211
D.4	Estimation Bayésienne sur l'échantillon non contraint . . . . .	212
D.5	Estimations fréquentistes . . . . .	214
D.6	Programme du modèle à deux effets aléatoires corrélés . . . . .	215
	Résumé . . . . .	230



# Introduction

## Sommaire

---

<b>1</b>	<b>Une petite histoire des modèles de durée . . . . .</b>	<b>4</b>
<b>2</b>	<b>Hétérogénéité observée et non observée . . . . .</b>	<b>6</b>
<b>3</b>	<b>Outils actuels et limites . . . . .</b>	<b>13</b>
<b>4</b>	<b>Méthodologie et Résultats . . . . .</b>	<b>15</b>

---

---

*“Laissons le temps au temps”*

Citation d’un doctorant en début de thèse

*“Le temps passe trop vite”*

Citation d’un doctorant en fin de thèse

QUANTIFIER le temps revient à lui associer un nombre et une unité pour en effectuer une mesure. Toutefois, cette connaissance est au mieux celle d’une représentation du temps : la mesure n’est pas le temps et n’apprend rien sur sa nature intime. Le temps est un objet ambiguë, dont la mesure permet de créer des repères mais pas de le définir, et appréhender la nature du temps semble au-delà de nos capacités, bien que tout individu ait l’intuition de ce qu’est le temps. Ce constat pousse Saint Augustin à écrire (vers 400) :

“Qu’est-ce donc que le temps ? Qui pourra le dire clairement et en peu de mots ? Qui pourra le saisir, même par la pensée, pour traduire cette conception en paroles ? Quoi de plus connu, quoi de plus familièrement présent à nos entretiens, que le temps ? Et quand nous en parlons, nous concevons ce que nous disons ; et nous concevons ce qu’on nous dit quand on nous en parle. Qu’est-ce donc que le temps ? Si personne ne m’interroge, je le sais ; si je veux répondre à cette demande, je l’ignore.”

Une multitude de rapports au temps se retrouve ainsi dans les expressions les plus courantes et en apparence les plus innocentes. La première citation ci-dessus fait référence à un “temps-cadre”, objectif et mesurable, au sein duquel une évolution se produit. Par analogie avec l’espace où un mouvement a lieu, par exemple une scène de théâtre, le temps est un milieu dans lequel se déroulent les actions. La seconde citation nous renvoie à un temps subjectif dont la perception immédiate ne correspond pas à nos références habituelles, où les heures s’accélèrent pour se transformer en secondes. Le temps se mesure cette fois à l’aune des étapes de l’évolution d’un phénomène, qui dans notre exemple semblent plus lentes que ça à quoi on s’attend. L’analogie se fait alors avec un flux, par exemple celui d’une rivière, où les accélérations et ralentissements du temps correspondent à des variations de débit. L’échelle temporelle n’est plus quelque chose d’extérieur et d’universel, elle est propre à l’évolution du phénomène et à son observateur.

Ces deux grandes approches se retrouvent en économétrie, avec les modèles de durée qui s’intéressent au temps passé dans un état avant de transiter vers un autre état, et les séries temporelles qui s’intéressent à l’évolution d’une quantité au fil du temps. La première considère le déplacement d’agents entre des états définis de manière exogène, tandis que la seconde est centrée sur



l'évolution d'une ou plusieurs variables. Dans cette thèse, nous nous intéressons plus particulièrement aux modèles de durée.

En continuité avec la première citation, le mot “durée” est entendu comme un espace de temps délimité par deux événements, et cette notion apparaît dans plusieurs champs d'application et sous de nombreuses dénominations : transitions ou durées en économie, mais aussi survies en biostatistique, pannes en ingénierie, risques en assurance ... Des contextes très divers se prêtent à l'utilisation de modèles de durée. A titre d'exemples, ils servent en économie du travail à l'étude du chômage, en économie industrielle à analyser les dynamiques d'apprentissage des firmes, en biométrie pour représenter le temps séparant le début d'un traitement médical de la rémission d'une maladie, en démographie pour analyser la mortalité ou la durée des mariages, en recherche opérationnelle pour évaluer le temps de fonctionnement d'une machine sans panne, en assurance pour identifier les individus à risques, en marketing pour mesurer le temps séparant les achats d'un bien, en finance pour représenter les transactions boursières, en sociologie pour l'étude du récidivisme, en sciences politiques pour décrire les périodes paix ou le temps séparant deux grèves, etc. A plusieurs reprises nous compléterons la littérature économique avec de nombreux travaux effectués en biostatistique, les outils connaissant un développement pluridisciplinaire.

## 1 Une petite histoire des modèles de durée

Les instruments utilisés pour étudier le temps sont eux-mêmes soumis à une évolution et on peut en esquisser un bref historique.<sup>1</sup> Les premiers travaux ont été menés dans des domaines très différents dans les années 50, et Cox (1959) en propose une première revue. On situe généralement la naissance des modèles de durée dans les années 70 car cette période voit la parution d'articles servant de points de départ à de nombreux développements encore en cours à l'heure actuelle. Ainsi, Cox (1972) présente les modèles de hasards proportionnels, aujourd'hui dits “de Cox”, qui seront parmi les plus utilisés dans l'analyse des durées. La vraisemblance partielle est exposée de manière formalisée dans Cox (1975), et Kalbfleisch (1978) la justifie dans un cadre Bayésien. On observe l'introduction d'un effet aléatoire dans le hasard, afin de prendre en compte une fragilité à une maladie, presque simultanément en biométrie chez Clayton (1978) et en démographie chez Vaupel *et al.* (1979).<sup>2</sup> Nelson (1969) propose un estimateur du hasard intégré, que Breslow

---

<sup>1</sup>Les références citées ultérieurement sont liées au sujet de la thèse, il n'est pas question ici de faire un historique exhaustif des modèles de durée.

<sup>2</sup>L'idée d'une susceptibilité inobservée accrue à certains événements, qui se traduit par

(1974) adapte au modèle de Cox, et Aalen (1978) en étudie les propriétés asymptotiques. Kalbfleisch et Prentice (1980) publient le premier ouvrage de synthèse dédié aux durées comprenant également des idées neuves, notamment sur l'introduction de régresseurs variant dans le temps. Cette période aura vu l'apparition du modèle de Cox et son extension afin de prendre en compte l'hétérogénéité non observée, ainsi que les premiers estimateurs non-paramétriques et semi-paramétriques du hasard de base.

Les années 80 voient un fort développement des travaux sur l'identification des modèles à un effet aléatoire, un approfondissement des connaissances sur les modèles de mélange et l'apparition de méthodes non-paramétriques pour estimer la distribution mélangeante. Elbers et Ridder (1982) sont les premiers à établir l'identification non-paramétrique du modèle de mélange de hasards proportionnels (ultérieurement abrégé modèle MPH). Heckman et Singer (1984b) obtiennent le même résultat sous des hypothèses légèrement différentes, excluant le cas où la distribution mélangeante est dégénérée. Le rôle des régresseurs dans l'identification des modèles à risques concurrents est étudié dans Heckman et Honoré (1989) et permet de dépasser le théorème de non-identification de Cox-Tsiatis (Cox, 1959 et Tsiatis, 1975) qui avait jusqu'alors bloqué le domaine. Du point de vue de l'inférence, Lindsay (1983a, 1983b) présente les résultats d'unicité et de convergence de l'estimateur du maximum de vraisemblance dans les modèles de mélange et Clayton et Cuzick (1985) étendent les premières réflexions de Clayton (1978), spécifiques au modèle de Cox. Heckman et Singer (1984a, 1984c) supposent une hétérogénéité discrète et proposent un estimateur non-paramétrique du maximum de vraisemblance (abrégé NPMLE). Kiefer (1988) fournit une revue de littérature très complète, couvrant les concepts de base aussi bien que les derniers développements d'alors, et Lancaster (1990) écrit un ouvrage de référence regroupant, entre autres, de nombreux résultats d'identification et méthodes d'estimation.

Depuis les années 90, on assiste à des progrès concernant l'identification pour des modèles plus élaborés que le modèle MPH à effet individuel. L'identification d'une famille de modèles englobant le cas MPH est établie dans Heckman (1991). Les démonstrations en présence d'épisodes uniques utilisent l'hypothèse de finitude de l'espérance de la distribution mélangeante, et Ridder (1990) montre qu'il existe toujours deux modèles observationnellement équivalents, l'un avec une espérance finie et l'autre avec une espérance infinie.

---

des fréquences plus élevées pour des groupes d'individus, se trouve déjà chez Greenwood et Yule (1920). Notons que seuls Vaupel *et al.* (1979) emploient le terme "frailty", Clayton (1978) ne l'utilise pas. Son article eut toutefois un rôle important dans la promotion du concept qu'il n'a pas nommé dans la mesure où il l'applique à un cadre multivarié, contrairement à Vaupel *et al.* (1979) dont le modèle est seulement univarié.

La présence d'épisodes multiples facilite l'identification et permet de relâcher cette hypothèse, comme le montre Honoré (1993). Ces différents résultats sont repris de manière unifiée dans Heckman et Taber (1994).

De nombreuses méthodes d'estimation ont été proposées durant les années 90. L'inférence voit une utilisation intensive de l'algorithme Espérance-Maximisation (abrégé EM) et de ses nombreuses adaptations, ainsi que l'utilisation dans l'étude des durées de méthodes connues ailleurs, comme la vraisemblance pénalisée ou le recours aux méthodes de Monte Carlo par Chaînes de Markov (abrégé MCMC). En présence d'un effet aléatoire normal et homoscédastique, le calcul du meilleur prédicteur linéaire sans biais est présenté dans McGilchrist et Aisbett (1991), et étendu dans McGilchrist (1993) pour en déduire l'estimateur du maximum de vraisemblance résiduel (REML). Yau et McGilchrist (1998) adaptent ces résultats à un effet aléatoire dont les réalisations sont corrélées, Yau (2001) à deux effets indépendants et Lee et Nelder (2003) à une hétérogénéité autre que normale. D'autres approches, comme Ripatti et Palmgren (2000) et Rondeau *et al.* (2003), utilisent la vraisemblance pénalisée qui est reliée à l'algorithme EM dans Therneau *et al.* (2003). L'hétérogénéité gamma est quant à elle généralement traitée au moyen de l'algorithme EM, par Klein (1992) en biométrie et Guo et Rodriguez (1992) en démographie. Ces travaux servent de point de départ à ceux spécifiant deux niveaux d'hétérogénéité, comme Sastry (1997) avec un modèle à deux effets aléatoires et Xue et Brookmeyer (1996) avec un hasard de base stratifié au premier niveau et un effet aléatoire au second.

L'algorithme EM a ensuite été adapté aux problèmes d'hétérogénéité non observée avec entre autres l'utilisation de l'échantillonnage de Gibbs lors de l'étape E (Vaida et Xu, 2000), ou en le faisant alterner avec un algorithme de Newton-Raphson (Maples *et al.*, 2002). Une formulation de l'algorithme indépendante de la spécification de la distribution mélangeante est proposée par Parner (1997). Clayton (1991) est un article particulièrement important dans l'approche Bayésienne car il montre comment utiliser l'échantillonnage de Gibbs dans un modèle MPH à hétérogénéité gamma et sert de point de départ aux études de Gustafson (1997) et Sargent (1998). L'approche non paramétrique laisse le hasard de base et la distribution mélangeante être de dimension infinie et est développée par Horowitz (1996, 1999). Horowitz et Lee (2004) y ajoutent le cas d'une censure non-indépendante des durées. Baker et Melino (1999) étudient le comportement du NPMLE sur des simulations et concluent à des difficultés lorsque le hasard de base et la distribution mélangeante ont tous deux des spécifications non paramétriques. Concernant les effets fixes, Ridder et Tunali (1999) détaillent l'approche de la vraisemblance partielle stratifiée et Lindeboom et Kerkhofs (2000) en fournissent une application en présence de risques concurrents.

On remarque le début d'une prise en compte plus fine de l'hétérogénéité non observée, avec notamment les travaux de Xue et Brookmeyer (1996), Sastry (1997) et Yau (2001) qui la spécifient à plusieurs niveaux, et nous allons voir ce qui motive de tels choix.

## 2 Hétérogénéité observée et non observée

Heckman (2001), dans son discours lors de la remise du prix Nobel, dit :

“the most important discovery is the evidence on the pervasiveness of heterogeneity and diversity in economic life”

Ce constat nous apparaît tellement naturel qu'il semble étrange de le considérer comme une découverte aussi importante. La même idée se retrouve dans le domaine très différent de la médecine, avec Aalen (1988) :

“It is a basic observation of medical statistics that individuals are dissimilar. Still, there is a tendency to regard this variation as a nuisance, and not as something to be considered seriously in its own right. Statisticians are often accused of being more interested in averages, and there is some truth to this.”

Autant la présence de différences entre les individus est naturelle et intuitive, autant l'habitude de travailler sur des comportements agrégés et des agents moyens fait perdre de vue l'hétérogénéité.

### 2.1 Qu'est ce que l'hétérogénéité ?

Plusieurs définitions de l'hétérogénéité existent, et nous considérons par la suite celle proposée dans Cunha *et al.* (2005) : **l'hétérogénéité** est la différence entre les facteurs pertinents lors de la prise de décision et connus des agents. Par exemple, nous sommes en présence d'hétérogénéité dès lors que les goûts, anticipations, capacités ou contraintes ne sont pas les mêmes d'un agent à l'autre.

Browning et Carro (2005) insistent sur la différence entre incertitude et hétérogénéité : les facteurs qui influencent le phénomène étudié ne sont pas encore connus en présence d'incertitude, tandis qu'ils sont déjà connus en présence d'hétérogénéité. Prenons l'exemple d'une population divisée en agents de type A et B, qui se distinguent par les critères pris en compte lors d'un choix. On parlera d'incertitude s'ils ne connaissent pas encore les critères pertinents pour eux. L'hétérogénéité est présente lorsque les agents savent s'ils sont de type A ou B, c'est-à-dire lorsqu'ils connaissent leurs critères importants et qu'un observateur extérieur distingue des différences entre ces

critères. Une telle situation se produit lorsque les agents ont déjà effectué un choix similaire par le passé.

Autant la différence entre les déterminants va créer de l'hétérogénéité, autant la connaissance par les agents de leurs facteurs personnels influençant le phénomène étudié est primordiale.

## 2.2 Qu'est ce que l'hétérogénéité non observée ?

Browning et Carro (2005) déduisent de la construction de l'hétérogénéité la définition suivante : nous sommes présence **d'hétérogénéité non observée** lorsque différents les facteurs pertinents et connus des agents, qui sont de plus inconnus de l'économètre. Il s'agit donc d'un type particulier d'hétérogénéité, caractérisé par le manque d'information du chercheur par rapport aux individus.

Remarquons que l'économètre peut avoir parfaitement conscience des facteurs pertinents et connus de l'agent, mais se trouver dans l'impossibilité de les prendre en compte dans son modèle. On peut dire que l'hétérogénéité observée renvoie aux différences entre les observations mesurées par les variables explicatives, et l'hétérogénéité non observée aux autres différences. Il est rarement raisonnable de penser que toutes les variables influençant les transitions sont observées : certaines peuvent ne pas être mesurables, codifiables ou encore être absentes des données. En prenant en compte l'hétérogénéité non observée, on accepte l'idée qu'il existe des déterminants inobservés par l'économètre du phénomène étudié et on reconnaît une certaine ignorance.

L'hétérogénéité non observée ne correspond pas à une erreur de mesure ou encore à une mauvaise spécification du modèle. Ces trois notions sont très différentes mais nous verrons dans la sous-section 1.3.2 qu'elles sont empiriquement difficiles à distinguer, car elles génèrent les mêmes modèles réduits. Ce phénomène ont en commun l'idée sous-jacente du manque d'information. L'omission de variables fait que les régresseurs ne véhiculent pas assez d'information pour expliquer les durées observées, de même qu'une variable entachée d'erreurs de mesure est "brouillée" et contient moins d'information qu'une variable qui en est exempte. Ainsi, l'omission ou une mesure imprécise de variables mène à des durées que les régresseurs ne peuvent entièrement expliquer, phénomène identique à celui de l'hétérogénéité non observée.

## 2.3 Où se trouve-t-elle ?

De manière générale, on peut se demander si les résultats relèvent de comportements individuels semblables ou de l'échantillonnage dans une population plus hétérogène que ce que le modèle peut décrire. Les problèmes soule-

vés par l'hétérogénéité non observée dépassent ainsi largement le seul cadre de cette thèse consacrée aux modèles de durée, et il existe une littérature très vaste à ce sujet. Parmi de nombreux articles, on peut citer notamment Gouriéroux et Peaucelle (1990) montrent que les modèles microéconomiques d'agents représentatifs, ainsi que les modèles agrégés macroéconomiques, sont mal spécifiés en présence d'hétérogénéité individuelle et que ceci conduit à des biais. Ces résultats sont illustrés dans un papier joint, Gouriéroux (1990), où l'hétérogénéité non observée amène à surestimer les effets de substitutions dans les modèles de consommation ou de production, ainsi que le taux de progrès technique. D'autres exemples sont présentés par Moulton (1986) sur les thèmes des prix hédoniques, de la demande de logements et des fonctions de revenus prédites dans la théorie du capital humain. Heitmüller (2005) présente une étude de Monte Carlo où l'omission de variables explicatives, avec des données de panel où les régresseurs ne varient pas dans le temps, conduit à des biais importants. Concernant les modèles de durée à hasards proportionnels, Gouriéroux (1989, p. 361-362) étudie la convergence et le biais de l'estimateur du maximum de vraisemblance en présence de durées non-censurées. Lancaster (1990, p. 264-267) illustre ces résultats par une étude de Monte Carlo. Toutes les études précitées nous amènent à conclure que l'hétérogénéité non observée peut avoir de nombreuses explications (typiquement des caractéristiques individuelles inobservées, des externalités, etc.) et l'ignorer mène à des biais non négligeables sur certains estimateurs.

La question de l'hétérogénéité non observée n'est centrale que pour la recherche appliquée : savoir si une variable n'est pas ou mal observable n'est pas une question des plus importantes pour l'économiste élaborant des modèles théoriques. Il revient avant tout à l'économètre d'en tenir compte pour éviter que les résultats d'études soient erronés du fait d'une mauvaise spécification. Par contre, on a vu plus haut que le problème de l'hétérogénéité non observée peut survenir dans toute étude, quels que soient son thème et la méthodologie retenue, et donc que tout domaine de recherche appliquée est susceptible de bénéficier des améliorations des outils existants.

## 2.4 Comment la gérer ?

Therneau et Grambsch (2000, p. 168-169) résument les différentes approches possibles. Une première méthode est de tenir compte de l'absence d'indépendance entre les observations d'un même individu pour en déduire un estimateur corrigé de la matrice de variance de l'estimateur. Cette manière de procéder est appelée **l'approche marginale**. Une seconde méthode, appelée **l'approche conditionnelle**, inclut spécifiquement un terme d'hétérogénéité dans le modèle et procède à l'inférence en conditionnant par rapport

à celui-ci. Selon le contexte, il peut être pertinent de supposer un terme d'hétérogénéité commun à plusieurs observations, ou encore prenant des valeurs corrélées. Au sein de l'approche conditionnelle, l'hétérogénéité non observée peut être prise en compte de deux manières : au moyen d'effets fixes ou d'effets aléatoires.

### 2.4.1 Modélisation par effets fixes

L'avantage des modèles à effets fixes est qu'ils sont relativement faciles à estimer et ne requièrent pas certaines hypothèses discutables, comme l'indépendance entre les variables observables et inobservables ou encore le choix d'une distribution mélangeante. Le moyen le plus simple de prendre en compte des effets fixes est encore de spécifier des indicatrices pour les observations dont on s'attend à ce qu'elles soient corrélées. Son principal inconvénient est d'ajouter un grand nombre de paramètres lorsque les groupes sont petits. De plus, Andersen *et al.* (1999) montrent que les tests de Wald, du score ou de rapport de vraisemblance utilisés pour détecter l'hétérogénéité non observée en présence d'effets fixes sont moins puissants que le niveau spécifié en présence d'échantillons de taille courante. Un test du score en présence d'un effet aléatoire n'admet pas ce travers.<sup>3</sup> Un autre moyen est de considérer une approche stratifiée, comme la vraisemblance partielle stratifiée décrite dans Kalbfleisch et Prentice (1980), Yamaguchi (1986) et Ridder et Tunalı (1999). La population est supposée divisée en groupes et les groupes définis à un même niveau sont appelés des "strates".<sup>4</sup> Un test de l'hypothèse d'effet fixe (population stratifiée) contre l'hypothèse d'effet aléatoire (population non stratifiée) est présenté dans Ridder et Tunalı (1999). Ils décrivent également les inconvénients de cette approche, à savoir qu'il n'est pas possible d'estimer l'ampleur de l'hétérogénéité non observée et que l'hétérogénéité observée définie au même niveau que la non observée n'est pas identifiée. De plus, la convergence requiert un nombre d'observations par groupe qui tend vers l'infini, faisant des effets fixes un outil peu adapté aux données contenant de petits groupes.

---

<sup>3</sup>Les effets fixes sont représentés par des indicatrices et les statistiques des tests suivent des Khi-deux à  $K - 1$  degrés de liberté, où  $K - 1$  est le nombre d'indicatrices dont on teste la nullité des coefficients. Pour l'effet aléatoire, la statistique de test suit un Khi-deux à 1 degré de liberté. L'influence du nombre de degrés de liberté sur la puissance des tests n'est pas discutée dans Andersen *et al.* (1999).

<sup>4</sup>La notion de niveau dans le contexte de l'hétérogénéité non observée est présentée en détails dans la sous-section 2.5.

### 2.4.2 Modélisation avec des effets aléatoires

Dans la pratique, on préfère souvent une spécification en termes d'effets aléatoires. On introduit alors une variable aléatoire latente dans le modèle, qui impliquerait des durées conditionnellement indépendantes si elle était observée. Sa variance mesure l'importance de ce type d'hétérogénéité : plus celle-ci est élevée et moins les groupes sont homogènes entre eux. L'utilisation d'effets aléatoires requiert deux hypothèses critiquables, à savoir l'indépendance entre hétérogénéité observée et non observée ainsi que le choix d'une distribution mélangeante. On peut en effet rester sceptique quant à l'omission d'une variable explicative importante mais indépendante des autres régresseurs. Nous verrons dans la sous-section 1.4.2 que l'hypothèse d'indépendance peut-être relâchée dès lors que plusieurs observations partagent une même réalisation du terme d'hétérogénéité. Une telle structuration des données est généralement considérée ici, et le cas d'épisodes uniques qui nécessite l'hypothèse d'indépendance pour garantir l'identification des modèles n'est présenté qu'à des fins pédagogiques ou illustratives dans la suite de cette thèse. Concernant le choix de la distribution mélangeante, plusieurs études (Heckman et Singer, 1984a et Gouriéroux, 1989, p. 382-383) montrent, sur la base d'exemples, que les estimations varient selon la spécification de la loi du terme d'hétérogénéité et concluent que ce choix n'est pas neutre dans le cadre d'une estimation par maximum de vraisemblance. Une revue de littérature assez complète sur le modèle MPH lorsque la distribution mélangeante est mal spécifiée est présentée dans Van den Berg (2001). Ces résultats sont établis en présence d'épisodes uniques et doivent eux aussi être relativisés dès lors que l'on dispose de plusieurs observations pour la même réalisation du terme d'hétérogénéité. En effet, de nombreuses études (Nielsen *et al.*, 1992, Guo et Rodriguez, 1992 et Gönül et Srinivasan, 1993) montrent que le choix de la distribution mélangeante n'a qu'une influence négligeable dans ce contexte. Il semble donc raisonnable de modéliser l'hétérogénéité non-observée au moyen d'effets aléatoires dès lors que les durées sont multivariées.

Le bien fondé de la prise en compte de l'hétérogénéité non observée au moyen d'effets aléatoires a été discuté à plusieurs reprises. Un argument fréquemment rencontré est que les biais des coefficients sont faibles si l'hétérogénéité non observée n'est pas significative. Comme le font remarquer Hausman et Woutersen (2004), les estimations de la variance des distributions mélangeantes sont généralement peu précises et il n'est pas évident de conclure à l'absence d'hétérogénéité avec ce seul résultat. Un autre argument courant est qu'une spécification suffisamment souple du hasard de base permet aussi de capter l'hétérogénéité et donc de ne pas avoir à la spécifier. Bijwaard et Ridder (2005) montrent que les biais dans les coefficients estimés sont indé-



pendants de la forme du hasard de base. Ainsi, il n'existe pas de spécification du hasard de base qui permette de se prémunir des biais dus à l'hétérogénéité non observée.

### 2.4.3 L'hétérogénéité non observée dans les modèles multivariés

De nombreuses études requièrent l'emploi de modèles multivariés et l'intérêt qui leur est porté est principalement motivé par des considérations empiriques. Évaluer les liens qui existent entre plusieurs durées est une question importante, et il n'est pas rare d'analyser des données comprenant plusieurs observations par individu. Des exemples d'applications en sont les travaux sur le chômage, l'absentéisme ou le renouvellement de biens de capital. De même, plusieurs durées peuvent être reliées entre elles du fait d'un traitement qui serait administré à une date aléatoire avant la fin de l'épisode. Une telle méthodologie est utilisée pour l'étude de l'effet de programmes de formation, ou de réduction d'allocations, sur le chômage ; remplacement de personnel sur la durée d'une grève, etc.

Dans les modèles multivariés, les durées sont indépendantes conditionnellement à l'hétérogénéité, observée et non observée. Conditionnellement aux variables explicatives, les durées ne sont dépendantes que si les termes d'hétérogénéité non observée sont supposés dépendants.<sup>5</sup> Le caractère multivarié d'un modèle peut s'expliquer de différentes manières, et nous supposons ici qu'il résulte uniquement de la prise en compte de l'hétérogénéité non observée. Les autres composantes du modèles n'introduisent pas de corrélation entre les observations, et seule la structure de l'hétérogénéité non observée fait du modèle qu'il est multivarié. Nous nous concentrons dans cette thèse sur une hétérogénéité non observée commune à plusieurs observations, cas de figure typique lorsque différentes durées concernent un même individu ou plus généralement une même unité. Toutefois, la structure des données peut-être plus complexe et les durées peuvent être reliées à plusieurs unités de types différents. Par exemple, les individus peuvent appartenir à des communautés, elle-mêmes faisant partie d'ensembles plus grands. Nous sommes alors naturellement amenés à considérer plusieurs termes d'hétérogénéité.

## 2.5 Pourquoi prendre en compte plusieurs termes d'hétérogénéité ?

L'hétérogénéité non observée peut-être située à différents **niveaux**, ce terme désignant la position d'une observation au sein d'une structure de

<sup>5</sup>Si on fait l'hypothèse de termes d'hétérogénéité non observée indépendants, le modèle multivarié se réduit alors à plusieurs modèles univariés indépendants.

groupes. La structure est hiérarchique lorsque les unités définies au niveau 1 sont contenues dans les unités du niveau 2, elles même incluses dans celles du niveau 3, etc. Un exemple en est les enfants (unités de niveau 1), qui sont issues de familles (unités de niveau 2) elles même appartenant à des communautés (unités de niveau 3). Les niveaux ne sont pas nécessairement hiérarchiques. Les enfants de l'exemple précédent peuvent être groupés selon la couleur de leurs yeux ou leur religion, et il y a un niveau "couleur des yeux" et un autre "religion". Il est possible de prendre en compte des niveaux d'hétérogénéité dans un modèle tant qu'on peut les distinguer, et on peut les modéliser avec des effets fixes, aléatoires ou un mélange des deux.

L'objectif d'une analyse multiniveaux est d'expliquer les réalisations d'une variable dépendante, mesurée au niveau de l'unité observée, en prenant en compte l'information située aux autres niveaux. Ne considérer qu'un seul niveau d'hétérogénéité commun à plusieurs observations contraint l'hétérogénéité non observée à être la même pour toutes les observations au sein du même groupe. Par exemple, on suppose que les facteurs inobservés spécifiques à un chercheur d'emploi restent identiques pour tous ses épisodes de chômage, ce qui est discutable lorsque les différents épisodes (unités de niveau 1) d'un individu (unité de niveau 2) sont très éloignés dans le temps. Si on désire prendre en compte de manière plus précise les relations de dépendance entre les observations au sein des groupes, on est amené à considérer un modèle avec plusieurs niveaux d'hétérogénéité. Ainsi, une étude sur la mortalité infantile avec un effet aléatoire au niveau des communautés et un autre au niveau des familles prend en compte les facteurs inobservés spécifiques aux familles et ne les suppose plus égaux pour tous les enfants de la communauté.

Une raison de poser plusieurs effets est ainsi la volonté de quantifier l'hétérogénéité non observée, pour savoir à quel niveau elle est localisée et donc où se situent des variables explicatives importantes et omises. De plus, l'effet des caractéristiques à un niveau peut être amplifié ou contrebalancé à un autre niveau, et les différentes influences peuvent être complémentaires ou antagonistes sur le phénomène observé. Un diagnostic précis de l'hétérogénéité nous permet ainsi de dégager des axes de recherche ultérieurs pour obtenir une meilleure compréhension du phénomène étudié.

La constance de l'hétérogénéité non observée fait que les durées au sein d'un groupe sont positivement corrélées. Cette propriété peut être inadaptée lorsqu'on étudie, par exemple, une situation où la transition d'un individu entraîne une externalité négative sur les individus du même groupe. Une façon d'y remédier est de spécifier, par exemple, un effet aléatoire au niveau individuel dont les réalisations ne sont pas indépendantes pour toutes les unités du groupe.

Bien qu'améliorant la capacité descriptive du modèle, la prise en compte

d'une hétérogénéité multiple a pour inconvénient d'introduire une complexité supplémentaire qui se traduit par des modèles moins parcimonieux et une inférence plus difficile. Il revient à l'économètre de choisir si l'augmentation de la puissance descriptive du modèle compense le coût induit par la difficulté supplémentaire.<sup>6</sup> La décision de modéliser plusieurs niveaux d'hétérogénéité non observée au moyen d'effets aléatoires ne doit pas être dogmatique et relever de l'a priori ; elle se doit au contraire d'être justifiée et pragmatique.

### 3 Outils actuels et limites

Les approches disponibles à l'heure actuelle pour estimer les modèles de durées avec hétérogénéité non observée peuvent être regroupés en famille de méthodes d'augmentation de données et en famille de mélanges généralisés.<sup>7</sup>

#### 3.1 Différentes approches possibles ...

L'idée de l'augmentation de données a été formulée pour mener l'inférence dans les modèles où la variable expliquée dépend de variables latentes. L'estimation serait facile si toutes les variables étaient observées, la complexité ne venant que de l'inobservabilité de certains déterminants. On est ainsi amené à approcher dans un premier temps les variables latentes, puis à les traiter comme données. Cette approche est d'abord utilisée dans Dempster *et al.* (1977) pour établir l'algorithme EM, une procédure d'approximation de l'estimateur du maximum de vraisemblance. Elle a été reprise dans un cadre Bayésien par Tanner et Wong (1987), où la loi a posteriori admet une décomposition faisant intervenir la loi a posteriori d'un modèle où tout serait observé ainsi que la distribution des variables latentes. Leur approche est présentée dans la sous-section 1.7.3 et donne suite, entre autres, à l'échantillonnage de Gibbs. L'idée d'augmenter les données donne lieu à des approches paramétriques que l'on raisonne dans un cadre fréquentiste ou Bayésien, avec Sastry (1997) (approche fréquentiste) et Bolstad et Manda (2001) (approche Bayésienne). Parmi les méthodes semi-paramétriques, l'approche Bayésienne

<sup>6</sup>Des résultats suffisamment précis peuvent être obtenus sans modélisation de l'hétérogénéité non observée. Par exemple, son omission ne prêle pas à conséquence sur le score asymptotique dans un modèle exponentiel sans censure, comme le montre Gouriéroux (1989), p. 365-366. Les procédures de test habituelles peuvent être utilisées dans ce cas.

<sup>7</sup>Xue et Brookmeyer (1996), Xue (1998) et Xue et Ding (1999) présentent l'inférence lorsqu'un niveau d'hétérogénéité est géré au moyen d'une stratification du hasard de base. Leurs études ne sont pas présentées ici, mais elles peuvent constituer une alternative intéressante lorsqu'une hétérogénéité réside à un niveau très agrégé et les autres à des niveaux plus fins.

de Clayton (1991) ne considère qu'un seul effet et celle de Manda et Meyer (2005) est spécifique à des durées discrètes. Au sein des méthodes fréquentistes semi-paramétriques, car s'inscrivant dans le cadre de la vraisemblance partielle, figurent Vaida et Xu (2000), Ripatti et Palmgren (2002) et Cortinas Abrahantes et Burzykowski (2004).

Le principe d'augmentation des données a été développé pour estimer des modèles dont le point commun est la présence de données incomplètes. L'approche est partie d'une famille générale de modèles pour ensuite être appliquée à différents cas de figure précis, comme les modèles de durée. D'autres outils ont suivi le parcours inverse, d'abord conçus pour un modèle précis avant d'être étendus peu à peu à une famille de plus en plus large. C'est le cas de l'estimateur du maximum de vraisemblance contraint (estimateur REML, de l'anglais "restricted maximum likelihood"), qui vient des raffinements de l'inférence dans les modèles linéaires. Le point de départ est meilleur prédicteur linéaire sans biais (BLUP, de l'anglais "best linear unbiased predictor" et introduit dans Henderson, 1975), dont est déduit l'estimateur REML (McGilchrist et Aisbett, 1991, McGilchrist, 1993, 1994), puis l'estimateur REML étendu (Lee et Nelder, 2003) pour un effet aléatoire qui n'est pas normalement distribué. L'estimateur REML est issu de la maximisation d'une fonction comprenant la vraisemblance partielle ainsi qu'une distribution mélangeante pour le terme d'hétérogénéité, et Yau (2001) le présente comme un estimateur du maximum de la vraisemblance partielle pénalisée.

La maximisation d'une vraisemblance partielle pénalisée permet à Therneau *et al.* (2003) d'obtenir la solution de l'algorithme EM dans un modèle semi-paramétrique à hétérogénéité gamma. Les approches d'augmentation des données et du REML étendu, initialement séparées et sans références communes car n'ayant pas le même point de départ, se sont ainsi rejointes.

### 3.2 ... et leurs restrictions

Les gains du REML par rapport au maximum de vraisemblance sont peu clairs (McGilchrist, 1993), et nous ne faisons qu'évoquer ici cette famille d'approches. Nous consacrons toute notre attention dans cette thèse aux méthode d'augmentation de données.

La méthode d'augmentation des données est très souple et a été étendue aux modèles à plusieurs effets aléatoires. Toutefois, ses implémentations dans le cadre fréquentiste se font au moyen de l'algorithme EM et souffrent de ses travers numériques, à savoir une convergence demandant généralement beaucoup d'itérations, sensible au choix des valeurs initiales, qui échoue parfois et

avec un temps de calcul élevé.<sup>8</sup> Les approches Bayésiennes sont quant à elles assez spécifiques, et le cas de plusieurs effets n'est traité qu'en temps discret.

**Les outils existants pour estimer les modèles de durée comprenant plusieurs effets aléatoires peuvent donc être améliorés, et l'objectif principal de cette thèse est de fournir des approches semi-paramétriques en présence de plusieurs effets aléatoires en partant du principe d'augmentation de données. Nous cherchons à fournir une méthode générale basée sur l'algorithme EM et exempte de ses difficultés numériques, ainsi qu'à compléter les méthodes Bayésiennes, encore peu développées sur ce point. Nous voulons également mettre en perspectives les approches fréquentistes et Bayésiennes.** Pour ce faire, nous allons nous concentrer sur la famille des modèles de mélange de hasards proportionnels (MPH).

Les modèles MPH n'émergent de la théorie économique qu'au prix de lourdes hypothèses. Considérant une approche de type "job-search", Eckstein et Van den Berg (2003) insistent sur les problèmes soulevés par une spécification MPH et Van den Berg (2001) étudie les hypothèses sous lesquelles le modèle structurel admet une forme réduite MPH. Cette spécification émerge de la théorie économique lorsque la stratégie optimale est myope (du fait de recherches répétées, ou bien de taux d'actualisation infini), c'est-à-dire que la spécification MPH peut difficilement se justifier théoriquement. Comme l'analyse structurelle appropriée ne surgit pas spontanément des applications auxquelles les économistes sont confrontés, le modèle MPH a souvent été utilisé de manière exploratoire, dans des domaines où il n'existe pas de théorie bien établie. La famille des modèles MPH est très largement utilisée dans les applications, et l'amélioration des outils existants en est donc d'autant plus nécessaire.

## 4 Méthodologie et Résultats

Le problème de l'hétérogénéité non observé n'est pertinent que d'un point de vue empirique, et cette thèse porte sur les **approches réduites**. Dans le Chapitre 1, nous présentons les principaux concepts de l'économétrie des modèles de durée. Nous introduisons la fonction de hasard et son rôle central dans la spécification, ainsi que sa reformulation en termes de processus de comptage. Puis nous nous intéressons aux principaux estimateurs non paramétriques du hasard intégré et de la fonction de survie. Nous introduisons ensuite le modèle de Cox avant de nous intéresser à sa généralisation qu'est

---

<sup>8</sup>Les difficultés numériques de l'algorithme EM sont présentées plus en détail dans le Chapitre 2.

le modèle MPH. Nous présentons en détail ses propriétés, qui sont étudiées et illustrées très précisément. Un grand avantage des modèles MPH est leur structure particulièrement souple, qui permet de prendre facilement compte des régresseurs variant dans le temps ou encore une hétérogénéité multiveaux. Nous représentons donc l'hétérogénéité non observée au moyen de **plusieurs effets aléatoires**.

Nous nous concentrons ensuite sur l'identification et l'inférence. L'identification est obtenue sous différentes hypothèses et nous nous intéressons plus particulièrement à celles portant sur la distribution mélangeante, que le modèle soit à épisodes uniques, multiples ou multivariés. Comme un aspect central de cette thèse est la présence de données manquantes, les méthodes d'estimation que nous étudions utilisent le principe d'augmentation des données. L'idée est d'ajouter des données à celles qui sont observées pour mener à bien l'inférence dans le modèle "complété". L'idée a été utilisée avec succès tout d'abord dans la conception de l'algorithme EM (Dempster *et al.*, 1977) puis dans l'analyse Bayésienne (Tanner et Wong, 1987). Nous revoyons les méthodes usuelles d'estimation utilisant le principe d'augmentation des données, en commençant par le maximum de vraisemblance. Elles servent d'introduction aux méthodes **méthodes semi-paramétriques** et plus particulièrement aux approches utilisant la vraisemblance partielle comme point de départ. Elles requièrent moins d'hypothèses que les approches complètement paramétriques, et le risque d'une spécification erronée en est réduit. Toujours guidés par des préoccupations empiriques, nous étudions ensuite les approches Bayésiennes utilisant l'augmentation des données. Nous considérons ainsi deux axes de recherche.

Un premier axe de recherche consiste à adapter l'algorithme EM aux modèles de durée multivariés comprenant plusieurs effets aléatoires. Nous étendons ainsi les travaux existants dans la mesure où notre reformulation ne dépend ni du nombre d'effets aléatoires ni d'une distribution mélangeante particulière. La méthode d'estimation proposée requiert l'estimation de modèles à un seul effet aléatoire, et nous montrons ensuite comment faire usage d'une vraisemblance partielle pénalisée au sein de ce nouvel algorithme. Nous obtenons ainsi une procédure itérative équivalente à l'algorithme EM et appelée algorithme EMPL (de l'anglais "Expectation Maximisation algorithm based on Penalised Likelihood), mais exempte de ses difficultés numériques car ne reposant pas sur les mêmes éléments, et bien plus rapide. Elle est illustrée par des simulations et une application à des données portant sur le rythme de ratification des conventions de l'Organisation Internationale du Travail (OIT) dans le Chapitre 2.

Un second axe consiste à adapter l'approche Bayésienne présentée dans Clayton (1991) à la présence de plusieurs effets aléatoires. Cette approche

n'est pas soumise aux mêmes difficultés numériques que l'algorithme EM et permet de gérer facilement les modèles de mélanges généralisés avec une structure de groupe au sein des données. Nous montrons dans le Chapitre 3 comment mener une inférence semi-paramétrique en présence de deux effets aléatoires et fournissons une application sur le rythme de ratification des conventions de l'OIT. Nos résultats indiquent des comportements de ratification très différents selon les pays et les conventions, les catégorisations usuelles en pays industrialisés et pays en voie de développement étant insuffisantes pour tenir compte de toute l'hétérogénéité, de même que le regroupement des conventions selon les thèmes qu'elles abordent. Les résultats Bayésiens sont comparés avec ceux fournis par l'EMPL, présenté dans le Chapitre 2. Ils indiquent une différence dans l'évaluation des effets aléatoires, l'algorithme EMPL sous-estimant l'importance d'un terme d'hétérogénéité non observée.<sup>9</sup>

Comme il est possible de gérer une hétérogénéité multiniveaux lorsqu'il y a deux effets indépendants, nous étudions dans le Chapitre 4 une structure d'hétérogénéité plus complexe faisant intervenir deux effets aléatoires corrélés. L'application porte sur la mobilité interprofessionnelle au Portugal et utilise des données appariées employeurs-employés. L'idée sous-jacente est que l'association des hétérogénéités non observées propres à une entreprise et à un individu peuvent intervenir dans la qualité de la relation d'emploi. Nous spécifions donc un modèle MPH en temps discret avec deux effets aléatoires corrélés que nous estimons avec une approche Bayésienne.

La thèse apporte donc, du point de vue **méthodologique**, des méthodes fréquentistes et Bayésiennes pour estimer les modèles MPH avec plusieurs effets aléatoires (Chapitres 2, 3 et 4) et met les différentes approches en perspective. Elle contribue également aux méthodes statistiques assistées par ordinateur et proposant un outil plus stable et plus rapide que ceux disponibles à l'heure actuelle (Chapitre 2). Du point de vue **empirique**, elle apporte des nouveaux éléments d'analyse du marché du travail. Dans un premier temps, nous l'appréhendons sous l'angle peu étudié de la régulation internationale par les conventions de l'OIT (Chapitre 3). Nous étudions dans un second temps les déterminants de la mobilité professionnelle en insistant sur l'appariement employeur-employé (Chapitre 4).

---

<sup>9</sup>Le même problème de sous estimation est constaté dans Rodriguez et Goldman (2001) et Therneau *et al.* (2003) pour d'autres méthodes déduites de l'estimateur du maximum de vraisemblance.

**Revue de littérature :  
propriétés, identification et  
inférence dans le modèle MPH**



## Résumé

Nous présentons un aperçu de la littérature sur les modèles de durée, en insistant plus particulièrement sur la spécification, et l'identification et l'inférence dans les modèles MPH. Après avoir introduit les concepts spécifiques à l'étude des durées, nous nous intéressons à l'hétérogénéité observée avec le modèle de Cox et son estimation. Il repose sur l'hypothèse de hasards proportionnels, que la présence de mélanges amène à rejeter.

Nous considérons l'extension du modèle de Cox au modèle MPH afin de prendre en compte l'hétérogénéité non observée. Les propriétés du modèle et différentes spécifications de l'hétérogénéité non observée sont discutées en détails. La structure des déterminants inobservés permet d'introduire des interdépendances précises entre les observations et caractérise les modèles MPH à épisodes multiples et MPH multivariés que nous présentons. L'identification de ces différents modèles est étudiée en détails. Nous présentons ensuite les méthodes récentes d'estimation, utilisant la méthode d'augmentation des données, par maximum de vraisemblance, maximum de vraisemblance partielle et l'approche Bayésienne.

## Sommaire

---

<b>Résumé . . . . .</b>	<b>20</b>
<b>1.1 Éléments de base des modèles de durée . . . . .</b>	<b>21</b>
<b>1.2 Le modèle de Cox . . . . .</b>	<b>34</b>
<b>1.3 Le modèle de Mélange de Hasards Proportionnels</b>	<b>49</b>
<b>1.4 Identification . . . . .</b>	<b>65</b>
<b>1.5 Inférence basée sur la vraisemblance . . . . .</b>	<b>78</b>
<b>1.6 Les approches basées sur la vraisemblance partielle</b>	<b>84</b>
<b>1.7 L'approche Bayésienne . . . . .</b>	<b>88</b>

---

## 1.1. Éléments de base des modèles de durée

L'OBJECTIF de cette section est de familiariser le lecteur avec les éléments de base des modèles de durée. Nous présentons en premier lieu deux concepts centraux que sont les fonctions de hasard et de survie. L'étude des durées a bénéficié des développements de la théorie des processus de comptage, dont on trouve un exposé chez Fleming et Harrington (1991). Les fondements théoriques de l'analyse des durées en ont été solidifiés, et certains outils ont ainsi pu être mis au point. C'est le cas de l'estimateur non paramétrique de Nelson-Aalen, sur lequel sont basés les estimateurs de la fonction de survie de Breslow et de Kaplan-Meier. À partir des résultats d'estimation obtenus au moyen de ces outils, on peut déduire certaines propriétés de la fonction de hasard que l'économètre devra prendre en compte en spécifiant un modèle. Nous aborderons ensuite la question de l'étude des durées au moyen de la spécification directe d'une forme fonctionnelle pour le hasard, avant de présenter la famille des modèles dits de "hasards proportionnels". Une alternative pour l'étude des durées est d'utiliser l'approche de régression, dont nous présenterons les fondements. Cette section s'appuie principalement sur les travaux d'horizon de Lancaster (1990), Van den Berg (2001), Therneau et Grambsch (2000), ainsi que Fleming et Harrington (1991) pour les questions relevant des processus de comptage.

### 1.1.1 Concepts fondamentaux : fonction de survie, fonction de hasard et censure

Commençons avec un exemple. Les modèles de durée servent souvent à étudier le chômage, et considérons un individu qui vient de perdre son emploi et en recherche un nouveau. La modélisation suppose que la durée passée au chômage par l'individu est aléatoire, et notons par  $T$  la variable aléatoire positive indiquant le temps passé dans cet état, et par  $t$  ses réalisations ( $0 \leq t < \infty$ ).  $T$  n'indique pas le temps calendaire, comme par exemple la date de sortie de l'individu étudié des fichiers de l'agence pour l'emploi, mais mesure le temps passé au chômage par le chercheur d'emploi, par exemple 185 jours ou bien 3 mois.

## Fonctions de survie et de hasard

On note  $f$  la fonction de densité de  $T$  et  $F$  sa fonction de répartition, définie par  $F(t) = P(T \leq t)$  où  $F(0) = 0$ . La **fonction de survie** est :

$$S(t) = 1 - F(t) = \int_t^{\infty} f(u)du = P(T \geq t). \quad (1.1)$$

Elle fournit la probabilité que la longueur de l'épisode de chômage excède  $t$ , c'est-à-dire la probabilité que l'individu soit toujours sans emploi après  $t$  unités de temps.

Nous allons maintenant définir la **fonction de hasard**. En temps discret, la fonction de hasard de la variable  $T$ , évaluée à la date  $t$ , est la probabilité de fin de l'épisode en  $t$  sachant qu'il ne s'est pas terminé plus tôt. Dans notre exemple, elle correspond à la probabilité que l'individu sorte du chômage à la date  $t$  sachant qu'il n'en est pas sorti avant. En temps continu, la fonction de hasard s'écrit :

$$\lambda(t) = \lim_{dt \rightarrow 0} \frac{1}{dt} P(T \in [t, t + dt] | T \geq t). \quad (1.2)$$

Le conditionnement par rapport à  $T \geq t$  implique que la probabilité est spécifique à une sous-population. Dans notre exemple, la fonction de hasard donne la probabilité qu'un individu appartenant à la sous-population de ceux qui chôment au moins 185 jours trouve un emploi au bout de ces 185 jours. La probabilité non conditionnelle serait quant à elle calculée pour tous les individus de la population, c'est-à-dire tous les chômeurs, sans égard aux temps qu'ils ont passé sans emploi. Le conditionnement devient particulièrement important lorsqu'un individu peut effectuer plusieurs transitions. Il s'agirait du cas où l'individu connaît plusieurs périodes de chômage durant l'étude et retrouve plusieurs fois du travail. Conditionner par rapport à  $T \geq t$  implique que chaque transition doit être considérée comme une nouvelle observation.

La valeur prise par la fonction de hasard à une date  $t$  donnée est appelée le hasard, ou encore le taux de sortie. Cette dernière expression souligne que la fin d'un épisode coïncide avec une transition vers un autre état, par exemple un nouvel emploi ou bien l'arrêt de la recherche de travail. A noter que l'expression "fonction de hasard" est une traduction littérale de l'anglais "hazard function", le mot "hazard" ayant une signification courante plus proche de risque naturel ou de risque lié à une activité professionnelle que son homologue français. Le mot "survie" indique quant à lui le risque naturel correspondant à la mort, et nous vient peut-être de la démographie ou de la biométrie.

## 1.1 Éléments de base des modèles de durée

---

Il est possible d'établir des liens entre la fonction de hasard et la fonction de survie. Une première relation est obtenue en explicitant la fonction de hasard au moyen de la formule de Bayes :

$$\begin{aligned} P(T \in [t, t + dt[ | T \geq t) &= \frac{P(T \in [t, t + dt[, T \geq t)}{P(T \geq t)} \\ &= \frac{P(T \in [t, t + dt[)}{P(T \geq t)}. \end{aligned} \quad (1.3)$$

On peut reformuler cette expression avec les fonctions de répartition et de survie :

$$P(T \in [t, t + dt[ | T \geq t) = \frac{F(t + dt) - F(t)}{S(t)}. \quad (1.4)$$

En divisant par  $dt$  et en prenant la limite lorsqu'il tend vers 0, on obtient :

$$\begin{aligned} \lambda(t) &= \frac{1}{S(t)} \lim_{dt \rightarrow 0} \frac{F(t + dt) - F(t)}{dt} \\ &= \frac{f(t)}{S(t)}. \end{aligned} \quad (1.5)$$

On a ainsi une relation entre le hasard, la densité de probabilité et la fonction de survie permettant de déduire l'une des fonctions à partir des deux autres.

Notons  $\Lambda(t)$  le hasard intégré  $\int_0^t \lambda(u) du$ . Sachant que  $f(t) = -dS(t)/dt$ , on en déduit :

$$\Lambda(t) = \int_0^t \frac{f(u)}{S(u)} du = - \int_0^t \frac{1}{S(u)} dS(u) = - \ln S(t). \quad (1.6)$$

D'où une seconde relation très utilisée :

$$S(t) = \exp \left( - \int_0^t \lambda(u) du \right). \quad (1.7)$$

On remarque que, grâce aux relations (1.5) et (1.7), la loi des durées peut être caractérisée par sa densité, sa fonction de répartition, sa fonction de survie, son hasard ou bien son hasard intégré.<sup>1</sup>

### Censure

Nous introduisons maintenant la notion de censure d'un épisode. Prenons l'exemple d'une étude sur le chômage dont l'objectif serait de voir si certaines

---

<sup>1</sup>Entre autres, puisque la fonction caractéristique ou la fonction génératrice des moments (si elle existe) peuvent aussi être utilisées.

mesures d’accompagnement de la recherche d’emploi augmentent le hasard. Lorsque l’étude se termine, certaines personnes sont toujours sans emploi et tout ce que nous savons à leur sujet est que la longueur de leur épisode de chômage est au moins aussi grande que la durée de l’enquête. Par ailleurs, certains individus peuvent sortir du champ de l’enquête avant que celle-ci prenne fin, que ce soit en ne répondant plus aux questionnaires qui leur sont envoyés, ou bien en ne venant plus aux entretiens. La notion de **censure** nous permet de tenir compte de ces observations et de l’information qu’elles contiennent. Soit une variable aléatoire  $C$ , indépendante de  $T$ , indiquant le temps qui sépare le début d’un épisode de chômage de la sortie du champ de l’enquête ou bien de la fin de l’étude.<sup>2</sup> Notons, pour l’individu  $i$  ( $i = 1, \dots, n$ ),  $c_i$  la réalisation de  $C$  et  $t_i$  la réalisation de  $T$ . Remarquons que la réalisation de  $C$  ne peut pas être observée pour les individus ayant trouvé un emploi avant la fin de l’étude. En effet, l’enquêteur ne dispose que de  $\min(t_i, c_i)$  et d’une variable indicatrice de non-censure définie comme  $\delta_i = I_{\{t_i \leq c_i\}}$  prenant la valeur 0 si l’observation  $i$  est censurée et 1 sinon.

Les concepts fondamentaux de hasard et de fonction de survie ont été présentés ci-dessus suivant l’approche “traditionnelle” des modèles de durée, que l’on retrouve notamment dans Lancaster (1990) ou Van den Berg (2001). Grâce au développement des processus de comptage, il est devenu possible de reformuler ces concepts fondamentaux.

### 1.1.2 Reformulation en termes de processus de comptage

Nous décrivons dans cette sous-section quelques notions des processus de comptage, afin d’éclairer les concepts de base des modèles de durée sous un jour nouveau et d’apporter des intuitions complémentaires. Des développements plus approfondis peuvent être trouvés dans Fleming et Harrington (1991), parmi d’autres. Par ailleurs, certains travaux comme Bijwaard *et al.* (2003), Florens et Fougère (1996) ou Therneau et Grambsch (2000) exposent les éléments nécessaires à l’analyse des modèles de durée.

---

<sup>2</sup>On parle alors de censure aléatoire. On distingue également entre censures de type I, survenant lorsque tous les individus sont observés du début à la fin d’une enquête dont la durée est fixe et connue à l’avance, et censures de type II lorsque l’enquête s’arrête après qu’un nombre prédéterminé de transitions ait eut lieu, les durées non échues étant censurées. Ces différents mécanismes de censure ont en commun d’être non informatifs et de ne pas entraîner de modification de la vraisemblance. Ils sont présentés en de plus amples détails dans Rodriguez (2001).

### Processus de comptage et de risque

L'approche par processus de comptage reconstitue l'information contenue dans les durées observées et les indicatrices de censure en utilisant le nombre de transitions effectuées et le nombre d'individus observés à chaque instant.

Plus précisément, un processus de comptage est un processus stochastique continu à droite dont la valeur initiale est 0 et dont les incréments sont de 1. Considérons un échantillon où les individus sont indicés par  $i$  ( $i = 1 \dots n$ ).  $N_i(t)$  est défini comme le nombre d'événements observés dans l'intervalle de temps  $[0, t]$  pour l'individu  $i$ . On note  $Y_i(t)$  la variable indicatrice prenant la valeur 1 si l'épisode de l'individu  $i$  est à la fois observé et susceptible de prendre fin en  $t$ , 0 sinon. On appelle  $R(t) = \{i \in \{1, \dots, n\}; Y_i(t) = 1\}$  l'ensemble des individus "au risque" à la date  $t$ , car il indique quels individus sont susceptibles d'effectuer une transition en  $t$ . On peut exprimer  $N_i(t)$  et l'ensemble des individus au risque à partir des  $t_i$  et  $\delta_i$  au moyen des relations suivantes :

$$N_i(t) = \mathbb{1}[t_i \leq t, \delta_i = 1], \quad (1.8)$$

$$Y_i(t) = \mathbb{1}[t_i \geq t]. \quad (1.9)$$

On remarque que  $N_i(t)$  définit bien un processus de comptage et comme les processus représentant des passages d'un état vers un autre produisent une suite de points suivant l'axe du temps, où chaque point correspondant au moment où une transition a été effectuée, on pressent déjà l'adéquation avec les processus stochastiques sous-tendant les modèles de durée. Des situations plus complexes, comme le cas d'individus connaissant des épisodes multiples ou bien des durées censurées à gauche, peuvent être modélisées à partir de  $N_i(t)$  et de  $Y_i(t)$ .

Un processus est prédictible lorsqu'il est mesurable par rapport à une  $\sigma$ -algèbre prédictible, c'est-à-dire une  $\sigma$ -algèbre générée par tout processus continu à gauche avec une limite à droite (abrégé caglad). Sans entrer dans des considérations se rapportant à la théorie de la mesure, prédictible signifie que la connaissance du processus à une date  $t - dt$  située infinitésimalement avant  $t$ , détermine la connaissance que nous en avons à la date  $t$ . Ou encore, la valeur d'un processus prédictible en  $t$  est uniquement influencée par des éléments observés à une date située infinitésimalement avant  $t$ .

L'observabilité est ici importante et peut rendre un processus prédictible ou non. Prenons l'exemple d'une personne en recherche d'emploi qui a prévu de déménager dans une autre région. Cette information peut affecter son comportement de recherche d'emploi, et donc sa fonction de hasard. Si cette information est connue de l'économètre, le processus est prédictible ; sinon, il ne l'est pas. Si la valeur actuelle d'un processus stochastique ne dépend que

du passé, alors il est prédictible. Un exemple souvent cité pour illustrer la notion de prédictibilité est celui d'un jeu de hasard. Soit  $Y(t)$  un processus indiquant si un individu participe à une  $t$ -ième loterie. Pour diverses raisons, il peut ne pas avoir participé à toutes les loteries précédentes. Par contre, juste avant le début de la loterie  $t$ , on sait si l'individu y prend part ou non. Il convient de remarquer que la prédictibilité d'un processus n'implique pas que la totalité de ses réalisations futures soit prévisible en un point donné du temps.

Soient les processus  $N = \{N_i(t) : t > 0\}$  et  $Y = \{Y_i(t) : t > 0\}$ . Une différence importante entre  $N$  et  $Y$  se situe au niveau de la continuité :  $N$  est continu à droite tandis que  $Y$  est continu à gauche, et  $Y$  est un processus caglad. La variable  $Y$ , indiquant l'observabilité, suit donc un processus prédictible et il n'en va pas de même pour  $N$ .

### Information disponible et réécriture du hasard

On précise l'information sur laquelle est basée le modèle en spécifiant l'histoire de l'enquête jusqu'à la date  $t$  incluse. Plus précisément, on appelle filtration une suite croissante de sous-tribus sur un espace probabilisé, que l'on note  $\{F_t; t \geq 0\}$ . Généralement, chaque sous-tribu  $F_t$  comprend l'état du processus de comptage  $N_i$ , du processus de risque  $Y_i$  et des variables explicatives  $X_i$  lorsqu'il y en a ; à la date  $t$ . Une sous-tribu est la formalisation de l'information disponible en  $t$ , et la filtration représente son évolution au fil du temps. Comme l'information disponible s'accroît au fur et à mesure que le temps passe, on a  $F_t \subseteq F_u$  pour  $u \geq t$ .

Il est possible de réécrire la fonction de hasard en termes de processus de comptage. Comme l'incrément du processus de comptage  $dN_i(t)$  ne peut prendre que la valeur 0 ou 1,<sup>3</sup> on peut faire l'approximation suivante :

$$E(dN_i(t)|F_{t-}) \approx P(dN_i(t) = 1|F_{t-}), \quad (1.10)$$

où  $F_{t-}$  représente l'information accumulée sur l'intervalle  $[0, t[$ . Du fait de l'indépendance des durées, seule l'information relative à l'individu  $i$  influence son épisode. On a donc, dans un cas sans variable explicative :  $P(dN_i(t) = 1|F_{t-}) = P(dN_i(t) = 1|Y_i(t))$ . Le cas  $Y_i(t) = 1$  correspond à une situation où l'individu est à la fois observé et susceptible d'effectuer une transition dans l'intervalle de temps  $[t, t + dt[$ , on a donc :

$$P(dN_i(t) = 1|Y_i(t) = 1) = P(t \leq T_i < t + dt | t \leq T_i, t \leq C_i). \quad (1.11)$$

---

<sup>3</sup>En effet, on suppose que l'individu  $i$  ne peut effectuer qu'une seule transition à l'instant  $t$ .

Or, les variables aléatoires  $T_i$  et  $C_i$  sont indépendantes dans le cas de censure non-informative. L'expression ci-dessus peut donc se simplifier, et on obtient la relation suivante entre les processus de comptage et le hasard défini en (1.2) :

$$P(dN_i(t) = 1 | Y_i(t) = 1) = \lambda(t)dt. \quad (1.12)$$

Le cas  $Y_i(t) = 0$  correspond à un individu dont l'épisode s'est terminé, ou bien à une personne sortie du champ de l'enquête et n'étant plus observée, donc son processus de comptage est incrémenté avec probabilité nulle. On en déduit la reformulation du hasard dans le cadre des processus de comptage :

$$E(dN_i(t) | F_{t-}) \approx P(dN_i(t) = 1 | Y_i(t)) = Y_i(t)\lambda(t)dt. \quad (1.13)$$

Les processus de comptage ont été utilisés dans de nombreuses approches, notamment pour l'estimation du hasard intégré  $\Lambda(t)$  ou de la fonction de survie  $S(t)$ . Bien que la fonction de hasard soit l'un des principaux centres d'intérêt en économétrie des modèles de durée, il est souvent utile de procéder de manière indirecte et d'estimer le hasard intégré ou bien la fonction de survie, et d'en déduire ensuite certaines caractéristiques de la fonction de hasard. La sous-section suivante présente d'abord l'estimateur de Nelson-Aalen de  $\Lambda(t)$ , à partir duquel les estimateurs de Kaplan-Meier et de Breslow pour  $S(t)$  ont été construits.

### 1.1.3 Estimation du hasard intégré : l'estimateur de Nelson-Aalen

L'estimateur de Nelson-Aalen a été initialement présenté dans l'article de Nelson (1969) étudiant la durée de fonctionnement de ventilateurs avant une panne. Une des idées avancées est que les ventilateurs les plus fragiles ou présentant des défauts vont tomber panne assez tôt, tandis que les autres fonctionneront plus longtemps. Comme le font remarquer Therneau et Grambsch (2000, p. 9), on trouve déjà ici l'idée d'une fragilité spécifique à une sous-population, dont nous reparlerons en plusieurs occasions. L'estimateur de Nelson-Aalen peut être obtenu de plusieurs manières différentes. Nous présentons ici celle utilisant les processus de comptage, mais on peut aussi l'obtenir par maximum de vraisemblance dans un modèle de hasard en temps discret (voir par exemple Zhou, 2002).

Soit un échantillon de  $n$  variables aléatoires indépendantes  $T_i$ , dont les réalisations respectives sont notées  $t_i$ . Les observations sont éventuellement censurées, et on définit  $Y_i(t) = \mathbb{1}[t_i \geq t]$ . On note par  $N_i(t)$  une indicatrice prenant la valeur 1 si l'épisode  $i$  est terminé à la date  $t$  ou avant, et 0 sinon. Soit le processus agrégé  $\bar{Y}(t) = \sum_{i=1}^n Y_i(t)$ . Il indique le nombre de durées



non échues et non censurées à chaque instant. Plus précisément, il s'agit du nombre de durées "au risque" de se terminer dans l'intervalle infinitésimal  $]t, t + dt]$ .<sup>4</sup> On définit le processus indiquant le nombre total de transitions effectuées à la date  $t$  et avant par  $\bar{N}(t) = \sum_{i=1}^n N_i(t)$ . Sur un intervalle de temps  $]s, s + ds]$  suffisamment petit, l'incrément du hasard intégré peut être approché par le hasard instantané en  $s$  multiplié par la longueur  $ds$  de l'intervalle :

$$\Lambda(s + ds) - \Lambda(s) \approx \lambda(s)ds. \quad (1.14)$$

Le membre de droite n'est autre que la probabilité instantanée qu'une transition ait lieu dans l'intervalle  $]s, s + ds]$  sachant les individus au risque à la date  $s$ . On peut d'estimer cette quantité par une fréquence, c'est-à-dire le rapport du nombre de durées échues sur le nombre total de durées pouvant se terminer dans l'intervalle  $]s, s + ds]$ , ou encore :

$$\hat{\lambda}(s)ds = \frac{\bar{N}(s + ds) - \bar{N}(s)}{\bar{Y}(s)}. \quad (1.15)$$

En sommant sur les intervalles de temps, on obtient :

$$\hat{\Lambda}(t) = \int_0^t \frac{d\bar{N}(s)}{\bar{Y}(s)}. \quad (1.16)$$

Habituellement, on note  $t_{(i)}$  les durées dans un échantillon ordonné avec  $(i) = 1$  l'indice de la durée la plus courte,  $(i) = 2$  la seconde durée la plus courte, et ainsi de suite ... D'où une formulation alternative utilisée dans les applications :

$$\hat{\Lambda}(t) = \sum_{(i):t_{(i)} \leq t} \frac{d\bar{N}(t_{(i)})}{\bar{Y}(t_{(i)})}. \quad (1.17)$$

Graphiquement, on obtient une fonction en escalier où chaque saut correspond à la fin d'un épisode. Il est possible de la lisser, et la pente en  $t$  de la courbe ainsi obtenue estime le hasard en  $t$ .

Il existe deux manières de calculer la variance de l'estimateur de Nelson-Aalen. La première méthode est basée sur un processus de Poisson, la seconde sur les martingales. Nous reviendrons ultérieurement sur cette dernière, et commençons par rappeler brièvement les principales caractéristiques du **processus de Poisson**.

---

<sup>4</sup>La forme de l'intervalle vient du fait que l'individu doit être au risque avant de pouvoir effectuer une transition.

### 1.1.4 Notions liées au processus de Poisson

Il existe de nombreuses explications détaillées de ce processus très simple et pouvant décrire de nombreuses situations. Le lecteur désirant plus de détails est renvoyé, par exemple, à Lancaster (1990, p. 85-88) ou bien à Gouriéroux et Monfort (1990, p. 913-915). Comme précédemment, le nombre de transitions se produisant dans l'intervalle  $]t, t + dt]$  est noté par  $d\bar{N}(t)$ . Les événements sont distribués selon un processus de Poisson de paramètre  $\lambda$  ( $\lambda > 0$ ) si :

$$P(d\bar{N}(t) = 0) = 1 - \lambda dt + o(dt), \quad (1.18)$$

$$P(d\bar{N}(t) = 1) = \lambda dt + o(dt). \quad (1.19)$$

et si l'incrément  $dN(t)$  est indépendant des incréments passés  $dN(u)$  ( $u < t$ ).<sup>5</sup> Les équations (1.18) et (1.19) impliquent que la réalisation de plusieurs événements dans l'intervalle  $]t, t + dt]$  a une probabilité de  $o(dt)$ . On peut montrer que si un nombre d'événements suit processus de Poisson de paramètre  $\lambda$ , les incréments  $N(t_2) - N(t_1)$  de ce nombre, où  $t_2 > t_1$ , suivent une loi de Poisson de paramètres  $(t_2 - t_1)\lambda$ . De plus, les durées séparant les événements sont distribuées selon une loi exponentielle de paramètre  $\lambda$ . La relation duale entre le processus de Poisson et la loi exponentielle, ainsi que l'indépendance de l'incrément  $dN(t)$  vis-à-vis des incréments passés, explique pourquoi cette loi est parfois appelée "loi sans mémoire".

On peut trouver la variance de l'estimateur de Nelson-Aalen au moyen d'un argument utilisant un processus de Poisson.

### 1.1.5 Variance de l'estimateur de Nelson-Aalen

Sur des intervalles de temps suffisamment petits, les incréments  $d\bar{N}(t)$  d'un processus de comptage sont de taille 1 et indépendants si les durées  $T_i$  sont elles-mêmes indépendantes. Il est dès lors naturel de modéliser les transitions par un processus de Poisson. Rappelons que pour un individu au risque, la probabilité d'une transition sur l'intervalle de longueur  $dt$  est de  $\lambda(t)dt$ . On en déduit que  $d\bar{N}(t)$  suit une loi de Poisson de paramètre  $\int_t^{t+dt} \bar{Y}(u)\lambda(u)du$ , que l'on approche par  $\bar{Y}(t)\lambda(t)dt$ . L'espérance et la variance de l'incrément peuvent donc être approchées par :

$$E(d\bar{N}(t)) = V(d\bar{N}(t)) \approx \bar{Y}(t)\lambda(t)dt. \quad (1.20)$$

L'estimateur de Nelson-Aalen, donné par (1.17), est la somme des incréments normalisés par le nombre d'individus au risque. Comme les transitions sont

---

<sup>5</sup>On rappelle que  $o(h)$  désigne toute fonction telle que  $\lim_{h \rightarrow 0} o(h)/h = 0$ .

supposées s'effectuer selon un processus de Poisson, les incréments du processus de comptage sont indépendants. On en déduit :

$$V \left[ \widehat{\Lambda}(t) \right] = \sum_{(i):t_{(i)} \leq t} \frac{1}{\overline{Y}^2(t_{(i)})} V [d\overline{N}(t_{(i)})] \approx \sum_{(i):t_{(i)} \leq t} \frac{\lambda(t_{(i)})}{\overline{Y}(t_{(i)})} dt_{(i)}. \quad (1.21)$$

En utilisant la relation (1.15), on obtient l'expression de l'estimateur de la variance de l'estimateur de Nelson-Aalen en temps discret :

$$V \left[ \widehat{\Lambda}(t) \right] = \sum_{(i):t_{(i)} \leq t} \frac{d\overline{N}(t_{(i)})}{\overline{Y}^2(t_{(i)})}. \quad (1.22)$$

### 1.1.6 Estimation de la fonction de survie : estimateurs de Breslow et de Kaplan-Meier

Disposant d'un estimateur du hasard intégré, il est désormais possible d'estimer la fonction de survie. La manière la plus simple de procéder est d'utiliser le théorème de Slutsky<sup>6</sup> et la relation (1.7) pour en déduire :  $\widehat{S}(t) = \exp \left( -\widehat{\Lambda}(t) \right)$ . L'estimateur de la fonction de survie proposé par Breslow (1972) utilise ce principe et a pour expression :

$$\widehat{S}_B(t) = \prod_{(i):t_{(i)} \leq t} \exp \left( -d\widehat{\Lambda}(t_{(i)}) \right). \quad (1.23)$$

Kaplan et Meier (1958) partent de l'idée que les durées supérieures à  $t$  correspondent à des épisodes qui ne se sont pas terminées avant  $t$  et en déduisent :

$$\widehat{S}_{KM}(t) = \prod_{(i):t_{(i)} \leq t} \left( 1 - d\widehat{\Lambda}(t_{(i)}) \right). \quad (1.24)$$

Les deux estimateurs donnent globalement des résultats très proches, mais ils ont tendance à diverger lorsque plusieurs durées se terminent dans le même intervalle de temps. Le problème posé par la présence d'éventuelles durées égales ("tied data" ou encore "ties" en anglais), et des solutions, sont décrits par Therneau et Grambsch (2000, p. 31-37). Un avantage de l'estimateur de Kaplan-Meier est de ne pas être affecté par des observations simultanées dues à une imprécision dans les mesures. En effet, considérons un échantillon de durées non censurées<sup>7</sup> et notons  $t_{(1)}$  la durée la plus courte,  $t_{(2)}$  la seconde

---

<sup>6</sup>On en trouve dans Gouriéroux et Monfort (1990) la formulation suivante : Si  $\widehat{\delta}_n$  est un estimateur faiblement (respectivement fortement) convergent de  $g(\theta)$  et  $h$  est une application continue sur l'image par  $g$  de l'ensemble de définition de  $\theta$ , alors  $h(\widehat{\delta}_n)$  est un estimateur faiblement (respectivement fortement) convergent de  $h[g(\theta)]$ .

<sup>7</sup>Cette restriction n'est faite que pour clarifier l'exposé, le raisonnement pouvant être étendu au cas censuré.

## 1.1 Éléments de base des modèles de durée

---

durée la plus courte et ainsi de suite jusqu'à  $t_{(n)}$ . Soit  $t_{(i)}$  une date à laquelle  $k$  sorties se produisent. On a la relation suivante entre les dates de transition :  $0 < t_{(1)} < \dots < t_{(i-1)} < t_{(i^1)} = \dots = t_{(i^k)} < t_{(i+k+1)} \dots < t_{(n)}$ . D'où l'expression de l'estimateur de Kaplan-Meier évalué entre  $t_{(i-1)}$  et  $t_{(i+1)}$  :

$$\prod_{s=t_{(i-1)}}^{t_{(i+1)}} \left(1 - d\widehat{\Lambda}(t_{(s)})\right) = \left(\frac{n-i-1}{n-i}\right) \left(\frac{n-i-k-1}{n-i-1}\right) \left(\frac{n-i-k-2}{n-i-k-1}\right). \quad (1.25)$$

Admettons que l'on dispose de mesures plus précises et que les données auparavant égales puissent s'écrire  $t_{(i)}^1 < t_{(i)}^2 < \dots < t_{(i)}^{k-1} < t_{(i)}^k$ . L'estimateur s'écrit alors :

$$\prod_{s=t_{(i-1)}}^{t_{(i+1)}} \left(1 - d\widehat{\Lambda}(t_{(s)})\right) = \left(\frac{n-i-1}{n-i}\right) \left(\frac{n-i-2}{n-i-1}\right) \dots \left(\frac{n-i-k}{n-i-k+1}\right) \left(\frac{n-i-k-1}{n-i-k}\right) \left(\frac{n-i-k-2}{n-i-k-1}\right). \quad (1.26)$$

Après simplification, on constate que (1.25) coïncide avec (1.26), et donc que l'estimateur de Kaplan-Meier n'a pas été influencé par les observations simultanées.

Les estimateurs non-paramétriques du hasard intégré nous permettent de déduire l'évolution de la fonction de hasard au fil du temps. De plus, ils peuvent être utilisés pour tester l'égalité des fonctions de survie estimées sur des sous-populations, par exemple au moyen d'un test des log-rangs (de l'anglais "log-rank test"). Cette statistique de test est calculée à partir des différences à chaque date entre le nombre de transitions observées et le nombre de transitions attendues dans une population homogène, et suit une loi du Khi-deux. Cependant, les méthodes non paramétriques sont caractérisées par des vitesses de convergences plus lentes que les méthodes paramétriques, et requièrent des échantillons de grande taille. L'économètre peut donc vouloir dans un premier temps obtenir des informations générales, par exemple sur la forme du hasard, *via* des estimateurs non paramétriques et spécifier ensuite un modèle paramétrique dont le hasard peut se comporter comme les résultats précédents le suggèrent. La manière la plus directe de traiter cette difficulté est de spécifier directement une fonction de hasard disposant de bonnes propriétés.

### 1.1.7 Spécification directe du hasard et modèles de hasards proportionnels

Les théories économiques cherchant à décrire les durées ont naturellement tendance à expliquer le hasard à la date  $t$  en fonction des décisions de l'agent et de son environnement. La théorie nous fournit donc parfois des informations sur la forme de cette fonction (monotone, ou bien croissante puis décroissante) que l'on utilise pour sa construction.

On peut spécifier le hasard afin que les durées suivent une loi connue (loi log-normale, log-logistique, weibull, etc...). D'autres modèles adaptent la transformation de Box-Cox aux durées, supposent des hasards constants par morceaux, log-linéaires ou encore dépendant de fonctions rationnelles du temps.<sup>8</sup>

Nous allons nous limiter ici aux modèles dits de **hasards proportionnels**. Cette famille englobe tous les modèles tels que, dans le cadre de régresseurs constants dans le temps, la fonction de hasard puisse s'écrire :

$$\lambda(t|X) = \lambda_0(t)\lambda_1(X), \quad (1.27)$$

où  $X$  dénote le vecteur des variables explicatives. Les fonctions  $\lambda_0$  et  $\lambda_1$  sont communes à tous les individus, et le rapport des hasards de deux individus caractérisés par des vecteurs  $X_1$  et  $X_2$  à une date  $t$  vaut  $\lambda_1(X_1)/\lambda_1(X_2)$  et est constant dans le temps. La fonction  $\lambda_0$  est appelée le **hasard de base**, et le modèle de hasard proportionnel est souvent désigné par l'abréviation PH (du fait de la littérature anglophone). L'hypothèse de hasards proportionnels permet une estimation semi-paramétrique assez simple, comme le montre Cox (1975). Nous reviendrons ultérieurement sur cette méthode qui ne nécessite pas la spécification d'une forme fonctionnelle pour le hasard de base. Par abus de langage, les modèles avec des régresseurs variant dans le temps définis par  $\lambda(t|X(t)) = \lambda_0(t)\lambda_1(X(t))$ , bien que le rapport des hasards ne soit plus constant, continuent d'être appelés des modèles PH. L'usage s'est établi de qualifier par "hasards proportionnels" des modèles où le hasard de base, que l'on laisse généralement non spécifié car une inférence semi-paramétrique peut facilement être mise en place dans ces modèles, est commun à tous les individus.

Une méthode alternative pour l'étude des durées est l'approche dite de **régression**. Elle diffère entre autres points de la méthode que nous venons de présenter par l'importance attribuée à la fonction de hasard, qui n'est plus l'élément central du modèle.

---

<sup>8</sup>Lancaster (1990, p. 40-50) présente un exposé détaillé de ces différents modèles.

### 1.1.8 Approche de régression

L'approche de régression, familière à l'économètre, a été principalement utilisée aux débuts de l'économétrie des durées. Bien que n'étant plus très répandue de nos jours, elle a donné lieu à des alternatives intéressantes aux modèles de hasard.

La fonction de survie peut être formulée de la façon suivante :

$$S(t) = P(T \geq t) = P\left(\Lambda(T) \geq \int_0^t \lambda(u) du\right). \quad (1.28)$$

En utilisant la relation (1.7), on obtient :

$$P\left(\Lambda(T) \geq \int_0^t \lambda(u) du\right) = \exp\left(-\int_0^t \lambda(u) du\right), \quad (1.29)$$

et on en déduit que le hasard intégré suit une loi exponentielle unitaire.<sup>9</sup>

En faisant un changement de variable, on vérifie qu'il est équivalent de dire que la variable aléatoire  $\epsilon = \ln\left(\int_0^T \lambda(u) du\right)$  est distribuée selon la loi "valeur extrême" de type 1 (EV1).<sup>10</sup> Rappelons que si on pose  $X = g(Z)$ , où  $Z$  est une variable aléatoire de densité  $f_Z$  et  $g$  est une application différentiable bijective, alors la densité  $f_X$  de  $X$  a pour expression :

$$f_X(x) = \frac{f_Z(g^{-1}(x))}{|g'[g^{-1}(x)]|}. \quad (1.30)$$

Soit la variable aléatoire  $Z = g^{-1}(X) = \exp(X) = \int_0^T \lambda(u) du$ , qui suit une loi exponentielle unitaire. Elle a donc pour densité  $f_Z(z) = \exp(-\int_0^z \lambda(u) du)$ . La densité de  $X$  obtenue par le lemme du changement de variable est :

$$\begin{aligned} f_X(x) &= \exp(-\exp(x)) \exp(x) \\ \Leftrightarrow f_X(x) &= \exp(\epsilon - \exp(\epsilon)), \end{aligned} \quad (1.31)$$

où  $\epsilon = x = \ln\left(\int_0^T \lambda(u) du\right)$ , qui est bien la densité d'une loi EV1.

L'approche par les modèles de régression, telle qu'elle est décrite par Lancaster (1990, p. 35-40), suppose que le hasard intégré est de la forme  $\Lambda(t, X) = \lambda_1(X)t$  avec  $\lambda_1(X)$  la partie du hasard qui dépend du vecteur de variables explicatives, noté  $X$ , et éventuellement d'une constante. On construit la régression linéaire :

$$\ln t = -\ln \lambda_1(X) + \epsilon, \quad (1.32)$$

---

<sup>9</sup>En effet, la loi exponentielle de paramètre  $\theta$  a comme densité  $f(t) = \theta \exp(-\theta t)$  et comme fonction de survie  $S(t) = \exp(-\theta t)$ . On a ainsi  $E(T) = 1/\theta$  et  $V(T) = 1/\theta^2$ .

<sup>10</sup>La loi EV1 n'a aucun paramètre et admet comme densité  $f(\epsilon) = \exp(\epsilon - \exp(\epsilon))$  et comme fonction de survie  $S(\epsilon) = \exp(-\exp(\epsilon))$ . On a alors  $E(\epsilon) = -0.5722$  et  $V(\epsilon) = \pi^2/6$ .

où on suppose que  $\epsilon$  est distribué suivant une loi adéquate. En effet, l'approche par les modèles de régression repose principalement sur des généralisations de la loi EV1 de  $\epsilon$ , comme par exemple la loi exponentielle ou la loi gamma. Cette méthode a cependant rapidement montré ses limites, notamment avec la forme très particulière du hasard intégré et l'impossibilité de prendre en compte des régresseurs variant dans le temps. L'idée des modèles de régression a ensuite été utilisée pour créer les modèles AFT (de l'anglais "Accelerated Failure Time") et le modèle de transformation de Horowitz (1996).

Comme le souligne Van den Berg (2001), il est important de noter que les approches de régression et de spécification directe du hasard ne tirent pas leur souplesse de la même source. En effet, il est possible d'écrire l'équation (1.27) de la manière suivante :

$$\ln \left( \int_0^t \lambda_0(u) du \right) = -\ln \lambda_1(X) + \epsilon, \quad (1.33)$$

avec  $\epsilon$  distribué selon une loi EV1. Si on compare avec la régression linéaire (1.32), on s'aperçoit que la flexibilité de cette dernière vient de la spécification de la loi de  $\epsilon$ . L'approche *via* la spécification directe du hasard implique quant à elle une transformation déterministe du temps, inconnue lorsqu'on opte pour une approche semi-paramétrique laissant libre la forme fonctionnelle de  $\lambda_0(t)$ , et tire donc sa souplesse du hasard de base. Ainsi, le modèle PH et le modèle de régression ne sont pas emboîtés, et leurs flexibilités respectives n'ont pas la même origine.

Dans la prochaine section, nous allons revenir à l'approche reposant sur la spécification directe du hasard et présenter le modèle de Cox qui est un cas important de modèle PH. Il est utilisé dans de nombreuses études, vraisemblablement du fait de sa simplicité et de son aptitude à capter de nombreuses propriétés des données.

## 1.2. Le modèle de Cox

Ce modèle a été développé pour la première fois par Cox (1972) et a connu par la suite un succès certain. Van den Berg (2001) compare sa popularité dans l'étude des durées à celle du modèle de régression linéaire dans l'approche de régression. Sa simplicité et la facilité avec laquelle il peut être étendu pour tenir compte de phénomènes complexes comme des régresseurs variant dans le temps ou bien de l'hétérogénéité non observée sont à l'origine de cet engouement.

Dans une première sous-section, nous présentons les notations utilisées ainsi que le modèle de Cox. Ensuite, nous indiquons de manière intuitive com-

ment prendre en compte les régresseurs variant dans le temps et quelles sont les hypothèses nécessaires à cela. Puis, nous discutons les différentes méthodes d'inférence, en commençant par l'inférence paramétrique avec l'estimation par maximum de vraisemblance puis l'inférence semi-paramétrique avec la vraisemblance partielle. Ces deux approches nécessitant le recours à des procédures numériques, nous présenterons rapidement l'algorithme EM qui a été fréquemment utilisé dans ce contexte, puis l'estimateur non-paramétrique de Breslow du hasard de base intégré. Nous mentionnerons également comment la vraisemblance partielle doit être adaptée pour tenir compte des observations simultanées. Une des hypothèses sous-tendant le modèle de Cox est l'hypothèse de proportionnalité, et nous allons voir comment il est possible de la tester. Enfin, nous présenterons les alternatives envisageables que sont les modèles additifs et AFT en cas de rejet de cette hypothèse.

### 1.2.1 Présentation du modèle

Soit  $X$  la matrice des régresseurs, de plein rang colonne. On note  $X_i$  la  $i$ -ème ligne de  $X$ , c'est-à-dire le vecteur ligne des valeurs prises par les variables explicatives pour l'unité  $i$ . La valeur de la  $j$ -ème ( $j = 1 \dots p$ ) variable explicative de l'unité  $i$  ( $i = 1 \dots n$ ) est notée  $X_{ij}$ .

Le hasard du modèle de Cox pour l'unité  $i$  s'écrit :

$$\lambda_i(t) = \lambda_0(t) \exp(X_i \beta), \quad (1.34)$$

où  $\lambda_0$  est une fonction positive commune à tous les unités, ne dépendant que du temps, et  $\beta$  est un vecteur  $p \times 1$  de coefficients. L'indiciage par  $i$  est ici facultatif et ne sert qu'à faciliter la comparaison avec les modèles plus complexes qui sont présentés ultérieurement. Le processus  $N$  a pour intensité :

$$\lambda_i(t) = Y_i(t) \lambda_0(t) \exp(X_i \beta), \quad (1.35)$$

cette dernière notation rappelant explicitement que, dans le cadre simple où les individus ne sont observés que pour un unique épisode, la probabilité d'effectuer une seconde transition est nulle, ainsi que la probabilité d'observer une transition dès lors que l'individu n'est plus observé. Ce modèle a été introduit par Aalen (1978), et est souvent utilisé dans la littérature sous le nom du modèle à intensités multiplicatives (abrégé MI) de Aalen.

Comme dans tout modèle PH, la fonction  $\lambda_0$  est appelée le hasard de base. La fonction exponentielle présente le double avantage d'être toujours positive<sup>11</sup> et de simplifier les calculs de log-vraisemblance. En contrepartie,

---

<sup>11</sup>Le hasard étant une probabilité limite, il ne peut pas être négatif.



chaque élément de  $X_i$  agit de façon multiplicative sur le hasard, point auquel il faut être attentif lors du choix des régresseurs. L'ajout d'un régresseur a pour seule conséquence de déplacer la fonction de hasard vers le haut ou le bas, sans en affecter la pente ou le mode.<sup>12</sup>

### 1.2.2 Prise en compte de régresseurs variant dans le temps

Dans de nombreuses applications, il n'est pas raisonnable de considérer les variables explicatives comme constantes dans le temps. On a parfois de bonnes raisons de penser que la fonction de hasard ne dépend pas que des valeurs prises par les variables explicatives au début de l'épisode, mais qu'il est également influencé leurs variations tout au long de l'épisode. Par exemple, il est vraisemblable que la diminution des allocations chômage a un impact sur le temps passé sans emploi, ou encore que la probabilité pour un réfrigérateur de tomber en panne évolue avec la température ambiante qui varie dans le temps. Nous allons voir comment traiter l'incorporation de régresseurs variant dans le temps dans le cadre du modèle de Cox, en suivant les développements de Fleming et Harrington (1991), Ridder et Tunalı (1999) et Van den Berg (2001). Ce problème est délicat et nécessite l'utilisation de la théorie de la mesure. La présentation que nous en ferons ici est informelle et vise principalement à donner l'intuition des résultats. Nous présentons ici le raisonnement dans le cadre d'un modèle PH, le cas particulier du modèle de Cox étant implicitement aussi traité.

Soit un processus aléatoire  $X = \{X(s) : s > 0\}$  dont les réalisations sont les valeurs prises par les variables explicatives à chaque instant. Notons  $X(t)$  l'ensemble des valeurs prises par les variables explicatives à la date  $t$ . Il peut évidemment contenir des constantes et des régresseurs dont l'évolution est déterministe. Les travaux s'intéressant à cette question supposent que  $X$  est un processus prédictible.<sup>13</sup> C'est-à-dire que la valeur  $X(t)$  dépend uniquement des événements observés qui se sont produits juste avant  $t$ . Ou encore, la réalisation  $X(t)$  n'est pas influencée par une éventuelle transition en  $t$ . Comme le font remarquer Lancaster (1990, p. 28) et Ridder et Tunalı (1999), la notion de prédictibilité pour les variables explicatives ressemble au concept d'exogénéité faible dans l'analyse des séries temporelles. Nous avons besoin d'une seconde hypothèse, à savoir que  $X(t)$  et  $\lambda_1(X(t))$  sont tous deux bornés. Cette hypothèse technique est expliquée par la théorie de la mesure

---

<sup>12</sup>Ce point est présenté plus en détail dans Kiefer (1988).

<sup>13</sup>Cette notion a déjà été présentée dans la section 1.1.2, au sujet du processus des durées lui-même. Il est ici abordé dans le cas des variables explicatives.

dans Fleming et Harrington (1991).

Il reste maintenant à faire le lien entre ces deux hypothèses et le modèle PH. Considérons le processus aléatoire  $P[T \leq t | \{X(s)\}_0^t]$ , qui est conditionnel à toutes les valeurs prises par  $X$  depuis le début de l'enquête. Pour pouvoir représenter le processus de comptage associé à  $T$  sous la forme d'un modèle de hasard de la famille PH, nous devons supposer que ce processus est continu. En plus de la prédictibilité, une condition suffisante à cela est que la durée n'ait pas, conditionnellement aux valeurs prises par  $X$  entre 0 et  $t$ , une probabilité strictement positive d'échoir à la date  $t$ . Autrement dit, une condition suffisante pour que le processus soit continu est qu'un individu n'ait pas une probabilité strictement positive d'effectuer une transition sur un intervalle de temps suffisamment petit, en plus de la prédictibilité de  $X$ . Avec ces trois hypothèses, il est possible de représenter le processus de comptage associé à  $T$  sous la forme d'un modèle PH. Réciproquement, un modèle PH peut être généré par un processus de comptage sous ces hypothèses.

En fondant une analyse sur un modèle de la famille PH, ce résultat nous indique qu'on peut alors faire des inférences en présence de régresseurs variant dans le temps au moyen des outils économétriques traditionnels que nous allons maintenant décrire.

### 1.2.3 Estimateur du Maximum de Vraisemblance

L'estimateur du maximum de vraisemblance (abrégé EMV) présente sous certaines conditions de régularité (voir Gouriéroux et Monfort, 1990, vol. 1, p. 173-174, ou encore Mittelhammer *et al.*, 2000, p. 140) des propriétés intéressantes comme la convergence forte, l'invariance par rapport à une reparamétrisation, l'efficacité asymptotique et la normalité asymptotique. Nous nous restreignons ici au cas simple où l'individu n'expérimente qu'un seul épisode et où il ne peut quitter son état que vers une unique destination.

Procédons par étapes. Avant le tirage de l'échantillon, on spécifie un modèle de hasard pour la variable aléatoire  $T$ , chaque réalisation  $t_i$  de la variable  $T_i$  dépendant des caractéristiques de l'individu  $i$  et d'un vecteur de paramètres  $\theta$ . On en déduit la densité jointe des variables aléatoires  $T_1, \dots, T_n$ , qui est conditionnelle à  $\theta$ . Les valeurs des paramètres étant inconnues, nous ne pouvons pas évaluer la probabilité de réalisation d'un échantillon quelconque. Par contre, une fois que l'on dispose d'un échantillon, la densité jointe des variables aléatoires peut se reformuler comme une fonction de  $\theta$ , conditionnelle aux observations et aux variables explicatives que l'on nomme "vraisemblance". Il est possible de déduire les valeurs des paramètres qui ont rendu maximale cette vraisemblance, que l'on appellera les estimations par maximum de vraisemblance.

A chaque date  $t$ , l'individu  $i$  ( $i = 1 \dots n$ ) peut soit être observé et avoir la possibilité d'effectuer une transition en  $t$ , soit ne plus être observé. Dans le premier cas, l'individu  $i$  contribue à la vraisemblance à hauteur de sa "probabilité" de changer d'état en  $t$ , c'est-à-dire  $\lambda(t)S(t)$ . Dans le second cas, son épisode est censuré et on sait juste qu'il n'a pas pris fin avant  $t$ . On a donc  $t_i > t$  et sa contribution en  $t$  est de  $S(t)$ . Remarquons que les deux cas ont en commun la fonction de survie  $S(t)$  car l'individu est observé au moins jusqu'en  $t$ . D'où l'expression de la contribution individuelle à la vraisemblance :

$$l_i(\theta) = \lambda_i(t_i)^{\delta_i} S(t_i) = \lambda_i(t_i)^{\delta_i} \exp\left(-\int_0^{t_i} \lambda_i(u) du\right), \quad (1.36)$$

où  $\delta_i$  est une indicatrice de non-censure, c'est-à-dire prenant la valeur 0 si l'observation est censurée et 1 sinon. La log-vraisemblance s'écrit :

$$\begin{aligned} \ln L(\theta) &= \sum_{i=1}^n \ln l_i(\theta) \\ &= \sum_{i=1}^n \left( \delta_i \ln [\lambda_i(t_i)] - \int_0^{t_i} \lambda_i(u) du \right). \end{aligned} \quad (1.37)$$

Si on considère le cas particulier du modèle de Cox en laissant le hasard de base sans spécification, le vecteur de paramètres  $\theta$  ne contient que  $\beta$  et la log-vraisemblance devient :

$$\ln L(\beta) = \sum_{i=1}^n \left( \delta_i [\ln \lambda_0(t_i) + X_i \beta] - \exp(X_i \beta) \int_0^{t_i} \lambda_0(u) du \right). \quad (1.38)$$

On a donc :

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^n X_i' \left( \delta_i - \exp(X_i \beta) \int_0^{t_i} \lambda_0(u) du \right) = 0, \quad (1.39)$$

$$\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = \sum_{i=1}^n X_i' X_i \left( -\exp(X_i \beta) \int_0^{t_i} \lambda_0(u) du \right). \quad (1.40)$$

La vraisemblance est concave car la matrice dans le membre de droite de l'équation (1.40) est semi-définie négative. L'EMV  $\hat{\beta}_{MV}$  du paramètre  $\beta$  est donc l'unique solution des équations de vraisemblance. Comme (1.39) n'admet pas de solution analytique en  $\beta$ , l'EMV doit être approché par des procédures numériques, par exemple au moyen d'un algorithme de Newton-Raphson.

Considérer des régresseurs variant dans le temps ne change rien aux calculs ci-dessus, hormis que la partie du hasard dépendant des variables explicatives ne peut plus être sortie de l'intégrale. Par ailleurs, il faut spécifier une forme fonctionnelle pour le hasard de base  $\lambda_0(t)$  afin de pouvoir l'estimer dans le cadre de cette approche totalement paramétrique. Une alternative est l'utilisation d'une approche semi-paramétrique ne demandant pas la spécification d'une forme fonctionnelle pour le hasard de base.

### 1.2.4 La Vraisemblance Partielle

Cox (1972) propose déjà, en même temps que le modèle qui prendra son nom, de baser l'inférence sur une fonction de vraisemblance ayant une forme différente de l'équation (1.37). Dans un travail ultérieur (Cox, 1975), il renforce les fondements de sa proposition à partir des statistiques de rang et d'ordre et fournit la méthode dite de la **vraisemblance partielle**.

#### Statistiques de rang et d'ordre

Comme dans le cas de la vraisemblance complète, considérons un échantillon où chaque individu  $i$  ( $i = 1 \dots n$ ) ne connaît qu'un seul épisode et où il ne peut quitter son état que vers une destination unique. La vraisemblance totale est basée sur l'information contenue dans le vecteur des durées. L'idée de départ de la vraisemblance partielle est de reconstituer cette information au moyen des statistiques de **rang** et **d'ordre**. On note  $s = (t_{(1)}, \dots, t_{(n)})$  la statistique d'ordre, qui n'est autre que le vecteur des durées classées par ordre croissant.<sup>14</sup> Soit  $r = (r_1, \dots, r_n)$  la statistique de rang, c'est-à-dire le vecteur des indices  $i$  classés suivant l'ordre de sortie des individus. Par exemple, soit le vecteur de durées  $t = (30, 17, 22)$ . La statistique d'ordre est  $s = (17, 22, 30) = (t_2, t_3, t_1) = (t_{(1)}, t_{(2)}, t_{(3)})$  et la statistique de rang  $r = (2, 3, 1) = (r_1, r_2, r_3)$ . Il convient ici d'être attentif à l'indigage et de remarquer que  $t_{(i)} = t_{r_i}$ ,  $i$  étant l'indice de l'observation dans l'échantillon initial et  $(i)$  l'indice dans l'échantillon ordonné.

Comme le couple  $(s, r)$  est en bijection avec  $t$ , il équivaut de considérer la densité jointe exprimée en fonction de  $t$  ou bien de  $(s, r)$  :

$$L = f(t_{(1)}, r_1, t_{(2)}, r_2 \dots, t_{(n)}, r_n). \quad (1.41)$$

---

<sup>14</sup>La notation  $t_{(i)}$  a déjà été introduite dans la section 1.1.3, les durées étant indicées par  $(i)$  dans l'échantillon ordonné.

Par conditionnements successifs, elle peut-être réécrite :

$$\begin{aligned} L &= f(t_{(1)}, r_1) f(t_{(2)}, r_2 | t_{(1)}, r_1) f(t_{(3)}, r_3 | t_{(1)}, r_1, t_{(2)}, r_2) \dots \\ &\quad \dots f(t_{(n)}, r_n | t_{(1)}, r_1, \dots, t_{(n-1)}, r_{n-1}) \\ &= \prod_{k=1}^n f(t_{(k)}, r_k | s^{(k-1)}, r^{k-1}), \end{aligned} \quad (1.42)$$

où  $s^{(k)} = (t_{(1)}, \dots, t_{(k)})$ ,  $r^k = (r_1, \dots, r_k)$  et le couple  $(x^{(0)}, s^{(0)})$  est connu. Chaque terme du produit peut se décomposer comme :

$$f(t_{(k)}, r_k, | t^{(k-1)}, r^{k-1}) = f(t_{(k)} | t^{(k-1)}, r^{k-1}) f(r_k | t^{(k)}, r^{k-1}). \quad (1.43)$$

La densité jointe peut donc être décomposée de la manière suivante :

$$L = \prod_{k=1}^n f(r_k | t^{(k)}, r^{k-1}) \prod_{k=1}^n f(t_{(k)} | t^{(k-1)}, r^{k-1}). \quad (1.44)$$

L'idée sous-jacente à la vraisemblance partielle est qu'il est plus facile de prédire l'ordre dans lequel les individus vont quitter leur état, plutôt que leurs dates de transition. On ignore donc l'information contenu dans les dates de sorties et on se base uniquement sur les rangs de sortie. La vraisemblance partielle est donc le premier terme de (1.44) :

$$L_p = \prod_{k=1}^n f(r_k | t^{(k)}, r^{k-1}). \quad (1.45)$$

Elle tire d'ailleurs son nom du fait qu'elle ne correspond qu'à une partie de la vraisemblance. Nous allons maintenant présenter la formulation de la vraisemblance partielle pour un modèle de hasards proportionnels.

### Expression de la vraisemblance partielle

La contribution à la vraisemblance partielle de la première observation est  $L_p^1 = f(r_1 | t_{(1)}) = f(r_1, t_{(1)}) / f(t_{(1)})$ . L'expression de  $L_p^1$  est donc le rapport entre la probabilité instantanée que l'individu 1 effectue une sortie en  $t_1$  et la probabilité instantanée qu'une transition ait lieu en  $t_1$ . La première probabilité instantanée n'est autre que  $\lambda_i(t)$ , la seconde est la somme sur les individus au risque des hasards en  $t_1$  (on suppose qu'il n'y a pas d'observations simultanées).<sup>15</sup> Le produit étant commutatif, il n'est pas nécessaire

---

<sup>15</sup>L'ensemble  $R(t)$  contient les indices des individus susceptible d'effectuer une transition en  $t$ . La définition de l'ensemble au risque  $R(t)$  a été donnée p. 25.

## 1.2 Le modèle de Cox

---

de considérer un échantillon ordonné. La vraisemblance partielle peut alors s'écrire :<sup>16</sup>

$$L_p(\beta) = \prod_{i=1}^n \frac{\lambda(t_i)}{\sum_{k \in R(t_i)} \lambda(t_k)}. \quad (1.46)$$

Dans le cadre du modèle de Cox, le hasard de base est commun à tous les individus et se simplifie :<sup>17</sup>

$$L_p(\beta) = \prod_{i=1}^n \frac{\exp(X_i\beta)}{\sum_{k \in R(t_i)} \exp(X_k\beta)}. \quad (1.47)$$

En maximisant (1.47) par rapport au paramètre  $\beta$ , on obtient l'estimateur du maximum de vraisemblance partielle (abrégé PL, de l'anglais Partial Likelihood)  $\hat{\beta}_{PL}$ .

La méthode peut aisément être étendue au cas de durées censurées à droite avec censure non informative. Les épisodes censurés n'interviennent pas dans le numérateur, car ils ne correspondent pas à une transition observée. Par contre, ils font partie de l'ensemble au risque jusqu'à leur date de censure, car ils sont susceptibles d'effectuer une transition jusque là. Ils figurent donc au dénominateur de la vraisemblance partielle :

$$L_p(\beta) = \prod_{i=1}^n \left[ \frac{\exp(X_i\beta)}{\sum_{k \in R(t_i)} \exp(X_k\beta)} \right]^{\delta_i}. \quad (1.48)$$

Alternativement, on peut reformuler cette équation avec les notations des processus de comptage. Pour simplifier, considérons l'échantillon ordonné  $t_{(1)} \dots t_{(n)}$  :

$$L_p(\beta) = \prod_{i=1}^n \prod_{(k)=1}^n \left[ \frac{Y_i(t_{(k)}) \exp(X_i\beta)}{\sum_{j=1}^n Y_j(t_{(k)}) \exp(X_j\beta)} \right]^{dN_i(t_{(k)})}. \quad (1.49)$$

Ici aussi, l'échantillon n'a pas besoin d'être ordonné. L'important est de considérer une "grille" temporelle, commune à tous les individus, suffisamment fine pour que les processus  $Y_i(t)$  et  $dN_i(t)$  prennent au moins une fois les valeurs 0 et 1 pour chaque individu.<sup>18</sup> Considérer l'échantillon ordonné n'est qu'une

<sup>16</sup>Lorsqu'il n'y a pas de censure, l'expression de la vraisemblance partielle est identique à la vraisemblance d'un Logit multivarié. Cette relation est présentée à plusieurs reprises dans la littérature, notamment dans So et Kuhfeld (1995, p. 412)

<sup>17</sup>La constante des variables explicatives n'est pas identifiée séparément du hasard de base.

<sup>18</sup>En présence de données censurées,  $dN_i$  peut être toujours nul.

manière d'obtenir la grille contenant juste assez de points pour ne pas perdre d'information. La prise en compte de régresseurs variant dans le temps se fait en gardant en mémoire les hypothèses détaillées dans la sous-section 1.2.2. Comme nous utilisons toutes les valeurs passées des variables explicatives, la vraisemblance partielle doit être réécrite pour tenir compte de l'histoire de  $X_i$ . Supposons que la durée de l'enquête soit de  $m$  périodes, les observations sont  $t_{ij}$  ( $j = 1 \dots m$ ) et la vraisemblance partielle devient :

$$L_p(\beta) = \prod_{i=1}^n \prod_{j=1}^m \left[ \frac{\exp(X_i(t_{ij})\beta)}{\sum_{k \in R(t_{ij})} \exp(X_k(t_{ij})\beta)} \right]^{\delta_{ij}}, \quad (1.50)$$

où  $\delta_{ij}$  est une indicatrice prenant la valeur 0 si l'observation est censurée et 1 sinon.

On note  $t_k$  ( $k = 1, \dots, K_j$ ) les observations jusqu'à la période  $j$ . La reformulation en termes de processus de comptage est :

$$L_p(\beta) = \prod_{i=1}^n \prod_{j=1}^m \prod_{k=1}^{K_j} \left[ \frac{Y_{ij}(t_k) \exp[X_{ij}(t_k)\beta]}{\sum_{l=1}^{K_j} Y_l(t_k) \exp[X_l(t_k)\beta]} \right]^{dN_{ij}(t_k)}. \quad (1.51)$$

Comme dans le cas de régresseurs indépendants du temps, il n'est pas nécessaire que les échantillons soient ordonnés sur chaque période.

Bien que n'étant pas une vraisemblance à part entière, la vraisemblance partielle peut être traitée comme telle dans le cadre de l'inférence. Breslow (1972) et Johansen (1983) montrent qu'il s'agit d'une vraisemblance  $f(\beta, \Lambda)$  concentrée par rapport à  $\Lambda$ , ce dernier étant estimé par Breslow. Dans la prochaine sous-section, nous présentons les équations de vraisemblance partielle.

### Equations de vraisemblance partielle

La log-vraisemblance partielle obtenue à partir de (1.48) s'écrit :

$$\ln L_p(\beta) = \sum_{i=1}^n \delta_i \left[ X_i \beta - \ln \left( \sum_{k \in R(t_i)} \exp(X_k \beta) \right) \right]. \quad (1.52)$$

On obtient :

$$\frac{\partial \ln L_p}{\partial \beta} = \sum_{i=1}^n \left[ \delta_i X_i' - \frac{\sum_{k \in R(t_i)} X_k' \exp(X_k \beta)}{\sum_{k \in R(t_i)} \exp(X_k \beta)} \right] = 0 \quad (1.53)$$

$$\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = - \sum_{i=1}^n \frac{\sum_{k \in R(t_i)} \exp(X_k \beta) Z_k' Z_k}{\sum_{k \in R(t_i)} \exp(X_k \beta)}, \quad (1.54)$$

où

$$Z_k = X'_k - \frac{\sum_{k \in R(t_i)} X'_k \exp(X_k \beta)}{\sum_{k \in R(t_i)} \exp(X_k \beta)}. \quad (1.55)$$

Le second terme de  $Z_k$  est une moyenne pondérée des variables explicatives  $X_k$  des individus au risque. Le Hessien est semi-défini négatif, et la log-vraisemblance partielle est bien concave. L'estimateur  $\widehat{\beta}_{PL}$  est donc l'unique solution des équations de vraisemblance partielle et ses propriétés asymptotiques ont été établies par Andersen et Gill (1982) aux moyen des martingales. Le hasard de base doit être estimé séparément dans cette approche. Si on suppose  $\lambda_0$  constant par morceaux sur des intervalles de temps infiniment petits, on obtient l'estimateur de Breslow.

Comme dans le cas de la vraisemblance complète, l'estimateur  $\widehat{\beta}_{PL}$  doit être approché au moyen de procédures numériques. Il y a plusieurs manière de procéder, certains logiciels estiment par exemple automatiquement la vraisemblance partielle avec l'algorithme de Newton-Raphson. Une autre possibilité est l'algorithme Espérance-Maximisation (en abrégé algorithme EM) que nous allons présenter. Son avantage est d'être relativement simple et surtout d'avoir été étendu à des modèles incluant de l'hétérogénéité non observée, que nous verrons par la suite.

### 1.2.5 L'algorithme EM

Nous commençons par présenter l'algorithme avant de passer à une application à la vraisemblance partielle dans le cadre du modèle de Cox. Des exemples plus détaillés sont fournis dans Dempster *et al.* (1977), qui proposent la première formulation générale de l'algorithme, et Gouriéroux et Monfort (1990, p. 424-428).

On considère un vecteur de variables latentes  $T^* = (T_1^*, \dots, T_n^*)'$  de densité  $l^*(t^*|X; \theta)$ . Ces variables latentes ne sont pas observées directement, on ne dispose que du vecteur de variables observables  $T$  définies comme  $T \equiv h(T^*)$  où  $h$  est une fonction connue. L'inobservabilité des variables latentes peut par exemple venir d'un phénomène de censure aléatoire. On définit la fonction :

$$Q(\theta, \widetilde{\theta}) = E_{\widetilde{\theta}}[\ln l^*(T^*|X; \theta)|T = t], \quad (1.56)$$

où  $\widetilde{\theta}$  correspond à une valeur quelconque du paramètre  $\theta$ .

Lors de la  $(q + 1)$ -ème itération, l'étape E de l'algorithme consiste à calculer la fonction  $Q(\theta, \theta^{(q)})$ . L'étape M correspond à la recherche de la valeur  $\theta^{(q+1)}$  maximisant cette fonction par rapport à  $\theta$ . Cet algorithme n'est intéressant que lorsque l'optimisation de  $Q(\theta, \theta^{(q)})$  est plus facile que celle de  $\ln l(T|X; \theta)$ .



Le score de la log-vraisemblance associée aux variables latentes est le membre de gauche de l'équation (1.39). On cherche à calculer son espérance par rapport à la loi des durées latentes, sachant l'information véhiculée par les transitions observées et que la variable latente est le hasard de base. D'où :

$$\frac{\partial Q(\beta, \beta^{(q)})}{\partial \beta} = \sum_{i=1}^n X_i' \left( \delta_i - \exp(X_i \beta) E_{\beta^{(q)}} \left[ \int_0^{t_i} \lambda_0(u) du \middle| T = t \right] \right). \quad (1.57)$$

Remarquons que, dans ce cas simple, le vecteur des paramètres inconnus ne contient que  $\beta$ . Pour continuer, nous avons besoin d'une expression de l'espérance conditionnelle du hasard intégré en fonction des observations. Lancaster (1990, p. 256) montre que :

$$E_{\beta^{(q)}} \left[ \int_0^{t_i} \lambda_0(u) du \middle| T = t \right] = \frac{1 - \delta_i}{\exp(X_i \beta^{(q)})} + \sum_{j:t_j \leq t_i} \left( \sum_{k \in R(t_i)} \exp(X_k \beta^{(q)}) \right). \quad (1.58)$$

Dans le cas non-censuré,  $\delta_i = 1$  pour toutes les observations et l'espérance conditionnelle du hasard intégré est égale au second terme de la somme.

L'étape E se déroule comme suit : on substitue (1.58) dans (1.57). Comme l'espérance du score est nulle, l'étape E nous donne :

$$\sum_{i=1}^n X_i' \left[ 2\delta_i - 1 - \exp(X_i \beta) \sum_{j:t_j \leq t_i} \left( \sum_{k \in R(t_i)} \exp(X_k \beta^{(q)}) \right) \right] = 0. \quad (1.59)$$

Il reste à résoudre cette équation en  $\beta$  au moyen d'une procédure numérique, et à itérer. La procédure a l'avantage d'être relativement simple ici, et nous la reverrons ultérieurement pour traiter des expressions plus complexes de la vraisemblance partielle, obtenues lorsque l'on prend en compte l'hétérogénéité non observée.

La vraisemblance partielle est une approche semi-paramétrique ne fournissant un estimateur que pour le vecteur de coefficients  $\beta$ , le hasard de base devant quant à lui être estimé séparément. Une première possibilité est d'utiliser l'estimateur non-paramétrique de Nelson-Aalen présenté en 1.1.3. Une alternative est l'estimateur du hasard de base de Breslow (1974), calculé à partir des martingales dont nous présentons ici quelques notions.

### 1.2.6 Éléments sur les martingales

Une martingale  $M(t)$  est un processus aléatoire défini conditionnellement à une filtration  $F_t$  et qui présente la caractéristique :

$$E[dM(t)|F_{t-}] = 0, \forall t > 0, \quad (1.60)$$

où  $dM(t)$  est l'incrément de  $M(t)$  sur  $[t, t + dt[$  et  $F_{t^-}$  désigne la filtration en  $t^-$ , c'est-à-dire l'information accumulée sur  $[0, t[$ . Certaines conditions supplémentaires sont requises, comme la continuité à droite et l'existence d'une limite finie lorsque  $t \rightarrow \infty$ . Une conséquence de la relation (1.60) est que pour tout  $0 \leq s < t$ ,  $E[M(t)|F_s] = M(s)$ , ou encore en temps discret  $E[M(t+1)|M(t), \dots, M(0)] = M(t)$ . La meilleure prédiction conditionnellement au passé d'une valeur future quelconque est donc la valeur présente de la martingale. A partir de cette dernière propriété et de l'équation (1.60), on déduit que toute martingale a une moyenne constante, et qu'il s'agit d'un processus sans dérive. Un exemple simple de martingale est une marche aléatoire définie à partir d'un bruit blanc symétrique. Un autre exemple est celui d'une marche aléatoire caractérisée par un déplacement d'une unité à droite avec probabilité de 80% et de 4 unités à gauche avec probabilité 20%. Le nom de "martingale" semble lié aux jeux de hasards.

Un élément important de la théorie des martingales est le théorème de Doob-Meyer, dont nous donnons ici une version informelle. Tout processus de comptage peut être décomposé de manière unique comme la somme d'une martingale et d'un processus appelé le compensateur. Le compensateur est un processus prédictible, continu à droite et valant 0 à la date  $t = 0$ . Ce théorème repose sur des hypothèses qui sont vérifiées dès lors que l'on considère des processus de comptage. Ainsi :

$$N_i(t) = M_i(t) + \int_0^t Y_i(s)\lambda_i(s)ds, \quad (1.61)$$

où le second terme présente bien les trois caractéristiques d'un compensateur. D'une part l'intégrande est le produit de deux processus prédictibles ; d'autre part les deux processus sont continus à gauche mais le second terme est continu à gauche et à droite par les propriétés de l'intégrale. Enfin, l'intégrale nous assure également que la quantité  $\int_0^t Y_i(s)\lambda_i(s)ds$  est nulle lorsque  $t = 0$ . A partir de ces quelques notions de martingales, nous avons la possibilité de justifier d'une autre manière l'estimateur de Breslow.

### 1.2.7 Estimateur de Breslow du hasard intégré

L'estimateur de Breslow peut être vu comme une extension au cas paramétrique de l'estimateur de Nelson-Aalen du haard intégré et nous en présentons ici une justification utilisant les martingales. Réécrivons la relation (1.61) avec le hasard d'un modèle de Cox :

$$M_i(t) = N_i(t) - \int_0^t Y_i(s) \exp[X_i(s)\beta] \lambda_0(s) ds. \quad (1.62)$$

On peut calculer les valeurs estimées de la martingale comme des résidus, en faisant la différence entre les incréments observés du processus de comptage et le compensateur estimé :

$$\widehat{M}_i(t) = N_i(t) - \int_0^t Y_i(s) \exp[X_i(s)\widehat{\beta}] d\widehat{\Lambda}_0(s). \quad (1.63)$$

En réarrangeant les termes et en différenciant par rapport à  $t$ , on a :

$$Y_i(t) \exp[X_i(t)\widehat{\beta}] d\widehat{\Lambda}_0(t) = d\left(N_i(t) - \widehat{M}_i(t)\right). \quad (1.64)$$

En faisant la somme sur les individus, comme l'espérance des incréments d'une martingale est nulle, on a  $\sum_{i=1}^n d\widehat{M}_i(t) \rightarrow 0$  lorsque  $n \rightarrow \infty$ . Il ne reste plus qu'à réordonner les termes et à intégrer par rapport à  $t$  pour obtenir l'estimateur de Breslow du hasard intégré :

$$\widehat{\Lambda}_0(t) = \int_0^t \frac{d\bar{N}(s)}{\sum_{i=1}^n Y_i(s) \exp[X_i(s)\widehat{\beta}]}. \quad (1.65)$$

Remarquons que si  $\widehat{\beta} = 0$ , l'estimateur du hasard intégré de Breslow se réduit à l'estimateur de Nelson-Aalen. Breslow (1972) montre que son estimateur et  $\widehat{\beta}_{PL}$  maximisent simultanément la vraisemblance.

Après avoir présenté la vraisemblance partielle et fait un détour nécessaire par les martingales pour obtenir un estimateur paramétrique du hasard de base, nous allons étudier les modifications à apporter à la vraisemblance partielle en présence d'observations simultanées.

### 1.2.8 Vraisemblance partielle et observations simultanées

Nous avons supposé dans la sous-section 1.2.4 qu'il n'y a pas d'observations simultanées pour pouvoir ensuite réduire le dénominateur de la vraisemblance partielle à une somme sur les individus au risque. Cependant, il arrive très souvent dans les applications que deux épisodes au moins se terminent en même temps.

Deux explications sont généralement avancées pour la présence d'observation simultanées : soit les mesures trop imprécises, soit plusieurs durées sont réellement identiques. Différentes adaptations de la vraisemblance partielle sont appropriées selon l'explication envisagée.

Considérons le cas où le manque de précision des mesures a entraîné l'illusion de la simultanéité. Soit quatre individus indicés par  $i$  tels que :  $t_1 = t_2 < t_3 <$

$t_4$ . Si on note  $a_i = \exp[X_i(t_i)\beta]$ , les deux premiers facteurs de la vraisemblance partielle devraient avoir l'une des deux formes suivantes :

$$\left( \frac{a_1}{a_1 + a_2 + a_3 + a_4} \right) \left( \frac{a_2}{a_2 + a_3 + a_4} \right);$$

$$\left( \frac{a_2}{a_1 + a_2 + a_3 + a_4} \right) \left( \frac{a_1}{a_1 + a_3 + a_4} \right).$$

Le produit des numérateurs est identique, par contre, le produit des dénominateurs diffère entre les 2 expressions. L'approximation la plus simple est celle de Breslow, qui utilise la somme complète  $\sum_{i=1}^4 a_i$  pour les deux observations. L'approche de Efron attribue des pondérations et remplace le dénominateur du second terme par  $0.5r_1 + 0.5r_2 + r_3 + r_4$ . S'il y avait eu  $k$  observations simultanées, les poids auraient de 1 pour le premier facteur de la vraisemblance partielle,  $(k-1)/k$  pour le second, et ainsi de suite jusqu'à  $1/k$  pour le  $k$ -ème terme. Après une étude de Monte Carlo, Hertz-Picciotto et Rockhill (1997) concluent que l'approximation d'Efron fournit des  $\beta$  moins biaisés que celle de Breslow, hormis lorsqu'il y a peu de durées égales où les deux approches donnent des résultats équivalents. Elle rallonge toutefois les temps de calcul.

Lorsque les durées sont effectivement simultanées, le problème doit être envisagé sous un angle différent. A notre connaissance, des procédures pour la vraisemblance partielle ont été développées pour le modèle logistique uniquement :

$$\frac{\lambda_i(t)}{1 + \lambda_i(t)} = \frac{\lambda_0(t)}{1 + \lambda_0(t)} \exp(X_i\beta). \quad (1.66)$$

Dans le cas du modèle de Cox, autant que nous le sachions, l'adaptation de la vraisemblance partielle au cas d'observations réellement simultanées reste encore à faire.

Ceci nous amène à nous interroger sur l'hypothèse de hasards proportionnel. Compte tenu des données, jusqu'à quel point est-il raisonnable de faire cette hypothèse? Plusieurs méthodes permettent de répondre à cette question.

### 1.2.9 Test de l'hypothèse de hasard proportionnel

Différentes approches sont possibles pour tester l'hypothèse de proportionnalité, la première repose sur une analyse graphique tandis que la seconde est basée sur des tests.

Il convient de noter que dans le cadre du modèle de Cox, l'hypothèse de hasard proportionnel se traduit par :

$$\frac{\lambda_0(t) \exp(X_i\beta)}{\lambda_0(t) \exp(X_j\beta)} = \exp[(X_i - X_j)\beta], \quad (1.67)$$

qui ne dépend pas du temps. Dans le cas de régresseurs variant dans le temps, le ratio est égal à  $\exp[(X_i(t) - X_j(t))\beta]$ . Il se modifie certes au fil du temps, mais le coefficient attribué à la différence de deux variables explicatives reste le même quelle que soit la date.

Une analyse graphique peut être basée sur l'équation (1.7) dans le cas du modèle de Cox, car :

$$\begin{aligned} S_i(t) &= \exp[-\exp(X_i\beta)\Lambda_0(t)] \\ \Leftrightarrow \ln[-\ln(S_i(t))] &= \ln[\Lambda_0(t)] + X_i\beta. \end{aligned} \quad (1.68)$$

Comme le hasard de base est commun à tous les individus, les représentations graphiques d'estimations de la fonction de survie (obtenues avec l'estimateur de Breslow ou de Kaplan-Meier) pour deux individus quelconques ne doivent différer en tout points du temps que par une constante, qui vaut  $(X_i - X_j)\beta$ . Cette approche n'est pas valide lorsque les régresseurs varient dans le temps, la décomposition  $\Lambda(t) = \exp(X_i\beta)\Lambda_0(t)$  n'étant alors plus possible.

Lancaster (1990, p. 323) propose un test du score fondé sur la vraisemblance partielle pour l'hypothèse de hasard proportionnel dans le cas où les variables explicatives peuvent être séparées en deux groupes : celles qui varient dans le temps et celles qui sont fixes. On suppose la forme suivante pour le hasard :

$$\lambda_i(t|X_i, \omega_i(t); \beta) = \lambda_0(t) \exp(X_i\beta_1 + \omega_i(t)\beta_2), \quad (1.69)$$

où  $\omega_i(t)$  est un régresseur variant dans le temps. La fonction de hasard devient celle d'un modèle PH lorsque  $\beta_2 = 0$ . En différentiant la log-vraisemblance par rapport à  $\beta_2$  au point  $\beta_2 = 0$ , on obtient une statistique proportionnelle à :

$$\xi_n^s = \sum_i^n \omega_i(t_i) \hat{e}_i, \quad (1.70)$$

où  $\hat{e}_i$  est le  $i$ -ème résidu obtenu après estimation *via* la vraisemblance partielle sous la contrainte  $\beta_2 = 0$ .

### 1.2.10 Causes de la non acceptation de l'hypothèse de hasards proportionnels

Une première explication du rejet de l'hypothèse de hasards proportionnels est qu'un modèle PH est inadapté à ces données. Parmi les spécifications alternatives figure le modèle AFT (de l'anglais "Accelerated Failure Time"), qui est une extension du modèle de régression présenté dans la sous-section 1.1.8.

On suppose dans les modèles AFT que les durées dépendent des caractéristiques individuelles de la façon suivante :

$$T = \frac{T_0}{\lambda_1(X_i)}. \quad (1.71)$$

Où encore :

$$\ln T = \ln T_0 - \ln \lambda_1(X_i). \quad (1.72)$$

Une régression apparaît en ajoutant un terme d'erreur à (1.72). Elle est linéaire si on suppose  $\lambda_1(X_i) = \exp(X_i\beta)$ , et les modèles de régression présentés dans la sous-section 1.1.8 apparaissent comme un cas particulier où  $T_0$  vaut 1. La variable aléatoire de durée  $T$  est accélérée ou ralentie (respectivement raccourcie ou allongée) par rapport à  $T_0$  suivant que la fonction  $\lambda_1(X_i)$  est plus grande ou plus petite que 1, d'où le nom de "Accelerated Failure Time".

L'absence de proportionnalité apparente peut également être motivée par le choix d'une mauvaise forme fonctionnelle pour certaines variables explicatives. Dans ce cas, ce n'est pas la forme générale du modèle de Cox qui doit être remise en cause mais le choix des  $X_i$  ainsi que la manière dont ils interviennent dans la fonction de hasard.

Supposons que les durées aient été générées par le modèle de Cox suivant :

$$\lambda(t|Z_i) = \lambda_0(t) \exp(Z_i^1\omega_1 + Z_i^2\omega_2 + \dots + Z_i^p\omega_p), \quad (1.73)$$

où  $Z_i^k$  est la valeur du  $k$ -ième régresseur pour l'individu  $i$ . On pose le modèle suivant, qui ne diffère que par la forme fonctionnelle de la première variable explicative :

$$\lambda(t|Z_i) = \lambda_0(t) \exp(\ln(Z_i^1)\beta_1 + Z_i^2\beta_2 + \dots + Z_i^p\beta_p), \quad (1.74)$$

La distance qui sépare les courbes estimant le hasard intégré de deux individus est de  $(Z_i^1 - Z_j^1)\omega_1 + (Z_i^2 - Z_j^2)\omega_2 + \dots + (Z_i^p - Z_j^p)\omega_p$ , qui n'a aucun raison d'être égale à  $(\ln Z_i^1 - \ln Z_j^1)\beta_1 + (Z_i^2 - Z_j^2)\beta_2 + \dots + (Z_i^p - Z_j^p)\beta_p$ . Ceci nous amènerait, à l'issue d'un examen graphique, à rejeter à tort l'hypothèse de hasards proportionnels, alors que c'est la forme fonctionnelle d'un seul régresseur qui pose problème.

D'autres raisons peuvent amener à reconsidérer l'hypothèse de proportionnalité, comme par exemple l'omission de certaines variables explicatives. L'idée sous-jacente est que la population mère peut être divisée en plusieurs sous-populations qui ne sont pas prises en compte dans le modèle. Ceci nous amène naturellement à considérer le problème sous l'angle des mélanges. Le modèle PH a ainsi été étendu de manière à tenir compte ce phénomène, et

c'est pourquoi nous allons considérer dans la prochaine section le modèle de mélange de hasards proportionnels.

## 1.3. Le modèle de Mélange de Hasards Proportionnels

Le modèle de mélange de hasards proportionnels (souvent abrégé MPH, de l'anglais "Mixed Proportional Hazard Model") est une extension du modèle de Cox. L'apport qui nous intéresse principalement dans cette thèse concerne la prise en compte de l'hétérogénéité non-observée, mais le modèle MPH gagne également en souplesse en ne forçant pas les régresseurs à intervenir dans le hasard par l'intermédiaire de la fonction exponentielle.

Dans les premières sous-sections, nous allons présenter les concepts généraux propres aux modèles de mélange et des exemples de situations pouvant les engendrer. Ensuite, nous présenterons le modèle MPH tel qu'il est défini par Van den Berg (2001) ainsi que ses propriétés, avant d'insister sur une question importante qui est le choix de la distribution mélangeante. Puis, nous aborderons les extensions apportées au modèle MPH pour tenir compte de schémas d'enquête plus complexes que sont les épisodes multiples et le modèle MPH multivarié.

### 1.3.1 Éléments d'algèbre des mélanges

Nous ne considérons ici que des mélanges continus, le lecteur intéressé par les mélanges discrets peut se référer à Marin *et al.* (2005), parmi d'autres. Supposons que le hasard dépende non seulement du temps et des variables explicatives, mais aussi d'une quantité  $v$ , indépendante du temps et inobservable. Supposons aussi que le terme d'hétérogénéité  $v$  soit la réalisation d'une variable aléatoire  $V$ . Le hasard devient alors  $\lambda(t|X(t), v)$  et s'interprète comme la probabilité instantanée qu'un individu dans une population homogène par rapport à  $X(t)$  et  $v$  effectue une transition dans l'intervalle de temps  $[t, t + dt[$ . Ce hasard est spécifique à la sous population considérée et sera appelé par la suite le **hasard conditionnel**, de par la manière dont  $v$  est pris en compte. Supposons de plus que  $V$  prend différentes valeurs dans la population mais n'est pas observé. On note  $H(v|X)$  la fonction de répartition sachant  $X$ ,  $h(v|X)$  la densité et  $\mathcal{V}$  le support de  $h(v|X)$ . La densité

non-conditionnelle à  $v$  des durées s'obtient par intégration :

$$\begin{aligned} f(t|X) &= \int_{\mathcal{V}} f(t|X, v) dH(v|X) \\ &= E[f(t|X, v)]. \end{aligned} \tag{1.75}$$

De même, la fonction de survie s'écrit :

$$\begin{aligned} S(t|X) &= \int_{\mathcal{V}} S(t|X, v) dH(v|X) \\ &= E[S(t|X, v)]. \end{aligned} \tag{1.76}$$

Le hasard défini par  $f(t|X)/S(t|X)$  est appelé le **hasard non conditionnel**, ou encore le **hasard moyen**. En effet, on supprime le conditionnement par rapport à  $v$  dans (1.75) et (1.76) en prenant l'espérance par rapport à la loi de l'hétérogénéité non observée, c'est-à-dire en faisant une "moyenne pondérée" sur les différentes sous-populations de l'échantillon. Des durées distribuées selon une loi caractérisée par les deux relations précédentes sont dites issues d'un modèle de mélange, et la loi de  $V$  est appelée la distribution mélangeante.

De nombreuses situations peuvent donner lieu à des modèles de mélange. Les démographes utilisent les effets aléatoires, entre autre, pour représenter une fragilité commune à plusieurs individus qui influence leur durée de vie. Nous présenterons quelques situations engendrant des mélanges dans la prochaine sous-section.

#### 1.3.2 Situations produisant des mélanges

Lancaster (1990, p. 59-61) explique les modèles de mélange comme résultants d'erreurs de mesure, soit sur les observations, soit sur les variables explicatives. Nous allons reprendre sa démarche et présenter ces deux situations dans le cadre du modèle PH.

##### Erreur de mesure de la variable expliquée

On note  $T^*$  la variable aléatoire dont les réalisations sont les durées, et  $T$  les durées observées. Supposons qu'il y ait des imprécisions dans la mesure, où encore que les erreurs soient les réalisations d'une variable aléatoire  $V$  indépendante de  $T^*$ . On note  $Z$  une fonction connue de  $V$  telle que  $Z(V)$  agisse sur les durées de manière multiplicative, c'est-à-dire :  $T = Z(V)T^*$ .



La fonction de répartition de  $T$  est :

$$\begin{aligned}
 S(t) &= P[T \geq t] \\
 &= P[Z(V)T^* \geq t] \\
 &= E[P(Z(v)T^* \geq t | V = v)] \\
 &= E[P(T^* \geq t/z | V = v)],
 \end{aligned} \tag{1.77}$$

avec  $z = Z(v)$ . Par l'indépendance de  $Z$  et  $T^*$  :

$$\begin{aligned}
 S(t) &= E[P(T^* \geq t/z)] \\
 &= E \left[ \exp \left( - \int_0^{t/z} \lambda(u|X) du \right) \right].
 \end{aligned} \tag{1.78}$$

Selon l'expression du hasard, la relation précédente peut se réécrire :

$$S(t) = E \left[ \exp \left( - \int_0^t \lambda(u|X, v) du \right) \right]. \tag{1.79}$$

Par exemple, considérons le modèle de Weibull où  $\lambda(t|X) = \alpha t^{\alpha-1} \lambda_1(X)$ . On a :  $\int_0^{t/z} \lambda(u|X) du = \lambda_1(X) \int_0^{t/z} \alpha u^{\alpha-1} du = \lambda_1(X) (t/z)^\alpha$ . Supposons que le hasard conditionnel à  $v$  a pour expression :  $\lambda(t|X, v) = v \alpha t^{\alpha-1} \lambda_1(X)$  où  $v = z^{-\alpha}$ . Alors,  $\int_0^{t/z} \lambda(u|X) du = \int_0^t \lambda(u|X, v) du$ , et on a :

$$S(t) = E[S(t|X, v)]. \tag{1.80}$$

La fonction de survie  $S(t)$  a une expression analogue à (1.76), et on retrouve bien un mélange. On obtient le même résultat avec le modèle exponentiel, qui est un cas particulier du modèle de Weibull lorsque  $\alpha = 1$ .

Des résultats similaires peuvent être obtenus dans le cas d'erreurs de mesure des variables explicatives, ou encore lorsque certaines d'entre elles n'ont pas été prises en compte dans la régression.

### Erreurs de mesure et omission de variables explicatives

Considérons le modèle PH de l'équation (1.27) dont la variable explicative  $X_1$  a été mesurée avec une erreur  $V$ , c'est-à-dire  $X = X_1 + v$ , où  $V$  est une variable aléatoire indépendante de  $X_1$ . Soit :

$$h(t, X_1, v) = \lim_{\substack{dX_1 \rightarrow 0 \\ dV \rightarrow 0}} p(T \geq t, X_1 \in [X_1, X_1 + dX_1[, V \in [v, v + dv]). \tag{1.81}$$

### 1.3 Le modèle de Mélange de Hasards Proportionnels

---

Si on note par  $f$  la densité jointe de  $X_1$  et de  $V$ , on a :

$$\begin{aligned} h(t, X_1, v) &= S[t|X_1, V] f(X_1, v) \\ &= \exp\left(-\lambda_1(X_1) \int_0^t \lambda_0(u) du\right) f(X_1, v). \end{aligned} \quad (1.82)$$

Après quelques opérations figurant dans l'annexe A.1, on obtient comme expression de la fonction de survie conditionnelle à la variable explicative observée :

$$\begin{aligned} S(t|X) &= \int_{\mathcal{V}} \exp\left(-\lambda_1(X-v) \int_0^t \lambda_0(u) du\right) f(v|X) dv \\ &= \int_{\mathcal{V}} \exp\left(-\int_0^t \lambda(t|X, v)\right) f(v|X) dv. \end{aligned} \quad (1.83)$$

Cette dernière relation est de même type que l'équation (1.76). Si le hasard conditionnel d'un modèle PH a la forme  $\lambda_0(t)\lambda_1(X-v)$ , il peut donc résulter d'erreurs de mesures sur la variable explicative  $X_1$ .

Cette approche montre également que l'omission de régresseurs mène à des modèles de mélanges. Supposons que les durées ont été générées selon le modèle PH caractérisé par  $\lambda(t) = \lambda_0(t)\lambda_1(X\beta)$  où  $X\beta = X_1\beta_1 + X_2\beta_2$ , mais que  $X_2$  ne soit pas observable. Alors, le hasard conditionnel est de la forme  $\lambda(t|X_1, v) = \lambda_0(t)\lambda_1(X\beta - v)$  où  $v \equiv X_2\beta_2$  et peut être justifié par l'omission de variables explicatives. Un exemple en est le modèle de Cox avec hétérogénéité multiplicative dont le hasard conditionnel est  $w\lambda_0(t) \exp(X\beta)$  où  $w \equiv \exp(-v)$ .

Lancaster (1990, p. 61) indique que la principale justification du terme d'erreur dans les premiers textes sur le modèle linéaire était justement la prise en compte d'erreurs de mesures dans les variables explicatives. Le même type d'argument peut-être utilisé dans les modèles de durée afin de justifier un élément  $v$  captant les erreurs de mesure. L'interprétation de  $v$  est alors différente de celle d'un terme d'hétérogénéité, même si les modèles réduits ont des expressions identiques.

Une façon de prendre en compte l'hétérogénéité non observée, qu'elle que soit son origine, est de mélanger avec une forme fonctionnelle du hasard un terme indépendant des variables (explicatives et expliquées) et du temps. Ceci nous amène à utiliser des modèles de mélange tels qu'ils ont été présentés dans la sous-section 1.3.1, dont un cas particulier est le modèle MPH.

#### 1.3.3 Modèle

Comme précédemment, on note  $X_{ij}$  la valeur de la  $j$ -ème ( $j = 1 \dots p$ ) variable explicative relative à l'unité  $i$  ( $i = 1 \dots n$ ). La matrice des variables

explicatives, de plein rang colonne, est notée  $X$  et  $X_i$  désigne le vecteur des valeurs prises par les variables explicatives pour l'unité  $i$ . Pour une unité entièrement caractérisé par  $X_i$  et la réalisation  $v$  d'un terme d'hétérogénéité non observée distribué suivant une loi quelconque, Van den Berg (2001) définit le modèle MPH de la manière suivante :

**Définition 1** *Il existe des fonctions  $\lambda_0$  et  $\lambda_1$  telles que pour tout  $t$ ,  $X_i$  et  $v$ , le hasard conditionnel a pour expression :*

$$\lambda_i(t|X_i, v) = v\lambda_0(t)\lambda_1(X_i). \quad (1.84)$$

La fonctions  $\lambda_0$  est appelée le **hasard de base** car elle est commune à tous les individus et ne dépend que du temps. Elle dénote l'évolution de la probabilité instantanée de transition alors que le temps passe, laquelle se trouve amplifiée ou atténuée par les caractéristiques individuelles. Les variables explicatives interviennent au travers de la fonction  $\lambda_1$ , appelée la **partie systématique** du hasard. Le terme d'hétérogénéité  $v$  peut être formulé de différentes manières, et nous présentons différentes spécifications dans la sous-section 1.3.4. Lorsque  $v$  est égal à 1, il n'y a pas d'hétérogénéité non observée et le modèle MPH devient un modèle PH comprenant le modèle de Cox comme cas particulier.

La formulation en termes de processus de comptage suppose que le processus  $N$  a comme intensité :

$$\lambda_i(t|X_i, v) = Y_i(t)v\lambda_0(t) \exp(X_i\beta). \quad (1.85)$$

Compte tenu de la manière dont  $Y_i(t)$  a été défini dans la sous-section 1.1.2, le modèle (1.85) ne représente que des situations où les individus ont été observés au cours d'un épisode unique et il est donc plus spécifique que (1.84).

Les conditions de régularité du modèle MPH varient d'une étude à l'autre selon les propriétés désirées. Van den Berg (2001) synthétise les hypothèses habituelles de la manière suivante :

**Hypothèse 1** *La fonction  $\lambda_0(t)$  est positive et continue sur  $[0, \infty[$ . La limite de  $\lambda_0(t)$  lorsque  $t$  tend vers 0 peut ne pas être finie. Pour tout  $t \geq 0$ , on a  $\int_0^t \lambda_0(u)du < \infty$  et  $\lim_{t \rightarrow \infty} \int_0^t \lambda_0(u)du = \infty$ .*

Le hasard de base correspond à la probabilité instantanée de transition dans un modèle sans hétérogénéité observée ou non observée, et doit donc être positif. La condition  $\lim_{t \rightarrow \infty} \int_0^t \lambda_0(u)du = \infty$  suppose que tous les épisodes se terminent par une transition à un moment ou un autre (au pire "à la fin des temps"). Le modèle MPH n'est donc pas conçu pour une population dont une partie des unités ne peuvent pas voir échoir leurs durées et le modèle de "split

population” (Schmidt, 1989) est alors approprié. Les autres conditions de l’hypothèse sont légèrement plus fortes que celles rencontrées habituellement dans la littérature, ce qui nous évite de reformuler ultérieurement (H1) à de nombreuses reprises dans la suite de cette thèse.

**Hypothèse 2** *Le vecteur  $X_i$  est de dimension  $p$  où  $1 \leq p < \infty$ . La fonction  $\lambda_1(X_i)$  est positive pour tout  $X_i \in \mathcal{X} \subset \mathbb{R}^p$ .*

La positivité de  $\lambda_1$  est nécessaire pour que le hasard, c’est-à-dire la probabilité instantanée de transition, soit positif. Dans les applications,  $\lambda_1$  est souvent spécifié comme la fonction exponentielle et chaque variable explicative supplémentaire entre alors de façon multiplicative dans le hasard, comme dans le modèle de Cox. Quelle que soit la fonction  $\lambda_1$ , la prise en compte de régresseurs variant dans le temps requiert l’hypothèse de prédictibilité présentée dans la sous-section 1.2.2.

**Hypothèse 3** *La loi de  $V$  satisfait la condition  $P(0 < V < \infty) = 1$ .*

Le support de la distribution mélangeante doit être dans l’intervalle  $[0, \infty[$ . La spécification d’un modèle MPH ne nécessite pas d’autres hypothèses sur la loi de  $V$ , le choix de la distribution mélangeante n’est nécessaire que pour l’estimation.

**Hypothèse 4** *Pour une unité de niveau 1,  $v$  ne dépend pas du temps.*

Nous verrons dans la sous-section 1.3.7 consacrée au modèle MPH multivarié que  $V$  peut prendre différentes valeurs dans le temps pour des unités de niveau supérieur à 1.<sup>19</sup>

#### 1.3.4 Effet(s) aléatoire(s)

La fonction de l’hétérogénéité non observée admet différentes spécifications, et nous présentons ici différentes formulations utilisées dans la littérature.

##### Formulations à un effet aléatoire

Soit  $i$ , ( $i = 1, \dots, I$ ) l’indice des unités de niveau 2 et  $j$ , ( $j = 1, \dots, J_i$ ) l’indice des unités de niveau 1. La spécification la plus courante est :

$$\lambda_j(t|X_j, v_j) = v_j \lambda_0(t) \lambda_1(X_j). \quad (1.86)$$

---

<sup>19</sup>La notion de niveaux est présentée dans la sous-section 2.5

L'effet aléatoire a une valeur propre à chaque unité de niveau 1. Les  $v_j$  sont généralement supposés indépendants (Clayton, 1978, McGilchrist et Aisbett, 1991), bien qu'ils puissent tout aussi bien être issus d'une loi multivariée.<sup>20</sup> On parle de modèles MPH à épisodes uniques,  $v_j$  est souvent un effet individuel qui capte les caractéristiques inobservées de l'agent influençant son unique durée observée. Des exemples d'application se trouvent dans la littérature sur le chômage (Lancaster, 1979), la mortalité infantile (Guo et Rodriguez, 1992), etc. On peut aussi adopter cette spécification lorsqu'on soupçonne des erreurs de mesure au niveau individuel, et  $v_j$  ne capte alors plus l'hétérogénéité non observée mais la surdispersion.

Lorsque l'individu est observé en plusieurs occasions, il devient une unité de niveau 2 tandis que chaque durée est une unité de niveau 1. Afin de capter l'hétérogénéité individuelle, on pose :

$$\lambda_{ij}(t|X_{ij}, v_i) = v_i \lambda_0(t) \lambda_1(X_{ij}). \quad (1.87)$$

Comme plusieurs durées correspondent à chaque unité de niveau 2 dans ces modèles MPH, on les appelle des modèles à épisodes multiples et ils sont présentés en détail dans la sous-section 1.3.7. La réalisation  $v_i$  est partagée par toutes les unités du groupe  $i$ , c'est-à-dire qu'il existe une hétérogénéité commune entre deux unités de niveau 1 qui appartiennent au groupe  $i$ . Les  $v_i$  sont généralement supposés indépendants, et les observations sont alors indépendantes conditionnellement aux variables explicatives et à l'effet groupe.

Il peut être raisonnable de penser que l'hétérogénéité se situe à deux niveaux : d'une part entre les groupes, et d'autre part entre les unités appartenant à un même groupe. La spécification adéquate est alors :

$$\lambda_{ij}(t|X_{ij}, v_{ij}) = v_{ij} \lambda_0(t) \lambda_1(X_{ij}). \quad (1.88)$$

Elle a été initialement proposée par McGilchrist (1993) pour l'étude de durées corrélées au niveau des individus et des familles du fait de facteurs génétiques. La fonction  $v$  prend simultanément en compte l'hétérogénéité non-observée à deux niveaux, entre les familles et entre les individus d'une même famille. Il s'agit d'une spécification hybride, où les  $v_i$  sont indépendants et où  $v_{ij}$  est corrélé avec  $v_{i'j'}$ . Ainsi, la matrice de variance du vecteur  $(v_{i1}, \dots, v_{iJ_i})$ , comprenant les termes d'hétérogénéité pour les membres du groupe  $i$ , n'est pas diagonale et le modèle MPH multivarié est présenté en détails dans la sous-section 1.3.7. La même fonction de l'hétérogénéité non observée peut être obtenue en utilisant deux effets aléatoires.

---

<sup>20</sup>Bonnal *et al.* (1997) et Fougère et Kamionka (2005) introduisent une relation de dépendance en posant  $v_j = \exp(a_j u_1 + b_j u_2)$ , où  $u_1$  et  $u_2$  sont des variables latentes indépendantes.

### Formulations à deux effets aléatoires

Yashin *et al.* (1995) considèrent la structure d'hétérogénéité suivante :

$$\lambda_j(t|X_j, v_0, w_j) = (v_0 + w_j)\lambda_0(t)\lambda_1(X_j), \quad (1.89)$$

où  $v_0$  est un effet commun à toutes les observations et  $w_j$  est spécifique à l'unité  $j$ . Les deux effets sont supposés indépendants et comme ils sont distribués selon des lois gamma de même paramètre d'échelle, la somme  $v_0 + w_j$  suit aussi une loi gamma et les conditions de régularité du modèle MPH sont satisfaites. Le modèle est similaire à celui défini par (1.88) où les effets sont issus d'une loi multivariée, et comprend le cas particulier du modèle à épisodes uniques lorsque  $\text{Var}(v_0)=0$ .

Une spécification légèrement différente est :

$$\lambda_{ij}(t|X_{ij}, v_0, w_j) = (v_0 + w_i)\lambda_0(t)\lambda_1(X_{ij}). \quad (1.90)$$

L'effet aléatoire  $w_i$  le plus fin est défini à un niveau plus agrégé que le niveau 1. Pour ne pas violer les conditions de régularité, le plus simple est supposer que les deux effets suivent une loi gamma. On obtient un modèle à épisodes multiples lorsque  $\text{Var}(v_0)=0$ .

Une autre spécification permet de décrire une hétérogénéité située à 3 niveaux et de s'affranchir de l'hypothèse d'hétérogénéité gamma. Soit  $i$ , ( $i = 1, \dots, I$ ) l'indice des unités de niveau 3,  $j$ , ( $j = 1, \dots, J_i$ ) l'indice des unités de niveau 2 et  $k$ ,  $k = 1, \dots, K_{ij}$  l'indice des unités de niveau 1. Sastry (1997) considère le modèle :

$$\lambda_{ijk}(t|X_{ijk}, v_i, w_j) = v_i w_j \lambda_0(t)\lambda_1(X_{ijk}). \quad (1.91)$$

Il étudie la mortalité infantile et considère deux niveaux d'hétérogénéité : une hétérogénéité au niveau des familles et une autre au niveau des communautés dans lesquelles résident les familles. Comme les niveaux sont hiérarchiques, les effets sont dits emboîtés : une unité de niveau quelconque n'appartient qu'à une seule unité de niveau supérieur. Les modèles (1.89) et (1.90) en sont des cas particuliers.

L'hétérogénéité n'est pas toujours hiérarchique et on peut étendre le modèle MPH en conséquence. Soit  $i$ , ( $i = 1, \dots, I$ ) l'indice des groupes définis par un premier effet,  $j$ , ( $j = 1, \dots, J$ ) l'indice des groupes définis par un second effet et  $k$ ,  $k = 1, \dots, n$  l'indice des observations. Le modèle (1.91) est alors un modèle à effets aléatoires non emboîtés, approprié par exemple pour un chercheur d'emploi  $k$  caractérisé par le diplôme  $i$  et la région géographique  $j$ .

D'autres formes plus générales pour la fonction d'hétérogénéité non observée sont présentées dans la sous-section 2.1, le modèle MPH étant très flexible. Combiné à la simplicité de sa forme générale et à la facilité avec laquelle il est possible de gérer les régresseurs variant dans le temps, il n'est pas surprenant qu'il soit utilisé dans de nombreuses applications.

### 1.3.5 Choix de la distribution mélangeante

L'utilisation de la loi gamma comme distribution mélangeante au sein de modèles durée a été introduite par Lancaster (1979). Peut-être a-t-il cherché à approfondir l'idée d'événements suivants chacun un processus de Poisson de paramètre distribué selon une loi gamma, c'est-à-dire un modèle Negbin, de Greenwood et Yule (1920). L'apport de Lancaster (1979) a ensuite été repris dans de nombreux travaux, où le succès de la loi gamma est principalement motivé par la simplicité de sa forme analytique. Abbring et Van den Berg (2001) lui donnent une justification *ex-post* plus intéressante. Considérons le cas où  $V$  est une variable aléatoire continue, de fonction de répartition  $H(v)$  dont le support est borné inférieurement par 0. Alors, sous certaines conditions de régularité, la loi de  $V$  parmi les survivants à la date  $t$  tend vers une loi gamma lorsque  $t$  tend vers l'infini. Ce résultat est valide lorsque la loi de  $V$  suit une distribution beta, uniforme, log-normale, inverse-gaussienne, etc... Par contre, la convergence n'a pas lieu si la vraie loi du terme d'hétérogénéité est discrète.

Le choix de la distribution mélangeante a des répercussions sur l'inférence. Le cas de l'estimation par maximum de vraisemblance a été largement traité dans la littérature, que l'inférence soit conduite sur données simulées (Ferreira et Garcia, 2002, parmi d'autres) ou réelles (par exemple Heckman et Singer, 1984a). Les conclusions sont que les estimations des paramètres varient considérablement selon la spécification de la loi de  $V$ . Une revue de littérature assez complète sur le modèle MPH lorsque la distribution mélangeante est mal spécifiée est présentée dans Van den Berg (2001). De manière générale, il semble qu'une spécification erronée pour la distribution mélangeante a de sérieuses conséquences sur les coefficients estimés.

Plusieurs approches sont dès lors possibles. Une première est de modéliser l'hétérogénéité au moyen d'une loi très souple. Ainsi, Hougaard (1984) propose une distribution admettant comme cas particuliers la loi gamma et la loi normale. Une seconde approche est d'opter pour une spécification non paramétrique de la loi de l'hétérogénéité non observée, c'est-à-dire une spécification où ni le terme d'hétérogénéité, ni l'appartenance à un groupe, ne sont observées. Heckman et Singer (1984a) proposent un estimateur non paramétrique du maximum de vraisemblance (abrégé NPMLE), qui suppose une loi

discrète avec un domaine de définition et un nombre de points de masse inconnus. Baker et Melino (1999) étudient le comportement du NPMLE par méthode de Monte Carlo. Leurs résultats indiquent que les estimateurs de la pseudo vraisemblance se comportent bien, même dans de petits échantillons, en présence d'une spécification non paramétrique de la loi de  $V$  ou bien d'une spécification non paramétrique du hasard de base. Par contre, lorsque la distribution mélangeante et le hasard de base sont simultanément spécifiés de manière non paramétrique, les estimateurs sont biaisés et ne convergent que très lentement au fur et à mesure que la taille de l'échantillon augmente.<sup>21</sup> On peut également spécifier une hétérogénéité non-paramétrique avec un mélange de processus de Dirichlet (Ferguson, 1973). L'approche hiérarchique Bayésienne se prête à l'estimation de tels modèles où la distribution mélangeante est elle-même générée par un processus de Dirichlet, c'est-à-dire où la loi de l'hétérogénéité non observée suit a priori un processus de Dirichlet. Florens *et al.* (1999) proposent une revue de la littérature Bayésienne des modèles de durée où le processus de Dirichlet est utilisé comme a priori. Campolieti (2001) estime un modèle en temps discret et Lau (2006) un modèle MPH munis d'une telle hétérogénéité. Poser un mélange de processus de Dirichlet amène à une loi de l'hétérogénéité continue, et le cas d'une hétérogénéité discrète et multivariée, associée à un hasard de base constant par morceaux, est estimé dans Van den Berg *et al.* (2002) et Zijl *et al.* (2004).

### 1.3.6 Propriétés

Nous présentons ici certaines propriétés du modèle MPH, qui ont été présentées à de nombreuses reprises dans la littérature.<sup>22</sup> Nous parlerons tout d'abord de la décroissance au fil du temps de la valeur moyenne de  $v$  parmi les survivants, qui a des conséquences directes sur le taux de variation (du fait du passage du temps ou de l'augmentation des variables explicatives) du hasard non conditionnel à  $v$ . Nous comparerons ensuite le taux de variation du hasard non conditionnel à  $v$  avec le taux de variation du hasard conditionnel pour voir de quelle manière l'omission de l'hétérogénéité non observée peut modifier les résultats d'une étude économétrique. Nous nous restreignons ici au cas où les régresseurs ne varient pas dans le temps et où l'hétérogénéité est indépendante des variables explicatives.

---

<sup>21</sup>Baker et Melino (1999) n'examinent pas spécifiquement le modèle MPH mais plutôt une version discrétisée de celui-ci. Les résultats sont donc donnés ici à titre indicatif, une réponse définitive sur le comportement du NPMLE dans ce modèle est une piste pour des applications ultérieures.

<sup>22</sup>Voir par exemple Lancaster (1990, p. 62-65), Van den Berg (2001) ou encore Gouriéroux et Jasiak (2001, p. 454-456).



### Décroissance de la moyenne de $v$ parmi les survivants

Nous reprenons ici principalement les développements de Van den Berg (2001) et de Lancaster (1990). A partir des relations (1.75) et (1.76), le hasard non conditionnel à l'hétérogénéité admet comme expression :<sup>23</sup>

$$\begin{aligned}\lambda(t) &= \frac{\int_{\mathcal{V}} f(t|v)dH(v)}{\int_{\mathcal{V}} S(t|v)dH(v)} \\ &= \int_{\mathcal{V}} \lambda(t|v) \frac{S(t|v)dH(v)}{\int_{\mathcal{V}} S(t|v)dH(v)} \\ &= \int_{\mathcal{V}} \lambda(t|v)f(v|t)dv.\end{aligned}\tag{1.92}$$

Ainsi, le hasard non conditionnel peut s'interpréter comme l'espérance du hasard conditionnel par rapport à la loi de densité  $f(v|t)$ , c'est-à-dire la densité conditionnelle de  $V$  sachant la survie à  $t$ . Le hasard non conditionnel à  $v$  est parfois appelé le "hasard moyen" car il s'agit d'une moyenne sur la population des survivants à  $t$ , laquelle peut être très différente de  $\lambda(t|v)$ . La valeur moyenne du terme d'hétérogénéité parmi les survivants à la date  $t$  est obtenue comme :

$$\begin{aligned}\mathbb{E}(V|T \geq t) &= \int_{\mathcal{V}} vf(v|t)dv \\ &= \int_{\mathcal{V}} v \frac{S(t|v)dH(v)}{\int_{\mathcal{V}} S(t|v)dH(v)} \\ &= \frac{\int_{\mathcal{V}} v \exp[-v\Lambda(t)]dH(v)}{\int_{\mathcal{V}} \exp[-v\Lambda(t)]dH(v)},\end{aligned}\tag{1.93}$$

où  $\Lambda(t) = \lambda_1(X) \int_0^t \lambda_0(u)du$ . On peut étudier l'évolution de cette espérance au fil du temps en calculant sa dérivée par rapport à  $t$ .

$$\begin{aligned}\frac{d\mathbb{E}(V|T \geq t)}{dt} &= \lambda_0(t)\lambda_1(X) \left[ -\frac{\int_{\mathcal{V}} v^2 S(t|v)dH(v)}{\int_{\mathcal{V}} S(t|v)dH(v)} + \left( \frac{\int_{\mathcal{V}} v S(t|v)dH(v)}{\int_{\mathcal{V}} S(t|v)dH(v)} \right)^2 \right] \\ &= -\lambda_0(t)\lambda_1(X) [\mathbb{E}(V^2|T \geq t) - [\mathbb{E}(V|T \geq t)]^2] \\ &= -\lambda_0(t)\lambda_1(X) \text{Var}[V|T \geq t] \leq 0.\end{aligned}\tag{1.94}$$

On voit ici un phénomène de sélection dans la population : plus le temps passe, plus les survivants auront en moyenne des petites valeurs  $v$ . En effet,

---

<sup>23</sup>Pour alléger les notations, le conditionnement par rapport aux variables explicatives a été omis.

### 1.3 Le modèle de Mélange de Hasards Proportionnels

---

des valeurs  $v$  élevées correspondent à des hasards, et donc des probabilités de transition, élevées et ainsi à des individus aux durées moyennes faibles. Cette propriété du modèle MPH est appelée le “weeding out” dans la littérature anglophone. Le modèle MPH fait partie de la famille des modèles “mover-stayer”, pouvant distinguer deux sous-populations : celle des individus motivés susceptibles d’effectuer une transition (dans le contexte de la recherche d’emploi) et celle des individus qui ne sont pas motivés et ne peuvent donc pas transiter. Dans ce type de modèles, la population des chômeurs ne comprend plus que d’individus non-motivés lorsque  $t \rightarrow \infty$ , les autres ayant déjà effectué une transition vers l’emploi.<sup>24</sup> La présence du “weeding out” explique pourquoi l’omission de l’hétérogénéité non observée entraîne des estimations biaisées. En effet, l’ensemble des individus au risque se modifie au fur et à mesure que le temps passe pour se restreindre à ceux caractérisés par les plus petites valeurs de  $v$ . Une estimation du hasard où l’effet aléatoire est omis sous-estimera donc l’influence des variables explicatives. Le “weeding out” a également des conséquences sur le taux d’accroissement du hasard en fonction du temps et des variables explicatives.

#### Taux d’accroissement du hasard dans le temps

Comparons maintenant l’évolution du hasard conditionnel à celle du hasard moyen. D’après l’équation (1.92), on a :

$$\begin{aligned}\lambda(t) &= \text{E} [\lambda(t|v)|T \geq t] \\ &= \lambda_0(t)\lambda_1(X)\text{E}[V|T \geq t].\end{aligned}\tag{1.95}$$

D’où, en utilisant la relation (1.94) :

$$\frac{d\lambda(t)}{dt} = \lambda_1(X)\frac{d\lambda_0(t)}{dt}\text{E}[V|T \geq t] - [\lambda_0(t)\lambda_1(X)]^2 \text{Var}[V|T \geq t].\tag{1.96}$$

Cette relation nous montre que si on observe un hasard moyen croissant, on peut déduire que le hasard conditionnel est lui aussi croissant dans le temps.

---

<sup>24</sup>Florens *et al.* (1996) montrent en détail le “weeding-out” se produisant dans un modèle “mover-stayer” à deux sous-populations.

En divisant par  $\lambda(t)$ , on obtient le taux de croissance du hasard moyen :

$$\begin{aligned}
 \frac{d \ln \lambda(t)}{dt} &= \frac{1}{\lambda(t)} \frac{d\lambda(t)}{dt} \\
 &= \frac{1}{\lambda_0(t)} \frac{d\lambda_0(t)}{dt} - \lambda_0(t)\lambda_1(X) \frac{\text{Var}[V|T \geq t]}{\text{E}[V|T \geq t]} \\
 &= \frac{d \ln \lambda_0(t)}{dt} - \lambda_0(t)\lambda_1(X) \frac{\text{Var}[V|T \geq t]}{\text{E}[V|T \geq t]} \\
 &= \frac{d}{dt} [\ln \lambda_0(t) + \ln \lambda_1(X) + v] - \lambda_0(t)\lambda_1(X) \frac{\text{Var}[V|T \geq t]}{\text{E}[V|T \geq t]} \\
 &< \frac{d \ln \lambda(t|v)}{dt}. \tag{1.97}
 \end{aligned}$$

Le hasard moyen  $\lambda(t)$  décroît donc plus vite que le hasard conditionnel  $\lambda(t|v)$ . Ce résultat est un corollaire du “weeding out” : le hasard moyen est faible car évalué sur une population peu susceptible d’effectuer une transition, tandis que le hasard conditionnel tient compte de cette faible mobilité. Si on ne prend pas en compte la présence d’hétérogénéité non observée dans les données alors qu’elle est présente, par exemple en posant un modèle PH alors que les durées sont issues d’un modèle MPH, le hasard PH fera état d’une décroissance plus rapide que celle du modèle MPH. On peut être amené à rejeter à tort l’hypothèse d’un hasard constant au fil du temps au profit de l’hypothèse de décroissance, du fait de l’omission de l’hétérogénéité non observée.

### Taux d’accroissement avec les variables explicatives

L’hétérogénéité non observée influence l’évolution du hasard en fonction des variables explicatives. Considérons la dérivée par rapport à  $X_j$  du logarithme de l’équation (1.95) pour calculer le taux de croissance du hasard non conditionnel. A noter que  $\text{E}[V|T \geq t]$ , évalué en (1.93), dépend des variables explicatives au travers du hasard intégré  $\Lambda(t)$ . En effectuant des calculs similaires à ceux qui ont permis d’obtenir la dérivée de  $\text{E}[V|T \geq t]$  par rapport au temps et en divisant le résultat par  $\text{E}[V|T \geq t]$ , on obtient :

$$\frac{d \ln \text{E}[V|T \geq t]}{dX_j} = - \frac{d\lambda_1(X)}{dX_j} \Lambda_0(t) \frac{\text{Var}(V|T \geq t)}{\text{E}[V|T \geq t]}. \tag{1.98}$$

D'où :

$$\begin{aligned} \frac{d \ln \lambda(t)}{dX_j} &= \frac{d\lambda_1(X)}{dX_j} \frac{1}{\lambda_1(X)} - \frac{d\lambda_1(X)}{dX_j} \Lambda_0(t) \frac{\text{Var}(V|T \geq t)}{\text{E}[V|T \geq t]} \\ &< \frac{d \ln \lambda(t|v)}{dX_j}. \end{aligned} \quad (1.99)$$

On en déduit que la présence d'hétérogénéité non observée diminue le taux de variation du hasard moyen. Par exemple, si l'élasticité par rapport à une variable est constante dans le vrai modèle, et que les données présentent de l'hétérogénéité non observée qui n'est pas prise en compte, alors l'économètre peut être amené à rejeter à tort un modèle à élasticité constante au profit d'un modèle supposant une élasticité décroissante. L'intuition est la même que pour le "weeding out".

### Effet croisé du temps et des variables explicatives

Nous nous intéressons à l'influence des variables explicatives sur l'évolution du hasard moyen dans le temps. Prenons un exemple : supposons que notre population soit composée de deux types d'individus distingués par une variable indicatrice. Supposons que les individus pour lesquels l'indicatrice vaut 0 aient en moyenne un hasard constant dans le temps. A l'inverse, les individus caractérisés par la valeur 1 de l'indicatrice ont en moyenne un hasard décroissant dans le temps. Dans ce cas, on dira que la variable explicative a un effet négatif sur l'évolution du hasard moyen dans le temps car il décroît pour la valeur "élevée" de la variable explicative.

Étudions plus précisément le signe de la dérivée croisée du hasard moyen par rapport au temps et aux variables explicatives. Pour simplifier, on suppose que les variables explicatives varient de manière continue afin que l'ordre de dérivation n'ait pas d'importance. Rappelons que sous les hypothèses de régresseurs constant dans le temps et d'effet aléatoire indépendant des variables explicatives, le hasard moyen a pour expression (1.95). La dérivée croisée par rapport au temps et aux variables explicatives du hasard moyen se réduit donc à la dérivée de  $\text{E}[V|T \geq t]$ . Après quelques calculs détaillés dans l'Annexe A.2, page 154, on obtient :

$$\begin{aligned} \frac{\partial^2 \ln \text{E}[V|T \geq t]}{\partial X_j \partial t} &= -\frac{d\lambda_1(X)}{dX_j} \lambda_0(t) \left\{ \frac{\text{Var}(V|T \geq t)}{\text{E}(V|T \geq t)} - \lambda_1(X) \Lambda_0(t) \right. \\ &\quad \left. \left[ 2\text{Var}(V|T \geq t) + \left( \frac{\text{Var}(V|T \geq t)}{\text{E}(V|T \geq t)} \right)^2 - \frac{\text{Var}(V^{3/2}|T \geq t)}{\text{E}(V|T \geq t)} \right] \right\}. \end{aligned} \quad (1.100)$$

Le taux de variation ne dépend que des trois premiers moments de la loi du terme d'hétérogénéité, c'est-à-dire que l'impact des variables explicatives sur l'évolution du hasard moyen au fil du temps ne dépend que du choix de  $H(v)$ .

Van den Berg (2001) s'intéresse à un modèle MPH où le terme d'hétérogénéité est distribué selon une loi discrète quelconque, et conclut que le signe de la dérivée croisée du hasard moyen par rapport au temps et aux variables explicatives est indéterminé. Il illustre cette propriété par un exemple où le taux d'accroissement du hasard moyen est nul jusqu'à ce que  $\lambda_1(X)\Lambda_0(t)$  franchisse un certain seuil au delà duquel il devient positif. Il fournit un second exemple où la loi de  $V$  ne comprend que deux points de support, le premier ayant une masse de probabilité très élevée par rapport au second, ce qui aboutit à une influence presque nulle de  $X$  sur l'évolution du hasard moyen dans le temps. Supposer une hétérogénéité discrète n'est donc pas suffisant pour connaître cette influence qui doit être étudiée pour chaque distribution mélangeante.

Il en va de même pour les lois continues. Lancaster (1990, p. 65-67) considère un modèle à hétérogénéité gamma. Il montre que sous cette hypothèse distributionnelle, le taux de variation du hasard moyen en fonction des variables explicatives est atténué par la présence d'hétérogénéité. En d'autres termes, le hasard moyen diminue lorsque  $X$  et  $t$  prennent des valeurs élevées.<sup>25</sup>

### 1.3.7 Adaptation à d'autres schémas d'enquêtes

Nous avons vu dans la sous-section 1.3.4 que la souplesse de la fonction de l'hétérogénéité non observée dans le modèle MPH permet de prendre en compte des données structurées de nombreuses manières.

#### Le modèle à épisodes multiples

Soit une unité de niveau 2 connaissant deux épisodes (unités de niveau 1). Comme les valeurs prises par les variables explicatives peuvent différer d'un épisode à l'autre, notons  $X_1$  leurs valeurs au début du premier épisode et  $X_2$  leur valeurs au début du second. La réalisation  $v$  est commune aux deux épisodes, comme dans le hasard 1.87, et les variables  $T_1$  et  $T_2$  sont indépendantes conditionnellement à  $V$ ,  $X_1$  et  $X_2$ . La densité jointe des durées est obtenue en généralisant l'équation (1.75) :

$$f(t_1, t_2 | X_1, X_2) = \int_{\mathcal{V}} f(t_1 | X_1, v) f(t_2 | X_2, v) dH(v). \quad (1.101)$$

---

<sup>25</sup>Le cas particulier de durées exponentielles est traité dans Gouriéroux et Jasiak (2001).

### 1.3 Le modèle de Mélange de Hasards Proportionnels

---

On suppose dans l'équation précédente que  $V$  est indépendant de  $X_i$ . La fonction de survie a pour expression en étendant (1.76) :

$$\begin{aligned} S(t_1, t_2 | X_1, X_2) &= \int_{\mathcal{V}} S(t_1 | X_1, v) S(t_2 | X_2, v) dH(v) \\ &= \int_{\mathcal{V}} \exp \left( -v \left[ \lambda_1(X_1) \int_0^{t_1} \lambda_0(u) du \right. \right. \\ &\quad \left. \left. + \lambda_1(X_2) \int_0^{t_2} \lambda_0(u) du \right] \right) dH(v). \end{aligned}$$

L'extension au cas où  $v_1$  caractérise le premier épisode et  $v_2$  le second relève du modèle MPH multivarié que nous présentons dans la sous-section suivante.

#### Le modèle MPH multivarié

Le modèle MPH multivarié est une généralisation du modèle à épisodes multiples. Il s'en distingue en autorisant  $\lambda_0$ ,  $\lambda_1$  et  $v$  à varier selon les épisodes. Considérons une unité  $i$  de niveau 2 dont les variables explicatives sont notées  $X_i$ , et effectuant deux épisodes (chaque épisode est une unité de niveau 1). Comme les variables explicatives peuvent différer d'un épisode à l'autre, on note  $X_{i,1}$  leurs valeurs au début du premier épisode et  $X_{i,2}$  leurs valeurs au début du second. Au premier épisode correspond la réalisation  $v_1$  et au second  $v_2$ , chacune pouvant être spécifique à l'unité  $i$  ou bien partagée avec d'autres unités de même niveau. Van den Berg (2001) définit le modèle MPH multivarié (MMPH) de la manière suivante :

**Définition 2** *Il existe des fonctions  $\lambda_{0,1}$ ,  $\lambda_{0,2}$ ,  $\lambda_{1,1}$  et  $\lambda_{1,2}$  telles que pour tout  $t$ ,  $X_{i,1}$ ,  $X_{i,2}$ ,  $v_1$ ,  $v_2$ , les hasards ont pour expression :*

$$\begin{aligned} \lambda_{i,1}(t | X_{i,1}, v_1) &= v_1 \lambda_{0,1}(t) \lambda_{1,1}(X_{i,1}), \\ \lambda_{i,2}(t | X_{i,2}, v_2) &= v_2 \lambda_{0,2}(t) \lambda_{1,2}(X_{i,2}). \end{aligned}$$

On suppose comme conditions de régularité les hypothèses de H2 à H4. Notons  $T_1$  (respectivement  $T_2$ ) la variable aléatoire indiquant la durée de l'épisode 1 (respectivement 2). Pour l'unité  $i$ ,  $T_1$  et  $T_2$  sont indépendantes conditionnellement à  $X_{i,1}$ ,  $X_{i,2}$ ,  $v_1$  et  $v_2$  si et seulement si  $v_1$  et  $v_2$  sont indépendants. La densité jointe de  $T_1$  et  $T_2$  est une généralisation de (1.101) et s'écrit, en omettant les indices  $i$  :

$$f(t_1, t_2 | X_1, X_2) = \int_{\mathcal{V}} \int_{\mathcal{V}} f(t_1 | X_1, v_1) f(t_2 | X_2, v_2) dH(v_1, v_2). \quad (1.102)$$

La fonction de survie jointe est :

$$S(t_1, t_2 | X_1, X_2) = \int_{\mathcal{V}} \int_{\mathcal{V}} S(t_1 | X_1, v_1) S(t_2 | X_2, v_2) dH(v_1, v_2). \quad (1.103)$$

Le MPH multivarié est très souple et permet d'introduire des structure d'hétérogénéité non observée complexes entre différents niveaux, hiérarchiques ou non. L'hétérogénéité n'est plus nécessairement fixe et peut évoluer au fur et à mesure des épisodes d'un même individu.

Nous voyons dans la Section suivante que l'identification des modèles à épisodes uniques, épisodes multiples et du modèle MPH multivarié ne requiert pas les même hypothèses.

## 1.4. Identification

Comme le note Lancaster (1990, p. 146), la question de l'identification du modèle MPH peut se formuler de la manière suivante : disposant des données et donc de la fonction de survie conditionnelle aux variables explicatives, sous quelles hypothèses peut-on en déduire une unique expression du hasard et de la loi du terme d'hétérogénéité ? En d'autres termes, sous quelles conditions la connaissance de  $S(t|X)$  telle qu'elle est définie dans l'équation (1.76) suffit-elle à déduire sans ambiguïté  $\lambda_0$ ,  $\lambda_1$  et  $H$  ?

Commençons par un exemple, tiré de Lancaster (1990, p. 146), montrant que cette question n'est pas triviale. Considérons un échantillon dont la fonction de survie est :

$$S(t) = (1 + \beta t)^{-\alpha}, \quad (1.104)$$

où  $\alpha$  et  $\beta$  sont positifs. On peut trouver au moins deux couples  $(\lambda, H)$  donnant l'expression précédente de  $S(t)$ .

Premièrement, si  $\int_0^t \lambda(u|v) du = v\alpha \ln(1 + \beta t)$  et  $H(v)$  est une densité dégénérée dont toute la masse de probabilité se trouve au point  $v = 1$ , on a :

$$\begin{aligned} S(t) &= \int_{\mathcal{V}} \exp[-u\alpha \ln(1 + \beta t)] dH(u) \\ &= (1 + \beta t)^{-\alpha}. \end{aligned} \quad (1.105)$$

Supposons maintenant que  $\int_0^t \lambda(u|w)du = w\alpha\beta t$  où  $v$  suit une loi  $\gamma(\alpha, \alpha)$ .  
On a :

$$\begin{aligned}
 S(t) &= \int_{\mathcal{W}} \exp(-u\alpha\beta t) \frac{\alpha^\alpha}{\Gamma(\alpha)} u^{\alpha-1} \exp(-\alpha u) du \\
 &= \frac{\alpha^\alpha}{\Gamma(\alpha)} \int_{\Xi} \exp[-u(\alpha + \alpha\beta t)] u^{\alpha-1} du \\
 &= \frac{\alpha^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{\alpha^\alpha (1 + \beta t)^\alpha} \\
 &= (1 + \beta t)^{-\alpha}.
 \end{aligned} \tag{1.106}$$

Sans faire d'hypothèses supplémentaires, les données ne nous permettent donc pas de discriminer entre  $\lambda(t|v)$  et  $\lambda(t|w)$ .

Nous présentons dans cette section les résultats relatifs à l'identification du modèle MPH en présence d'épisodes uniques, d'épisodes multiples et les résultats concernant le modèle MPH multivarié.

### 1.4.1 Identification en présence d'épisodes uniques

Nous présentons dans la sous-section suivante les résultats généraux portant sur l'identification du modèle MPH. Nous verrons ensuite dans quelle mesure l'omission de variables explicatives ou la prise en compte de régresseurs variant dans le temps modifient ces conclusions.

#### Présentation des hypothèses

Une littérature importante existe sur l'identification *non-paramétrique* du modèle MPH, c'est-à-dire l'identification sans spécification de formes fonctionnelles pour  $\lambda_0$ ,  $\lambda_1$  et  $H$ . Elle a été initiée par Elbers et Ridder (1982) et Heckman et Singer (1984b), avant de connaître de nombreux développements.

Elbers et Ridder (1982) sont les premiers à prouver l'identification non-paramétrique du modèle MPH. Ils se basent sur des hypothèses que Van den Berg (2001) reformule de la manière suivante :

**Hypothèse 5**  $V$  est indépendante de  $X_i$ .

**Hypothèse 6** L'ensemble  $\mathcal{X}$  contient aux moins deux valeurs différentes, et  $\lambda_1(X_i)$  n'est pas constante sur  $\mathcal{X}$ .

**Hypothèse 7** Pour un couple donné  $(t_0, X_0)$ , on a  $\int_0^{t_0} \lambda_0(u)du = 1$  et  $\lambda_1(X_0) = 1$ .



**Hypothèse 8**  $E(V) < \infty$ .

La cinquième hypothèse est relativement forte, car elle suppose que les déterminants inobservés des transitions ne sont pas liés aux déterminants observés. L'hypothèse H6 signifie qu'il nous faut disposer d'au moins deux valeurs différentes de  $X_i$  pour assurer l'identification, donc qu'un modèle ne comprenant qu'une constante, comme dans l'exemple précédent, n'est pas identifié. L'hypothèse H7 est une simple normalisation des fonctions  $\lambda_0$  et  $\lambda_1$ . La dernière hypothèse suppose que  $V$  admet une espérance finie, hypothèse elle aussi relativement lourde dans la mesure où des modèles où  $E(V) = \infty$ , comme le cas où  $V$  suit une loi positive stable (Hougaard, 1984), sont tout à fait plausibles.

La démonstration de Elbers et Ridder (1982) est présentée dans l'annexe A.3. Elle n'est pas constructive dans la mesure où elle n'exprime pas le hasard de base et la distribution mélangeante en fonction de quantités observées. Une preuve constructive, suggérant donc un estimateur, est proposée dans Hausman et Woutersen (2004).

Heckman et Singer (1984b) proposent une alternative à l'hypothèse H8. Au lieu de considérer une loi  $H$  d'espérance finie, ils obtiennent l'identification du modèle MPH sous l'hypothèse :

**Hypothèse 9** Soit  $g$  la densité d'une variable aléatoire  $V$  continue telle que :

$$\lim_{v \rightarrow \infty} v^{1+\epsilon} g(v) = S(v), \quad (1.107)$$

où  $\epsilon \in ]0, 1[$  et  $S(v)$  est telle que :

$$\lim_{u \rightarrow \infty} \frac{S(uv)}{S(u)} = 1, \forall v > 0. \quad (1.108)$$

Heckman et Singer (1984b) montrent que sous cette hypothèse sur l'épaisseur de la queue de distribution,  $E(V) = \infty$  pour une valeur quelconque de  $\epsilon \in ]0, 1[$ . Ils proposent également une reformulation de cette hypothèse dans le cas où  $V$  est discrète, qui entraîne que  $V$  n'a besoin d'avoir aucun moment. Cette hypothèse permet donc d'étudier les cas de figure qui ne sont pas caractérisés par  $E(V) < \infty$ , c'est-à-dire les cas où la masse de probabilité des grandes valeurs de  $V$  est trop élevée pour que l'espérance soit finie. Van den Berg (2001) souligne qu'elle élimine la possibilité que  $V$  suive une loi dégénérée, et donc que le cas particulier d'un modèle MPH se réduisant à un modèle PH par ce moyen n'est pas considéré dans Heckman et Singer

(1984b). Par contre, leur spécification inclut le cas d'une loi positive stable, entraînant un hasard  $\lambda(t|X)$  de type PH.<sup>26</sup>

L'identification non-paramétrique d'une classe générale de modèles englobant le cas MPH est présentée dans Heckman (1991) et Honoré (1993), ce dernier s'intéressant plus particulièrement à une hétérogénéité commune à plusieurs unités de niveau 1. Ils considèrent la fonction de hasard :

$$\lambda_i(t|X_i, v) = v\lambda_0(t|X_i). \quad (1.109)$$

Comme on peut spécifier la fonction  $\lambda_0(t|X_i) = \lambda_0(t)\lambda_1(X_i)$ , le modèle MPH est donc un cas particulier de celui caractérisé par (1.109). Heckman (1991) montre que sous des hypothèses similaires à celles de Elbers et Ridder (1982), le modèle est identifié. Une hypothèse importante est que l'hétérogénéité non observée soit séparable de l'hétérogénéité observée : le résultat ne tient plus si le hasard ne peut pas être exprimé comme le produit d'une fonction de  $V$  avec une fonction de  $X_i$ .

### Identification sans variables explicatives et identification avec des régresseurs variant dans le temps

L'hypothèse H6 indique que les variables explicatives sont nécessaires à l'identification non-paramétrique du modèle. Heckman et Taber (1994) présentent un aperçu de l'identification dans les modèles paramétriques sans que  $X_i$  soit pris en compte. Ainsi, la famille des modèles de hasards de Box-Cox, dont les modèles de Weibull et de Gompertz sont des cas particuliers, est identifiée sans les hypothèses H2 et H6 concernant les variables explicatives.<sup>27</sup> La démonstration de ce résultat se trouve dans Heckman et Singer (1984b) et est présentée dans l'Annexe A.4. Ce résultat peut éventuellement être obtenu avec d'autres spécifications, mais Heckman et Singer (1984b) se limitent à l'étude de la forme fonctionnelle du hasard de Box-Cox.

La présence de régresseurs variant dans le temps facilite l'identification. Pour prendre explicitement en compte l'évolution de  $X_i$ , les hypothèses H5 et H6 doivent être reformulées :

**Hypothèse 10** *Le processus aléatoire  $\{X_i(u)\}_0^t$  est indépendant de  $V$ .*

**Hypothèse 11**  *$\lambda_1(X_i(s)) \neq \lambda_1(X(u))$  pour au moins deux dates  $s$  et  $u$ .*

---

<sup>26</sup>La loi positive stable est présentée dans Lancaster (1990, p. 80-81). Elle admet comme transformée de Laplace  $\mathcal{L}(s) = \exp(-s^\alpha)$ , où  $\alpha \in ]0, 1]$ . Pour  $\alpha = 1$ , la distribution est dégénérée et pour  $\alpha < 1$ ,  $E(X) = \infty$ .

<sup>27</sup>Heckman et Singer (1984b) définissent le modèle de hasard de Box-Cox par  $\lambda(t|X) = \lambda_1(X) \exp(\gamma t^{(\alpha)})$  où  $\alpha \geq 0$  et  $t^{(\alpha)} = (t^\alpha - 1)/\alpha$  dénote la transformation de Box-Cox.

Honoré (1990) considère un modèle MPH comprenant à la fois des variables explicatives ne dépendant pas du temps, et des variables dépendant de  $t$  satisfaisant les deux dernières hypothèses ainsi que celles présentées dans la section 1.2.2. Le modèle est alors identifié sans qu'il soit nécessaire d'imposer des restrictions sur la loi de  $V$ , comme  $E(V)$  finie ou bien l'alternative proposée par Heckman et Singer (1984b).

Les conclusions précédentes laissent penser qu'il est possible de relâcher l'hypothèse de finitude de  $E(V)$  moyennant une forme spécifique (au moins séparable) du hasard ou bien une structure particulière des régresseurs. Lever l'hypothèse arbitraire  $E(V) < \infty$  demande donc soit de faire d'autres hypothèses tout aussi arbitraires sur la loi des durées, soit de pouvoir extraire plus d'information des données. Nous verrons ultérieurement que ce constat relativement pessimiste est atténué lorsqu'on dispose de plusieurs observations pour un individu, donc lorsqu'on dispose de plus d'information.

Jusqu'ici, nous avons considéré l'identification non-paramétrique du modèle MPH. Considérant un modèle semi-paramétrique, où les formes fonctionnelles de  $\lambda_0$  et de  $\lambda_1$  sont connues, Hahn (1994) montre que la matrice d'information du modèle de Weibull à épisode unique est singulière.<sup>28</sup> En utilisant le second théorème de Chamberlain (1986), il en déduit qu'il ne peut exister une suite d'estimateurs réguliers pour les paramètres du hasard de base, et donc que ceux ci ne peuvent pas être estimés à la vitesse  $\sqrt{n}$ .<sup>29</sup> Ishwaran (1996) arrive à un résultat similaire sans utiliser le théorème de Chamberlain. Bien que le modèle de Weibull, comme tout modèle MPH, soit identifié à partir d'une condition sur le moment d'ordre 1, Ishwaran (1996) montre qu'il ne peut être estimé qu'à la vitesse  $\ln n$  sous cette hypothèse. La vitesse de convergence ne tend vers  $\sqrt{n}$  qu'en imposant des contraintes sur tous les moments jusqu'à ceux d'ordre  $\infty$ , et même ainsi elle lui reste inférieure.

Ridder et Woutersen (2003) posent une hypothèse sur le hasard de base du modèle MPH assurant l'inversibilité de la matrice d'information. Supposons :

**Hypothèse 12**  $\lim_{t \rightarrow 0} \lambda_0(t) = \lambda_0(0) \in ]0, \infty[.$

Sous des hypothèses similaires à H2-H7 et avec H12, Ridder et Woutersen (2003) montrent l'identification semi-paramétrique du modèle MPH (leur dé-

---

<sup>28</sup>En présence d'épisodes multiples, Hahn (1994) montre que la matrice d'information n'est pas singulière.

<sup>29</sup>Chamberlain (1986) définit un estimateur régulier de la manière suivante : soit un modèle de paramètres  $\theta$  et  $\delta$  et soient  $\theta_n = \theta_0 + \alpha_1 \sqrt{n}$  et  $\delta_n = \delta_0 + \alpha_2 \sqrt{n}$ . La suite d'estimateurs  $\{\hat{\theta}_{n,1}\}$  est régulière pour  $\theta_0$  si la distribution de  $\sqrt{n}(\hat{\theta}_{n,1} - \theta_{n,1})$  converge faiblement vers une distribution unique pour toutes valeurs réelles de  $\alpha_1$  et  $\alpha_2$ .

monstration se trouve dans l'annexe A.6). Grâce à H12, il n'est plus nécessaire de supposer  $E(V) < \infty$  et il est possible d'estimer le modèle à la vitesse  $\sqrt{n}$ . Supposer H12 revient à exclure les modèles dont le hasard de base intégré est stable par l'opération puissance, c'est-à-dire les modèles satisfaisant la relation suivante : si  $\Lambda(t) \in \mathcal{A}$  alors  $\Lambda^\alpha(t) \in \mathcal{A}$  pour tout  $\alpha > 0$ , où  $\mathcal{A}$  est l'ensemble des  $\Lambda(t)$  possibles pour le modèle considéré. Le modèle de Weibull est stable par l'opération puissance, le modèle de Gompertz et le hasard normal ne le sont pas. Sous certaines conditions, le hasard log-normal ou le hasard de base constant par morceaux satisfont H12. Un avantage de l'hypothèse H12 est qu'elle peut être facilement testée, par exemple à partir de l'estimateur de Breslow décrit en 1.2.7.<sup>30</sup>

### Le cas particulier des régresseurs endogènes variant dans le temps

Le modèle MPH est également identifié en présence de régresseurs endogènes variant dans le temps. Plus précisément, Abbring et Van den Berg (2003b) s'intéressent à la causalité reliant un événement quelconque à la fin de l'épisode en cours. Ils indiquent dans leur article qu'il peut s'agir, entre autres, du lien entre une diminution des allocations chômage ou la fin d'une formation avec le temps passé sans emploi. Nous sommes également en présence de régresseurs endogènes variant dans le temps lorsqu'on tient compte de l'influence d'une réforme fiscale sur l'emploi, ou plus généralement de l'impact d'une politique économique. Comme les biomètres considèrent souvent l'influence d'un traitement sur la longueur d'une maladie, l'événement dont on s'attend à ce qu'il influence l'épisode en cours est par analogie appelé le "traitement".

Nous présentons ici la forme générale d'un modèle de traitement que l'on peut trouver, entre autres, dans Abbring et Van den Berg (2003b). Notons  $s$  ( $s \in \mathbb{R} \cup \{\infty\}$ ) la date d'assignation du traitement (et par extension le traitement lui-même) et  $t^*(s)$  la durée passée par un individu dans un état donné lorsqu'il a reçu le traitement  $s$ . L'assignation des traitements est déterminée par la variable aléatoire continue  $S$ , et on note  $t(s)$  la durée observée.<sup>31</sup> On suppose  $S$  indépendante de l'ensemble des  $t^*$  possibles, c'est-à-dire que la date d'assignation d'un traitement n'est pas liée à ses conséquences sur les durées.

L'hypothèse d'absence d'anticipation se traduit par :

**Hypothèse 13** *Pour tout  $s, y \in \mathbb{R} \cup \{\infty\}$ ,  $\int_0^t \lambda^s(u) du = \int_0^t \lambda^y(u) du, \forall t \leq \min(s, t)$ .*

---

<sup>30</sup>Tester  $\Lambda_0 < \infty$  revient à tester  $1/\Lambda_0 > 0$ .

<sup>31</sup>Les traitements sont supposés mutuellement exclusifs, un individu ne peut recevoir qu'un seul traitement durant la période d'observation.

Elle assure que toutes les durées ont la même loi tant que le premier traitement n'a pas été assigné. Qualifier cette hypothèse d'absence d'anticipation est réducteur, dans la mesure où elle permet aux agents de faire des anticipations tant qu'elles ne modifient pas le choix du traitement qui leur sera affecté.

Soit  $Q_S(t, s) = P[T(S) > t, S > s, T(S) > S], \forall (t, s) \in \mathbb{R}_+^2$ , c'est-à-dire la probabilité que la durée observée soit supérieure à  $t$  et que la date d'assignation du traitement soit supérieure à  $s$  et que l'individu ait été traité avant d'effectuer sa transition. Soit  $Q_T(t) = P[T(S) > t, T(S) < S], \forall (t, s) \in \mathbb{R}_+^2$  la probabilité que la durée observée soit supérieure à  $t$  et que l'épisode s'achève avant l'application du traitement. On note  $Q_S^0(s)$  la probabilité jointe que le traitement soit assigné après une date  $s$  quelconque et que l'individu soit traité durant la période d'enquête, c'est-à-dire  $Q_S^0(s) = Q_S(-\infty, s)$ . Le couple  $(Q_S^0, Q_T)$  caractérise un modèle à risques concurrents, emboîté dans le modèle plus général caractérisé par  $(Q_S, Q_T)$ . Abbring et Van den Berg (2003b) spécifient les hasards suivants :

$$\lambda_T(t) = \begin{cases} v_T \lambda_0^T(t) \lambda_1^T(X) & \text{si } t \leq S \\ v_T \delta(t|s, X) \lambda_0^T(t) \lambda_1^T(X) & \text{si } t > S \end{cases}, \quad (1.110)$$

où  $\lambda_T(t)$  désigne le hasard que l'épisode prenne fin en  $t$  et où  $\delta(t|s, X)$  est une fonction à valeur dans  $\mathbb{R}$ , introduisant formellement la dépendance entre  $T$  et  $S$ , conditionnellement à  $X$  et  $v$ . La probabilité instantanée qu'un traitement soit assigné en  $t$  est définie par :

$$\lambda_S(t) = v_S \lambda_0^S(t) \lambda_1^S(X). \quad (1.111)$$

Le problème de l'identification se divise en deux parties. Dans un premier temps, Abbring et Van den Berg (2003b) montrent que le modèle, hormis  $\int_s^t \delta(u|s, X) du$ , est identifié à partir de  $(Q_S^0, Q_T)$ . En effet, il s'agit alors d'un modèle à risques concurrents dont les risques d'un même individu partagent le même  $v$ . Cette partie de la démonstration sera d'ailleurs réutilisée dans le cas du modèle MMPH. Ensuite, Abbring et Van den Berg (2003b) prouvent que le modèle complet est identifié à partir de  $(Q_S, Q_T)$ .

En résumé, l'identification non-paramétrique du modèle MPH modifié pour prendre en compte le traitement est obtenue sous des hypothèses semblables à celles de Elbers et Ridder (1982). Il faut faire l'hypothèse H13, c'est-à-dire que soit les agents n'ont pas accès à d'autres informations sur le traitement que celles contenues dans son historique, soit ils n'en tiennent pas compte. En effet, si les agents font des anticipations sur le mécanisme d'allocation du traitement, alors la perception d'une modification de ce mécanisme peut influencer leur comportement, c'est-à-dire la fonction de hasard, sans

qu'aucun traitement n'ait été alloué. Cette hypothèse peut cependant être problématique, par exemple dans le cas de l'étude d'une réforme fiscale, où elle revient à supposer que les agents n'anticipent pas la réforme. De plus, on considère que tous les éléments influençant les durées sont pris en compte dans les variables observées et inobservées.

L'identification, dont la démonstration se trouve dans l'annexe A.5, repose ici sur la séparabilité du modèle en deux fonctions, l'une dépendant de  $t$  et l'autre de  $X_i$ . Cette hypothèse peut être relâchée en présence d'épisodes multiples, c'est-à-dire lorsque plusieurs observations sont caractérisées par une même valeur  $v$ . Disposer de plusieurs observations partageant le même  $v$  permet d'obtenir l'identification sous des conditions moins restrictives. Nous allons voir maintenant comment obtenir l'identification dans le modèle MPH à épisodes multiples.

### 1.4.2 Identification en présence d'épisodes multiples

Van den Berg (2001) met en avant deux approches pour l'identification non-paramétrique d'un modèle MPH en présence d'épisodes multiples. La première s'intéresse à l'identification complète du modèle sans variable explicative. La seconde prend en compte les variables explicatives et traite de l'identification de la partie du hasard qui en dépend.

#### Identification sans variable explicative

Honoré (1993) étend les résultats obtenus par Heckman (1991) au contexte d'épisodes multiples. Il étudie l'identification d'un modèle dont la fonction de hasard est :

$$\lambda_i(t|X_i, v) = v\lambda_0(t|X_i), \quad (1.112)$$

où  $v|X_i \sim g(v|X_i)$ . Son approche est conditionnelle à une valeur donnée de  $X_i$ . La fonction  $\lambda_0$  peut dépendre de  $X_i$  de manière non-spécifiée et Honoré (1993) autorise  $\lambda_0$  à varier d'un épisode à l'autre. Il considère donc un modèle MMPH, plus général que celui des épisodes multiples, mais ses résultats peuvent s'adapter directement au contexte qui nous intéresse ici. Le terme d'hétérogénéité peut ne pas être indépendant des variables explicatives.

L'identification non-paramétrique est obtenue sous les hypothèses H2 à H4. Ainsi, il n'est plus nécessaire de supposer que  $E(V)$  est finie (ni même la reformulation proposée par Heckman et Singer (1984b) et notée ici par H9) dès lors qu'au moins deux observations sont disponibles pour une même valeur  $v$ . On peut également se passer de l'hypothèse d'indépendance entre les variables explicatives et le terme d'hétérogénéité. La démonstration se trouve dans l'annexe A.7.

L'identification d'un modèle MPH à épisodes multiples prenant en compte la durée de l'épisode précédent est établie dans Frijters (2002) sous les mêmes conditions. La fonction de hasard pour de l'individu  $i$  à l'épisode  $m$  est :

$$\lambda_{i,m}(t|X_i, v) = v h(t_{m-1}) \lambda_0(t|X_i), \quad (1.113)$$

où  $h$  est une fonction positive captant l'influence de la longueur de l'épisode antérieur. Ce résultat ne requiert pas non plus  $E(V) < \infty$ .

Considérons maintenant un modèle à plusieurs effets aléatoires. Comme nous sommes dans le contexte d'épisodes multiples, les réalisations des termes d'hétérogénéité sont identiques pour toutes les observations d'une unité de niveau 2. Supposons qu'on observe deux épisodes par individu  $i$ , on a alors les fonctions de hasard suivantes :

$$\lambda_{i,1}(t|X_i, v, w) = v_i w_1 \lambda_0(t) \lambda_1(X_i), \quad (1.114)$$

$$\lambda_{i,2}(t|X_i, v, w) = v_i w_2 \lambda_0(t) \lambda_1(X_i), \quad (1.115)$$

où  $V$  et  $W$  sont deux termes d'hétérogénéité non observée. Comme le terme d'hétérogénéité non observée peut se réécrire sous la forme  $z_{ij} = v_i w_j, \forall j = 1, 2$ , on peut montrer très facilement en reprenant la démarche de Honoré (1993) que  $\lambda_0, \lambda_1$  et la loi de  $Z$  sont identifiées. Cependant, l'identification de la loi de  $Z$  n'assure pas l'identification des lois de  $V$  et  $W$  sans restrictions identifiantes supplémentaires. En effet, supposons que l'on puisse écrire  $v_i = \exp(u_i)$  et  $w_j = \exp(a_j)$  où  $u_i$  et  $a_j$  suivent chacun une loi normale. Si l'identification de  $z_{ij} = \exp(u_i + a_j)$  nous permet de dire que la somme  $u_i + a_j$  suit une loi normale, on ne pourra pas, sans contraintes supplémentaires, en obtenir une déconvolution unique.

Dans l'annexe C.1, je reprends la démarche de Honoré (1993) et montre que  $H_V$  et  $H_W$  sont bien identifiées, sans changer les hypothèses standard des épisodes multiples. Ce résultat, contre-intuitif de prime abord, repose sur la structure de l'hétérogénéité : elle peut être exprimée sous la forme d'un produit de deux effets et la réalisation de chaque terme d'hétérogénéité est commune à plusieurs épisodes. C'est donc la structure de l'hétérogénéité qui permet de se passer de restrictions identifiantes sur la distribution mélangeante. L'approche de Honoré (1993) se passe de variable explicative et l'identification est donc obtenue sans les hypothèses H2, H5 et H6.

### Identification en présence de variables explicatives

La seconde approche prend explicitement en considération les variables explicatives dans le hasard. Kalbfleisch et Prentice (1980) montrent l'identification non-paramétrique du modèle MPH sous les hypothèses H2-H4. Leur

## 1.4 Identification

---

approche est basée sur la vraisemblance partielle stratifiée. Nous rappelons qu'une strate est définie comme l'ensemble des individus partageant la même réalisation d'au moins un terme d'hétérogénéité non observée. L'idée de la vraisemblance partielle stratifiée est assez simple : au lieu de construire directement une vraisemblance partielle sur toute la population, on en définit une sur chaque strate avant de les combiner pour obtenir la vraisemblance de la population. Ils distinguent deux types de variable explicative : celles qui varient parmi les individus appartenant à une même strate et celles qui ne varient que d'une strate à l'autre. Notons les respectivement par  $X_i$  et par  $X_i^*$ . Leur approche s'applique à une famille englobant le modèle MPH. Supposons que le hasard a pour expression :

$$\lambda(t|X_i^*, X_i, v) = \lambda_0(t|X_i^*, v)\lambda_1(X_i), \quad (1.116)$$

où  $v|X_i^*, X_i \sim g(v|X_i^*, X_i)$ . Il convient de remarquer que  $v$  n'intervient plus nécessairement de manière multiplicative dans le hasard, et qu'il est, par construction, commun à tous les individus de la strate. Comme précédemment, on peut supposer que  $V$  n'est pas indépendant de  $X_i^*$  et de  $X_i$ . Considérons les observations ordonnées  $t_{(1)}, \dots, t_{(n_k)}$  de la strate  $k$ . La contribution à la vraisemblance partielle de la strate  $k$  est :

$$L_p^k = \prod_{(i)=1}^{n_k} \frac{\lambda_1(X_{(i)})}{\sum_{l \in R(t_{(i)})} \lambda_1(X_l)}. \quad (1.117)$$

Comme les éléments qui ne varient pas au sein de la strate  $k$  se simplifient,  $L_p^k$  ne dépend que de  $\lambda_1$ . La vraisemblance partielle  $L_p = \prod_{k=1}^K L_p^k$  ne permet donc d'estimer que  $\lambda_1$ . La fonction  $\lambda_0$  n'est pas identifiée avec cette approche, mais elle peut toutefois être estimée de manière non-paramétrique, par exemple avec l'estimateur de Breslow défini sur chaque strate. Remarquons qu'il faut au minimum deux épisodes par strate, sinon  $L_p^k = 1$  et  $\lambda_1$  n'est plus identifiée. Ainsi, l'identification de  $\lambda_1$  n'est assurée sans l'hypothèse de finitude de  $E(V)$  et d'indépendance entre  $V$  et les variables explicatives observées que si on dispose d'au moins deux épisodes par strate (chacun étant caractérisé par un  $X_i$ ).

En résumé, la première approche ne nécessite pas de variable explicative et conclut à l'identification complète du modèle. La seconde repose sur la vraisemblance partielle stratifiée et tient compte de régresseurs différents pour chaque épisode pour aboutir à l'identification de  $\lambda_1$ . Les conclusions sont les mêmes dans les deux cas : le modèle MPH est identifié sans l'hypothèse que le terme d'hétérogénéité non observé soit indépendant des variables explicatives et que son espérance soit finie dès qu'on dispose de deux observations pour une réalisation de  $v$ .



### 1.4.3 Identification du modèle MPH multivarié

Nous nous intéressons toujours à l'identification non-paramétrique. Nous verrons dans un premier temps les résultats généraux pour le modèle MMPH, et détaillerons le cas où le terme d'hétérogénéité non observée admet la même réalisation pour tous les épisodes d'un même individu. Des résultats concernant un modèle de type "competing risks" où le hasard est de type MPH, ce qui inclut le modèle MMPH comme cas particulier, sont également présentés.

#### Identification du modèle MMPH

Honoré (1993) montre que le modèle MMPH de la Définition 2, page 65, où  $v_{i,1}, v_{i,2}, \lambda_{0,1}, \lambda_{0,2}, \lambda_{1,1}$  et  $\lambda_{1,2}$  diffèrent d'un épisode à l'autre est identifié sous les 8 hypothèses de Elbers et Ridder (1982) (H8 pouvant être remplacée par H9). Ce résultat est intuitif : le modèle ainsi obtenu correspond à un modèle MPH à épisode unique où  $\lambda_0$  et  $\lambda_1$  varient selon les individus. La démonstration est donc la même que celle de Elbers et Ridder (1982), présentée dans l'annexe A.3. De même, la démonstration avec  $v_{i,1} = v_{i,2}$  et  $\lambda_{0,1}, \lambda_{0,2}, \lambda_{1,1}$  et  $\lambda_{1,2}$  tous différents, qui correspond à une extension des résultats en présence d'épisodes multiples, est dans l'annexe A.7.

#### L'approche des risques concurrents

Le modèle MMPH peut être vu comme un cas particulier de modèle de risques concurrents à hasard MPH, et les résultats d'identification de ce dernier s'appliquent donc ici. Nous voyons sous quelles conditions l'hypothèse  $E(V) < \infty$  n'est plus nécessaire.

Dans un modèle de risques concurrents, on considère plusieurs durées commençant à la même date pour un agent, lequel est observé jusqu'à l'échéance de la durée la plus courte. Le modèle MPH en présence de données censurées peut être vu comme un modèle de risques concurrents à deux risques, l'un correspondant à la durée qui nous intéresse et l'autre à la durée à l'issue de laquelle survient la censure. Plus généralement, on obtient un modèle MMPH si on suppose que le premier risque est caractérisé par  $\lambda_{i1}(t|X_i, v_{i,1}) = v_{i,1}\lambda_{0,1}(t)\lambda_{1,1}(X_i)$  et le second par  $\lambda_{i2}(t|X_i, v_{i,2}) = v_{i,2}\lambda_{0,2}(t)\lambda_{1,2}(X_i)$ . Le théorème de Cox-Tsiatis, initialement formulé dans Cox (1959) pour le cas de deux risques et étendu par Tsiatis (1975) dans le cas d'un nombre fini de risques, nous dit qu'un modèle de risques concurrents n'est pas identifié à partir de l'observation des durées et des états. En effet, un modèle dont les risques sont dépendants est observationnellement équivalent à un modèle de risques indépendants. L'identification n'est donc obtenue qu'en faisant des

hypothèses supplémentaires, généralement en supposant que les risques sont indépendants.<sup>32</sup>

Abbring et Van den Berg (2003a) montrent que le MPH à risques concurrents est identifié sous les hypothèses H2-H5 et H7-H8 de Elbers et Ridder (1982), en considérant une alternative plus restrictive à H6 :

**Hypothèse 14** *Les fonctions  $\lambda_{1,1}(X_i)$  et  $\lambda_{1,2}(X_i)$  prennent toutes les valeurs possibles dans un sous-ensemble non vide de  $(0, \infty)^2$  lorsque  $X_i$  varie dans  $\mathcal{X}$ .*

La démonstration a déjà été utilisée pour montrer l'identification du modèle à régresseurs endogènes et se trouve dans l'annexe A.5.1. Comme le souligne Van den Berg (2001), une condition suffisante à cela lorsque  $\lambda_{1,k}(X_i) = \exp(X_i\beta_k)$ , est que  $X_i$  contiennent deux variables explicatives continues non-colinéaires dont les coefficients diffèrent selon le risque  $k$  considéré. Heckman et Honoré (1989) considèrent un modèle plus général et, moyennant une version plus lourde de H14 (ils supposent que  $\lambda_{1,1}(X_i)$  et  $\lambda_{1,2}(X_i)$  prennent toutes les valeurs dans  $(0, \infty)^2$  et non plus dans un sous-ensemble), ils obtiennent l'identification de leur modèle. En contrepartie, ils se passent de l'hypothèse  $E(V) < \infty$ . Des hypothèses restrictives sur les variations de la partie déterministe du hasard sont nécessaires à l'identification des modèles de risques concurrents, et marquent les limites actuelles de cette approche pour l'identification du modèle MMPH.

En résumé, le modèle MMPH est identifié sous les 8 hypothèses de Elbers et Ridder (1982). Pour pouvoir relâcher l'hypothèse de finitude de  $E(V)$ , il faut utiliser la relation d'emboîtement entre le modèle MMPH et le MPH à risques concurrents et spécifier l'hypothèse que  $\lambda_{1,1}(X_i)$  et  $\lambda_{1,2}(X_i)$  sont à valeurs dans  $(0, \infty)^2$ . Relâcher l'hypothèse arbitraire de finitude de l'espérance de  $V$  demande donc de faire une hypothèse tout aussi arbitraire sur la manière dont varient les régresseurs.

Comme nous l'avons vu, le modèle MPH est identifié sous les hypothèses H2-H8, cette dernière pouvant être remplacée par H9. S'il contient des régresseurs variant dans le temps, il faut également supposer H10-H11 (cette dernière n'est qu'une adaptation de H6), et H13 s'ils sont endogènes. L'hypothèse H12 est destinée au modèle de Weibull et vise à concilier l'identification avec une vitesse de convergence de  $\sqrt{n}$ . Quant à H14, elle ne concerne que le modèle à risques concurrents et hasard MPH. Les hypothèses H5 et H8, c'est-à-dire d'indépendance entre  $V$  et  $X_i$  et de finitude de  $E(V)$ , fonctionnent de pair et ne sont pas toujours nécessaires à l'identification. Pour récapituler, ces deux hypothèses sont nécessaires dans les cas :

---

<sup>32</sup>Le théorème de Cox-Tsiatis est une des raisons expliquant pourquoi on suppose les durées indépendantes du mécanisme de censure.

- à épisode unique (avec ou sans variables explicatives) ;
- multivarié lorsque tous les éléments du hasard varient d'un épisode à l'autre.

Ces hypothèses ne sont pas nécessaires dans les cas :

- à épisode unique lorsque l'on dispose de régresseurs variant dans le temps ;
- à épisodes multiples, avec ou sans variables explicatives ;
- lorsque l'on dispose de deux termes d'hétérogénéité dont chaque réalisation est commune à aux moins deux épisodes.

Il convient de noter que certains choix effectués lors de la spécification, même s'ils visent uniquement à assurer l'identification, influencent les résultats. Prenons l'exemple d'un modèle à épisodes uniques et sans variables explicatives, qui requiert donc le choix d'une forme fonctionnelle pour la hasard de base afin d'être identifié. Yashin *et al.* (2001) estiment deux modèles à hétérogénéité gamma, où le hasard de base vaut  $\lambda_0(t) = a \exp(bt)$  dans le premier modèle et  $\lambda_0(t) = \frac{\exp t}{1+\exp t}$  dans le second. Les variances estimées des distributions mélangantes sont respectivement de 0.5 et de 0.

## 1.5. Inférence basée sur la vraisemblance

Maintenant que nous savons sous quelles hypothèses le modèle MPH est identifié, nous pouvons nous intéresser à l'inférence. Le principe général est de faire des généralisations sur les caractéristiques de la population à partir d'un échantillon. Savoir ce qu'est l'inférence ne signifie pas que l'on sache comment inférer, et comme le soulignent Mittelhammer, Judge, et Miller (2000, p.22), une question centrale est : comment peut-on extraire l'information contenue dans les données pour en déduire les valeurs des paramètres du modèle ?

Plusieurs réponses ont été apportées à cette question. Nous allons dans un premier temps présenter les méthodes basées sur la vraisemblance complète. Les approches développées à partir de la vraisemblance partielle et Bayésiennes sont décrites dans des sous-sections ultérieures. Les approches reposant sur la vraisemblance complète ont été historiquement les premières utilisées. Du fait du résultat de Breslow (1972) établissant que l'estimateur des coefficients est identique selon qu'on maximise une vraisemblance partielle ou bien une vraisemblance complète concentrée par rapport au hasard de base, de nombreuses approches semi-paramétriques utilisent la vraisemblance complète comme point de départ. L'argument de Breslow (1972) est ensuite utilisé pour fournir un estimateur du hasard de base et établir que la méthode est semi-paramétrique.

Nous présenterons dans un premier temps l'expression de la vraisem-

blance, en utilisant les éléments d'algèbre des mélanges présentés dans la sous-section 1.3.1. Nous verrons ensuite, d'un point de vue théorique et empirique, comment utiliser l'algorithme EM pour obtenir l'estimateur du maximum de vraisemblance avant de présenter son extension aux épisodes multiples. Les limites cette procédure numérique seront ensuite discutées.

### 1.5.1 Estimateur du Maximum de Vraisemblance

Les approches que nous allons développer ici sont basées sur la vraisemblance complète. On rappelle qu'à chaque date  $t$ , l'agent peut soit effectuer une transition, soit ne plus être observé. Les probabilités de ces deux événements sont respectivement de  $\lambda(t)S(t)dt$  et de  $S(t)dt$ . On déduit l'expression de la densité conditionnelle à  $v$  de l'échantillon pour le modèle MPH dont il faut prendre l'espérance par rapport à la distribution mélangeante pour en déduire la densité non conditionnelle. Dans un modèle à épisode unique, la contribution à la vraisemblance de l'épisode  $i$  est :

$$L_i(\beta) = \int_{\mathcal{V}} [v\lambda_0(t_i)\lambda_1(X_i(t_i))]^{\delta_i} \exp\left(-\int_0^{t_i} v\lambda_0(u)\lambda_1(X_i(u))du\right) dH(v). \quad (1.118)$$

La vraisemblance a pour expression :

$$L(\beta) = \int_{\mathcal{V}} \cdots \int_{\mathcal{V}} \prod_{i=1}^n [v_i\lambda_0(t_i)\lambda_1(X_i(t_i))]^{\delta_i} \exp\left(-v_i \int_0^{t_i} \lambda_0(u)\lambda_1(X_i(u))du\right) dH(v_1, \dots, v_n). \quad (1.119)$$

Les procédures d'optimisations habituelles, telles que l'algorithme de Newton-Raphson, ne peuvent pas être utilisées car les équations de vraisemblance n'admettent pas de solution analytique. Il faut alors recourir à d'autres méthodes pour approcher  $\hat{\beta}$ , comme par exemple l'algorithme EM.

Il est possible d'écrire la log-vraisemblance de manière simple. Comme :

$$\begin{aligned}\ln \lambda(t_i|X_i) &= \int_{\mathcal{V}} \ln \lambda(t_i|X_i, v) dH(v) \\ &= \int_{\mathcal{V}} \ln [v \lambda_0(t_i) \lambda_1(X_i(t_i))] dH(v) \\ &= E_V(\ln v_i) + \ln [\lambda_0(t_i) \lambda_1(X_i(t_i))],\end{aligned}\tag{1.120}$$

$$\begin{aligned}\ln S(t_i) &= \int_{\mathcal{V}} \ln S(t_i|v) dH(v) \\ &= \int_{\mathcal{V}} -v \left[ \int_0^{t_i} \lambda_0(u) \lambda_1(X_i(u)) du \right] dH(v) \\ &= -E_V(v_i) \int_0^{t_i} \lambda_0(u) \lambda_1(X_i(u)) du,\end{aligned}\tag{1.121}$$

et la relation (1.37) devient en présence de mélange :

$$\begin{aligned}\ln L(\beta) &= \sum_{i=1}^n \left[ \delta_i E_V(\ln v_i) + \delta_i \ln [\lambda_0(t_i) \lambda_1(X_i(t_i))] \right. \\ &\quad \left. - E_V(v_i) \int_0^{t_i} \lambda_0(u) \lambda_1(X_i(u)) du \right].\end{aligned}\tag{1.122}$$

### 1.5.2 Algorithme EM et modèle MPH : théorie...

Nous présenterons l'approche générale dans un premier temps, puis nous verrons plus précisément son adaptation au contexte d'épisodes multiples.

On remarque en considérant l'expression (1.122) que le modèle MPH, tout comme le modèle de Cox, se prête bien à une estimation *via* l'algorithme EM. Son fonctionnement dans le cas général a été présenté dans la sous-section 1.2.5.

Si les  $v_i$  étaient observés, il ne serait plus nécessaire de les évacuer de la vraisemblance en calculant  $E_V[L_i(\beta|v_i)]$ . La vraisemblance d'un échantillon où les  $v_i$  sont observés est le produit de  $L_i(\beta|v_i)$  avec  $h(v_i)$  et la log-vraisemblance s'écrit :

$$\begin{aligned}\ln L_V(\beta) &= \sum_{i=1}^n \left[ \delta_i \ln [v_i \lambda_0(t_i) \lambda_1(X_i(t_i))] - \left( \int_0^{t_i} v_i \lambda_0(u) \lambda_1(X_i(u)) du \right) \right. \\ &\quad \left. + \ln h(v_i) \right].\end{aligned}\tag{1.123}$$

Comme le souligne Lancaster (1990, p.197), il est impossible de calculer  $\ln L_V(\beta)$  puisqu'on ne connaît pas les  $v_i$ , mais il est possible de l'estimer. La  $q$ -ème itération de l'algorithme EM se déroule de la manière suivante : on calcule l'espérance conditionnelle aux observations et à la valeur courante des paramètres de l'équation (1.123), qui est ensuite maximisée par rapport au paramètre  $\beta$  et à ceux de la loi de  $V$ . L'expression obtenue lors de l'étape E à l'itération  $q$  est :

$$Q(\beta, \beta^{(q)}) = \sum_{i=1}^n \left[ \delta_i E_{\beta^{(q)}}(\ln V|T = t, \delta) + \delta_i \ln [\lambda_0(t_i)\lambda_1(X_i(t_i))] - E_{\beta^{(q)}}(V|T = t, \delta) \int_0^{t_i} \lambda_0(u)\lambda_1(X_i(u))du + E_{\beta^{(q)}}(\ln h(V)|T = t, \delta) \right]. \quad (1.124)$$

La maximisation par rapport à  $\beta$  annule la quantité :

$$\frac{\partial Q(\beta, \beta^{(q)})}{\partial \beta} = \sum_{i=1}^n \left[ \delta_i \frac{\partial \ln \lambda_1(X_i(t_i))}{\partial \beta} - E_{\beta^{(q)}}(V|T = t, \delta) \frac{\partial}{\partial \beta} \left( \int_0^{t_i} \lambda_0(u)\lambda_1(X_i(u))du \right) \right] \quad (1.125)$$

Les autres conditions du premier ordre dépendent du choix de la distribution mélangeante. Les coefficients estimés sont utilisées pour évaluer l'espérance de (1.123) lors de la  $(q + 1)$ -ème itération.

### 1.5.3 ... et applications

De nombreuses études utilisent l'algorithme EM. Nous présenterons tout d'abord l'article de Guo et Rodriguez (1992), qui l'ont appliqué à un modèle MPH où  $V$  suit une loi gamma, puis celui de Sastry (1997) qui l'a étendu au cas de deux effets aléatoires emboîtés. Ces deux études ont pour objet la mortalité infantile, et le lecteur intéressé par la biométrie trouvera en Klein (1992) ou Clayton et Cuzick (1985) des alternatives à Guo et Rodriguez (1992).

#### Un effet aléatoire

Guo et Rodriguez (1992) étudient un échantillon stratifié. La population est divisée en groupes d'individus, et tous les individus appartenant au même groupe partagent la même réalisation du terme d'hétérogénéité. On indice par

$i$  ( $i = 1 \dots I$ ) les strates et  $j$  ( $j = 1 \dots J_i$ ) les individus. Lorsque  $V$  suit une loi  $\gamma(\eta, \eta)$ , Lancaster (1990, p. 198) montre que :

$$E_{\beta^{(q)}, \eta^{(q)}}(V|T = t, \delta) = \frac{\eta^{(q)} + \sum_{j=1}^{J_i} \delta_{ij}}{\eta^{(q)} + \sum_{j=1}^{J_i} \Lambda_{ij}^{(q)}}, \quad (1.126)$$

$$E_{\beta^{(q)}, \eta^{(q)}}(\ln V|T = t, \delta) = \psi \left( \eta^{(q)} + \sum_{j=1}^{J_i} \delta_{ij} \right) - \ln \left( \eta^{(q)} + \sum_{j=1}^{J_i} \Lambda_{ij}^{(q)} \right), \quad (1.127)$$

où  $\Lambda_{ij}^{(q)} = \int_0^{t_{ij}} \lambda_1(X_i(u)) \lambda_0(u) du$  et  $\psi$  est la fonction digamma.<sup>33</sup> A noter que le hasard intégré dépend de  $\beta^{(q)}$  et des estimations des paramètres du hasard de base à l'itération ( $q$ ). L'équation (1.124) devient :

$$\begin{aligned} Q(\beta, \eta, \beta^{(q)}, \eta^{(q)}) = \sum_{i=1}^I \left[ (\eta + \delta_{ij} - 1) \left( \psi \left( \eta^{(q)} + \sum_{j=1}^{J_i} \delta_{ij} \right) \right. \right. \\ \left. \left. - \ln \left( \eta^{(q)} + \sum_{j=1}^{J_i} \Lambda_{ij}^{(q)} \right) \right) - (\eta + \Lambda_{ij}) \frac{\eta^{(q)} + \sum_{j=1}^{J_i} \delta_{ij}}{\eta^{(q)} + \sum_{j=1}^{J_i} \Lambda_{ij}^{(q)}} \right. \\ \left. + \delta_{ij} \ln \lambda_0(t_{ij}) \lambda_1(X_{ij}(t_{ij})) + \eta \ln \eta - \ln \Gamma(\eta) \right]. \end{aligned} \quad (1.128)$$

Il est important de distinguer dans l'équation (1.128) les  $\eta$  des  $\eta^{(q)}$ .

L'étape M consiste à maximiser (1.128) par rapport à  $\beta$  et à  $\eta$ . Guo et Rodriguez (1992) effectuent l'optimisation par rapport à  $\eta$  au moyen d'une procédure de Newton-Raphson utilisant les résultats suivants :

$$\begin{aligned} \frac{\partial Q(\beta, \eta, \beta^{(q)}, \eta^{(q)})}{\partial \eta} = \psi \left( \eta^{(q)} + \sum_{j=1}^{J_i} \delta_{ij} \right) - \ln \left( \eta^{(q)} + \sum_{j=1}^{J_i} \Lambda_{ij}^{(q)} \right) \\ - \left( \frac{\eta^{(q)} + \sum_{j=1}^{J_i} \delta_{ij}}{\eta^{(q)} + \sum_{j=1}^{J_i} \Lambda_{ij}^{(q)}} \right) + 1 + \ln \eta - \psi(\eta). \end{aligned} \quad (1.129)$$

$$\frac{\partial^2 Q(\beta, \eta, \beta^{(q)}, \eta^{(q)})}{\partial \eta^2} = \frac{1}{\eta} - \frac{d \ln \psi(\eta)}{d \eta}. \quad (1.130)$$

La maximisation par rapport à  $\beta$  nous amène à utiliser un algorithme de Newton-Raphson basé sur les résultats présentés dans la sous-section 1.2.5, la seule différence étant qu'il faut prémultiplier  $\lambda_1$  par  $E_{\beta^{(q)}}(V|T = t, \delta)$ .

<sup>33</sup>La fonction digamma est  $\psi(\alpha) = d \ln \Gamma(\alpha) / d \alpha$ , où  $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp(-x) dx$ .

## Deux effets aléatoires

Sastry (1997) étend l'approche de Guo et Rodriguez (1992) au cas de deux effets aléatoires emboîtés. Leur application porte sur la mortalité infantile, et ils considèrent des communautés, subdivisées en familles, dans lesquelles vivent les enfants. Nous avons dès lors deux niveaux de stratification, emboîtés dans la mesure où chaque famille n'appartient qu'à une seule communauté. On indice par  $i$  ( $i = 1 \dots I$ ) les communautés, par  $j$  ( $j = 1 \dots J_i$ ) les familles et  $k$  ( $k = 1 \dots K_{ij}$ ) les individus. Le hasard admet comme expression :

$$\lambda_{ijk}(t) = v_i w_{ij} \lambda_0(t) \exp(X_{ijk}(t)\beta), \quad (1.131)$$

où  $v_i$  est l'effet communauté et  $w_{ij}$  l'effet famille. Tous deux sont distribués respectivement selon les lois  $\gamma(\alpha, \alpha)$  et  $\gamma(\eta, \eta)$ .

Le calcul des quantités  $E_{\beta^{(q)}, \alpha^{(q)}, \eta^{(q)}}(W|T = t, \delta)$ ,  $E_{\beta^{(q)}, \alpha^{(q)}, \eta^{(q)}}(\ln W|T = t, \delta)$ ,  $E_{\beta^{(q)}, \alpha^{(q)}, \eta^{(q)}}(VW|T = t, \delta)$ ,  $E_{\beta^{(q)}, \alpha^{(q)}, \eta^{(q)}}(V|T = t, \delta)$  et  $E_{\beta^{(q)}, \alpha^{(q)}, \eta^{(q)}}(\ln V|T = t, \delta)$ , nécessaires à l'étape E, est présenté dans l'annexe A.9.

D'où :

$$\begin{aligned} Q &= \sum_{i=1}^I [\alpha \ln \alpha + (\alpha - 1) E_{\beta^{(q)}, \alpha^{(q)}, \eta^{(q)}}(\ln V|T = t, \delta) \\ &\quad - \alpha E_{\beta^{(q)}, \alpha^{(q)}, \eta^{(q)}}(V|T = t, \delta) - \ln \Gamma(\alpha)] \\ &\quad + \sum_{i=1}^I \sum_{j=1}^{J_i} [\eta \ln \eta + (\eta - 1) E_{\beta^{(q)}, \alpha^{(q)}, \eta^{(q)}}(\ln W|T = t, \delta) \\ &\quad - \eta E_{\beta^{(q)}, \alpha^{(q)}, \eta^{(q)}}(W|T = t, \delta) - \ln \Gamma(\eta)] \\ &\quad + \sum_{i=1}^I \sum_{j=1}^{J_i} \sum_{k=1}^{K_{ij}} [\delta_{ijk} \ln \lambda_{ijk}(t) - E_{\beta^{(q)}, \alpha^{(q)}, \eta^{(q)}}(VW|T = t, \delta) \Lambda_{ijk}] \\ &= Q(\alpha, \alpha^{(q)}) + Q(\eta, \eta^{(q)}) + Q(\beta, \beta^{(q)}). \end{aligned} \quad (1.132)$$

L'étape M revient à maximiser (1.132) par rapport à  $\beta, \alpha$  et  $\eta$ , ces deux derniers intervenant dans les espérances. Le problème se divise en 3 parties. L'optimisation des deux premiers termes, chacun ne dépendant que d'une unique variable de contrôle, peut se faire au moyen d'un algorithme de Newton-Raphson. Le troisième terme doit être optimisé par rapport à  $\beta$  et aux paramètres de  $\lambda_0$ , que Sastry (1997) spécifie constant par morceaux.

L'algorithme EM peut également être combiné avec d'autres techniques lors de l'inférence. Maples *et al.* (2002) proposent de l'utiliser dans un premier temps pour se rapprocher du maximum, et d'atteindre ensuite ce dernier avec une procédure de Newton-Raphson qui a l'avantage de converger plus



rapidement. De même, les espérances de l'étape E peuvent être approchées en utilisant par exemple avec la quadrature de Gauss-Hermite lorsque  $V$  suit une loi normale (Friedl et Kauermann, 2000, Maples *et al.*, 2002; Vaida et Xu, 2000) ou encore en simulant les résultats de l'étape E avec des procédures de Monte Carlo (Ripatti et Palmgren, 2002).<sup>34</sup>

Bien qu'il soit possible de baser son raisonnement sur la vraisemblance complète et d'en déduire une approche semi-paramétrique du fait du résultat de Breslow (1972), une approche plus directe consiste à travailler directement avec la vraisemblance partielle. Nous voyons comment procéder dans la sous-section suivante.

## 1.6. Les approches basées sur la vraisemblance partielle

Les approches basées sur la vraisemblance partielle sont semi-paramétriques car elles n'imposent pas au hasard de base une forme déterminée par un nombre fini de paramètres. Le reste du modèle est spécifié de manière complètement paramétrique.

Nous décrivons ici les approches fondées sur la vraisemblance partielle dont nous indiquons la formulation en présence d'un hasard MPH. Nous présentons ensuite l'approche récente et prometteuse de la vraisemblance partielle pénalisée que nous réutilisons aux Chapitre 2.

### 1.6.1 Adaptation de la vraisemblance partielle

Nous avons vu dans la sous-section 1.2.4 que la contribution de l'individu  $i$  à la vraisemblance partielle s'interprète comme le rapport entre la probabilité instantanée qu'il effectue une transition en  $t_i$  et la probabilité instantanée qu'un des individus au risque effectue une transition en  $t_i$ . Cette contribution s'écrit :

$$L_p^i(\beta) = \int_{\mathcal{V}} \cdots \int_{\mathcal{V}} \frac{v_i \lambda_1(X_i)}{\sum_{k \in R(t_i)} v_k \lambda_1(X_k)} dH(v_1, \dots, v_n). \quad (1.133)$$

---

<sup>34</sup>Vaida et Xu (2000) n'utilisent la quadrature de Gauss-Hermite que pour des groupes de petite taille, l'échantillonnage de Gibbs, dont nous reparlerons ultérieurement, lui étant préféré autrement car demandant moins de temps de calcul.

L'estimateur du maximum de vraisemblance partielle est donc solution de la maximisation de :

$$L_p(\beta) = \prod_{i=1}^n \int_{\mathcal{V}} \dots \int_{\mathcal{V}} \frac{v_i \lambda_1(X_i)}{\sum_{k \in R(t_i)} v_k \lambda_1(X_k)} dH(v_1, \dots, v_n). \quad (1.134)$$

Comme pour la vraisemblance,  $\hat{\beta}_{PL}$  doit être approché par des procédures numériques.

### 1.6.2 La vraisemblance partielle pénalisée

Nous commencerons par présenter la vraisemblance pénalisée dans le cas général, avant de voir des applications plus spécifiques aux modèles de durées.

#### Présentation générale

Comme le fait remarquer Green (1999), baser les approches semi-paramétriques sur la vraisemblance est toujours délicat dans la mesure où un élément de dimension potentiellement infinie ne peut pas être identifié à partir d'un nombre fini d'observations. La vraisemblance pénalisée évite ce problème, et constitue un compromis entre les approches paramétriques et non-paramétriques. L'idée est de retrancher une fonction positive, appelée "pénalité", à la log-vraisemblance. La fonction objectif du programme de maximisation est composée de deux éléments, le premier mesurant l'adéquation entre le modèle et les données et la pénalité, qui à la fois restreint les valeurs possibles des quantités estimées et mesure le compromis effectué entre l'ajustement aux données et l'aspect lisse de la fonction objectif. Généralement, on choisit comme pénalité une fonction quadratique, mais cette spécification n'est en rien obligatoire et nous en verrons d'autres par la suite. L'approche de la vraisemblance pénalisée semble avoir été introduite par Good et Gaskins (1971) pour l'estimation non-paramétrique de densités de probabilité, tandis que Anderson et Senthilselvan (1980) semblent avoir été les premiers à l'utiliser dans les modèles de hasard.

Dans un modèle MPH, la vraisemblance partielle pénalisée s'écrit :

$$\ln L_{PPL}(v, \theta) = \ln L_{PL}(v) - g(v, \theta), \quad (1.135)$$

où  $\theta$  est un vecteur de paramètres,  $g(v, \theta)$  est une fonction positive quelconque et  $\ln L_{PL}$  est la vraisemblance partielle d'expression :

$$\ln L_{PL}(v) = \sum_{i=1}^I \sum_{j=1}^{J_i} \delta_{ij} \left[ \ln [v_i \lambda_1(X_{ij}(t_{ij}))] - \ln \left( \sum_{k \in R(t_{ij})} v_k \lambda_1(X_k(t_k)) \right) \right]. \quad (1.136)$$

Comme la pénalité ne dépend pas des paramètres intervenant dans  $L_{PL}$ , l'estimateur du maximum de la vraisemblance partielle pénalisée (abrégé estimateur PPL) est identique à l'estimateur du maximum de vraisemblance partielle (abrégé PL) pour les paramètres de  $\lambda_1$ . Seul l'estimateur des  $v_i$  est différent, et s'obtient à partir de :

$$\frac{\partial \ln L_{PPL}}{\partial v_i} = \frac{\partial \ln L_{PL}}{\partial v_i} - \frac{\partial g}{\partial v_i} = 0. \quad (1.137)$$

On note  $R^i(t_{ij})$  l'ensemble des unités de niveau 2 au risque à la date  $t_{ij}$  et partageant un même  $v_i$ . On a :

$$\frac{\partial \ln L_{PL}}{\partial v_i} = \sum_{j=1}^{J_i} \delta_{ij} \left[ \frac{1}{v_i} - \frac{\sum_{l \in R^i(t_{ij})} \lambda_1(X_l(t_l))}{\sum_{k \in R(t_{ij})} v_k \lambda_1(X_k(t_k))} \right]. \quad (1.138)$$

L'estimateur de Breslow du hasard de base intégré a pour expression :<sup>35</sup>

$$\widehat{\Lambda}_0(t_{ij}|\beta, v, X) = \sum_{l:t_l < t_{ij}} \frac{\delta_l}{\sum_{k \in R(t_l)} v_k \lambda_1(X_k(t_k))}. \quad (1.139)$$

On a ainsi :

$$d\widehat{\Lambda}_0(t_{ij}|\beta, v, X) = \frac{\delta_{ij}}{\sum_{k \in R(t_{ij})} v_k \lambda_1(X_k(t_k))}. \quad (1.140)$$

Le dénominateur du membre de droite de (1.138) est donc l'inverse de  $d\widehat{\Lambda}_0(t_{ij}|\beta, v)$ , et on peut écrire le score dans un échantillon ordonné :

$$\begin{aligned} \frac{\partial \ln L_{PL}}{\partial v_i} = \sum_{j=1}^{J_i} \left[ \frac{\delta_{ij}}{v_i} - \left( \lambda_1(X_{i1}(t_{i1})) d\widehat{\Lambda}_0(t_{i1}|\beta, v, X_{i1}) + \dots \right. \right. \\ \left. \left. + \lambda_1(X_{ij}(t_{ij})) d\widehat{\Lambda}_0(t_{ij}|\beta, v, X_{ij}) \right) \right]. \quad (1.141) \end{aligned}$$

On en déduit :

$$\frac{\partial \ln L_{PL}}{\partial v_i} = \sum_{j=1}^{J_i} \left[ \frac{\delta_{ij}}{v_i} - \int_0^{t_{ij}} \lambda_1(X_{ij}(u)) \lambda_0(u) du \right]. \quad (1.142)$$

---

<sup>35</sup>Une explication de cette formule utilisant les processus de comptage est donnée dans la sous-section 1.2.7.

### Estimation de $(\beta, v)$ en présence d'hétérogénéité gamma

Un résultat intéressant, obtenu par Therneau *et al.* (2003), est que la solution de l'estimation par l'algorithme EM du modèle à un effet aléatoire gamma dont chaque réalisation est commune à plusieurs observations peut être approchée par une méthode de vraisemblance pénalisée. Nous reprenons ici leur raisonnement et l'étendons au cas de régresseurs variant dans le temps.

Considérons la fonction de pénalité :

$$g(v, \eta) = -\eta \sum_{i=1}^I (\ln v_i - v_i). \quad (1.143)$$

Comme  $g$  ne dépend pas de  $\beta$ , l'estimateur de  $\beta$  de la vraisemblance partielle pénalisée est identique à celui de la vraisemblance partielle. Or, comme celui-ci est équivalent à l'estimateur de la vraisemblance complète lorsque le hasard de base est approché par l'estimateur de Breslow, les problèmes de maximisation de la vraisemblance complète et de la vraisemblance partielle pénalisée admettent la même solution en  $\beta$ , fonction de  $\hat{v}$ . Si  $\hat{v}$  est le même dans les deux problèmes, on aura alors la même estimation de  $\beta$  quelle que soit l'approche retenue.

Notons  $(\hat{\beta}_{EM}(\eta), \hat{v}_{EM}(\eta))$  l'argument de la maximisation de la vraisemblance complète par l'algorithme EM. L'expression de  $\hat{v}_{EM}(\eta)$  à l'optimum, présentée dans l'équation (1.126), est équivalente à :

$$\sum_{j=1}^{J_i} \int_0^{t_{ij}} \lambda_1(X_{ij}(u)) \lambda_0(u) du = \frac{\eta + \sum_{j=1}^{J_i} \delta_{ij}}{\hat{v}_{EM}^i(\eta)} - \eta. \quad (1.144)$$

En substituant cette expression dans (1.142), on obtient grâce à (1.143) :

$$\begin{aligned} \frac{\partial \ln L_{PPL}}{\partial \hat{v}_{EM}^i(\eta)} &= \frac{\sum_{j=1}^{J_i} \delta_{ij}}{\hat{v}_{EM}^i(\eta)} - \frac{\eta + \sum_{j=1}^{J_i} \delta_{ij}}{\hat{v}_{EM}^i(\eta)} + \eta - \frac{\partial g(\hat{v}_{EM}^i(\eta), \eta)}{\partial \hat{v}_{EM}^i(\eta)} \\ &= -\frac{\eta}{\hat{v}_{EM}^i(\eta)} + \eta + \frac{\eta}{\hat{v}_{EM}^i(\eta)} - \eta \\ &= 0 \end{aligned} \quad (1.145)$$

On a ainsi que l'estimation de  $v$  fournie par l'algorithme EM annule le gradient de la vraisemblance partielle pénalisée, et donc que  $(\hat{\beta}_{EM}, \hat{v}_{EM}) = (\hat{\beta}_{PPL}, \hat{v}_{PPL})$ .

### Estimation des paramètres de la distribution mélangeante en présence d'hétérogénéité gamma

Therneau *et al.* (2003) proposent d'estimer  $\eta$  à partir de la vraisemblance complète. Après quelques calculs détaillés dans l'annexe A.10, on obtient comme expression de la vraisemblance concentrée par rapport à  $\eta$  :

$$\ln L_V(\eta) = PPL(\eta) + \sum_{i=1}^I \left[ \eta - (\eta + d_i) \ln(\eta + d_i) + \eta \ln(\eta) - \ln \Gamma(\eta) + \ln \Gamma(\eta + d_i) \right]. \quad (1.146)$$

L'algorithme proposé par Therneau *et al.* (2003) est le suivant : dans une première étape, la vraisemblance partielle pénalisée est maximisée par rapport à  $(\beta, v)$  au moyen d'une procédure de Newton-Raphson pour une valeur fixe de  $\eta$ .<sup>36</sup> Dans une seconde étape, la vraisemblance concentrée (1.146) est maximisée par rapport à  $\eta$ . Une fois la convergence atteinte, le programme retourne à la première étape et itère.

Ce résultat très intéressant montre que l'estimation du modèle à un effet aléatoire gamma *via* la vraisemblance partielle et l'algorithme EM peut être obtenue au moyen de la vraisemblance partielle pénalisée, permettant au passage un gain important de vitesse.

Ripatti et Palmgren (2000) étudient un modèle avec un effet aléatoire gaussien dont les réalisations sont corrélées. Leur approche n'est cependant pas aussi clairement reliée à la vraisemblance pénalisée que dans le cas d'une hétérogénéité gamma. En effet, d'une part ils utilisent une approximation de Laplace de la loi marginale, et d'autre part seule une partie de la vraisemblance marginale obtenue est traitée comme vraisemblance pénalisée lors de la maximisation. Les auteurs indiquent que la perte d'information résultant de l'omission d'une partie de la vraisemblance est faible.

## 1.7. L'approche Bayésienne

Les modèles économétriques fournissent une description d'un phénomène observé, où on explique la différence entre les réalisations de variables endogènes et leur prévision par des variables exogènes au moyen d'une perturbation aléatoire. Comme le souligne Robert (1996), l'expérience ne se répétera jamais exactement dans les mêmes conditions et la nature aléatoire de la dif-

---

<sup>36</sup>La procédure numérique pour  $\beta$  est identique à celle de la vraisemblance partielle et celle pour  $v$  est obtenue à partir de la relation (1.142).

férence importe peu. Un phénomène n'a nul besoin de correspondre effectivement à des tirages dans une loi de probabilité, du fait de la non répétabilité de l'expérience. L'écriture probabiliste des modèles est donc un outil simplificateur, mais efficace, de représentation d'événements. Ce qui est acceptable dans l'approche fréquentiste pour les observations est acceptable dans l'approche Bayésienne pour les paramètres. L'interprétation du modèle en est changée, car supposer une distribution de probabilité pour les paramètres revient à probabiliser l'inconnu, c'est-à-dire à passer de la notion d'inconnu à la notion d'aléatoire. Cette modification permet d'envisager les problèmes habituels d'une manière qui peut se révéler plus simple à l'usage. C'est pourquoi cette littérature est en pleine effervescence actuellement, d'autant plus que l'approche Bayésienne fait souvent appel à des procédures numériques lourdes que l'augmentation de la puissance de calcul des ordinateurs rendent actuellement possibles.

Nous présenterons dans une première sous-section le résultat de Kalbfleisch (1978), qui reformule la vraisemblance partielle dans une optique Bayésienne. Nous évoquerons ensuite brièvement les méthodes de simulation que sont l'échantillonnage de Gibbs et l'algorithme Acceptation-Rejet, du fait des nombreuses applications qui y ont recours pour approcher l'estimateur Bayésien. Les travaux portant sur l'estimation de modèles MPH avec un ou plusieurs effets aléatoires sont présentés dans la dernière sous-section.

### 1.7.1 Justification de la vraisemblance partielle dans l'optique Bayésienne

Kalbfleisch (1978) formule un argument permettant de retrouver la vraisemblance partielle dans une approche Bayésienne. Il suppose :

$$d\Lambda_0(t) \sim \gamma(cd\Lambda_0^*(t), c), \quad (1.147)$$

où  $\gamma$  désigne la densité de la loi gamma,  $c$  est un paramètre et  $\Lambda_0(t)^*$  une fonction connue. Dans une approche Bayésienne,  $\Lambda_0(t)^*$  est un a priori sur le hasard de base, dont  $c$  serait le niveau de confiance. À une faible valeur de  $c$  correspond une forte variance de  $d\Lambda_0(t)$  et donc un a priori non-informatif. Spécifier des a priori indépendants pour le hasard de base est discutable, mais raisonnable dans la mesure où l'intérêt principal réside dans l'estimation des coefficients et non pas celle de  $\lambda_0$ .

Utilisons la formulation du modèle de Cox en termes de processus de comptage, c'est-à-dire définissons le processus  $N_i(t)$  comptant le nombre de transitions observées pour l'individu  $i$  ( $i = 1, \dots, n$ ) sur l'intervalle  $[0, t]$ . On

suppose que l'intensité du processus admet comme expression :

$$E(dN_i(t)|F_{t-}) = Y_i(t)v\lambda_1(X_i(t))d\Lambda_0(t), \quad (1.148)$$

où  $F_{t-}$  est l'information accumulée sur l'intervalle  $[0, t[$  et  $Y_i(t)$  un processus prenant la valeur 1 si l'individu est observé en  $t$ , 0 sinon. La vraisemblance est :

$$L(N_i(t), Y_i(t), X_i|v, \beta, d\Lambda_0(t)) = \prod_{i=1}^n \prod_{t \geq 0} \left\{ [Y_i(t)v\lambda_1(X_i)d\Lambda_0(t)]^{dN_i(t)} \exp[-Y_i(t)v\lambda_1(X_i)d\Lambda_0(t)] \right\}. \quad (1.149)$$

Il convient de noter qu'il est équivalent de spécifier l'intensité du processus de comptage sous une forme de hasards proportionnels ou bien de supposer que les incréments du processus de comptage suivent une loi de Poisson de paramètre  $Y_i(t)v\lambda_1(X_i)d\Lambda_0(t)$ . En multipliant (1.149) par la loi a priori de  $d\Lambda_0(t)$  et en intégrant ensuite par rapport à celui-ci, on a :

$$\begin{aligned} L(\beta|N_i(t), Y_i(t), X_i, v) &\propto \prod_{i=1}^n \prod_{t \geq 0} \int_{\Lambda_0} \left\{ [Y_i(t)v\lambda_1(X_i)d\Lambda_0(t)]^{dN_i(t)} \right. & (1.150) \\ &\exp[-Y_i(t)v\lambda_1(X_i)d\lambda_0(t)] d\Lambda_0(t)^{cd\Lambda_0^*(t)-1} \\ &\left. \exp[-cd\Lambda_0(t)] \right\} d\Lambda_0(t) \\ &\propto \prod_{i=1}^n \left\{ \prod_{t \geq 0} [Y_i(t)v\lambda_1(X_i)]^{dN_i(t)} \right\} \\ &\left\{ \prod_{t \geq 0} \int_{\Lambda_0} d\Lambda_0(t)^{dN_i(t)+cd\Lambda_0^*(t)-1} \right. \\ &\left. \exp[-Y_i(t)v\lambda_1(X_i)d\Lambda_0(t) - cd\Lambda_0(t)] d\Lambda_0(t) \right\} \\ &\propto \prod_{i=1}^n \prod_{t \geq 0} \frac{[Y_i(t)v\lambda_1(X_i)]^{dN_i(t)}}{\sum_{t \geq 0} [Y_i(t)v\lambda_1(X_i) + c]^{dN_i(t)+cd\Lambda_0^*(t)}}. \end{aligned}$$

Lorsque l'a priori sur le hasard de base est non-informatif, la loi marginale a posteriori de  $\beta$  tend vers la vraisemblance partielle. L'inférence Bayésienne requiert le calcul de  $E[\beta|N_i(t), Y_i(t), X_i, v]$ , qui est approché par la moyenne

de tirages dans la densité (1.150). Ils peuvent être simulés de différentes manières, et nous présentons brièvement la méthode d'échantillonnage de Gibbs.

### 1.7.2 L'échantillonnage de Gibbs et l'algorithme Acceptation-Rejet

L'estimation Bayésienne de modèles de durée multivariés a souvent recours aux méthodes d'Acceptation-Rejet, à l'échantillonnage de Gibbs ou à des méthodes qui en sont dérivées, et nous voyons ici brièvement le fonctionnement de ces techniques de simulation.

Les méthodes d'Acceptation-Rejet reposent sur l'approximation de la densité que l'on veut simuler par une autre dans laquelle il est facile d'effectuer des tirages. Soit  $f$  la densité de la loi dans laquelle on veut tirer un échantillon et  $g$  une autre densité telle que  $f(x) \leq Mg(x)$  où  $M$  est une constante. L'algorithme est le suivant :

1. Simuler  $x \sim g$  et  $u \sim U_{[0,1]}$ ,
2. Accepter  $x$  si  $u \leq f(x)/Mg(x)$ , sinon reprendre à l'étape 1.

Comme  $f$  et  $g$  sont deux densités,  $M$  est nécessairement plus grand que 1. Une tentative d'optimisation de l'algorithme revient à choisir  $g$  telle que  $M$  soit le plus proche possible de 1, limitant ainsi le nombre de tirages rejetés.

L'échantillonnage de Gibbs, introduit par Geman et Geman (1984), a été repris et présenté dans de nombreux travaux et ouvrages, comme Robert (1996) entre autres. Considérons un vecteur  $X=(x_1, \dots, x_n)$ , de  $n$  ( $n > 1$ ) variables aléatoires, distribué selon une loi  $f$ . On suppose qu'il existe une densité  $g$  telle que :

$$\int_{\mathbf{z}} g(x, z) dz = f(x). \quad (1.151)$$

Hormis dans les méthodes d'augmentation des données, introduites par Tanner et Wong (1987), où  $z$  s'interprète comme une variable latente ; elle n'a pas de signification particulière dans la plupart des cas. Il s'agit d'une variable utilitaire que l'on prend en compte pour simplifier l'analyse.

On peut évaluer  $g_1(x_1|x_2, \dots, x_p), \dots, g_p(x_p|x_1, \dots, x_{p-1})$ . L'échantillonnage de Gibbs consiste à effectuer des tirages dans la loi de  $x_i$  conditionnellement aux valeurs courantes des  $x_j$ , avec  $j \neq i$ , et la répétition de cette opération produit une chaîne de Markov de loi  $f$ . Plus précisément, l'itération ( $q$ ) de l'algorithme est construite selon la règle suivante :

1. Simuler  $x_1^{(q)} \propto g_1(x_1|x_2^{(q-1)}, \dots, x_p^{(q-1)})$ ,
2. Simuler  $x_2^{(q)} \propto g_2(x_2|x_1^{(q)}, x_3^{(q-1)}, \dots, x_p^{(q-1)})$ ,



3. Procéder ainsi pour les autres variables aléatoires, jusqu'à :

$$x_p^{(q)} \propto g_p(x_p | x_1^{(q)}, \dots, x_{p-1}^{(q)}). \quad (1.152)$$

Alors que les lois marginales ne suffisent pas à résumer la loi jointe, les lois conditionnelles sont suffisantes pour simuler un échantillon distribué suivant  $f$ . Remarquons qu'il n'est pas nécessaire de compléter  $f$  en  $g$ , comme le décrit la relation (1.151), pour mettre en oeuvre l'algorithme de Gibbs. En effet, il suffit de connaître les densités conditionnelles pour procéder aux tirages et de tels cas de figures sont présentés dans Robert (1996, p. 167).

### 1.7.3 Liens avec l'algorithme EM

La méthode d'augmentation des données de Tanner et Wong (1987) est l'un des premiers avatars de l'échantillonnage de Gibbs et fait explicitement apparaître la complétion de  $f$  en  $g$ .<sup>37</sup> En effet, la loi a posteriori s'écrit :

$$f(\theta|T, \delta) = \int_{\mathcal{Z}} p(\theta|T, \delta, z)p(z|T, \delta)dz, \quad (1.153)$$

où  $\theta$  est le vecteur des paramètres,  $T$  et  $\delta$  les vecteurs des durées et des indicatrices de transitions et  $z$  une variable latente. La fonction  $g$  qui complète  $f$  est alors le produit de la densité des paramètres conditionnelle aux données augmentées avec la densité des variables latentes conditionnelles aux observations. Tanner et Wong (1987) en déduisent les densités conditionnelles qui seront utilisées pour les tirages selon la procédure traditionnelle de l'échantillonnage de Gibbs.

On peut ainsi résumer l'approche Bayésienne de l'augmentation des données de la manière suivante. Supposons que la loi a posteriori des données augmentées, c'est-à-dire  $p(\theta|T, \delta, z)$ , puisse être aisément calculée, mais que la loi des données observées,  $p(\theta|T, \delta)$ , soit difficile à évaluer. On peut simuler des valeurs de  $z$  dans une distribution  $p(z|T, \delta)$  et approcher la densité conditionnelle aux observations en faisant la moyenne des  $p(\theta|T, \delta, z)$  sur les simulations. Cependant,  $p(z|T, \delta)$  est paramétrique et dépend donc de  $p(\theta|T, \delta)$ . Il y a ainsi une relation de dépendance mutuelle entre les deux densités, ce qui amène à considérer naturellement une approche itérative basée sur des substitutions successives.

La loi a posteriori (1.153) a une expression très proche de la relation servant de point de départ à l'algorithme EM et représentée dans l'équation

---

<sup>37</sup>La première utilisation de l'échantillonnage de Gibbs est due à Geman et Geman (1984) et tire son nom de l'application considérée. Il n'y a toutefois pas de complétion dans le modèle.

(1.56). En effet, l'idée sous-jacente est la même : la vraisemblance comprenant les variables latentes est plus facile à évaluer que la vraisemblance ne comprenant que les observations. On pose une forme fonctionnelle pour la distribution mélangeante, ce qui nous permet d'évaluer lors de l'étape E les espérances des paramètres et des variables latentes. Elles sont ensuite injectées dans la vraisemblance où les espérances des variables latentes sont considérées comme des observations lors de l'étape M. De nouveau il y a une relation d'interdépendance entre la distribution mélangeante et la vraisemblance, ce qui nous amène à un algorithme itératif, utilisant aussi des substitutions successives. L'idée sous-tendant l'algorithme EM et la méthode d'augmentation des données est la même.

### 1.7.4 Inférence dans les modèles à effet(s) aléatoire(s)

Clayton (1991) estime un modèle de Cox à un effet aléatoire en utilisant l'argument de Kalbfleisch (1978) pour obtenir l'estimation par la vraisemblance partielle des coefficients et l'échantillonnage de Gibbs pour les simulations, fournissant ainsi une alternative à l'algorithme EM de Klein (1992). Clayton (1991) spécifie une hétérogénéité gamma et le cas de l'hétérogénéité log-normale est traité dans Gauderman et Thomas (1994). Des évolutions dans les procédures numériques ont ensuite été produites par, entre autres, Korsgaard *et al.* (1998), qui utilisent l'échantillonnage de Gibbs ou l'algorithme Acceptation-Rejet selon les paramètres, et Sargent (1998) qui utilise un algorithme de Metropolis pour les paramètres définis dans  $\mathfrak{R}$  et l'échantillonnage de Gibbs ou l'algorithme de Hasting pour les paramètres strictement positifs.

Le modèle de Cox à 2 effets aléatoires a été estimé par Bolstad et Manda (2001) lorsque les deux effets sont emboîtés, et Gustafson (1997) ajoute encore un niveau de stratification au hasard de base. L'hétérogénéité est gamma dans la première étude et log-normale dans la seconde. Certains travaux utilisent différents simulateurs selon les paramètres à estimer. Ainsi, Bolstad et Manda (2001) utilisent l'échantillonnage de Gibbs pour estimer les termes d'hétérogénéité, l'algorithme Acceptation-Rejet pour les variances des distributions mélangeantes et un algorithme de Metropolis pour les coefficients. Quant à Gustafson (1997), il utilise une méthode Monte Carlo hybride (Neal, 1993 et Neal, 1994) pour tous les paramètres. L'échantillonnage de Gibbs est utilisé dans un modèle en temps discret à 2 effets aléatoires log-normaux par Manda et Meyer (2005).

## 1.8. Conclusion

Le temps que passe un individu dans un état est au coeur de nombreuses questions économiques et nous présentons dans le Chapitre 1 les principales notions de l'économétrie des modèles de durée. Après avoir introduit la fonction de hasard et sa reformulation en termes de processus de comptage, nous nous intéressons aux principaux estimateurs non paramétriques du hasard intégré et de la fonction de survie. Puis, nous introduisons le modèle de Cox avant de nous intéresser au modèle MPH qui en est l'extension. Nous présentons en détails ces propriétés et les principaux résultats d'identification. Les méthodes usuelles d'estimation, basées sur la vraisemblance et la vraisemblance partielle, sont revues ainsi que les méthodes Bayésiennes.

Il ressort de la revue de littérature qu'il n'existe pas de méthode semi-paramétrique basée sur la vraisemblance partielle pour les modèles à 2 effets aléatoires ou plus. Ainsi, on ne peut pas estimer de modèles où l'hétérogénéité non observée est multiniveaux. De plus, les procédures basées sur l'algorithme EM souffrent de nombreux travers numériques. Dans le Chapitre 2, nous nous intéressons à l'inférence dans les modèles à plusieurs effets aléatoires. La structure du problème permet de dégager un algorithme EM qui est à la fois plus rapide et plus stable.

# Estimation par vraisemblance partielle de modèles MPH à effets aléatoires : un algorithme EM basé sur la vraisemblance pénalisée

---

<sup>0</sup>Ce chapitre est tiré de Horny, G. (2006) : “Partial Likelihood Estimation of a Cox Model with Random Effects : an EM Algorithm based on Penalized Likelihood,” Working Paper 2006-10, BETA.

## Résumé

Nous présentons un algorithme EM général, estimant les modèles MPH à  $I$  effets aléatoires. L'approche est valable pour toute distribution mélangeante admettant une transformée de Laplace. Nous pouvons ainsi mener l'inférence dans une famille comprenant les modèles MPH à hétérogénéité gamma, log-normal, etc.

Nous montrons également comment l'inférence dans un modèle complexe à  $I$  effets peut se simplifier en l'estimation de  $I$  modèles simples, comprenant chacun un seul effet. Nous décrivons ensuite comment la vraisemblance partielle pénalisée peut être utilisée au sein de l'algorithme EM proposé pour obtenir des gains en temps de calcul et en stabilité. Le comportement de l'estimateur est illustré avec une étude Monte Carlo et nous l'appliquons à la ratification des conventions du BIT. Les résultats indiquent une importante diminution des temps de calcul, une convergence améliorée et une approche numériquement stable.

## Sommaire

---

<b>Résumé . . . . .</b>	<b>96</b>
<b>Introduction . . . . .</b>	<b>98</b>
<b>2.1 Le modèle MPH à effets aléatoires . . . . .</b>	<b>99</b>
<b>2.2 Inférence avec l'algorithme EM . . . . .</b>	<b>101</b>
<b>2.3 Estimation de modèles MPH à un effet aléatoire</b>	<b>104</b>
<b>2.4 Simulations par méthodes de Monte Carlo . . .</b>	<b>105</b>
<b>2.5 Ratification des conventions du BIT . . . . .</b>	<b>108</b>
<b>2.6 Conclusion . . . . .</b>	<b>109</b>

---

## Introduction

**N**OUS présentons une méthode générale d'estimation de modèles MPH, où l'hétérogénéité non observée réside à différents niveaux. La population mère est supposée stratifiée selon différents critères, et l'hétérogénéité non observée peut être représentée au moyen de plusieurs effets aléatoires. Aucune structure particulière d'hétérogénéité n'est présupposée, les effets peuvent être emboîtés ou non.

De nombreux travaux, comme par exemple Clayton et Cuzick (1985), Gill (1985) et Parner (1997), spécifient un modèle MPH où l'hétérogénéité non observée est représentée par un seul effet aléatoire. Ils sont par contre peu nombreux à considérer une hétérogénéité multiple, et elle se limite alors à deux effets aléatoires. C'est le cas de Manda et Meyer (2005) qui approchent une modèle MPH en temps discret par un modèle logit. En temps continu, Yau (2001) et Sastry (1997) étudient respectivement des modèles à deux effets aléatoires emboîtés distribués selon une loi log-normal ou gamma. Ce dernier procède à l'estimation au moyen de l'algorithme EM, que nous présentons pour le modèle de Cox dans la sous-section 1.5.2 et pour le modèle MPH en 1.5.3.

Comme le problème de l'hétérogénéité non observée est un problème de données manquantes, l'algorithme EM est un outil approprié de par sa parenté avec les méthodes d'augmentation des données.<sup>1</sup> Cette attractivité théorique est toutefois à mettre en perspective de nombreuses difficultés numériques. En effet, la convergence est sensible au choix des valeurs initiales, demande un grand nombre d'itérations, un temps de calcul élevé, et échoue parfois. Tout d'abord, Ng *et al.* (2004) mettent en avant l'importance des valeurs initiales lorsqu'elles sont proches de la frontière de l'espace des paramètres et que la vraisemblance n'est pas bornée. Ensuite, Dempster *et al.* (1977) montrent que les itérations de l'algorithme EM n'utilisent que les dérivées premières et qu'il donc plus lent que les méthodes usuelles, comme les procédures de Newton-Raphson, impliquant des approximations à l'ordre 2. De plus, la vitesse de convergence est croissante avec la quantité d'information disponible dans l'échantillon, qui est généralement faible car les données sont habituellement incomplètes. Deux cas de non convergence sont référencés dans la littérature. Bolstad et Manda (2001) indiquent qu'une variance trop élevée de l'effet aléatoire peut créer des difficultés numériques. Le second cas de non convergence figure dans Lancaster (1990, p. 267), avec une vraisemblance qui n'est plus bornée lorsque la variance du terme d'hétérogénéité

---

<sup>1</sup>Les liens entre l'algorithme EM et les méthodes d'augmentation des données sont présentées dans la sous-section 1.7.3.

tend vers 0. Bien que le nombre d'itérations requis pour atteindre la convergence soit élevé, le temps total de calcul peut être faible si chaque itération est rapide. Ceci n'est généralement pas le cas, car l'étape E n'admet de solution analytique que dans des cas très particuliers et les expressions doivent être évaluées par intégration numérique (Maples *et al.*, (2002), et Vaida et Xu, 2000) ou par des méthodes de Monte Carlo. Un algorithme EM par méthodes de Monte Carlo (abrégé MCEM) est présenté dans Wei et Tanner (1990), et les simulations pèsent lourdement sur le temps de calcul, en plus d'introduire une erreur de Monte Carlo.

Tous ces inconvénients ont amenés Ripatti et Palmgren (2000), et Therneau *et al.* (2003) à estimer les paramètres d'un modèle MPH moyen de la vraisemblance partielle pénalisée. En présence d'un effet gamma, Therneau *et al.* (2003) obtiennent un estimateur équivalent à celui fournit par l'algorithme EM qui n'est pas tributaire de ses difficultés numériques.

Nous présentons dans la Section 2.1 une approche générale pour estimer les modèles MPH à plusieurs effets aléatoires au moyen de l'algorithme EM. Notre méthode a l'avantage de transformer l'estimation d'un modèle MPH à  $I$  effets en l'estimation de modèles à un effet, comme nous le montrons dans la Section 2.2. Abbring et Van den Berg (2001) montrent que la distribution mélangeante parmi les individus au risque tend vers une loi gamma, et nous décrivons donc comment utiliser la vraisemblance partielle pénalisée proposée par Therneau *et al.* (2003) dans la Section 2.3.<sup>2</sup> Ainsi, nous restons dans le cadre théorique commode de l'algorithme EM tout en bénéficiant des gains de vitesse et de stabilité de la vraisemblance partielle pénalisée. Notre approche est illustrée par une étude Monte Carlo dans la Section 2.4 et appliquées aux données du BIT, présentées dans le chapitre 3, dans la Section 2.5. Nous cherchons à voir dans ces 2 cas les gains par rapport à l'algorithme EM accéléré présenté par Sastry (1997).

## 2.1. Le modèle MPH à effets aléatoires

Considérons un modèle MPH dont la fonction de l'hétérogénéité non observée s'écrit comme le produit de  $I$  effets aléatoires ( $I \geq 2$ ). Il y a plusieurs termes d'hétérogénéité car ils captent les relations de dépendance, situées à plusieurs niveaux, qui existent entre les durées. Les effets aléatoires peuvent

---

<sup>2</sup>Le résultat de Abbring et Van den Berg, 2001, suppose une distribution mélangeante variant régulièrement à 0, comme défini dans Feller (1971). Ce pré-requis n'est guère contraignant car satisfait par la plupart des distributions usuelles comme la loi exponentielle, uniforme, beta et plus généralement toutes les distributions avec un point de masse en 0.

## 2.1 Le modèle MPH à effets aléatoires

---

être emboîtés (Bolstad et Manda, 2001, Sastry (1997) et Yau (2001)) ou indépendants comme dans le chapitre 3.

Nous spécifions la fonction de hasard :

$$\lambda_{ik}^j(t) = \left( \prod_{i=1}^I v_i^j \right) \lambda_0(t) \lambda_1 [X_k(t), \beta], \quad (2.1)$$

où  $i$  ( $i = 1, \dots, I$ ) est l'indice de niveau,  $j$  ( $j = 1, \dots, J_i$ ) l'indice des groupes définis au niveau  $i$ , et  $k$  l'indice d'unité ( $k = 1, \dots, N$ ). Le terme "niveau" fait référence à une source de corrélation possible entre les durées.

Le modèle MPH satisfait les hypothèses habituelles : les fonctions  $\lambda_0$  et  $\lambda_1$  sont positives, le processus  $X_k(t)$  est absolument continu et chaque  $v_i^j$  est tiré dans une loi  $h_i(v_i^j; \alpha_i)$  de support  $[0, \infty[$ . Elles sont présentées en détail dans la sous-section ???. Nous supposons de plus que les distributions mélangeantes admettent une transformée de Laplace.<sup>3</sup>

Cette spécification flexible s'adapte à de nombreuses situations :

- Modèle à un effet individuel

$$\lambda_{1k}^j(t) = v_1^j \lambda_0(t) \lambda_1 [X_k(t), \beta], \quad (2.2)$$

Les durées sont corrélées au travers de  $v_1^j$ . Il n'y a qu'un seul niveau et le modèle est approprié lorsque chaque individu  $j$  connaît plusieurs épisodes. Le cas limite de l'effet individuel est obtenu pour  $j = k$ , où le terme d'hétérogénéité ne capte plus une dépendance entre les durées et traduit une surdispersion.

- Modèle à deux effets aléatoires

$$\lambda_{ik}^j(t) = v_1^j v_2^{j'} \lambda_0(t) \lambda_1 [X_k(t), \beta], \quad (2.3)$$

Nous étendons le modèle (2.2) de manière à prendre en compte deux niveaux de dépendance. Le modèle (2.3) est utilisé en démographie pour relier les durées de vie à la famille de l'individu, et à la communauté où il réside. Les effets aléatoires sont alors emboîtés, et le cas complémentaire est obtenu lorsque chaque individu appartient à plusieurs groupes qui ne sont pas reliés. On retrouve cette situation dans des études sur le comportement d'achat d'un consommateur qui fréquente plusieurs centres de distributions. Comme dans le modèle (2.2), un effet peut être spécifié pour prendre en compte la surdispersion.

---

<sup>3</sup>Dans le cas contraire, elles n'ont pas d'espérance et il est impossible d'estimer le modèle par l'algorithme EM.



Les durées sont indépendantes conditionnellement à l'hétérogénéité observée et non observée. On note par  $T$  et  $d$  les vecteurs des durées et des indicatrices de transition, respectivement. Lorsque les effets aléatoires sont indépendants, la vraisemblance est obtenue en faisant le produit des distributions mélangées avec la vraisemblance conditionnelle :

$$L(\alpha, \beta; T, d, X, v) = \prod_{k=1}^N \left\{ \left[ \left( \prod_{i=1}^I v_i^j \right) \lambda_0(t_k) \lambda_1 [X_k(t_k), \beta] \right]^{d_k} \exp \left[ - \left( \prod_{i=1}^I v_i^j \right) \int_0^{t_k} \lambda_0(u) \lambda_1 [X_k(u), \beta] du \right] \prod_{i=1}^I h_i(v_i^j; \alpha_i) \right\}. \quad (2.4)$$

La log-vraisemblance s'écrit :

$$\ln L(\alpha, \beta; T, d, X, v) = \ln L_1(\alpha; v, d) + \ln L_2(\beta; T, d, X, v), \quad (2.5)$$

où

$$\ln L_1(\alpha; v) = \sum_{k=1}^N \left[ \sum_{i=1}^I \ln h_i(v_i^j, \alpha_i) + \delta_k \sum_{i=1}^I \ln v_i^j \right], \quad (2.6)$$

$$\ln L_2(\beta; T, d, X, v) = \sum_{k=1}^N \left[ \delta_k \ln \{ \lambda_0(t_k) \lambda_1 [X_k(t_k), \beta] \} - \left( \prod_{i=1}^I v_i^j \right) \int_0^{t_k} \lambda_0(u) \lambda_1 [X_k(u), \beta] du \right]. \quad (2.7)$$

## 2.2. Inférence avec l'algorithme EM

Comme les effets aléatoires sont indépendants, l'étape E à l'itération  $q$  est l'évaluation de :

$$Q(\alpha, \beta; \alpha^{(q)}, \beta^{(q)}) = Q_1(\alpha; \alpha^{(q)}, \beta^{(q)}) + Q_2(\beta; \alpha^{(q)}, \beta^{(q)}), \quad (2.8)$$

where

$$Q_1(\alpha; \alpha^{(q)}, \beta^{(q)}) = \sum_{k=1}^N \left( \sum_{i=1}^I E_{\beta^{(q)}, \alpha^{(q)}} [\ln h_i(v_i^j, \alpha_i) | T, d] + \delta_k \sum_{i=1}^I E_{\beta^{(q)}, \alpha^{(q)}} [\ln v_i^j | T, d] \right), \quad (2.9)$$

$$Q_2(\beta; \alpha^{(q)}, \beta^{(q)}) = \sum_{k=1}^N \left( \delta_k \ln \lambda_0(t_k) \lambda_1[X_k(t_k), \beta] - \prod_{i=1}^I E_{\beta^{(q)}, \alpha^{(q)}} [v_i^j | T, d] \int_0^{t_k} \lambda_0(u) \lambda_1[X_k(u), \beta] du \right). \quad (2.10)$$

Nous devons donc évaluer  $E_{\beta^{(q)}, \alpha^{(q)}} [v_i^j | T, d]$  pour tout  $(i, j)$ . En généralisant l'approche proposée dans Parner (1997), on montre dans l'Annexe B.1 le résultat suivant :

$$E[v_i^j | T, d] = \frac{E[(v_i^j)^{1+l_{ij}} \xi_{(-i)}]}{E[(v_i^j)^{l_{ij}} \xi_{(-i)}]}, \quad (2.11)$$

où :

$$\xi_{(-i)} = E \left[ (v_i^j)^{l_{ij}} \dots E \left[ (v_{(i+1)}^j)^{l_{(i+1)ij}} E \left[ (v_{(i-1)}^j)^{l_{(i-1)ij}} \dots E \left[ (v_2^j)^{l_{2ij}} \mathcal{L}_1^{(l_{1ij})} \right] \right] \right] \right], \quad (2.12)$$

où  $l_{i'i'j}$  est le nombre de transitions observées dans le sous-groupe défini comme l'intersection des groupes contenant l'individu  $j$  au niveaux de stratifications  $i$  et  $i'$ . L'équation 2.11 admet une solution analytique en présence d'un effet aléatoire gamma (voir Clayton et Cuzick, 1985). En présence de plusieurs effets gamma ou d'hétérogénéité log-normale, les espérances doivent être évaluées au moyen de procédures numériques (Sastry, 1997 et Vu et Knui-man, 2002).

L'étape M requiert la maximisation de 2.8. On note  $R_k$  l'ensemble des épisodes qui ne sont pas terminés à la date  $t_k$ . Johansen (1983) montre que l'estimateur du hasard intégré, obtenu en concentrant la vraisemblance complète, est :

$$\hat{\Lambda}_0^{(q)}(t) = \sum_{t_k < t} \frac{\delta_k}{\sum_{m \in R_k} \left( \prod_{i=1}^I v_i^{l_i^{(q)}} \right) \lambda_1[X_m(t_m), \beta^{(q)}]}. \quad (2.13)$$

On peut montrer avec un argument similaire que la maximisation de 2.8 lors de l'étape M ne nécessite que l'optimisation par rapport à  $\alpha$  et  $\beta$  pour obtenir l'estimateur de la vraisemblance partielle (Cox, 75 et 1.2.4).

Nous considérons donc l'algorithme suivant. Les valeurs initiales  $(\alpha^{(0)}, \beta^{(0)})$  correspondent aux estimations dans un modèle de Cox standard, où  $\text{Var}(v_i^j) = 0, \forall i$ . L'itération  $(q)$  se déroule comme suit :

**Étape 1** Pour  $i=1$ , évaluer  $Q_1(\alpha_1; \alpha^{(q)}, \beta^{(q,1)})$  et l'optimiser pour obtenir  $\alpha_1^{(q+1)}$ ,

**Étape 2** Evaluer  $Q_2(\beta; \alpha_1^{(q+1)}, \alpha_2^{(q)}, \dots, \alpha_I^{(q)}, \beta^{(q,1)})$  et l'optimiser pour obtenir  $\beta^{(q+1,1)}$ ,

**Étape 3** Itérer entre les étapes 1 et 2 jusqu'à la convergence,

**Étape 4** Pour  $i = 2, \dots, I$ , effectuer les étapes de 1 à 3,

**Étape 5** Evaluer  $Q_2(\beta; \alpha_1^{(q+1)}, \dots, \alpha_I^{(q+1)}, \beta^{(q)})$  et la maximiser pour en déduire  $\beta^{(q+1)}$ ,

**Étape 6** Itérer entre les étapes de 1 à 5 jusqu'à la convergence.

On peut penser à aller directement de l'étape 1 à l'étape 4. Toutefois, comme  $Q_1$  est une fonction de  $\beta^{(q,i)}$ , les étapes 2 et 3 sont nécessaires pour obtenir des estimations stables des  $\alpha_i^{(q+1)}$  et donc pour réduire le nombre total d'itérations de l'algorithme.

En concentrant par rapport aux  $\alpha_{i'}^{(q)} \forall i' \neq i$ , notre approche est équivalente à l'estimation successive de  $I$  modèles impliquant chacun un effet aléatoire, les autres étant traités comme des variables explicatives de coefficient 1. On peut donc reformuler l'algorithme de la manière suivante :

**Étape A** Pour  $i=1$ , estimer le modèle :

$$\lambda_{1k}^j(t) = v_1^j \left( \prod_{i=2}^I v_i^{j,(q)} \right) \lambda_0(t) \lambda_1 [X_k(t), \beta^{(q,1)}]. \quad (2.14)$$

Récupérer  $v_1^{j,(q+1)}$  et  $\beta^{(q+1,1)}$ ,

**Étape B** Effectuer l'étape A pour  $i = 2, \dots, I$ ,

**Étape C** Estimer le modèle :

$$\lambda_{ik}^j(t) = \left( \prod_{i=1}^I v_i^{j,(q+1)} \right) \lambda_0(t) \lambda_1 [X_k(t), \beta^{(q)}]. \quad (2.15)$$

Récupérer  $\beta^{(q+1)}$ ,

**Étape D** Itérer entre les étapes A et C jusqu'à la convergence.

Les  $\alpha_i$  sont constants durant l'estimation et les écarts-type des  $\beta$  obtenus lors des étapes 5 et C sont sous-estimés. Nous utilisons l'approche de Louis (1982) pour évaluer la matrice d'information à l'issue de la dernière itération de l'algorithme EM. Elle a pour expression :

$$I_{obs} = E [-H(\alpha, \beta)|T, d, v_1, \dots, v_I] - \text{Var} [S(\alpha, \beta)|T, d], \quad (2.16)$$

où  $I_{obs}$  est la matrice d'information du modèle observé,  $H$  le Hessien et  $S$  le score. Les détails des calculs se trouvent dans l'Annexe 2.16.

L'approche proposée à deux avantages : l'étape E est facilitée car elle implique moins d'intégrales et l'étape M est accélérée car les optimisations sont effectuées sur des espaces de dimensions plus faibles. Ceci vient du fait que nous passons de l'estimation d'un modèle complexe en l'estimation de  $I$  sous-modèles, chacun ne comprenant qu'un seul effet aléatoire. Nous présentons dans la section suivante une manière d'estimer les modèles à un effet.

## 2.3. Estimation de modèles MPH à un effet aléatoire

La spécification  $\lambda_1 [X_k(t_k), \beta] = \exp[X_k(t_k)\beta]$  est fréquemment utilisée et nous allons l'adopter par la suite. Abbring et Van den Berg (2001) montrent que la distribution mélangeante parmi les survivants tends vers une loi gamma pour un grand nombre de lois pour l'hétérogénéité non observée. Le choix de la loi gamma est donc justifié par des arguments plus solides que sa simplicité analytique.<sup>4</sup> Klein (1992) estime le modèle 2.2 en présence d'hétérogénéité gamma avec l'algorithme EM. Parner (1997) étend son approche aux modèles à épisodes multiples, avec un effet aléatoire dont la transformée de Laplace de la distribution mélangeante est connue. Nous adaptons dans la sous-section suivante l'approche de Therneau *et al.* (2003) au modèle 2.1.

### 2.3.1 Inférence en présence d'hétérogénéité gamma

Pour une unité  $i$  donnée, supposons que les  $v_i^j$  sont distribués selon une loi gamma, d'espérance unitaire et de variance  $1/\alpha_i$ , et que les autres effets aléatoires sont considérés comme exogènes. Therneau *et al.* (2003) montrent que l'estimateur proposé par Klein (1992) est également obtenu en maximisant

---

<sup>4</sup>Supposer un nombre fini de groupes amène aux modèles de mélange finis. Heckman et Singer (1984b), (1984c) étudient un modèle semi-paramétrique où les groupes sont latents. Lorsque l'appartenance aux différents groupes est connue, on peut utiliser à effets fixes avec des variables indicatrices.

la vraisemblance pénalisée suivante :

$$\ln L_{PPL}(\beta, v, \alpha_i) = \ln L_{PL}(\beta, v) - \frac{1}{\alpha_i} \sum_{j=1}^{J_i} (\ln v_i^j - v_i^j), \quad (2.17)$$

où  $L_{PL}(\beta, v)$  est la vraisemblance partielle :

$$L_{PL}(\beta, v) = \prod_{k=1}^N \left[ \frac{v_i^j \left( \prod_{i' \neq i} v_{i'}^j \right) \exp [X_k(t_k)\beta]}{\sum_{m \in R_k} v_i^l \left( \prod_{i' \neq i} v_{i'}^l \right) \exp [X_m(t_m)\beta]} \right]^{\delta_k}. \quad (2.18)$$

Dans les approches de vraisemblance pénalisée,  $1/\alpha_i$  est un paramètre de lissage indiquant l'arbitrage entre l'adéquation aux données et l'aspect lisse de la vraisemblance pénalisée. La solution au problème de maximisation de 2.17 est donc la même que celle obtenue en appliquant l'algorithme EM à un modèle à hétérogénéité gamma. L'obtention de ce résultat repose sur le choix de la fonction de pénalité, et ne tient plus pour d'autres spécification telles que  $\int_0^\infty \left( \lambda_0^{(2)}(u) \right)^2 du$  où  $\lambda_0^{(2)}(u)$  est la dérivée seconde du hasard de base. Rondeau *et al.* (2003) utilisent cette fonction de pénalité car elle garantit l'existence et l'unicité de l'estimateur, comme le montrent de Montricher *et al.* (1975).

Nous utilisons la vraisemblance partielle pénalisée 2.17 pour effectuer les étapes de 1 à 3, ou de manière équivalente les étapes A et B. Nous évaluons la matrice d'information avec l'approche de Louis (1982), présentée dans l'Annexe B.3 en présence d'hétérogénéité gamma, une fois que l'algorithme EM a convergé.

## 2.4. Simulations par méthodes de Monte Carlo

Le principal objectif de cette étude de Monte Carlo est de comparer le temps de calcul de l'algorithme que nous proposons avec l'algorithme EM accéléré présenté dans Sastry (1997), pour différentes tailles de groupes et de sous-groupes en présence de deux effets aléatoires. Notre procédure est appelée par la suite "algorithme EM basé sur la vraisemblance pénalisée" et abrégé EMPL.

### 2.4.1 Un algorithme EM accéléré

Avec un hasard de base constant par morceaux, Sastry (1997) montre que le résultat de l'étape E peut être séparé en 3 fonctions, les deux premières

dépendant des paramètres des distributions mélangeantes et la dernière de  $\beta$  et  $\Lambda_0(t)$ . Il en déduit l'algorithme EM suivant : après l'étape E de l'itération ( $q$ ), les fonctions des paramètres de lois des termes d'hétérogénéité non observée sont optimisées au moyen de leur propre algorithmes EM. Lorsque ces sous-routines ont convergé, l'étape M de l'itération ( $q$ ) est effectuée pour tous les paramètres et ensuite l'itération ( $q + 1$ ) commence. L'utilisation de sous-routines entraîne des gains de vitesse importants, ce qui permet à Sastry (1997) de ne pas avoir à mettre en place d'autres techniques d'accélération tel celles proposées par Louis (1982) ou Meilijson (1989). La méthode de Sastry (1997) est présentée plus en détails dans la sous-section 1.5.3. Nous étendons cet algorithme à la vraisemblance partielle grâce au résultat de Johansen (1983) et considérons ainsi une procédure similaire à celle décrite dans Horny (2001).

### 2.4.2 Échantillons simulés et valeurs initiales

Pour chaque taille de groupes et de sous-groupe, nous simulons 200 échantillons de 2000 observations, avec 2 niveaux d'hétérogénéité. Les groupes ont des tailles allant de 10 à 100 observations. Au delà de 100 observations, la stratification du hasard de base est souvent préféré dans la pratique et en deçà de 10 observations, il est peu vraisemblable qu'un second effet aléatoire soit défini à un niveau plus fin. Les sous-groupes sont de tailles allant de 1 à 5, cette dernière étant la valeur la plus élevée permettant d'avoir plusieurs sous-groupes au sein d'un groupe de taille 10.

Nous spécifions  $E(v_{1j})=E(v_{2j})=1$  et  $\text{Var}(v_{1j})=\text{Var}(v_{2j})=0.5$ . Le hasard de base est supposé constant et il n'y a pas de censure. Deux régresseurs normaux, centrés et réduits, sont simulés et leurs coefficients sont  $\beta_1 = 1$  et  $\beta_2 = -1$ . Il n'y a pas de constante car elle n'est pas identifiée dans une approche de vraisemblance partielle.

Comme l'indiquent Ng *et al.* (2004), le choix des valeurs initiales affecte la convergence et donc le temps de calcul. Nous prenons comme valeurs initiales les résultats d'estimation du modèle de Cox sans hétérogénéité, c'est-à-dire qu'elles sont de 1 pour  $v_i$  et  $w_i^j$ , de 0 pour leurs variances et de  $\hat{\beta}$  pour les coefficients. Les routines d'estimation sont programmées pour le logiciel R 2.0.1 (Team, 2005) et utilisent la fonction 'coxph' du package 'survival'. Le code de l'EMPL et de la méthode de Louis (1982) se trouvent dans l'Annexe B.4, l'algorithme EM accéléré inspiré de Sastry (1997) figure dans l'Annexe B.5.

### 2.4.3 Résultats

L'algorithme EM accéléré et l'EMPL obtiennent les mêmes résultats pour tous les échantillons, c'est pourquoi nous n'indiquons que les temps de calcul. Le tableau 2.4.3 contient les 25ème, 50ème et 75ème centiles des temps de calcul. Même pour deux échantillons ayant la même structure d'hétérogénéité, les temps de calcul sont très différents pour un même estimateur. Dans la plupart des cas, le rapport de la durée la plus élevée sur la plus courte pour est de 100 pour l'algorithme EMPL et de 500 pour l'algorithme EM accéléré.

TAB. 2.1 – Computing time ( $\text{Var}(v_1^j)=\text{Var}(v_2^j)=0.50$ )

Number of spells			Computing Time in seconds					
Total	Per group	Per subgroup	EMPL			EM		
			$Q_{25}$	Median	$Q_{75}$	$Q_{25}$	Median	$Q_{75}$
2000	100	5	18	22	29	55	110	187
		4	22	24	35	88	176	561
		3	30	32	48	31	76	201
		2	55	74	83	30	87	213
		1	120	173	181	30	41	56
	50	5	19	27	34	98	163	313
		4	24	35	38	93	184	467
		3	33	49	58	72	159	401
		2	68	78	84	31	94	227
		1	156	175	181	37	45	60
	40	5	16	22	30	87	159	385
		4	26	39	46	126	218	574
		3	34	51	52	72	158	401
		2	57	72	83	39	105	216
		1	144	155	171	30	37	44
20	5	20	29	37	174	284	510	
	4	38	44	58	260	445	723	
	3	45	52	61	225	359	583	
	2	58	66	79	120	251	457	
	1	138	150	158	48	73	136	
10	5	18	32	45	511	722	1028	
	4	48	56	79	727	937	1403	
	3	54	58	77	676	955	1457	
	2	74	80	89	521	859	1485	
	1	151	162	181	137	229	311	

Les tailles de groupes et de sous-groupes ont un effet ambiguë sur les temps de calcul de l'algorithme EMPL. Ils sont monotones décroissants avec les tailles des sous-groupes et ne semblent pas affectés par le nombre d'observations au sein d'un groupe. Ainsi, un échantillon avec des groupes de 10 épisodes et des sous-groupes de 5 observations sera estimé bien plus rapidement qu'un échantillon où les groupes sont de 100 épisodes et les sous-groupes de 2 observations. Les temps de calcul de l'algorithme EM accéléré ont un profil différent, en forme de  $\cup$  inversé, dont le maximum est situé aux alentours de 4 épisodes par sous-groupe. Plus de sous-groupes implique plus de  $v_{2j}$  à évaluer à chaque itération, ce qui augmente le temps de calcul jusqu'à un certain seuil. Mais plus de  $v_{2j}$  implique aussi plus d'information, ce qui accélère la convergence une fois ce seuil dépassé.

L'algorithme EMPL est globalement rapide, avec des centiles pour les temps de calcul inférieurs à ceux de l'algorithme EM. Ce phénomène est particulièrement important lorsqu'il y a moins de 20 épisodes par groupe, le 75ème centile de l'EMPL est toujours en dessous du 25ème centile de l'algorithme EM. Ce résultat ne tient plus dès lors qu'il y a un épisode par sous-groupe, hormis lorsqu'il y a 10 épisodes par groupe.

## 2.5. Ratification des conventions du BIT

Nous appliquons l'algorithme EMPL à des données relatives à la ratification des conventions du BIT. L'échantillon est présenté dans le Chapitre 3 et dans Boockmann (2001). Les durées sont le temps séparant l'adoption des conventions du BIT lors des conférences annuelles de leur ratification par un pays membre et en voie de développement sur la période 1975-1995. Elles concernent 80 pays et 29 conventions pour un total de 228 ratifications. Nous considérons la fonction de hasard  $\lambda_{2jk}(t) = v_{1j}v_{2j}\lambda_0(t) \exp[X_k(t)\beta]$ , où  $v_{1j}$  est l'effet convention et  $v_{2j}$  l'effet pays. Dans le Chapitre 3, ce modèle est estimé avec une approche Bayésienne basée sur la vraisemblance partielle.

Le modèle est estimé à la fois avec l'algorithme EM accéléré et l'EMPL. L'algorithme EM ne converge pas. Le résultat de l'étape E se scinde en deux parties, comme expliqué dans la Section 2.3, la première comprenant les paramètres des distributions mélangeantes et le second les coefficients  $\beta$ . La première partie est optimisée avec ses propres sous algorithmes EM et les variances des distributions mélangeantes croissent au fil des itérations jusqu'à se stabiliser. L'algorithme s'arrête ensuite lors de l'étape M, après avoir retourné des  $\beta$  tendant vers  $-\infty$ . Le processus complet dure 8 heures et demi. Inversement, l'algorithme EMPL converge. Les résultats, obtenus en 66 secondes, sont reportés dans l'Annexe B.6 et sont proches de ceux obtenus



dans le Chapitre 3.

## 2.6. Conclusion

Nous proposons dans ce chapitre un algorithme EM pour les modèles MPH à plusieurs effets aléatoires dont les distributions mélangeantes admettent une transformée de Laplace. Nous accordons ensuite une attention particulière à l'hétérogénéité gamma. Choisir cette distribution mélangeante est une hypothèse faible dans la mesure où Abbring et Van den Berg (2001) montrent que la distribution mélangeante parmi les survivants tend vers elle. On décrit ensuite comment utiliser la vraisemblance pénalisée pour éviter les difficultés numériques inhérentes à l'algorithme EM. Nous présentons ainsi un algorithme EM modifié qui est non seulement rapide, mais aussi simple et stable. Notre approche est semi-paramétrique car basée sur la vraisemblance partielle. Elle ne requiert donc pas la spécification d'un hasard de base et étend Sastry (1997). Nous présentons également une étude de Monte Carlo et une application sur les données du BIT, présentées plus en détails dans le Chapitre 3, pour illustrer le comportement de notre procédure avec l'algorithme EM accéléré.

Les temps de calcul sont fortement réduits par l'utilisation de l'algorithme EMPL, rendant superflue l'utilisation de des routines d'accélération décrites dans Louis (1982) et Meilijson (1989) pour des échantillons de taille modeste. Le cas d'un effet aléatoire définissant des groupes d'une unique observation est une exception et la vitesse dépend également de la taille des groupes définie par les autres effets aléatoires. De plus, l'algorithme EMPL converge dans des situations où l'algorithme EM échoue.

Therneau, Grambsch, et Pankratz (2003) indiquent que la vraisemblance partielle pénalisée qu'ils proposent fournit un estimateur biaisé du paramètre de la distribution mélangeante, leur approche étant centrée sur les coefficients  $\beta$ . Les résultats de nos études de Monte Carlo vont dans le même sens. Or, Rodriguez et Goldman (2001) comparent différentes méthodes d'estimation pour un modèle Logit à deux termes d'hétérogénéité non observée emboîtés. Ils concluent que l'estimateur Bayésien est moins biaisé que les méthodes basées sur la vraisemblance en présence d'une forte hétérogénéité. C'est pourquoi nous intéressons dans le Chapitre suivant aux méthodes bayésiennes en présences de deux effets aléatoires.

# Une approche Bayésienne du comportement de ratification des conventions de l' OIT

---

<sup>0</sup>Ce chapitre est tiré de Horny, G., D. Boockmann, D. Djurdjevic, et F. Laisney (2005) : “Bayesian Estimation of Cox Models With Non-Nested Random Effects : An Application to the Ratification of ILO Conventions by Developing Countries,” Discussion Paper 05-23, ZEW.

## Résumé

Nous spécifions un modèle MPH multivarié pour étudier les temps de ratification des conventions de l'OIT. La fonction de hasard contient deux effets aléatoires, l'un défini au niveau des pays et l'autre au niveau des conventions. Après avoir établi l'identification, nous estimons une approche Bayésienne semi-paramétrique, basée sur la vraisemblance partielle. Nos résultats confirment ceux obtenus lors d'études précédentes, à savoir que la décision de ratifier dépend de facteurs économiques et politiques. Toutefois, les résultats diffèrent entre l'approche Bayésienne et l'approche fréquentiste quant à l'importance de l'hétérogénéité non observée au niveau des conventions et des pays. L'approche fréquentiste semble sous-estimer l'influence de l'hétérogénéité entre les différentes conventions.

## Sommaire

---

<b>Résumé . . . . .</b>	<b>112</b>
<b>Introduction . . . . .</b>	<b>114</b>
<b>3.1 Les données . . . . .</b>	<b>115</b>
<b>3.2 Les modèles . . . . .</b>	<b>118</b>
<b>3.3 L'inférence Bayésienne . . . . .</b>	<b>120</b>
<b>3.4 Resultats . . . . .</b>	<b>121</b>
<b>Conclusion . . . . .</b>	<b>125</b>

---

## Introduction

**N**OUS proposons une approche Bayésienne pour les modèles MPH dont hétérogénéité non observée est multiniveaux. Les déterminants inobservés n'ont pas de structure hiérarchique et sont situés à la fois au niveau des unités effectuant des transitions et des destinations. Notre application concerne la ratification des conventions de l'Organisation Internationale du Travail (OIT). Les données comprennent 184 conventions de thèmes variés, que les états membres de l'OIT sont libres de ratifier ou non. Dans ce contexte, il est raisonnable de penser qu'une partie des différences entre les rythmes de ratification relève de l'hétérogénéité non-observée spécifique aux conventions et aux pays.

L'hétérogénéité non observée est représentée au moyen d'effets fixes ou aléatoires. Le principal avantage des effets fixes est la facilité avec laquelle on peut les introduire dans le modèle, sans ajouter d'hypothèse sur la distribution jointe des hétérogénéités observées et non observées. La vraisemblance partielle stratifiée (Kalbfleisch et Prentice, 1980, Yamaguchi, 1986 et Ridder et Tunalı (1999)) est une manière d'estimer les modèles à effets fixes. Les observations sont divisées en groupes au sein desquels l'hétérogénéité non observée est supposée constante, et chaque groupe est appelé une strate. Ridder et Tunalı (1999) comparent l'utilisation d'un effet fixe avec celle d'un effet aléatoire dans leur application à la mortalité infantile. Boockmann (2001) modélise l'hétérogénéité non observée au niveau des pays membres avec un effet fixe et utilise la vraisemblance partielle stratifiée. Un test d'Hausman rejette le modèle stratifié pour les pays industrialisés tandis qu'il est accepté pour les pays en voie de développement. C'est l'une des raisons qui nous poussent à utiliser des effets aléatoire pour les pays en voie de développement, en plus des inconvénients habituels de la vraisemblance partielle stratifiée présentés dans Therneau et Grambsch (2000), à savoir qu'elle ne quantifie pas l'impact de l'hétérogénéité non observé et que la précision de l'estimateur décroît avec le nombre de strates.

Par l'estimation des paramètres des distributions mélangeantes et la possibilité de comparer l'importance de l'hétérogénéité non observée aux différents niveaux, les effets aléatoires nous apportent une meilleure compréhension des caractéristiques inobservées. Ce choix n'est toutefois pas dépourvu de conséquences déplaisantes, à savoir que nous devons spécifier la loi jointe des effets aléatoires. Le cas d'un effet aléatoire à hétérogénéité gamma est présenté dans Guo et Rodriguez (1992), dont l'approche est présentée dans la sous-section 1.5.3, qui procèdent à l'inférence lorsque le hasard de base est constant par morceaux, ou encore dans Meyer (1990), Dolton et Van der Klaauw (1995),

et Fortin *et al.* (2001) parmi d'autres. L'estimation en présence de deux effets aléatoires emboîtés est présentée dans Gustafson (1997), Sastry (1997) et Milcent (2003) ainsi que dans la sous-section 1.5.3. Guo et Rodriguez (1992) utilisent l'algorithme EM (dempster *et al.*, ? et Sastry (1997) étend leur approche à deux effets aléatoires. Une approche Bayésienne pour un modèle à deux aléatoires emboîtés et hasards de base constant par morceaux est présenté dans Bolstad et Manda (2001).

Djurdjevic (2000) mène l'inférence dans un modèle MPH avec un estimateur similaire à celui de Klein (1992), la principale différence étant qu'elle base son analyse sur la vraisemblance partielle tandis que ce dernier considère la vraisemblance complète comme point de départ avant de la concentrer par rapport au hasard de base. Horny (2001) généralise le travail de Djurdjevic (2000) à deux effets aléatoires emboîtés, étendant par la même occasion les travaux de Sastry (1997) présentés dans la sous-section 1.5.3 à la vraisemblance partielle. Tous ces travaux utilisent l'algorithme EM, et Djurdjevic (2000) et Horny (2001) l'appliquent à la ratification des conventions de l'OIT. Leurs résultats sont mitigés : bien que fonctionnant convenablement sur des données simulées, l'algorithme ne converge pas avec ces données.<sup>1</sup> Une version plus stable et plus rapide de l'algorithme EM est présentée au Chapitre 2.

Toutefois, Therneau, Grambsch, et Pankratz (2003) présentent une étude de Monte Carlo indiquant une sous-estimation de l'ampleur de l'hétérogénéité non observée lors d'une telle procédure EM. Rodriguez et Goldman (2001) comparent le maximum de vraisemblance, la quasi-vraisemblance marginale, la quasi-vraisemblance pénalisée et l'approche Bayésienne avec de données groupées dans le cadre d'une régression logistique. Ils concluent à la sous-estimation de l'hétérogénéité observée et non observée, hormis avec l'estimateur du maximum de vraisemblance et Bayésien. Bien que les biais puissent être enlevés par une procédure de bootstrap, la convergence n'est pas assurée et le temps de calcul final est plus élevé que lors de l'estimation par Méthodes de Monte Carlo par Chaînes de Markov.

Ce chapitre propose une approche Bayésienne semi-paramétrique, basée sur la vraisemblance partielle, pour estimer un modèle MPH à deux effets aléatoires non emboîtés. Nous montrons que le modèle est identifié dès lors que chaque réalisation de chaque effet est partagée par au moins deux ob-

---

<sup>1</sup>Lancaster (1990, p. 267) présente un exemple de non-convergence lorsque la vraisemblance est bornée et la variance de l'effet aléatoire tend vers 0. Le problème est ici différent : les variances croissent au fur et à mesure des itérations jusqu'à ce que les coefficients retournés tendent vers  $-\infty$ . Bolstad et Manda (2001) signalent des difficultés similaires, l'augmentation de la variance causant chez eux des erreurs d'arrondis et entraînant des difficultés numériques.

servations. Ce résultat reprend une démarche similaire à celle proposée par Honoré (1993) en présence à épisodes multiples. Nous n'avons donc pas besoin de supposer que les effets sont indépendants des variables explicatives, ni qu'ils ont une espérance finie, par opposition aux applications usuelles en économétrie des modèles de durée.

Dans le prolongement de Boockmann (2001), nous utilisons une approche réduite pour étudier les déterminants observés et inobservés qui influencent la durée séparant l'adoption d'une convention de sa ratification. Dans le prolongement de ses conclusions, nous spécifions un premier effet aléatoire au niveau des pays et un second au niveau des conventions. L'idée sous-jacente est que certaines conventions sont plus faciles à ratifier que d'autres, non seulement du fait du sujet traité mais aussi car elles offrent plus de flexibilité aux pays signataires, ou sont par exemple moins complexes. Par conséquent, les observations sont corrélées entre elles pour deux raisons différentes et nous évaluons l'importance de chacune d'elles.

Lors de l'inférence, nous reprenons l'approche développée dans Kalbfleisch (1978) qui justifie la vraisemblance partielle dans le cadre Bayésien. L'estimateur est ensuite approché au moyen de l'échantillonnage de Gibbs, introduit par Gelfand et Smith (1990) et présenté dans la sous-section 1.7.1.

La première Section de ce chapitre décrit les données. Nous présentons les modèles et leur identification dans la seconde Section. Le choix des lois a priori et l'algorithme de Gibbs sont présentés dans la troisième Section. Les résultats sont présentés dans la dernière Section.

## 3.1. Les données

Les durées sont présentées et analysées dans Boockmann (2001). Les épisodes sont collectés par échantillonnage de flux sur la période 1975-1995. Les durées de la base originale sont le nombre de jours séparant l'adoption d'une convention de sa ratification. Pour faciliter les calculs, elles ont été mise en forme de la même manière que si les différents processus de ratification avaient été observées tous les 15 jours. Il est peu probable que cela affecte significativement les résultats, vu la longueur des durées présentées ci-dessous.

L'échantillon comprend 80 pays, aucun ne faisant parti de l'OCDE, que nous appelons "pays en voie de développement" pour simplifier. Les données couvrent 29 conventions pour un total de 228 ratifications. La durée moyenne avant ratification est de 8 ans, mais les épisodes sont très différents : 20% des ratifications ont lieu en moins de 3 ans et 20% au delà de 13 ans. Les variables explicatives ne restent pas constantes sur des épisodes aussi long, et

les données sont par conséquent mises en forme pour tenir compte de régresseurs variant dans le temps. Chaque épisode est scindé en sous-épisodes d'une année sur lesquels les caractéristiques observées ne changent pas (pour plus d'informations, voir Hosmer et Lemeshow, 1999, p.248-253). La distribution du nombre de ratification par pays est récapitulé dans le Tableau 3.1 :

TAB. 3.1 – Nombre de ratification par pays

Nombre de ratifications	Nombre de pays
Plus de 12	3
9-10	3
7-8	1
5-6	10
3-4	17
1-2	20
0	26
228	80

Les trois états à avoir ratifié plus de douze conventions sont le Brésil, le Mexique et l'Uruguay. On constate que 6 pays ont ratifié plus de 8 conventions tandis que 26 n'ont rien signé.

Le nombre de pays ratificateurs est très variable. La convention numéro 144 dans la classification de l'OIT, au sujet des consultations tripartites relatives aux normes internationales du travail, a été ratifiée 38 fois et la convention numéro 159, sur la réadaptation professionnelle et l'emploi des personnes handicapées, 25 fois. A l'inverse, les conventions 143 (sur les migrations dans des conditions abusives et la promotion de l'égalité de chances et de traitement des travailleurs migrants) et 157 (sur la conservation des droits en matière de sécurité sociale) n'ont été respectivement signées que par 6 et 1 pays. Nous estimons les fonctions de survie avec l'estimateur non-paramétrique de Kaplan-Meier, présenté dans la sous-section 1.1.5, et testons leur égalité avec un test du logrank.<sup>2</sup> Nous interprétons le rejet de l'hypothèse nulle comme un indice d'hétérogénéité et considérons deux effets aléatoires : un effet pays et un effet convention, coiffant l'hétérogénéité observée décrite ci dessous.

Les variables explicatives prennent différentes valeurs selon les pays, les conventions et la date. Dans le prolongement de la littérature économique et politique sur le comportement de ratification, nous répartissons les dé-

---

<sup>2</sup>Notons  $O_i$  le nombre de ratifications observées et  $E_i$  le nombre de ratifications théoriques pour le pays  $i$  ( $i = 1, \dots, I$ ). La statistique de test vaut  $\sum_i (O_i - E_i)^2 / E_i$  est suit une loi de Khi-deux à  $I - 1$  degrés de liberté.

### 3.1 Les données

---

terminants en 3 groupes : les variables relatives aux coûts économiques et administratifs, à la situation politique intérieure et aux pressions extérieures.

Le premier groupe de variables comprend le PIB par tête (tiré des Penn Tables au prix internationaux de 1985, en \$-US constants) et la proportion des exportations dans le PIB. Les travailleurs des pays les plus riches peuvent exprimer une demande pour des normes et des conditions de travail élevées, ce qui augmente la probabilité de ratification. Le PIB est pris en compte de manière linéaire et quadratique pour permettre une influence non-monotone sur le hasard. Les termes de plus haut degré ne sont pas significatifs. Deux autres variables de ce groupe sont des indicatrices. La première précise si une convention est une mise à jour d'une convention pré-existante (cette information figure dans le texte de la convention). La seconde indique si l'état membre a ratifié la convention pré-existante dont la convention étudiée est une mise à jour. On s'attend à ce qu'une ratification antérieure rende la ratification actuelle moins coûteuse. De plus, l'absence de ratification d'une convention pré-existante, si elle existe, rend la ratification moins probable comparativement au cas où aucune convention n'est pré-existante. La dernière variable de ce groupe est la population, utilisée comme approximation du coût par habitant de la ratification.

Le second groupe de variables explicatives est composé d'indicatrices du contexte politique intérieur influençant la ratification. La première, obtenue à partir de Alvarez *et al.* (1996) et réactualisée par Boockmann (2001), est une indicatrice de démocratie. La seconde, issue de différentes sources électroniques, vaut 1 si la majorité parlementaire est un parti "de gauche". Chaque État envoie des délégués nationaux (représentant le gouvernement, les employés et le patronat) aux Conférences du Bureau International du Travail où a lieu l'adoption des conventions. Deux autres variables font référence au vote des représentants nationaux du gouvernement et des employeurs lors de la conférence d'**adoption** de la convention. Elles valent 1 si les délégués ont voté contre une convention ou se sont abstenus. Le vote du représentant des syndicats n'a pas été pris en compte ici car favorable à l'adoption pour plus de 99% des observations.

Les variables captant les pressions extérieures se justifient car, en dépit du caractère volontaire de la ratification, certains pays peuvent être influencés par d'autres états ou bien des institutions internationales telles que la Banque Mondiale ou le Fond Monétaire International (FMI). Plus la dépendance à l'égard de ces pays ou organisations est élevée, plus les pressions sont possibles. Le montant de l'aide au développement reçue chaque année par un pays est la première variable explicative de ce groupe. Nous prenons également en compte les prêts du FMI et les crédits de la Banque Mondiale. Les exportations sont prises en compte, car une plus grande ouverture peut



rendre la ratification plus délicate en créant une dépendance vis-à-vis des partenaire commerciaux. Les exportations en direction des pays industrialisés sont étudiées car elles indiquent une vulnérabilité à l'égard de sanctions commerciales. En effet, ce moyen de rétorsion est vraisemblablement plus utilisé par les pays industrialisés que par les pays en voie de développement.<sup>3</sup> Comme des sanctions à l'encontre de pays exportateurs de pétrole sont moins probables, on inclut une indicatrice d'appartenance à l'OPEP que l'on multiplie avec les exportations en direction des pays industrialisés. Toutes les variables sont en pourcent du PIB.

Un problème d'endogénéité survient si les politiques commerciales, les prêts du FMI ou de la Banque Mondiale, sont utilisés pour récompenser ou sanctionner la décision de signer la convention. Pour éviter ça, nous utilisons une moyenne mobile sur 3 ans décalée d'une année pour ces variables.

Nous créons également un ensemble d'indicatrices pour prendre en compte les différences des **sujets** des conventions. Ils sont présentés dans le Tableau 3.2 , avec le nombre de conventions correspondantes. Le lecteur intéressé

TAB. 3.2 – Nombre de conventions par sujet

Sujet	Nombre de conventions
Libertés fondamentales	4
Emploi	3
Administration du travail	3
Conditions de travail	5
Travailleurs migrants	2
Travail spéciaux	3
Total	29

par de plus amples informations sur le choix des variables explicatives et leur effet attendu sur le hasard est prié de consulter Boockmann (2001), qui présente également des statistiques descriptives et d'autres estimations de Kaplan-Meier.

---

<sup>3</sup>Les pays industrialisés sont : Allemagne, Australie, Autriche, Belgique, Canada, Danemark, Espagne, Etats-Unis, Finlande, France, Grande-Bretagne, Grèce, Islande, Irlande, Italie, Japon, Luxembourg, Nouvelle Zélande, Norvège, Pays-Bas, Portugal, Suède, Suisse.

## 3.2. Les modèles

Nous considérons trois modèles MPH, en commençant par le modèle de Cox à deux effets aléatoires dont nous déduisons le modèle de Cox à un effet aléatoire et le modèle de Cox. Le modèle avec deux effets prend en compte l'hétérogénéité non-observée de la manière la plus précise possible. Le modèle avec un seul effet le spécifie au niveau des pays, car Boockmann (2001) a déjà considéré l'hétérogénéité non observée au niveau des pays et nous pouvons comparer nos résultats avec ceux qu'il a obtenu avec des effets fixes. De plus, la partition définie au niveau des pays est plus fine que celle au niveau des conventions et nous permet donc de disposer de plus d'information.

Les pays sont indicés par  $i$  ( $i = 1, \dots, I$ ) et les conventions par  $j$  ( $j = 1, \dots, J_i$ ).<sup>4</sup> Le temps séparant l'adoption de la convention  $j$  de sa ratification par le pays  $i$  est noté  $t_{ij}$ .

La fonction de hasard s'écrit :

$$\lambda_{ij}(t_{ij}|x_{ij}, v_i, w_j) = \lambda_0(t_{ij}) \exp[\beta'x_{ij}(t_{ij}) + v_i + w_j], \quad (3.1)$$

où  $v_i$  est le logarithme de l'effet aléatoire défini au niveau des pays et  $w_j$  le logarithme de l'effet au niveau des conventions. Nous étudions donc un modèle MPH multivarié, dont la formulation générale est présentée dans la sous-section 1.3.7. Les deux effets sont supposés indépendants de par l'universalité des conventions car aucun texte n'est destiné à certaines régions en particulier.

Elbers et Ridder (1982) montrent que le modèle MPH à épisodes uniques est identifié sous les hypothèses de finitude de l'espérance de la distribution mélangeante et d'indépendance entre l'hétérogénéité observée et non observée (voir la sous-section 1.4.1). Honoré (1993), dont l'approche est présentée dans la sous-section 1.4.2, montre que ces deux hypothèses peuvent être relâchées dans un modèle MPH multivarié lorsque les réalisations de l'effet aléatoire sont communes à au moins deux observations.<sup>5</sup> L'identification du modèle (3.1) est prouvée dans l'Annexe C.1, sans avoir à supposer  $E(V) < \infty$  ou  $E(W) < \infty$ , ni l'indépendance entre l'hétérogénéité observée et non observée. Intuitivement, l'un des effets aléatoires peut être tenu pour constant selon que le modèle est formulé pour un pays ou une convention donnée. En

<sup>4</sup>Du fait du mode d'échantillonnage, les conventions étudiées diffèrent d'un pays à l'autre. Une notation complète ferait appel à une fonction  $j_i$  indiquant les conventions pertinentes pour le pays  $i$  de la manière suivante :  $j_i(1), \dots, j_i(J_i)$ . Nous pensons toutefois que l'indication retenue dans ce chapitre est suffisamment explicite pour ne pas créer de confusion.

<sup>5</sup>L'approche de Honoré (1993) est conditionnelle aux variables explicatives.

alternant ces deux points de vue, on peut obtenir l'identification avec un raisonnement similaire à celui de Honoré (1993). Notre résultat est obtenu sans variables explicatives et en laissant libre la partie du hasard qui ne dépend pas de l'hétérogénéité non-observée, c'est-à-dire dans un contexte bien plus général que le modèle étudié ici. Comme le fait remarquer Van den Berg (2001), les régresseurs facilitent l'identification dans les modèles de durée (et particulièrement ceux qui varient dans le temps). Une fois le modèle identifié sans variable explicative, nous n'avons donc pas besoin de plus de structure si nous voulons ajouter ultérieurement des régresseurs variant dans le temps.

En annulant un terme d'hétérogénéité non observé, on obtient le modèle de Cox à un effet aléatoire (utilisé entre autres par Clayton, 1978, 1991, Klein, 1992 et Guo et Rodriguez, 1992). L'effet restant peut représenter l'hétérogénéité au niveau des conventions ou de pays. Comme la partition définie au niveau des pays est plus fine, nous considérons le modèle contraint :

$$\lambda_{ij}(t_{ij}|x_{ij}, v_i) = \lambda_0(t_{ij}) \exp[\beta'x_{ij}(t_{ij}) + v_i]. \quad (3.2)$$

Le modèle de Cox standard, présenté dans la sous-section 1.2.1, est obtenu en annulant les deux termes d'hétérogénéité. Les durées sont indépendantes conditionnellement aux caractéristiques observées, et la fonction de hasard est :

$$\lambda_{ij}(t_{ij}|x_{ij}) = \lambda_0(t_{ij}) \exp[\beta'x_{ij}(t_{ij})]. \quad (3.3)$$

Comme le modèle à un effet et le modèle de Cox standard sont des cas particuliers du modèle MPH à deux effets, ils sont aussi identifiés.

L'approche de la vraisemblance partielle, présentée dans la sous-section 1.6.1 en présence d'hétérogénéité non-observée, permet d'éviter tout risque de mauvaise spécification au niveau du hasard de base. La ratification de la convention  $j$  par le pays  $i$  à la date  $t_{ij}$  contribue à la vraisemblance partielle par :

$$L_{ij}(t_{ij}|v_i, w_j, \beta) = \frac{\exp[\beta'x_{ij}(t_{ij}) + v_i + w_j]}{\sum_{(k,l) \in R_{ij}} \exp[\beta'x_{kl}(t_{ij}) + v_k + w_l]}, \quad (3.4)$$

où  $R_{ij}$  est l'ensemble des individus au risque à la date  $t_{ij}$ . Le hasard de base se simplifie mais pas les effets aléatoires, car ils prennent plusieurs valeurs pour les individus au risque. La vraisemblance partielle des modèles emboîtés se déduit de 3.4 en ignorant le ou les effet(s) absent(s). La vraisemblance partielle complète est obtenue en prenant le produit des  $L_{ij}$  sur  $i$  et  $j$ .

La vraisemblance partielle est conditionnelle aux effets aléatoires et il nous reste à spécifier les distributions mélangeantes. Un grand nombre de choix sont possibles pour  $\ln v_i$  et  $\ln w_j$ , parmi lesquels figurent la loi gamma,

inverse-gaussienne, log-normale et positive stable.<sup>6</sup> On suppose que  $v_i$  et  $w_j$  suivent tous deux une loi normale, comme l'ont fait McGilchrist (1993), Sargent (1998), Yau (2001), Ripatti et Palmgren (2000), Vaida et Xu (2000) et d'autres. Deux raisons motivent ce choix. D'une part, nous choisissons une loi symétrique car nous n'avons aucune raison de penser que les effets aléatoires ont un impact positif ou négatif sur le hasard. D'autre part, la spécification d'une loi normale correspond à l'idée que l'hétérogénéité non-observée est due à un grand nombre des variables explicatives inobservées spécifiques aux pays ou aux conventions.<sup>7</sup> On suppose que les  $\{v_i\}_{i=1}^I$  et  $\{w_j\}_{j=1}^{J_i}$  sont indépendants et :

$$v_i \sim N(0, \tau^2), w_j \sim N(0, \alpha^2). \quad (3.5)$$

Les lois sont centrées pour que les effets représentent des déviations par rapport à la moyenne. Une extension possible est de supposer des variances spécifiques aux pays et aux conventions.

La vraisemblance partielle peut facilement comprendre les observations censurées, tant que le mécanisme de censure est non informatif (voir la sous-section 1.2.4). L'ensemble des durées au risque au point  $t_{ij}$  contient les épisodes censurés à la date  $t_{ij}$  ou après. Près de 90% des épisodes sont censurés et contribuent à la vraisemblance partielle au travers de son dénominateur.

## 3.3. Inférence Bayésienne

Les paramètres ont été estimés au moyen d'une approche Bayésienne. L'information véhiculée par les observations est complétée par des croyances a-priori, avant que l'estimateur Bayésien ne puisse être approché au moyen de méthodes de Monte Carlo par Chaînes de Markov (MCMC). Nous expliquons dans cette Section le choix des a-priori, la forme de la loi a posteriori et les méthodes d'échantillonnage utilisées.

### 3.3.1 Lois a priori et a posteriori

La vraisemblance partielle est justifiée dans le paradigme Bayésien par Kalbfleisch (1978), dont l'approche est présentée dans la sous-section 1.7.1. Une reformulation avec la fonction de hasard traitée dans ce chapitre figure dans l'Annexe C.2. En utilisant les processus de comptage, il montre qu'en

---

<sup>6</sup>Voir respectivement Clayton (1978), Milcent (2003), Gustafson (1997) et Hougaard pour des illustrations. Le choix d'une distribution mélangeante est également abordé dans la sous-section 1.3.5.

<sup>7</sup>Nous avons également estimé ce modèle avec des distributions mélangeantes gamma et la convergence n'a pas eut lieu.

posant un a priori gamma diffus sur le hasard de base amène à une loi a posteriori proportionnelle à la vraisemblance partielle. Ce résultat utilise un argument limite sur la loi gamma et nous supposons dans notre application que ces deux paramètres valent 0.001.<sup>8</sup> Elle admet donc une espérance de 1 et une variance de 1000, ce qui ne contraint pas le hasard car le terme  $\exp[\beta'x_{ik}(t_{ik}) + v_i + w_j]$  n'est pas borné.

L'étape suivante d'une analyse Bayésienne hiérarchique requiert l'assignation de lois a priori aux paramètres du modèle. Pour faciliter les simulations, nous utilisons des a priori propres et non informatifs, supposés indépendants pour les coefficients  $\beta$ . Une alternative est de considérer une distribution normal multivariée, mais cela induirait plus de complexité dans le modèle vu que nous aurions à spécifier une matrice de covariance pour 22 paramètres. Nous avons donc choisi une loi normale de moyenne nulle et de variance  $10^6$  pour chaque coefficient.

La dernière étape de la spécification concerne le choix des a priori sur les paramètres des distributions mélangeantes. Les précisions des lois gamma (c'est-à-dire  $\tau^{-2}$  et  $\alpha^{-2}$ ) sont supposés distribuées selon des lois gamma d'espérance unitaire et de variance  $10^3$ . La loi gamma est l'a priori conjugué pour la précision de la loi normale (voir par exemple Gouriéroux et Monfort, 1990, pp. 381-382) et ce choix accélère les calculs.

Soit  $M$  le nombre de variables explicatives observées. La loi a posteriori s'écrit :

$$\pi(\beta, \alpha, \tau | t_{ij}) \propto \left[ \prod_{i=1}^I \prod_{j=1}^{J_i} L_{ij}(t_{ij} | v_i, w_j, \beta) f(w_j | \alpha) \right] \prod_{i=1}^I f(v_i | \tau) \prod_{m=1}^M f(\beta_m) f(\alpha) f(\tau). \quad (3.6)$$

On en déduit les lois a posteriori du modèle de Cox à un effet et du modèle de Cox standard.

### 3.3.2 Estimation des paramètres

L'estimateur Bayésien de l'équation (3.6) n'admet de solution analytique. Cependant, il peut être approché au moyen des MCMC et plus particulièrement pas échantillonnage de Gibbs. Parmi les textes de référence sur les MCMC figurent Robert (1996), Neal (1997) and Robert et Casella (1999).

L'échantillonnage de Gibbs est décrit dans la sous-section 1.7.2. Soit un vecteur  $X$ , composé de  $n$  ( $n > 1$ ) variables aléatoires. L'algorithme tire des

---

<sup>8</sup>Nous faisons ici référence à la paramétrisation  $\gamma(\alpha, \lambda)$  qui a pour espérance  $\alpha/\lambda$  et pour variance  $\alpha/\lambda^2$ .

$x_i$  de la loi conditionnelle  $f_i(x_i|\{x_j\}_{j \neq i})$ , sachant la valeur courante  $x_j$ , pour  $i = 1 \dots n$ . La suite  $x_i$ , obtenue en répétant cette procédure, est une chaîne de Markov. Une fois la convergence atteinte, l'échantillonnage de Gibbs produit un échantillon simulé à partir de la loi a posteriori, à partir duquel les lois des quantités d'intérêt peuvent être estimées.

L'échantillonnage de Gibbs requiert des simulations dans les lois marginales conditionnelles, elles même utilisant des procédures numériques. Les coefficients sont simulés au moyen de l'algorithme Acceptation-Rejet (ARS), présenté dans la sous-section 1.7.2 ainsi que dans Gilks et Wild (1992) et Robert et Casella (1999). Les variances  $\tau^2$  et  $\alpha^2$  sont simulées directement à partir de tirages dans des lois inverse gamma. Le programme d'estimation est écrit pour le logiciel WinBUGS et se trouve dans l'Annexe C.3.<sup>9</sup>

## 3.4. Résultats

Nous présentons les résultats obtenus pour les trois modèles : le modèle de Cox, le modèle de Cox avec un effet aléatoire et le modèle de Cox avec deux effets. Deux chaînes de Markov avec des valeurs initiales différentes ont été simulées pour chaque modèle. Des simulations antérieures ont montré que les variances convergent plus lentement que  $\beta$ , c'est pourquoi nous fixons les valeurs initiales des coefficients à 0 pour les deux chaînes et choisissons des valeurs différentes pour  $\tau^2$  et  $\alpha^2$ . Les variances ont été fixées à 1 pour la première chaîne et à 50 pour la seconde. 11000 itérations ont été effectuées pour le modèle de Cox, 12000 pour le modèle avec l'effet pays et 35000 pour le modèle aux deux effets. La convergence a été déterminée avec les outils présentés dans Gelman et Rubin (1992) et les graphiques des quantiles, nous permettant de conclure que 5000 itérations sont nécessaires pour le "burn-in" des deux premiers modèles et 10000 pour le dernier. Les tirages dans la loi a posteriori correspondent aux itérations des deux chaînes une fois le "burn-in" terminé.

Nous avons également estimé les modèles avec des approches fréquentistes. Le modèle de Cox standard a été estimé par maximisation de la vraisemblance partielle, le modèle à un effet par maximisation de la vraisemblance partielle pénalisée présentée dans Therneau, Grambsch, et Pankratz (2003), et le modèle à deux effets par l'algorithme EMPL présenté au Chapitre 2.<sup>10</sup> Cette

---

<sup>9</sup>WinBUGS est un logiciel libre disponible à l'adresse <http://www.mrc-bsu.cam.ac.uk/bugs/>. Spiegelhalter et al. (2000) fournit le manuel et des exemples.

<sup>10</sup>Les résultats sont obtenus avec le logiciel R 2.0.1 et le package "survival" est requis pour les vraisemblances pénalisées. R est un logiciel libre disponible à <http://www.r-project.org/>.

dernière approche est stable et rapide, mais nécessite l'hypothèse d'une hétérogénéité gamma pour bénéficier pleinement de la vraisemblance pénalisée. De ce point de vue, l'approche Bayésienne est plus flexible. Toutefois, nous n'avons pas obtenu la convergence des chaînes de Markov avec des hétérogénéités gamma.

Le Tableau 3.3 montre les valeurs estimées des paramètres des distributions mélangeantes.

TAB. 3.3 – Estimations des écarts-type des distributions mélangeantes

Structure d'hétérogénéité	Paramètre	Médiane	2.5%	97.5%
Approche Bayésienne (effets gaussiens)				
Simple : effet pays	$\tau$	0.49	0.42	0.57
Double : effet pays	$\tau$	0.49	0.42	0.57
effet convention	$\alpha$	0.85	0.66	1.12
Approche fréquentiste (effets gammas)				
Simple : effet pays	$\tau$	0.53	0.45	0.68
Double : effet pays	$\tau$	0.53	0.45	0.68
effet convention	$\alpha$	0.26	0.23	0.33

Note : les résultats fréquentistes sont basés sur l'approximation asymptotique normale de la loi des paramètres estimés.

Les résultats obtenus avec une approche fréquentiste sont similaires à ceux de l'approche Bayésienne au niveau de l'effet pays. Une différence sensible apparaît avec l'effet convention. Tandis que l'approche Bayésienne nous indique un effet convention deux fois plus important que l'effet pays, l'approche fréquentiste nous donne un effet convention deux fois plus faible que l'effet pays. Nous confirmons donc les résultats de Rodriguez et Goldman (2001) et de Therneau *et al.* (2003) quant à la sous-estimation de l'hétérogénéité non observée. Ce point mérite des travaux futurs.

Le Tableau 3.4 indique les médianes et intervalles de confiance au seuil de 5% des tirages dans la loi a posteriori pour les coefficients  $\beta$  des 3 modèles, les résultats obtenus par maximisation d'une vraisemblance partielle (pénalisée lorsqu'il y a au moins un effet aléatoire) sont dans le Tableau 3.5. L'entrée "Log-vraisemblance avec pénalité" du Tableau 3.4 sert au calcul du Critère d'Information Bayésien (de l'anglais "Bayesian Information Criteai" et abrégé BIC), tandis que l'entrée "Log-vraisemblance pénalisée" du Tableau 3.5 correspond à la valeur à l'optimum de la fonction objectif.

Le premier modèle a moins de paramètres significatifs que le second et troisième modèle pour ce qui est des variables de coûts. En particulier, l'indicatrice de la ratification d'une convention dont la révision est considérée a une

### 3.4 Résultats

TAB. 3.4 – Estimations Bayésiennes des coefficients  $\beta$

Variables	Cox			Cox : 1 effet			Cox : 2 effets		
	Med.	2.5%	97.5%	Med.	2.5%	97.5%	Med.	2.5%	97.5%
<b>Coûts</b>									
PIB réel par tête <sup>a</sup>	2.01	-0.76	4.92	<b>4.00</b>	1.49	6.71	<b>3.81</b>	1.12	6.64
PIB réel par tête, au carré	-2.24	-5.46	0.77	<b>-3.32</b>	-6.25	-0.69	<b>-3.19</b>	-0.33	-6.29
Pas de mise à jour	0.53	-0.05	1.15	<b>0.98</b>	0.45	1.58	<b>1.39</b>	0.88	1.95
Ratification antérieure si mise à jour	<b>1.50</b>	0.77	2.23	<b>1.37</b>	0.64	2.10	<b>1.62</b>	0.91	2.34
Population <sup>b</sup>	-0.05	-0.14	0.03	-0.01	-0.09	0.08	-0.02	-0.11	0.07
<b>Pressions Intérieures</b>									
Démocratie	0.12	-0.18	0.43	<b>0.36</b>	0.06	0.67	<b>0.34</b>	0.04	0.64
Majorité de gauche Vote contre :	-0.39	-1.04	0.19	<b>-0.72</b>	-1.37	-0.14	<b>-0.69</b>	-0.10	-1.33
Gouvernement	-0.15	-0.59	0.33	-0.18	-0.62	0.28	-0.22	-0.66	0.24
Employeurs	-0.15	-0.56	0.28	0.25	-0.14	0.68	<b>0.38</b>	0.01	0.79
<b>Pressions Extérieures</b>									
Aides au développement <sup>c</sup>	<b>-0.06</b>	-0.11	-0.03	<b>-7.43</b>	-11.69	-3.55	<b>-7.65</b>	-3.81	-11.85
Prêts de la Banque Mondiale <sup>c</sup>	<b>-4.20</b>	-7.63	-0.82	2.55	-0.59	5.71	2.00	-1.11	4.97
Crédits du FMI <sup>c</sup>	<b>6.23</b>	2.03	10.10	3.61	-0.24	7.29	3.96	-0.09	7.62
Exportations <sup>c</sup>	-0.76	-3.64	1.45	-0.78	-3.54	1.12	-0.79	-3.76	1.21
Exportations vers les pays industrialisés <sup>c</sup>	-0.07	-6.37	6.29	0.22	-6.20	6.61	-0.18	-7.43	6.96
Exportations vers les pays industrialisés (hors pays de l'OPEP) <sup>c</sup>	-0.55	-6.13	5.28	-1.10	-7.14	5.00	-0.77	-7.51	6.07
Pays de l'OPEP	0.19	-0.91	1.30	0.26	-0.87	1.50	0.25	-1.04	1.65
Log-vraisemblance avec pénalité		-1728			-1630			-1536	
Sujet des conventions		Oui			Oui			Non	
Nombre d'épisodes		2349			2349			2349	
Nombre de ratifications		228			228			228	

Note : Les entrées en gras sont significatives au seuil de 5%. *a.* Prix internationaux de 1985, en 10 000\$.

*b.* Centaines de millions. *c.* Pourcents du PIB. Les autres variables sont des indicatrices, et des indicatrices des sujets des conventions sont incluses pour les 2 premiers modèles.

influence positive significative comme nous l'attendions. De même, l'absence de ratification d'une convention ayant fait l'objet d'une mise à jour influence négativement le hasard pour la convention révisée. Ces deux résultats laissent penser que le processus de ratification suit une dynamique d'une convention à



3 Une approche Bayésienne du comportement de ratification des conventions de l'OIT

TAB. 3.5 – Estimations fréquentistes des coefficients  $\beta$

Variables	Cox			Cox : 1 effet			Cox : 2 effets		
	Coef.	2.5%	97.5%	Coef.	2.5%	97.5%	Coef.	2.5%	97.5%
<b>Coûts</b>									
PIB réel par tête <sup>a</sup>	<b>3.84</b>	1.10	6.58	<b>3.94</b>	1.20	6.67	<b>3.06</b>	0.55	5.57
PIB réel par tête, au carré	<b>-3.14</b>	-6.08	-0.20	<b>-3.19</b>	-0.27	-6.12	-2.47	-5.29	0.35
Pas de mise à jour	<b>0.96</b>	0.40	1.52	<b>0.90</b>	0.33	1.46	<b>0.90</b>	0.37	1.43
Ratification antérieure si mise à jour	<b>1.37</b>	0.66	2.08	<b>1.38</b>	0.67	2.10	<b>1.38</b>	0.68	2.09
Population <sup>b</sup>	0.01	-0.08	0.09	0.01	-0.08	0.10	-0.03	-0.13	0.07
<b>Pressions Intérieures</b>									
Démocratie	<b>0.37</b>	0.06	0.67	<b>0.39</b>	0.09	0.69	0.27	-0.03	0.57
Majorité de gauche	<b>-0.68</b>	-1.28	-0.07	<b>-0.69</b>	-1.29	-0.08	<b>-0.62</b>	-1.23	-0.01
Vote contre :									
Gouvernement	-0.19	-0.66	0.27	-0.13	-0.60	0.33	-0.09	-0.54	0.36
Employeurs	0.23	-0.17	0.64	0.26	-0.15	0.67	0.24	-0.15	0.63
<b>Pressions Extérieures</b>									
Aides au développement <sup>c</sup>	<b>-7.22</b>	-11.25	-3.19	<b>-7.22</b>	-11.26	-3.18	<b>-8.55</b>	-4.41	-12.69
Prêts de la Banque Mondiale <sup>c</sup>	2.62	-0.41	5.65	2.74	-0.30	5.78	<b>3.27</b>	0.21	6.33
Crédits du FMI <sup>c</sup>	3.65	-0.14	7.44	3.49	-0.29	7.27	3.61	-0.25	7.47
Exportations <sup>c</sup>	-0.16	-2.87	2.56	-0.09	-2.72	2.54	0.23	-1.89	2.35
Exportations vers les pays industrialisés <sup>c</sup>	-0.84	-8.10	6.43	-0.85	-8.10	6.39	-2.50	-9.93	4.93
Exportations vers les pays industrialisés (hors pays de l'OPEP) <sup>c</sup>	-0.67	-7.43	6.10	-0.78	-7.57	6.01	0.56	-4.08	5.20
Pays de l'OPEP	0.13	-1.19	1.44	0.16	-1.16	1.48	-0.02	-0.82	0.86
Log-vraisemblance pénalisée		-2203			-2156			-2143	
Sujet des conventions		Oui			Oui			Non	
Nombre d'épisodes		2349			2349			2349	
Nombre de ratifications		228			228			228	

Note : Les entrées en gras sont significatives au seuil de 5%. *a.* Prix internationaux de 1985, en 10 000\$.

*b.* Centaines de millions. *c.* Pourcents du PIB. Les autres variables sont des indicatrices, et des indicatrices des sujets des conventions sont incluses pour les 2 premiers modèles.

l'autre. Les modèle à effets aléatoires indique une influence non-monotone du PIB réel par tête.<sup>11</sup> En comparant avec les résultats de Boockmann (2001),

<sup>11</sup>Pour les deux modèles avec hétérogénéité non-observée, l'influence du PIB est une  $\cup$  inversée avec un maximum à un PIB par tête de 6000\$. La relation est ainsi croissante et

### 3.4 Résultats

---

on constate que les variables significatives de coûts sont les même dans les deux études.<sup>12</sup>

Concernant les variables de pressions internes, aucune d'entre elles n'est significative dans la spécification sans hétérogénéité non-observée. Prendre en compte au moins un effet aléatoire affecte les résultats. L'indicatrice de démocratie a un impact positif, ce qui est plausible car les agents dont les conditions de travail sont améliorées par la convention étudiée, comme les exploitants agricoles ou les ouvriers, sont vraisemblablement mieux représentés dans une démocratie que dans un régime autoritaire. De façon surprenante, l'indicatrice d'une majorité de gauche a une influence négative. Ce résultat vient du fait que la distinction entre les partis de gauche et de droite est souvent inadaptée pour capter les politiques intérieures des pays en voie de développement. Le vote des délégués du gouvernement lors de l'adoption n'a aucune influence sur la ratification. A l'inverse, un vote négatif du représentant des employeurs augmente la probabilité de ratification. Les conventions concernées sont vraisemblablement celles qui concernent les normes de travail les plus avancées. Ainsi, les syndicats peuvent s'organiser pour peser plus lourdement en faveur de la ratification.<sup>13</sup> A contrario, Boockmann (2001) ne trouve pas de coefficient significatif parmi les variables de pressions internes lorsqu'il prend en compte l'hétérogénéité non-observée.

Le groupe suivant de variables détecte les pressions externes. L'aide au développement a une influence négative sur le hasard, en particulier lorsque l'hétérogénéité non-observée est prise en compte. Une explication est peut-être que les pays recevant une aide importante ont également des problèmes économiques qui n'apparaissent pas dans les autres variables. Tout comme dans l'étude de Boockmann, les prêts de la Banque Mondiale désincitent à la ratification dans le premier modèle avant de devenir non-significatif dans les modèles à effets aléatoires.<sup>14</sup> Les crédits du FMI ne sont pas non plus significatifs. Enfin, il n'y a pas d'impact de l'exposition à d'éventuelles sanctions commerciales mesurée au travers des exportations en direction des pays industrialisés. Globalement, les résultats Bayésiens suggèrent que d'éventuelles pressions extérieures n'interviennent pas dans le processus de ratification.

---

concave, mais la concavité n'a ici qu'une influence réduite dans la mesure où les pays en voie de développement de l'échantillon ont un PIB par tête moyen de 2280\$.

<sup>12</sup>Les résultats pour le modèle de Cox standard diffèrent d'avec Boockmann (2001) par la modélisation de l'hétérogénéité et la présence du temps calendaire parmi les régresseurs chez ce dernier.

<sup>13</sup>Remarquons que la variable correspondant au vote du représentant des syndicats a été omise, pour les raisons précisées dans la Section 3.1.

<sup>14</sup>On remarque ici une différence avec les estimations fréquentistes, car le Tableau 3.5 indique un coefficient non significatif pour les prêts de la Banque Mondiale, hormis dans le modèle à 2 effets où il est positif.

Les résultats du modèle de Cox sont étonnamment différents selon qu'ils aient été obtenus par l'approche Bayésienne ou maximum de vraisemblance. Une explication possible est la présence de durées simultanées, gérées par l'approximation de Breslow (Breslow, 1974) dans le cadre Bayésien et l'approximation d'Efron (Efron, 1977) dans le cadre fréquentiste.<sup>15</sup> Hertz-Picciotto et Rockhill (1997) montrent l'approximation de Breslow sous-estime les coefficients  $\beta$  dans le modèle de Cox standard tandis que l'approximation d'Efron est moins biaisée mais plus coûteuse en temps de calcul. Nous n'observons pas une telle différence pour les autres modèles, les résultats Bayésiens n'étant pas sous-estimés par rapport aux fréquentistes en présence d'hétérogénéité non observée.

Afin de comparer les modèles au moyen du Critère d'Information Bayésien (BIC, Schwarz, 1978), nous avons calculé des log-vraisemblances avec une pénalité égale à la log-vraisemblance moins la moitié du nombre de paramètres du modèle, multiplié par le logarithme du nombre d'observations. Le BIC a été développé pour trouver le modèle le plus probable compte tenu des données tout en prenant en compte le rasoir d'Occam, c'est-à-dire que le modèle le plus parcimonieux est choisi lorsque deux modèles correspondent aussi bien aux données. Wasserman (2000) indique que sous conditions de régularité peu contraignantes, le BIC fournit une approximation du facteur de Bayes. Si les a priori de différents modèles sont identiques, le facteur de Bayes est le ratio a posteriori d'un modèle sur l'autre. Le BIC s'obtient comme les différences des log-vraisemblances pénalisées reportées dans le Tableau 3.4, et vaut 98 lorsque l'on compare le modèle de Cox au modèle avec un effet, 94 lorsqu'on compare le modèle à 2 effets avec celui qui n'en a qu'un seul.<sup>16</sup> Le facteur de Bayes entre les deux derniers modèles est donc de  $\exp(94)$ , ce qui fournit une forte évidence en faveur du modèle à deux effets. Ce résultat n'est pas évident ex ante car le modèle à deux effets ne prend pas en compte l'indicatrice du thème de la convention considérée, qui figure dans les autres modèles.

## Conclusion

Nous utilisons une approche Bayésienne pour estimer un modèle MPH avec différentes spécification de l'hétérogénéité non observée, la plus détaillée comprenant deux effets aléatoires. Pour éviter tout risque de mauvaise spé-

---

<sup>15</sup>Le problème des durées simultanées, leur impact sur la vraisemblance partielle et les solutions proposées sont présentés dans la sous-section 1.2.8.

<sup>16</sup>La formule est :  $BIC_{ij} = \ln L_i - \ln L_j + \frac{d_j - d_i}{2} \ln n$  où  $i$  et  $j$  sont les indices du modèle,  $L$  la vraisemblance,  $d$  le nombre de paramètres et  $n$  la taille de l'échantillon.

### 3.4 Conclusion

---

cification du hasard de base, nous utilisons la vraisemblance partielle de Cox au lieu des spécifications paramétriques usuelles. Elle a justifiée dans un cadre Bayésien par Kalbfleisch (1978). Après avoir vérifié l'identification du modèle, nous procédons à l'inférence avec l'échantillonnage de Gibbs. Pour effectuer des tirages dans les lois marginales conditionnelles, nous utilisons également d'autres procédures comme l'algorithme d'Acceptation-Rejet.

Nos résultats confirment l'importance de prendre en compte l'identité du pays lors de la ratification. Certains déterminants inobservés spécifiques au pays semblent influencer grandement le comportement de ratification. Les résultats Bayésiens indiquent la présence d'une hétérogénéité non observée encore plus forte au sein des conventions. Ils diffèrent des résultats fréquentistes, où l'hétérogénéité entre les conventions est plus faible qu'entre les pays. Certaines sont très concensuelles parmi les états membres de l'OIT ou n'impliquent pas de coûts économique ou politique élevé. D'autres conventions sont ratifiées plus difficilement, car plus ambitieuses et coûteuses. Passer d'une approche fréquentiste à une approche Bayésienne n'influence pas les autres résultats, car les conclusions que nous obtenons confirment celles de précédents travaux, et en particulier Boockmann (2001). Toutefois, les approches fréquentistes n'évaluent pas précisément la quantité d'hétérogénéité non observée et seule l'approche Bayésienne semble pouvoir le faire à l'heure actuelle.<sup>17</sup>

Les approches Bayésiennes permettent de gérer facilement les modèles à plusieurs effets aléatoires. On peut alors spécifier une structure d'hétérogénéité plus complexe, permettant des interactions plus riches. Nous avons supposés dans ce Chapitre que les deux effets aléatoires sont indépendants. Nous étudions dans le Chapitre suivant un modèle MPH où les deux termes d'hétérogénéité sont corrélés, que nous appliquons à la mobilité professionnelle sur des données d'appariement employeur-employé.

---

<sup>17</sup>Une comparaison précise des approches fréquentistes et Bayésiennes demande la même spécification des distributions mélangeantes. Ce n'est pas malheureusement possible avec les données étudiées ici, car l'estimateur Bayésien ne converge pas en présence d'hétérogénéité gamma tandis que les méthodes de vraisemblances pénalisées reposent lourdement sur l'hypothèse d'une hétérogénéité gamma.



# La mobilité professionnelle au Portugal : une approche Bayésienne avec données d'appariement employeurs-employés

---

<sup>0</sup>Ce chapitre est tiré de Horny, G., R. Mendes, et G. Van den Berg (2006) : “Job mobility in Portugal : a Bayesian empirical study with matched worker-firm data,” Mimeo.

## Résumé

Nous analysons la mobilité professionnelle avec un modèle MPH multivarié en temps discret. Il comprend deux effets aléatoires corrélés, l'un au niveau de l'entreprise et l'autre au niveau des individus. L'inférence est Bayésienne et nous utilisons des données d'appariement employeurs-employés.

Nos résultats confirment l'importance de l'hétérogénéité non observée au niveau des individus et indiquent une hétérogénéité toute aussi importante au niveau des firmes. De plus, la capacité descriptive du modèle s'accroît lors de la prise en compte d'un mécanisme d'appariement basé sur les inobservés.

## Sommaire

---

<b>Résumé . . . . .</b>	<b>128</b>
<b>Introduction . . . . .</b>	<b>128</b>
<b>4.1 Le modèle . . . . .</b>	<b>128</b>
<b>4.2 Les données . . . . .</b>	<b>128</b>
<b>4.3 Inférence Bayésienne . . . . .</b>	<b>128</b>
<b>4.4 Résultats . . . . .</b>	<b>128</b>
<b>Conclusion . . . . .</b>	<b>128</b>

---

## Introduction

L'EMPLOI et la mobilité professionnelle sont caractérisés par de nombreux faits stylisés : l'emploi de longue durée est courant, de nombreux nouveaux emplois disparaissent rapidement et la probabilité de fin d'un emploi décroît avec le temps passé en poste. Farber (1999) présente une revue de la littérature empirique pour l'Europe et les pays de l'OCDE, et le cas du Portugal est étudié en particulier par Lima (2004) avec les données que nous utilisons ici.

La présence d'hétérogénéité non observée dans la probabilité de transition peut expliquer ces faits-stylisés. Supposons que la population soit divisée en différents types de travailleurs, plus moins susceptibles d'être en emploi du fait de caractéristiques observées par la firme mais inobservées de l'économètre. Les travailleurs les plus mobiles connaissent des épisodes courts et nombreux, tandis que les autres connaissent des relations d'emploi moins nombreuses mais plus longues. La fin rapide des nouveaux emplois s'explique alors par la présence de travailleurs très mobiles, et les emplois de longue durée par les individus stables. Le déclin de la probabilité de transition observée tient à un effet d'agrégation : les individus aux longues relations d'emplois sont ceux aux caractéristiques les moins associées à la mobilité professionnelle, les autres ayant déjà effectué une transition.<sup>1</sup>

L'importance de l'hétérogénéité non observée entre les employé est bien établie à l'heure actuelle (Farber, 1999, Bellmann *et al.*, 2000 et Del Boca et Sauer, 2006, parmi d'autres). Il n'en va pas de même pour l'hétérogénéité au niveau des entreprises, la littérature empirique à ce sujet étant plus réduite. L'étude récente de Abowd *et al.*, 2006 est à notre connaissance la seule à inclure simultanément un effet au niveau des employeurs et un second des employés. Les résultats indiquent une forte hétérogénéité, traduisant des politiques de gestion du personnel très différentes au sein des entreprises françaises.

Comme la mobilité professionnelle dépend à la fois de l'employeur et de l'employé, elle peut être simultanément affectée par l'hétérogénéité non observée de chacune des parties prenante. Il s'agit alors, entre autres, d'une propension constante de l'individu à changer d'emploi et d'une préférence de l'entreprise pour une main d'oeuvre plus ou moins stable. Nous décrivons la mobilité professionnelle avec un modèle de mélange de hasards proportion-

---

<sup>1</sup>Le modèle où l'hétérogénéité non observée suffit à expliquer les deux faits stylisés est une extension du modèle "mover-stayer" introduit par Blumen *et al.*, 1955 et étendu par Goodman (1961) et Spilerman (1972). Les travaux empiriques sont nombreux et comptent Dormont, Fougère, et Prieto (2001) et Fougère et Kamionka (2003) pour l'étude du cas de la France.



nels (modèle MPH, voir la Section 1.3) où l'hétérogénéité non observée a une spécification très souple. Les transitions dépendent des caractéristiques inobservées des firmes et des individus, représentées par des effets aléatoires. De plus, comme le processus d'appariement peut ne pas être uniquement aléatoire et dépendre de caractéristiques inobservées par l'économètre, nous permettons à l'effet firme d'être corrélé avec l'effet individuel. Une entreprise est reliée longitudinalement et en coupe transversale à plusieurs travailleurs, mais un individu est relié longitudinalement à plusieurs firmes. Notre approche permet une structure d'interdépendance souple car les effets ne sont ni emboîtés, ni indépendants. Du fait des interactions au sein de l'hétérogénéité non observée, nous utilisons une approche Bayésienne dans le prolongement de Manda et Meyer (2005) et du Chapitre 3. Ce travail apporte une contribution méthodologique en montrant comment gérer cette construction complexe de l'hétérogénéité non observée durant l'inférence.

La modélisation de l'hétérogénéité non observée dans l'étude de la mobilité professionnelle a progressé avec la disponibilité de données riches sur le marché du travail et plus précisément des données d'appariement employeur-employé. Il existe de vastes littératures portant sur le rôle du capital spécifique d'une entreprise ou l'hétérogénéité non observée parmi les employés. Un courant récent s'intéresse à l'estimation d'équations simultanées de salaires et de durées d'emploi (Buchinsky *et al.*, 2005, Dostie, 2005, Abowd *et al.*, 2006 et Beffy *et al.*, 2006), utilisant les spécifications les plus flexibles à l'heure actuelle. Tous ces travaux incluent un effet employé dans l'équation représentant la mobilité. Dostie (2005) inclue un effet au niveau de l'épisode de travail et Abowd *et al.* (2006) un effet entreprise, ce dernier captant la préférence de l'entreprise à employer un personnel plus ou moins stable.

Le chapitre est organisé en cinq sections. Les données sont présentées dans la Section 4.1. Nous discutons le modèle en temps discret à deux effets aléatoires corrélés et ses cas particuliers (c'est-à-dire les modèles à deux effets indépendants, un effet ou aucun), dans la Section 4.2. Le choix des a-priori est présenté dans la Section 4.3. Les résultats sont discutés dans la Section 4.4.

## 4.1. Les données

Les données sont extraites des *Quadros de Pessoal* (littéralement "Tableaux du personnel"), une base de données collectée par le ministère portugais du travail et de la solidarité portant sur les employeurs et employés. Les données sont collectées *via* des questionnaires que les entreprises avec des employés sont contraintes par la loi de remplir tous les ans. Les données

portent sur tous les salariés de l'entreprise lors du retour des questionnaires (fixé au mois de Mars jusqu'en 1993, puis Octobre après 1994). Le secteur agricole est peu couvert de par le faible nombre de salariés y opérant, et les administrations publiques ainsi que les services à domiciles ne sont pas couverts du tout. Par contre, les secteurs de l'industrie et des services sont couverts de façon presque exhaustive.

Un identifiant est assigné à toute entreprise la première fois qu'elle reçoit le questionnaire, et l'identifiant de l'employé est une modification de son numéro de sécurité social. Avec ces deux numéros, on peut associer employeurs et employés et les suivre au fil du temps afin d'identifier les transitions d'un emploi vers un autre. Pour éviter tout problème de condition initiales, nous avons reconstruit les données comme si elles avaient été collectées par échantillonnage de flux, c'est-à-dire en ne conservant que les épisodes d'emploi dont le commencement est observé.

Le ministère a modifié la procédure d'attribution d'identifiants aux firmes en 1994, et nous sommes limités à la période 1994-2000. Les données fournissent toutefois un cliché comprenant près de 77 000 entreprises et 750 000 employés.

Nous avons extrait un échantillon de 7307 employeurs et 10 139 employés, dont les caractéristiques sont identiques à celles de la population mère, pour simplifier les calculs. Les épisodes d'emploi terminés par une transition observée sont présentés en années dans le Tableau 4.1. Les durées sont courtes :

TAB. 4.1 – Durées des épisodes d'emploi

Durée	Pourcentages
4 et plus	7
3	9
2	20
1	64
Total	100

Note : les durées sont en années.

95% des épisodes font 2 ans ou moins. On retrouve ici le fait stylisés des nouveaux emplois se terminent rapidement et que les taux de transition diminuent avec le temps en poste.

L'enquête ne comprend pas d'information sur l'histoire de l'employé entre deux questionnaires, pas plus que la date de départ d'une entreprise. Les transitions sont identifiées sur un intervalle d'une année, sans exclure des épisodes de courtes durées (d'emploi, de chômage ou de non participation) dans l'intervalle. Le Tableau 4.2 indique le nombre d'épisodes par individu.

On observe peu de transitions : seul 2% des employés ont connu 3 employeurs ou plus. En effet, l'enquête ne porte pas sur les travailleurs temporaires.

TAB. 4.2 – Nombre d'épisodes par individu

Nombre d'épisodes	Pourcentage
3 et plus	2
2	13
1	85
Total	100

Un individu est caractérisé par l'âge, le sexe, le niveau d'études et le temps de travail hebdomadaire. L'âge capture les effets de cycles au court de la vie, le nomadisme professionnel se trouvant plutôt concentré chez les jeunes peu informés de leurs propres capacités ou bien des caractéristiques du marché du travail (Johnson, 1978). Les individus sont regroupés en classes : les 16-25, 26-35 et 36-55 ans. Les employés de plus de 55 ans sont omis à cause des transitions emploi-retraite, qui sont dehors du champs de l'étude. La prise en compte du genre peut traduire différents comportements sur le marché du travail, résultant d'arbitrages entre vie privée et vie professionnelle par exemple. Nous incluons donc une indicatrice pour les femmes. Nous contrôlons de plus pour les niveaux d'études, agrégés en trois groupes : études primaires, secondaires et post-baccalauréat. Une indicatrice de travail à temps partiel est définie, car on peut penser qu'une entreprise confrontée à un choc réduira en premier lieu le nombre de travailleurs à temps partiels pour réduire la perte de capital humain.

Nous ne prenons pas en compte les entreprises qui disparaissent de l'enquête, pour éliminer les transitions dues à une cessation d'activité. Parmi les caractéristiques observées d'une entreprise figurent le secteur d'activité, la localisation et une indicatrice d'implantation sur plusieurs sites. Nous cherchons à capturer les spécificités d'un secteur ou d'une région. Le salaire est inclus dans les déterminants observés de la mobilité professionnelle. Il peut être vu comme une caractéristique de la firme et donc une variable exogène. A contrario, le salaire peut être influencé par la mobilité professionnelle et donc endogène. Nous avons donc estimé les modèles avec et sans le salaire comme variable explicative.

## 4.2. Modèles

Les employés peuvent entrer et sortir de la main d'oeuvre d'une entreprise à tout moment et les durées sont continues. Du fait du mode d'échantillonnage, nous ne les observons toutefois que sous forme discrète. Nous spécifions donc un modèle MPH en temps discret, c'est-à-dire que nous utilisons la fonction de lien log-log comme décrit dans Kalbfleisch et Prentice (1980). La fonction de lien logit constitue une alternative, inappropriée ici dans la mesure où le processus sous-jacent est continu et la discrétisation du temps ne vient que du mode de collecte des données. De plus, les résultats du modèle Logit sont sensibles à un changement des intervalles constituant l'axe du temps (Firth *et al.*, 1999). Spécifier une fonction de lien log-log nous amène à considérer différentes formulations du modèle MPH, comme présenté dans Grilli (2005), qui diffèrent par la paramétrisation du hasard de base.

Nous présentons différents modèles appartenant à la famille des modèles MPH en temps discret. Le plus simple ne prend en compte que l'hétérogénéité observée, tandis qu'un second comprend un effet aléatoire au niveau de l'employé. Nous l'étendons ensuite de manière à inclure deux effets. Chaque terme d'hétérogénéité non observée a ses réalisations communes à plusieurs épisodes et capte les relations de dépendances possibles entre les durées. Le modèle à deux effets prend en compte les interdépendances de la manière la plus fine possible, tandis que le modèle sans effet aléatoire suppose les durées reliées uniquement au travers des observables.

Dans notre application, une entreprise est reliée longitudinalement et en coupe transversale à plusieurs salariés, tandis que chaque individu n'est relié à plusieurs firmes que longitudinalement. Il n'y a pas de structure hiérarchique figée dans la population : une entreprise a plusieurs employés à la date courante, mais ils changent d'employeurs alors que le temps passe. Nous utilisons donc les notations suivantes : soit  $i = 1, \dots, I$  l'indice d'une entreprise et  $j = 1, \dots, J$  l'indice d'un individu.

Les principaux modèles de durée en temps discret sont présentés, entre autres, dans Hosmer et Lemeshow (1999, p. 241-270). Supposons que l'axe du temps soit divisé en intervalles  $[a_{k-1}, a_k[$  où  $0 = a_0 < a_1 < \dots < a_K < \infty$ . Les durées en temps discret  $t_{ijk}$  sont dans l'ensemble  $\{1, \dots, K\}$  et indiquent une transition dans l'intervalle de temps  $[a_{k-1}, a_k[$ . La fonction de hasard est une probabilité conditionnelle, à l'inverse du temps continu où il s'agit d'une probabilité limite, et peut s'écrire :

$$\lambda [t_{ijk} | x_{ij}(t_{ij(k-1)}), v_i, w_j] = p[a_{k-1} < T < a_k | T \geq a_{k-1}, x_{ij}(t_{ij(k-1)}), v_i, w_j], \quad (4.1)$$

où  $x_{ijk}(t_{ij(k-1)})$  représente les variables explicatives observées au niveau des

individus et des entreprises,  $v_i$  un effet aléatoire au niveau des entreprises et  $w_j$  un effet aléatoire au niveau des individus.<sup>2</sup>

Le hasard d'un modèle MPH sans hétérogénéité non observée a pour expression :<sup>3</sup>

$$\lambda [t_{ijk}|x_{ij}(t_{ij(k-1)}), \beta_0, \beta_1] = 1 - \exp \left( - \exp[\beta_{0(k-1)} + x_{ij}(t_{ij(k-1)})'\beta_1] \right), \quad (4.2)$$

où  $\beta_{0(k-1)}$  est le hasard de base intégré sur l'intervalle  $[a_{k-1}, a_k[$ . Le modèle incluant l'effet employé a pour fonction de hasard :

$$\lambda [t_{ijk}|x_{ij}(t_{ij(k-1)}), \beta_0, \beta_1, w_j] = 1 - \exp \left( - \exp[\beta_{0(k-1)} + x_{ij}(t_{ij(k-1)})'\beta_1 + w_j] \right). \quad (4.3)$$

Un modèle en temps discret à deux effets aléatoires est défini comme :

$$\lambda [t_{ijk}|x_{ij}(t_{ij(k-1)}), \beta_0, \beta_1, v_i, w_j] = 1 - \exp \left( - \exp[\beta_{0(k-1)} + x_{ij}(t_{ij(k-1)})'\beta_1 + v_i + w_j] \right). \quad (4.4)$$

Le départ de l'individu  $j$  de l'entreprise  $i$  à la date  $t_{ijk}$  contribue à la vraisemblance par :

$$L_{ij}^d(t_{ijk}|\beta_0, \beta_1, v_i, w_j) = \lambda_{ijk} \prod_{s=1}^{k-1} (1 - \lambda_{ijs}). \quad (4.5)$$

Un épisode censuré à la date  $t_{ijk}$  contribue à la vraisemblance :

$$L_{ij}^c(t_{ijk}|\beta_0, \beta_1, v_i, w_j) = \prod_{s=1}^k (1 - \lambda_{ijs}). \quad (4.6)$$

---

<sup>2</sup>Les régresseurs variant dans le temps de l'épisode  $ijk$  sont mesurés à la date  $t_{ij(k-1)}$ , afin de constituer un processus prédictible. Leurs valeurs ne sont influencées que par ce qui s'est produit juste avant  $t_{ijk}$ , afin d'éviter un problème potentiel d'endogénéité (voir Section 1.2.2).

<sup>3</sup> Le modèle Logit est souvent utilisé comme alternative (voir Firth *et al.*, 1999, Biggeri *et al.*, 2001, Manda et Meyer, 2005, parmi d'autres). MacCullagh et Nelder (1996, p. 107-110) indique que les deux modèles donnent des résultats similaires lorsque les probabilités de transition sont de moins de 0.15. Le modèle logit n'est toutefois pas invariant à une modification de la longueur des intervalles de temps, et un changement dans les dates de collecte des données ou même dans l'échelle du temps influence les résultats (Rodriguez, 2001).

La vraisemblance complète est :

$$L(t|\beta_0, \beta_1, v, w) = \prod_{i=1}^I \prod_{j=1}^J \prod_{k=1}^K \lambda_{ijk}^{\delta_{ijk}} (1 - \lambda_{ijk})^{1-\delta_{ijk}}, \quad (4.7)$$

où  $\delta_{ijk}$  est une indicatrice de transition. La vraisemblance (4.7) est équivalente à celle d'un modèle où les  $\delta_{ijk}$  sont des variables aléatoires de Benouilli. On déduit de la relation (4.7) les contributions des sous-modèles en omettant les effets aléatoires correspondants.

### 4.2.1 Spécification des hétérogénéités non observées corrélées

La fonction de hasard est conditionnelle à l'hétérogénéité non observée et nous spécifions les distributions mélangeantes. Elles sont supposées continues et étendent ainsi l'hétérogénéité dichotomique du modèle "mover-stayer" originel de Blumen *et al.* (1955).

Les modèles à un effet aléatoire supposent généralement les réalisations issues de tirages indépendants dans une distribution gaussienne ou log-gamma (Firth *et al.*, 1999 et Conaway, 1990, respectivement). Lindeboom et Van den Berg (1994) montrent que la distribution mélangeante influence l'évolution du hasard, et que son évolution peut être contrainte par une spécification malheureuse. Les lois marginales des durées n'ont pas de restriction sur leur support lorsque l'hétérogénéité est issue d'une loi normale multivariée. Nous supposons donc que l'effet employé du modèle (4.3) tel que :

$$w_j \underset{\text{i.i.d}}{\sim} N(0, \sigma_w^2). \quad (4.8)$$

Chaque  $w_j$  est commun à tous les épisodes de l'employé  $j$ . Nous estimons le modèle le modèle (4.4) avec deux effets aléatoires indépendants, distribués selon les lois normales univariées de la relation (4.8) et :

$$v_i \underset{\text{i.i.d}}{\sim} N(0, \sigma_f^2). \quad (4.9)$$

Les deux effets ne sont vraisemblablement pas indépendants. En effet, Mendes *et al.* (2005) montrent qu'il existe un fort appariement dans ces données, c'est-à-dire que les employeurs et les employés tendent à s'associer sur la base de facteurs inobservés. Nous supposons donc que les deux effets sont issus d'une loi normale multivariée et peuvent être corrélés. Nous estimons un modèle défini par la fonction de hasard (4.4) et la distribution mélangeante

obtenue comme le produit de la densité marginale de l'effet entreprise avec la densité de l'effet employé conditionnelle à l'effet employeur :

$$v_i \underset{\text{i.i.d.}}{\sim} N(0, \sigma_f^2), \quad (4.10)$$

$$w_j | v_1, \dots, v_I \underset{\text{i.i.d.}}{\sim} N\left(\frac{\rho\sigma_w}{\sigma_f} \left(\sum_{l=1}^I \delta_{lj} v_l\right), \sigma_w^2 \left(1 - \rho^2 \sum_{l=1}^I \delta_{lj}\right)\right), \quad (4.11)$$

où  $\sigma_w^2$  est la variance de la loi marginale de l'effet individu,  $\sigma_f^2$  la variance de la loi marginale de l'effet entreprise et  $\rho$  le coefficient de corrélation entre  $v_i$  et  $w_j$ . La variable  $\delta_{ij}$  est une indicatrice prenant la valeur 1 si l'individu  $j$  est à un moment employé par l'entreprise  $i$ , 0 sinon. L'équation (4.11) est justifiée dans l'Annexe D.1.

Au fil du temps, un individu peut travailler dans plusieurs entreprises et une entreprise peut changer son personnel. Supposons qu'un individu connaisse plusieurs firmes et que ses caractéristiques inobservées sont fortement associées à celles de l'entreprise. Dans ce cas, les déterminants inobservés au niveau des entreprises sont reliés. Inversement, si les caractéristiques inobservées des entreprises ne sont pas corrélées, l'individu ne peut pas être relié aux déterminants inobservés des entreprises où il a travaillé. Une formulation plus générale du modèle suppose 3 corrélations : une première entre les  $v_i$ , une seconde entre les  $w_j$  et une troisième reliant les effets employeurs et employés. Il faut tirer les effets dans une loi multivariée dont la dimension est égale à la somme du nombre d'entreprises avec le nombre d'individus, et dont la matrice de variance est structurée car ne dépendant que de deux variances et trois corrélations. Chaque élément doit satisfaire un nombre de contraintes croissant avec la dimension de la matrice afin qu'elle soit définie positive. Compte tenu du nombre d'individus et de firmes dans notre échantillon, on conclut à partir d'une étude numérique que les corrélations doivent être comprises entre 0 et 0.1.

Pour conserver les hypothèses d'indépendance des effets entre les firmes et entre les individus, c'est-à-dire de  $v_i$  indépendants et de  $w_j$  indépendants, nous limitons le nombre d'individus par entreprise et le nombre d'entreprises connues par un individu. Nous extrayons donc un sous-échantillon, que nous l'échantillon contraint, qui ne contient pas d'épisode au delà du troisième tiré pour un individu donné et pas d'employé au delà du troisième tiré pour une entreprise donnée. Nous limitons donc les épisodes à au plus 3 pour un individu et au plus 3 individus par entreprise. L'échantillon contraint compte 7307 entreprises et 10 139 individus et l'échantillon non contraint 7325 entreprises et 8468 employés. Les Tableaux D.1 et D.2, de l'Annexe D.2, présentent respectivement les durées et le nombre d'épisodes dans tous les échantillons,

et les Tableaux D.4 et D.3 de l'Annexe D.3 présentent les statistiques descriptives des variables explicatives des employeurs et employés pour tous les échantillons. Comme peu de transitions sont observées, cette restriction ne modifie pas les statistiques des caractéristiques des individus. Toutefois, se limiter à trois employés ou moins par entreprise réduit la représentation des grandes entreprises et en particulier de celles implantées sur plusieurs unités de production. Le problème ne se serait pas posé si les données contenaient un identifiant par site de production au lieu d'un identifiant par entreprise.

### 4.3. Inférence Bayésienne

L'approche Bayésienne complète le modèle avec des croyances a priori sur les paramètres. Nous spécifions des a priori propres mais non informatifs. Manda et Meyer (2005) supposent un hasard de base en escalier (Lancaster, 1990, p.172-181), dont les termes suivent un processus auto-régressif à l'ordre 1, et Grilli (2005) utilise une spécification polynomiale. Du fait du petit nombre d'intervalles de temps, les durées ne dépassent pas 6 ans, on suppose le hasard de base comme constant par morceaux et de paramètres indépendants.<sup>4</sup> Les coefficients ont des a priori gaussiens de moyenne nulle et de variance 1000.

Les précisions des effets aléatoires, c'est-à-dire  $\sigma_f^{-2}$  et  $\sigma_w^{-2}$ , sont tirées dans des lois gamma dont le paramétrage est déduit de statistiques descriptives. Les taux de transitions par individu sont de 3.5% pour le 5ème quantile de la distribution des durées et de 0.9% pour le 95ème quantile. Les probabilités varient donc dans un rapport de 1 à 4 pour 90% de la population. L'intervalle de confiance correspondant est de longueur 3, ce qui implique  $\sigma_w = 0.5$ . Nous spécifions donc que la loi a priori de  $\sigma_w^{-2}$  admet une espérance de 2 et une variance de 4. De même, les taux de transition par entreprise vont de 1.3% à 4% pour 90% de la population, ce qui implique une loi gamma d'espérance 3 et de variance 9 pour  $\sigma_f^{-2}$ . Le paramètre  $\rho$  est pourvu d'une loi a priori uniforme sur  $[-1, 1]$ , c'est-à-dire un a priori diffus sur l'ensemble de définition de  $\rho$ .

On note par  $T$  le vecteur des durées et  $M$  le nombre de variables expli-

---

<sup>4</sup>Nous avons estimé un modèle avec une spécification polynomiale et sommes arrivés à un polynôme de degré 6, c'est-à-dire une spécification moins parcimonieuse qu'un hasard de base constant par morceaux.



catives. La densité jointe des observations et des paramètres s'écrit :

$$f(T, \beta_0, \beta_1, v, w) = f(\sigma_f^2)f(\sigma_w^2)f(\rho) \left[ \prod_{s=1}^{K-1} f(\beta_{0s}) \right] \left[ \prod_{m=1}^M f(\beta_{1m}) \right] \left[ \prod_{i=1}^I f(v_i|\sigma_f^2) \right] \left[ \prod_{j=1}^J f(w_j|\sigma_w^2, \rho) \right] \left[ \prod_{i=1}^I \prod_{j=1}^J \prod_{k=1}^K L_{ijk}(t_{ijk}|\beta_0, \beta_1, v_i, w_j) \right] \quad (4.12)$$

La loi a posteriori correspond à la division de (4.12) par son intégrale sur l'espace des paramètres. Même en choisissant des loi a priori univariées, la loi a posteriori comprend des intégrales n'admettant pas de solution analytiques. On peut toutefois construire une chaîne de Markov dont les éléments sont distribués selon la a posteriori et approcher l'estimateur Bayésien par une méthode de Monte Carlo par Chaînes de Markov (MCMC).<sup>5</sup> Les quantités d'intérêts sont approchées par échantillonnage de Gibbs (Gelfand et Smith, 1990, voir la sous-section 1.7.2). L'échantillonnage de Gibbs procède à des tirages dans les lois des paramètres conditionnelles aux autres paramètres pertinents.

## 4.4. Résultats

Nous calculons deux chaînes par modèle. Sur des essais précédents, on remarque que les chaînes de Markov des paramètres  $\sigma_f^{-2}$  et  $\sigma_w^{-2}$  convergent plus lentement que celles des  $\beta$  (un constat similaire est déjà obtenu dans la Section 3.4) et de  $\rho$ . Les valeurs de départ sont donc fixées aux résultats d'estimation par maximum de vraisemblance d'un modèle sans hétérogénéité non observée pour les deux chaînes. Les paramètres  $\sigma_f^{-2}$  et  $\sigma_w^{-2}$  ont eut des chaînes initiées à 1 et 50. La convergence du modèle n'est pas influencée par la valeur initiale de  $\rho$ , et nous la fixons à 0 pour les deux modèles. Nous calculons 50 000 itérations pour le modèles à deux effets. Au vu des graphiques des échantillons simulés et des statistiques de Gelman et Rubin (1992), 20 000 itérations sont suffisantes pour la période "rodage" de l'échantillonnage de Gibbs. Les quantités d'intérêts sont calculées à partir des itérations suivant la convergence.

Comme indiqué plus haut, nous avons estimé les modèles à effets corrélés avec et sans salaire. Il peut en effet être déterminé simultanément avec le changement d'emploi et donc être endogène. Inversement, dans les théories de recherche d'emploi, le salaire est une caractéristique de l'entreprise et donc exogène à la décision changer d'entreprise. Le coefficient du salaire n'est pas

---

<sup>5</sup>Robert et Casella, 1999, est un ouvrage complet sur les méthodes de MCMC.

## 4.4 Résultats

significatif et les estimations des  $\beta$  ne diffèrent que de quelques centièmes d'un modèle à l'autre. Inclure le salaire diminue toutefois de 7% les estimations de  $\sigma_f$  et de 3% les estimations de  $\sigma_w$ , laissant penser qu'il est légèrement corrélé avec l'effet employeur. Seul un modèle structurel peut toutefois gérer précisément ce problème d'endogénéité potentielle, et comme les coefficients  $\beta$  des autres variables ne sont pas influencés par son exclusion, nous indiquons les résultats obtenus sans les salaires.

Les estimations des paramètres des distributions mélangeantes sont présentées dans le Tableau 4.3. Les résultats sont similaires pour les deux modèles et on retrouve une conclusion similaire à celle du Tableau 3.3 de la Section 3.4, à savoir que l'ajout d'un élément à la structure de l'hétérogénéité ne modifie pas les résultats des autres paramètres dans une approche Bayésienne.

TAB. 4.3 – Ecarts-types estimés sur l'échantillon contraint des distributions mélangeantes

Type d'hétérogénéité	Paramètre	Moyenne	2.5%	97.5%
<b>Effets corrélés</b>				
effet entreprise	$\sigma_f$	0.40	0.28	0.61
effet employé	$\sigma_w$	0.40	0.28	0.54
corrélation	$\rho$	0.17	-0.33	0.56
<b>Effets indépendants</b>				
effet entreprise	$\sigma_f$	0.41	0.27	0.64
effet employé	$\sigma_w$	0.42	0.30	0.55
<b>Effet unique</b>				
effet employé	$\sigma_w$	0.41	0.30	0.57

Le coefficient de corrélation est de 0.17 et non significatif à l'ordre de 5%.<sup>6</sup> L'association entre employeurs et employés ne dépend pas de manière claire des caractéristiques inobservés, telles la politique de renouvellement de personnel ou une forme de nomadisme professionnelle. La corrélation est toutefois spécifiques aux éléments inobservés par l'économètre et ne prend pas en compte les critères observés. Nos résultats n'excluent donc pas la présence d'un mécanisme d'appariement dans les relations d'emploi, dépendant aussi d'éléments observés. Ce point mérite clairement des travaux ultérieurs.

Les médianes a posteriori des coefficients  $\beta$  figurent dans le Tableau 4.4.

<sup>6</sup>Les estimations sur l'échantillon non contraint des écarts types des distributions mélangeantes figurent dans le Tableau D.5, dans l'Annexe D.4. La corrélation est positive et significative, toutefois, les hypothèses d'indépendance des  $v_i$  et d'indépendance des  $w_j$  ne sont vraisemblablement pas respectées sur l'échantillon non contraint.

4 La mobilité professionnelle au Portugal : une approche Bayésienne avec données d'appariement employeurs-employés

Les estimations des  $\beta$  sur l'échantillon non contraint sont reportées dans le Tableau D.6 de l'Annexe D.4. Le temps passé en poste réduit significativement la probabilité de transition dans les modèles à deux effets aléatoires, qu'ils soient indépendants ou corrélés. La présence d'hétérogénéité non observée n'explique pas à elle seule la décroissance du hasard au fil du temps (cette propriété est présentée dans la sous-section 1.3.6), qui dépend directement du temps passé en emploi dans l'entreprise.

TAB. 4.4 – Estimations Bayésiennes des coefficients  $\beta$

Variable	Aucun	Employé	Effet(s) aléatoire(s)	
			Indépendants	Corrélés
<b>Années en poste</b>				
2	<b>-0.18</b>	<b>-0.17</b>	<b>-0.16</b>	<b>-0.16</b>
3	<b>-0.52</b>	<b>-0.49</b>	<b>-0.47</b>	<b>-0.46</b>
4 et plus	<b>-0.98</b>	<b>-0.94</b>	<b>-0.91</b>	<b>-0.90</b>
<b>Caractéristiques de l'individu</b>				
Femme	<b>-0.25</b>	<b>-0.26</b>	<b>-0.26</b>	<b>-0.26</b>
Age :				
16 - 25	<b>0.55</b>	<b>0.55</b>	<b>0.54</b>	<b>0.54</b>
26 - 35	<b>0.33</b>	<b>0.32</b>	<b>0.31</b>	<b>0.31</b>
Niveau d'études :				
primaires	-0.02	-0.01	-0.02	-0.02
secondaires	0.07	0.07	0.06	0.06
Temps-partiel	<b>0.27</b>	<b>0.27</b>	<b>0.27</b>	<b>0.27</b>
<b>Caractéristiques de l'entreprise</b>				
Plusieurs sites	<b>0.28</b>	<b>0.29</b>	<b>0.29</b>	<b>0.29</b>
Région :				
Centre	0.16	0.16	0.16	0.16
Lisbonne et Vallée du Tage	<b>0.27</b>	<b>0.27</b>	<b>0.27</b>	<b>0.27</b>
Alentejo, Algarve et Isles	<b>0.25</b>	<b>0.25</b>	<b>0.25</b>	<b>0.25</b>
Secteur :				
Construction	0.15	0.15	0.14	0.14
Commerce	0.17	0.17	0.17	0.17
Finance	<b>0.34</b>	<b>0.35</b>	<b>0.35</b>	<b>0.35</b>
Constante	<b>-3.22</b>	<b>-3.31</b>	<b>-3.37</b>	<b>-3.39</b>
Log-vaisemblance	-3628	-3547	-3460	-3448
DIC	7515	7366	7215	7211
Number d'employés	8468	8468	8468	8468
Number d'entreprises	7325	7325	7325	7325

Note : les coefficients en caractères gras sont significatifs au seuil de 5%.

Concernant les caractéristiques des employés, on remarque que les femmes

sont moins enclines à changer d'emploi que les hommes. Une explication est que notre étude porte sur la période 1994-2000 et que les différences de mobilité dues au genre s'estompent dans le temps. Light et Ureta (1992) concluent que la mobilité professionnelle des femmes aux États-Unis évolue : les cohortes les plus anciennes sont plus mobiles que celles des hommes, et le résultat s'inverse pour les cohortes récentes. L'explication avancée étant que les femmes sont de plus en plus intégrées au marché du travail.

La catégorie de référence pour l'âge est la tranche des 36-55 ans. Les résultats indiquent une mobilité plus élevée pour les moins de 35 ans et plus particulièrement les moins de 25 ans. Lorsque le niveau d'étude est pris en compte, l'âge capte l'expérience acquise sur le marché du travail. Ces résultats indiquent l'influence de l'expérience, et peuvent s'interpréter à la lumière des modèles "on-the-job search" ou de "job shopping". Les premiers prédisent que, lorsque la qualité de la relation d'emploi est connue à l'avance, les individus les plus expérimentés sont moins mobiles car ils ont déjà eut des opportunités de changer d'entreprise pour des appariements de meilleur qualité. Les seconds concluent à la décroissance de la mobilité avec l'âge, au fur et à mesure que l'individu accumule de l'information sur ses propres compétences et les caractéristiques du marché du travail.

Le niveau d'étude n'a pas d'influence significative sur la mobilité professionnelle, comme dans Buchinsky *et al.* (2005). Les auteurs indiquent que ce résultat est conforme aux prédictions de la théorie du capital humain, car les études constituent un capital très général qui n'est pas spécifique à une entreprise, que l'individu peut réutiliser chez un autre employeur.

L'influence du travail à temps partiel confirme un fait stylisé de la littérature empirique : le temps partiel a un effet positif sur la probabilité de séparation.

En ce qui concerne les caractéristiques des entreprises, le secteur d'activité et la région d'implantation influence la probabilité de transition vers un nouvel emploi. Le Nord est utilisé comme région de référence et est la région la moins mobile, à l'inverse de Lisbonne. Le secteur de la finance et des assurances se distingue par une mobilité plus élevée que tous les autres, tandis que l'industrie (prit comme secteur de référence) est le plus statique.

La significativité des coefficients dans le Tableau 4.4 évolue avec la structure de l'hétérogénéité non observée. Avoir passé moins de deux ans dans l'entreprise devient de moins en moins important jusqu'à ne plus être significatif, de même que travailler dans le secteur de la construction ou du commerce. A l'inverse, avoir un emploi à temps partiel augmente la probabilité de transition et devient significatif lorsque les effets peuvent être corrélés.

L'accroissement des log-vraisemblances alors que la structure d'hétérogénéité non observée devient de plus en plus détaillée laisse penser que les

modèles sont très différents en terme d'adéquation aux données. Afin de comparer les modèles sur des critères formels, nous avons calculé les critères d'information de la déviance (DIC, Spiegelhalter *et al.*, 2002). Fixer un seuil au delà duquel la différence des DIC est importante, et plus généralement la différence de deux critères d'information, est délicat et nous suivons les recommandations de Burnham et Anderson (1998) et Spiegelhalter *et al.* (2002). On préfère le modèle avec le plus petit DIC, une différence de 1 à 2 indique que le modèle alternatif mérite un certain intérêt tandis qu'une différence de 3 à 7 lui est très défavorable. La différence est ici de 4 en faveur du modèle à effets corrélés, qui est le plus apte à générer des données ayant la même structure que celle observée.<sup>7</sup> A partir des DIC, nous pouvons conclure que le modèle à deux effets aléatoires corrélés est le plus performant.

Le calcul des chaînes de Markov est demandeur en temps de calcul et nous estimons également les modèles sans effets corrélés par maximum de vraisemblance, en utilisant la quadrature de Gauss-Hermite pour approcher les distributions mélangeantes. Le gain de vitesse nous permet d'estimer le modèle sur l'échantillon complet et les résultats sont présentés dans l'Annexe D.5, Tableau D.7. Ils sont proches de résultats Bayésiens bien que plus de coefficients soient significatifs, du fait de l'utilisation complet et donc de plus d'information.

## Conclusion

Nous estimons un modèle MPH en temps discret avec une approche Bayésienne. Il comprend plusieurs structures d'hétérogénéité, la plus fine contenant deux effets aléatoires corrélés : l'un au niveau des entreprises, l'autre au niveau des individus. La corrélation capte un éventuel mécanisme d'appariement, basé sur les caractéristiques inobservées par l'économètre. L'hétérogénéité non observée a une structure complexe car une entreprise est reliée longitudinalement et en coupe transversale à plusieurs employés, tandis qu'un individu n'est que longitudinalement relié à plusieurs entreprises. Nous montrons comment procéder à l'inférence en utilisant l'échantillonnage de Gibbs.

Nos résultats confirment la présence d'une forte hétérogénéité non observée au niveau des employés, et indiquent une hétérogénéité toute aussi influente entre les employeurs. Ce point est important car peu d'études prennent en compte les déterminants inobservés situés au niveau des entreprises et aucune, à notre connaissance, ne comprend simultanément un

---

<sup>7</sup>Les résultats sur l'échantillon non contraint figurent dans le Tableau D.6 de l'Annexe D.4. La différence des DIC est de 38, en faveur du modèle à deux effets aléatoires corrélés.

#### 4.4 Conclusion

---

effet individuel et un effet firme dans un modèle réduit. Supposer que l'hétérogénéité non observée sous jacente aux processus de transition ne relève que des caractéristiques inobservées des employés est insuffisant. Intuitivement, la mobilité professionnelle dépend des propensions individuelles inobservées à changer d'emploi, ainsi que des politiques de gestion des ressources humaines des entreprises. La première est très différente d'un employé à l'autre, ainsi que la seconde, et même autoriser deux effets ne suffit pas à représenter les interactions complexes qui ont lieu entre employeurs et employés. De plus, la capacité descriptive du modèle s'améliore grandement dès que l'on autorise un mécanisme d'appariement basé sur les caractéristiques inobservées des employeurs et des employés. Les résultats vont dans le sens des modèles à hétérogénéité non observée comme explication des faits stylisés du marché du travail, et impliquent que le temps passé en entreprise n'a qu'un effet secondaire sur la mobilité professionnelle. Les modèles à hétérogénéité non observée ne fournissent qu'une explication partielle mais pertinente, qui n'a pas encore été étudiée en détail.

Toutefois, nos résultats ne prennent pas en compte directement l'influence des observables dans l'appariement. Une corrélation dépendant des hétérogénéités observées et non observées, possible sans remettre en cause l'identification d'un modèle en temps continu et en présence d'épisodes multiples, fournit une meilleure indication de l'existence d'un processus d'appariement. Ce point mérite clairement des travaux ultérieurs, utilisant des effets fixes ou bien un modèle à épisodes multiples.

4 La mobilité professionnelle au Portugal : une approche Bayésienne avec  
données d'appariement employeurs-employés

---

CHAPITRE 5

# Conclusion



LES mots “Combien de temps . . .” sont au début de nombreuses questions économiques, ne serait-ce que dans l’étude du marché du travail ou du comportement du consommateur. La diversité des comportements des agents est une évidence, et cette thèse fournit un aperçu de la prise en compte d’une hétérogénéité non observée multinationaux au sein des modèles de durée multivariés.

Nous avons vu les éléments à la base de modèles de durée dans le Chapitre 1, en insistant sur la fonction de hasard. Son comportement détermine la distribution des durées, et les processus de comptage nous apportent un nouvel éclairage sur cette clef de voûte. Nous avons ensuite vu l’estimateurs non-paramétriques du hasard intégré (estimateur de Nelson-Aalen) et la fonction de survie (estimateurs de Breslow et Kaplan-Meier). Puis le modèle de Cox a été présenté, avant de passer à son extension qu’est le modèle MPH. La forme fonctionnelle de l’hétérogénéité non observée y est très souple et nous pouvons ainsi spécifier facilement une hétérogénéité multinationaux au moyen de plusieurs effets aléatoires. Nous voyons ensuite les conséquences d’une spécification MPH sur le taux d’accroissement du hasard, ainsi que les modèles MPH à épisodes multiples et multivariés.

L’identification du modèle MPH donne lieu à une vaste littérature, et nous nous intéressons principalement aux hypothèses relatives à la loi de l’hétérogénéité non observée. On voit notamment que la présence d’épisodes multiples à un niveau suffit à se passer des hypothèses de finitude de l’espérance de la distribution mélangeante et d’indépendance entre hétérogénéité observée et non observée. Nous commençons notre revue de littérature avec les modèles à épisodes uniques, avant de passer aux modèles à épisodes multiples et multivariés. Une fois l’identification assurée, nous voyons les principales méthodes d’estimation. Les approches par maximum de vraisemblance sont étroitement apparentées à l’approche de la vraisemblance partielle, centrale dans les modèles MPH. Elle sert de point de départ à des nombreux travaux, y compris dans la littérature Bayésienne.

Dans le Chapitre 2, nous reprenons le principe d’augmentation des données pour montrer comment estimer par maximisation de la vraisemblance partielle un modèle à plusieurs effets aléatoires. L’approche s’applique à toute distribution mélangeante admettant une transformée de Laplace, étendant ainsi les travaux antérieurs spécifiques à une loi de l’hétérogénéité non observée. Nous montrons comment l’inférence dans un modèle à plusieurs effets peut se simplifier en l’estimation de plusieurs modèles impliquant chacun un effet. A l’issue d’une étude de Monte Carlo et d’une application à la ratification des conventions de l’OIT, nous concluons que la procédure itérative ainsi constituée est à la fois plus rapide que l’algorithme EM et numériquement plus stable. Elle a toutefois l’inconvénient, comme l’algorithme EM, de four-

---

nir des estimations biaisées des paramètres des distributions mélangeante, tendant à sous-estimer l'importance de l'hétérogénéité non observée. Cette procédure est donc à privilégier lorsque des résultats rapides sont désirés, où l'estimation précise des paramètres de la distribution mélangeante n'est pas d'une importance cruciale. Les résultats de Rodriguez et Goldman (2001) laissent penser que l'approche Bayésienne fournit des résultats plus fiables, et nous examinons cet axe de recherche dans le chapitre suivant.

Nous montrons dans le Chapitre 3 comment estimer un modèle MPH à deux effets aléatoires indépendants au sein d'une approche Bayésienne. Nous relient ce travail à la vraisemblance partielle en étendant à deux effets aléatoires les travaux de Kalbfleisch (1978). En plus de la contribution méthodologique, ce chapitre apporte de nouveaux éléments à la littérature sur la régulation internationale du marché du travail. Nous constatons notamment que le terme de "pays en voie de développement" recouvre des situations extrêmement diverses en termes de comportements de ratification qu'il est difficile d'agrèger, et plus encore que les conventions élaborées par l'OIT sont extrêmement variées. Les textes ne demandant qu'un accord de principe sont ratifiés facilement, tandis que les documents plus précis et plus contraignants sont plus difficilement signés. En général, l'adéquation des systèmes juridiques et économiques en place avec le texte de la convention est un facteur déterminant, bien plus que l'orientation politique des dirigeants ou d'éventuelles sanctions commerciales. En comparant l'approche Bayésienne avec l'algorithme présenté dans le Chapitre 2, nous corroborons l'indication de Rodriguez et Goldman (2001) quant à la sous-estimation de l'importance de l'hétérogénéité non observée par les approches fréquentistes.

L'approche Bayésienne permet de gérer facilement des interactions complexes entre les effets aléatoires et, dans le Chapitre 4, nous permettons aux termes d'hétérogénéité non observée d'être corrélés. Le modèle est utilisé pour étudier la mobilité professionnelle, sur des données portugaises d'appariements employeurs-employés. Les travaux du domaine ne prennent en compte habituellement qu'un effet au niveau de l'employé. Nous ajoutons un second effet au niveau de la firme, ce qui nous permet de conclure que les politiques de rétention du personnel sont très hétérogènes d'une firme à l'autre et qu'elles influencent la rotation du personnel. De plus, nous permettons aux caractéristiques inobservées de l'individu d'être corrélées avec les caractéristiques inobservées de la firme et les résultats laissent penser qu'il existe une association entre les caractéristiques des deux parties. Ce point mérite clairement des travaux ultérieurs, notamment afin de prendre également en compte les variables explicatives observées.

A l'issue de cette thèse, de nombreuses pistes apparaissent comme dignes d'intérêt. Nous avons supposé dans le Chapitre 2 que les termes d'hétérogé-

néité sont indépendants. Cette hypothèse est parfois discutable, voir inappropriée à l'analyse de certains contextes comme celui du Chapitre 4. Un premier axe de recherche future serait donc d'étendre l'EMPL à des effets potentiellement corrélés, ou encore de transposer l'approche présentée au Chapitre 4 dans un cadre fréquentiste. Les méthodes MCMC sont en effet gourmandes en temps de calcul, et une extension de l'EMPL serait peut-être plus rapide. De plus, elle fournirait une alternative fréquentiste.

Après les Chapitres 2 et 3, il apparaît que l'estimation des paramètres de la distribution mélangeante par l'EMPL n'est pas précise, amenant à sous-estimer l'importance de l'hétérogénéité non observée. Les résultats de Rodriguez et Goldman (2001) indiquent que ce travers est commun à plusieurs méthodes utilisant le maximum de vraisemblance dans les modèles Logit à deux effets emboîtés. Une étude précise des biais apporterait des éléments d'explication à ce phénomène, de même qu'une comparaison systématique avec les estimateurs Bayésiens fournirait à l'économétrie appliquée des éléments d'arbitrage entre les deux approches.

Le Chapitre 4 ouvre à lui seul plusieurs pistes de recherches. La prise en compte d'une corrélation entre l'hétérogénéité observée et non observée, possible en présence d'épisodes multiples, permettrait de mieux capturer les appariements complexes qui peuvent se constituer entre employeurs et employés. Le problème de l'endogénéité potentielle du salaire nécessiterait lui aussi d'être étudié, ce qui nous entraînerait vers les approches structurelles.

Enfin et plus généralement, nous avons supposés tout au long de cette thèse une hétérogénéité non observée paramétrique. L'estimateur du NPMLE (Heckman et Singer, 1984c) existe déjà et est souvent retrouvé en utilisant un algorithme EM. Il semble toutefois instable (Campolieti, 2001), en particulier en présence d'un hasard de base flexible. Des alternatives Bayésiennes, utilisant les mélanges de processus de Dirichlet (Florens *et al.*, 1999) sont disponibles pour le modèle de hasards proportionnels de Weibull (Ondrich et Prasad, 1997) et Probit (Campolieti, 2001). Lau (2006) fournit une approche Bayésienne demandant une spécification du hasard de base, et une extension à la vraisemblance partielle serait intéressante. Il serait ainsi possible de comparer les résultats des modèles à hétérogénéité paramétrique et non-paramétrique. Les pistes de recherches ouvertes à l'issue de cette thèse sont ainsi au moins aussi nombreuses que celles qui étaient ouvertes avant son commencement, et le domaine est loin d'être épuisé.

Cette thèse ne prétend pas à l'exhaustivité quant à la prise en compte de l'hétérogénéité non observée dans les modèles de durée. Elle montre toutefois que l'hétérogénéité non observée mérite parfois d'être traitée de manière plus approfondie que comme un simple paramètre de nuisance. Bien qu'inobservée, elle peut être quantifiée et nous fournir ainsi des éléments d'analyse

---

venant élégamment compléter ceux collectés par l'étude de l'hétérogénéité observée. Curieusement, ce que l'économètre n'observe pas lui en apprend parfois beaucoup sur ce qu'il observe.

...



ANNEXE A

# Annexes du Chapitre 1

## A.1. Calcul de la fonction de survie conditionnelle aux variables explicatives observées dans le cadre d'erreurs de mesure sur les régresseurs

Les calculs présentés dans cette annexe sont tirés de Lancaster (1990, p.60). Considérons le changement de variable par la fonction  $g$  :

$$(X_1, v) \xrightarrow{g} (X_1 + v, v) = (X, v). \quad (\text{A.1})$$

On peut utiliser le lemme du changement de variable pour trouver la loi du couple  $(X, v)$ . Ainsi,  $g^{-1}(X, v) = (X - v, v)$  et :

$$\det(J_{g^{-1}}) = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1. \quad (\text{A.2})$$

D'où  $f(X, v) = f(X - v, v)$ . On en déduit la densité de  $v$  conditionnellement à  $X$  :

$$f(v|X) = \frac{f(X - v, v)}{\int_{\mathcal{V}} f(X - v, v) dv}, \quad (\text{A.3})$$

où  $\mathcal{V}$  est le domaine de définition de la variable  $v$ . On peut alors réécrire :

$$P[T \geq t, v|X] = \exp\left(-\lambda_1(X - v) \int_0^t \lambda_0(u) du\right) f(v|X). \quad (\text{A.4})$$

L'expression précédente doit encore être intégrée par rapport à  $v$  pour que la fonction de survie soit conditionnelle à la variable explicative observée :

$$\begin{aligned} P[T \geq t|X] &= \int_{\mathcal{V}} \exp\left(-\lambda_1(X - v) \int_0^t \lambda_0(u) du\right) f(v|X) dv \\ &= \int_{\mathcal{V}} \exp\left(-\int_0^t \lambda(t|X, v)\right) f(v|X) dv. \end{aligned} \quad (\text{A.5})$$

## A.2. Calcul de l'effet croisé du temps et des variables explicatives sur la hasard moyen

Une manière d'étudier l'évolution du hasard moyen en fonction du temps et des variables explicatives est de dériver par rapport à  $t$  le taux de variation

A.2 Calcul de l'effet croisé du temps et des variables explicatives sur la  
hasard moyen

---

de  $E[v|T \geq t]$  (équation (1.98), p.62) en fonction de  $X_j$ .<sup>1</sup> Rappelons que :

$$\frac{d \ln E[V|T \geq t]}{dX_j} = -\frac{d\lambda_1(X)}{dX_j} \Lambda_0(t) \frac{Var(V|T \geq t)}{E[V|T \geq t]}. \quad (\text{A.6})$$

D'où :

$$\begin{aligned} \frac{d^2 \ln E[V|T \geq t]}{dX_j dt} &= -\frac{d\lambda_1(X)}{dX_j} \lambda_0(t) \frac{Var(V|T \geq t)}{E[V|T \geq t]} + \\ &\quad - \frac{d\lambda_1(X)}{dX_j} \Lambda_0(t) \frac{d}{dt} \left[ \frac{Var(V|T \geq t)}{E[V|T \geq t]} \right]. \end{aligned} \quad (\text{A.7})$$

Or :

$$\begin{aligned} \frac{d}{dt} \left[ \frac{Var(V|T \geq t)}{E[V|T \geq t]} \right] &= \frac{1}{E[V|T \geq t]} \frac{d}{dt} Var(V|T \geq t) + \\ &\quad - \frac{Var(V|T \geq t)}{E[V|T \geq t]^2} \frac{d}{dt} E[V|T \geq t]. \end{aligned} \quad (\text{A.8})$$

Et :

$$\begin{aligned} \frac{d}{dt} \left[ \frac{Var(V|T \geq t)}{E[V|T \geq t]} \right] &= \frac{1}{E[V|T \geq t]} \frac{d}{dt} Var(V|T \geq t) \\ &\quad + \lambda_0(t) \lambda_1(X) \left[ \frac{Var(V|T \geq t)}{E[V|T \geq t]} \right]^2. \end{aligned} \quad (\text{A.9})$$

Et :

$$\frac{d}{dt} Var(V|T \geq t) = \frac{d}{dt} E[V^2|T \geq t] - \frac{d}{dt} (E[V|T \geq t]^2), \quad (\text{A.10})$$

Avec :

$$\begin{aligned} \frac{d}{dt} E[V^2|T \geq t] &= \frac{d}{dt} \frac{\int_{\mathcal{V}} v \exp[-v^2 \lambda_1(X) \Lambda_0(t)] dH(v)}{\int_{\mathcal{V}} \exp[-v \lambda_1(X) \Lambda_0(t)] dH(v)} \\ &= -\lambda_1(X) \lambda_0(t) \frac{\int_{\mathcal{V}} v^3 \exp[-v \lambda_1(X) \Lambda_0(t)] dH(v)}{\int_{\mathcal{V}} \exp[-v \lambda_1(X) \Lambda_0(t)] dH(v)} + \\ &\quad - \lambda_1(X) \lambda_0(t) \frac{\int_{\mathcal{V}} v^3 [\exp(-v \lambda_1(X) \Lambda_0(t))]^2 dH(v)}{\left[ \int_{\mathcal{V}} \exp[-v \lambda_1(X) \Lambda_0(t)] dH(v) \right]^2} \\ &= -\lambda_1(X) \lambda_0(t) \{ E[(V^{3/2})^2|T \geq t] - (E[V^{3/2}|T \geq t])^2 \} \\ &= -\lambda_1(X) \lambda_0(t) Var[V^{3/2}|T \geq t]. \end{aligned}$$

---

<sup>1</sup>En effet, on a supposé que les variables explicatives ne dépendent pas du temps.



En remplaçant les dérivées par leur expression analytique, on peut réécrire le taux de variation de  $E[V|T \geq t]$  en fonction de  $t$  et de  $X_j$  comme :

$$\frac{d^2 \ln E[V|T \geq t]}{dX_j dt} = -\frac{d\lambda_1(X)}{dX_j} \lambda_0(t) \left\{ \frac{\text{Var}(V|T \geq t)}{E(V|T \geq t)} - \lambda_1(X) \Lambda_0(t) \right. \\ \left. \left[ 2\text{Var}(V|T \geq t) + \left[ \frac{\text{Var}(V|T \geq t)}{E(V|T \geq t)} \right]^2 - \frac{\text{Var}(V^{3/2}|T \geq t)}{E(V|T \geq t)} \right] \right\}. \quad (\text{A.11})$$

### A.3. Épisodes uniques et identification

On rappelle qu'en présence de régresseurs constants dans le temps, la fonction de survie non conditionnelle à  $v$  s'écrit :

$$S(t|X) = \int_{\mathcal{V}} \exp[-v\Lambda_0(t)\lambda_1(X)] dH(v). \quad (\text{A.12})$$

Notons  $\mathcal{L}_v$  la transformée de Laplace d'une variable aléatoire  $v$  définie comme :

$$\mathcal{L}_v(s) = \int_{\mathcal{V}} \exp(-sv) dH(v), \quad (\text{A.13})$$

avec  $s \geq 0$  et  $V$  une variable aléatoire positive ou nulle. Les propriétés des transformées de Laplace sont présentées, entre autres, dans Lancaster (1990, p.330-332). Une relation importante est que la transformée de Laplace caractérise la distribution. Ainsi, il suffit de connaître  $\mathcal{L}_v$  pour en déduire  $H(v)$ .

Comme  $P[0 < V < \infty] = 1$  et  $\Lambda_0(t)\lambda_1(X) \geq 0$ , on peut réécrire :

$$S(t|X) = \mathcal{L}_v[\Lambda_0(t)\lambda_1(X)]. \quad (\text{A.14})$$

Elbers et Ridder (1982) commencent leur preuve en dérivant (A.12) :

$$s(t|X) = -\lambda_0(t)\lambda_1(X) \int_{\mathcal{V}} v \exp[-v\Lambda_0(t)\lambda_1(X)] dH(v). \quad (\text{A.15})$$

En divisant par  $s(t|X_0)$  :

$$\frac{s(t|X)}{s(t|X_0)} = \frac{\lambda_1(X)}{\lambda_1(X_0)} \frac{\int_{\mathcal{V}} v \exp[-v\Lambda_0(t)\lambda_1(X)] dH(v)}{\int_{\mathcal{V}} v \exp[-v\Lambda_0(t)\lambda_1(X_0)] dH(v)} \quad (\text{A.16})$$

$$= \frac{\lambda_1(X)}{\lambda_1(X_0)} \frac{E(V|T > t, X)}{E(V|T > t, X_0)}. \quad (\text{A.17})$$

### A.3 Épisodes uniques et identification

---

Comme  $E(V) < \infty$ , on en déduit en prenant la limite lorsque  $t \rightarrow 0$  :<sup>2</sup>

$$\lambda_1(X) = \lambda_1(X_0) \lim_{t \rightarrow 0} \frac{s(t|X)}{s(t|X_0)}. \quad (\text{A.18})$$

Comme on a normalisé  $\lambda_1(X_0) = 1$ ,  $\lambda_1$  est obtenue en faisant varier  $X$  sur son support. En notant  $\mathcal{L}_v^{-1}$  l'application réciproque de la transformée de Laplace, on déduit de (A.14) :

$$\Lambda_0 = \frac{\mathcal{L}_v^{-1}[S(t|X)]}{\lambda_1(X)}. \quad (\text{A.19})$$

Pour alléger les notations, supposons que  $X$  soit un scalaire. Dérivons cette expression par rapport à  $X$  et prenons-en la limite lorsque  $X \rightarrow X_0$ . Comme  $\Lambda_0(t) = \int_0^t \lambda_0(u) du$  qui ne dépend pas de  $X$  :

$$\lambda_1(X_0) \mathcal{L}'_v^{-1}[S(t|X_0)] \frac{\partial S(t|X_0)}{X_0} - \mathcal{L}_v^{-1}[S(t|X_0)] \lambda'_1(X_0) = 0 \quad (\text{A.20})$$

Notons  $a = S(t|X_0)$ , on a donc  $a \in [0, 1]$ . Par le théorème de fonctions implicites, on peut écrire  $t = K(a, X_0)$  (car  $S'(t|X_0) = -\lambda(t|X_0)f(t) < 0$  sous les hypothèses H2, H1 et H3). Cette équation différentielle se résout en  $\mathcal{L}_v^{-1}$  et admet comme solution :

$$\mathcal{L}_v^{-1}(a) = C \exp \left[ \frac{d \ln \lambda_1(X_0)}{dX} \int_{1/2}^a \frac{1}{\frac{\partial S}{\partial X}[K(u, X_0)|X_0]} du \right]. \quad (\text{A.21})$$

Comme nous raisonnons dans le cadre de modèles de hasards proportionnels, on peut réécrire sans perte de généralité  $S(t|X)$  sous la forme  $\mathcal{L}_w(\Psi_0(t)\lambda_1(X))$  où  $W$  est une variable aléatoire positive ou nulle de même espérance que  $V$ ,  $\Psi_0$  satisfaisant l'hypothèse H1 et  $\lambda_1$  la même fonction que précédemment. Alors :

$$\mathcal{L}_w^{-1}(a) = \tilde{C} \exp \left[ \frac{d \ln \lambda_1(X_0)}{dX} \int_{1/2}^a \frac{1}{\frac{\partial S}{\partial X}[K(u, X_0)|X_0]} du \right]. \quad (\text{A.22})$$

D'où :

$$\mathcal{L}_w^{-1}(a) = \frac{\tilde{C}}{C} \mathcal{L}_v^{-1}(a). \quad (\text{A.23})$$

Comme  $\mathcal{L}'_v^{-1}(1) = 1/\mathcal{L}'_v(\mathcal{L}_v^{-1}(1))$ ,  $\mathcal{L}_v^{-1}(1) = 0$  par propriété des transformées de Laplace et  $\mathcal{L}'_v(0) = -E(V)$ , on a  $\mathcal{L}'_v^{-1}(1) = -1/E(V)$ . On en déduit que

---

<sup>2</sup>Plus précisément, Elbers et Ridder (1982) normalisent l'espérance de  $V$  à 1, mais le raisonnement est valide pour toute quantité finie.

$C = \tilde{C}$ , donc que  $\mathcal{L}_v^{-1}(a) = \mathcal{L}_w^{-1}(a)$  et que  $V$  et  $W$  ont même loi. Il vient que  $\Lambda_0$  est déterminé de manière unique et peut être obtenu par (A.19).

## A.4. Preuve de l'identification sans variable explicative du modèle de Box-Cox

Le modèle de Box-Cox est caractérisé par  $\lambda_0(t) = \exp \left[ \gamma \left( \frac{t^\alpha - 1}{\alpha} \right) \right]$ . On obtient le modèle de Weibull lorsque  $\alpha = 0$ , le modèle de Gompertz pour  $\alpha = 1$  et les durées suivent une loi exponentielle pour  $\gamma = 0$ . Heckman et Singer (1984b) montre que le modèle est identifié en présence d'épisode unique sans variable explicative.

Leurs résultats supposent que  $\gamma$  est borné et peuvent se résumer de la manière suivante : lorsque  $\alpha \leq 0$ , l'hypothèse  $E(V) < \infty$  assure l'identification. Lorsque  $0 < \alpha < j$ , où  $j \in \mathbb{N}^*$ ,  $H(v)$  doit avoir la même moyenne finie pour toutes les observations et le moment d'ordre  $j + 1$  doit lui aussi être fini (mais pas nécessairement commun à toutes les observations).

### Preuve pour $\alpha \leq 0$

Ils obtiennent leur premier résultat de la manière suivante : supposons qu'il existe un triplet  $(\gamma_1, \alpha_1, H_1)$ , différent du vrai  $(\gamma_0, \alpha_0, H_0)$ , tel que le modèle ne soit pas identifié. L'absence d'identification implique :

$$\begin{aligned} \frac{s_1(t|X)}{s_0(t|X)} &= \frac{\lambda_0^1(t) \int_{\mathcal{Y}} v \exp[-v\Lambda_0^1(t)] dH_1(v)}{\lambda_0^0(t) \int_{\mathcal{Y}} v \exp[-v\Lambda_0^0(t)] dH_0(v)} \\ &= \frac{\exp \left[ \gamma_1 \left( \frac{t^{\alpha_1} - 1}{\alpha_1} \right) \right] \int_{\mathcal{Y}} v \exp[-v\Lambda_0^1(t)] dH_1(v)}{\exp \left[ \gamma_0 \left( \frac{t^{\alpha_0} - 1}{\alpha_0} \right) \right] \int_{\mathcal{Y}} v \exp[-v\Lambda_0^0(t)] dH_0(v)} \\ &= \frac{\exp \left[ \gamma_1 \left( \frac{t^{\alpha_1} - 1}{\alpha_1} \right) \right] E_0(V|T > t)}{\exp \left[ \gamma_0 \left( \frac{t^{\alpha_0} - 1}{\alpha_0} \right) \right] E_1(V|T > t)} = 1 \quad \forall t. \end{aligned} \tag{A.24}$$

Comme  $E(V) < \infty$ , on en déduit en prenant la limite lorsque  $t \rightarrow 0$  :

$$\lim_{t \rightarrow 0} \frac{s_1(t|X)}{s_0(t|X)} = \lim_{t \rightarrow 0} \exp \left[ \gamma_1 \left( \frac{t^{\alpha_1} - 1}{\alpha_1} \right) - \gamma_0 \left( \frac{t^{\alpha_0} - 1}{\alpha_0} \right) \right]. \tag{A.25}$$

Rappelons qu'on se limite ici à  $\alpha_0 \leq 0$  et qu'on suppose  $E(V) < \infty$ .

Si  $\underline{\gamma_0} \neq 0$ , il faut alors distinguer deux sous-cas. Si  $\alpha_1 > 0$  :

$$\lim_{t \rightarrow 0} \frac{s_1(t|X)}{s_0(t|X)} \rightarrow \begin{cases} \infty & \text{si } \gamma_0 > 0 \\ 0 & \text{si } \gamma_0 < 0 \end{cases} \tag{A.26}$$

#### A.4 Preuve de l'identification sans variable explicative du modèle de Box-Cox

---

Si  $\alpha_1 \leq 0$  :

$$\lim_{t \rightarrow 0} \frac{s_1(t|X)}{s_0(t|X)} \rightarrow \begin{cases} \infty & \text{si } \gamma_1 < 0, \gamma_0 > 0 \\ & \text{ou } \gamma_1 < 0, \gamma_0 < 0, \alpha_1 < \alpha_0 \\ & \text{ou } \gamma_1 > 0, \gamma_0 > 0, \alpha_1 > \alpha_0 \\ 0 & \text{si } \gamma_1 > 0, \gamma_0 < 0 \\ & \text{ou } \gamma_1 < 0, \gamma_0 < 0, \alpha_1 > \alpha_0 \\ & \text{ou } \gamma_1 > 0, \gamma_0 > 0, \alpha_1 < \alpha_0 \end{cases} \quad (\text{A.27})$$

En résumé, la limite tend vers 0 ou l'infini si  $\alpha_1 \neq \alpha_0$ . L'absence d'identification nécessite donc, si  $\gamma_0 \neq 0$ , que  $\alpha_1 = \alpha_0$ . On obtient alors, pour  $\gamma_1 \neq \gamma_0$  :

$$\lim_{t \rightarrow 0} \frac{s_1(t|X)}{s_0(t|X)} = \lim_{t \rightarrow 0} \exp \left[ (\gamma_1 - \gamma_0) \left( \frac{t^{\alpha_0} - 1}{\alpha_0} \right) \right] \rightarrow \begin{cases} \infty & \text{si } \gamma_1 < \gamma_0 \\ 0 & \text{si } \gamma_1 > \gamma_0 \end{cases} \quad (\text{A.28})$$

L'absence d'identification nécessite donc  $\gamma_1 = \gamma_0$ . On a ainsi  $\alpha_1 = \alpha_0$  et  $\gamma_1 = \gamma_0$ . Par (A.24) et l'unicité des transformées de Laplace,  $H_0 = H_1$ , et on déduit qu'il ne peut pas y avoir deux triplets distincts produisant la même fonction de survie (c'est-à-dire la même distribution des durées).

Si  $\underline{\gamma_0 = 0}$ ,  $\alpha_0$  n'apparaît pas dans le modèle et n'est donc pas identifié. Il vient que  $\Lambda_0(t) = t$ , et en reprenant le même raisonnement que pour  $\gamma_0 \neq 0$ , on arrive à la conclusion  $\gamma_0 = \gamma_1 = 0$ . Ainsi,  $H$  est identifié à partir de l'unicité de la transformée de Laplace.

#### Preuve pour $0 < \alpha < j$

Nous commencerons par la démonstration dans le cas  $0 < \alpha < 1$ . Ensuite, nous présenterons les arguments avancés par Heckman et Singer (1984b) pour la généralisation au cas  $j \in \mathbb{N}^*$ . Cette démonstration suit globalement la même démarche que la précédente, c'est à dire que l'on montre qu'il ne peut pas exister un triplet  $(\gamma_1, \alpha_1, H_1)$  différent du vrai triplet  $(\gamma_0, \alpha_0, H_0)$  et menant à la même loi des durées.

Si  $\underline{\gamma_0 \neq 0}$ , il y a deux sous-cas à traiter. Le premier est  $\alpha_1 \leq 0$  et se prouve de la même manière que dans la sous-section précédente. Le second cas,  $\alpha_1 > 0$ , demande un traitement spécifique. Comme toutes les lois possibles de  $V$  ont la même espérance, la relation (A.24) nous donne :

$$\lim_{t \rightarrow 0} \frac{s_1(t|X)}{s_0(t|X)} = \exp \left( \frac{\gamma_0}{\alpha_0} - \frac{\gamma_1}{\alpha_1} \right) = 1 \quad (\text{A.29})$$

et  $\gamma_0/\alpha_0 = \gamma_1/\alpha_1$ . L'absence d'identification implique aussi que, pour tout  $t$ ,  $s'_1(t|X) = s'_0(t|X)$ . Ainsi :

$$\frac{s'_1(t|X)}{s'_0(t|X)} = \frac{a}{b}, \quad (\text{A.30})$$

où :

$$a = \gamma_1 t^{\alpha_1 - 1} \lambda_0^1(t) \int_{\mathcal{V}} v \exp[-v \Lambda_0^1(t)] dH_1(v) - [\lambda_0^1(t)]^2 \int_{\mathcal{V}} v^2 \exp[-v \Lambda_0^1(t)] dH_1(v), \quad (\text{A.31})$$

$$b = \gamma_0 t^{\alpha_0 - 1} \lambda_0^0(t) \int_{\mathcal{V}} v \exp[-v \Lambda_0^0(t)] dH_0(v) - [\lambda_0^0(t)]^2 \int_{\mathcal{V}} v^2 \exp[-v \Lambda_0^0(t)] dH_0(v). \quad (\text{A.32})$$

Comme on suppose  $E(V^2) < \infty$ , on obtient pour  $0 < \alpha_0 < 1$  :

$$\lim_{t \rightarrow 0} \frac{s_1'(t|X)}{s_0'(t|X)} \rightarrow \begin{cases} 0 & \text{si } \alpha_1 > \alpha_0 \\ \pm \infty & \text{si } \alpha_1 < \alpha_0 \end{cases} \quad (\text{A.33})$$

Ainsi, ce rapport ne vaut 1 que lorsque  $\alpha_1 = \alpha_0$ . On en déduit, par  $\gamma_0/\alpha_0 = \gamma_1/\alpha_1$ , que  $\gamma_0 = \gamma_1$ , et par l'unicité de la transformée de Laplace que  $H$  est identifiée.

Si  $\gamma = 0$ , l'argument est exactement le même que celui de la sous-section précédente pour ce cas particulier.

L'extension au cas  $j \in \mathbb{N}^*$  est assez directe. Tous les détails ne sont pas donnés ici car les calculs sont longs et peu utiles, l'idée étant que :

$$\lim_{t \rightarrow 0} \frac{s_1^{(l)}(t|X)}{s_0^{(l)}(t|X)} \rightarrow 0 \text{ ou } \pm \infty \text{ pour } \alpha_1 \neq \alpha_0, \quad (\text{A.34})$$

où l'exposant  $(l)$  désigne la dérivée à l'ordre  $l$ , avec  $l \leq j$ . Ainsi,  $\alpha_1 = \alpha_0$  et  $\gamma_0 = \gamma_1$  car  $\gamma_0/\alpha_0 = \gamma_1/\alpha_1$ . Il en résulte que le modèle est identifié.

## A.5. Preuve de l'identification en présence de régresseurs endogènes

Dans cette section, nous présentons les démonstrations de Abbring et Van den Berg (2003b) concernant l'identification lorsque au moins une variable explicative est endogène. Nous verrons dans un premier temps les résultats dans le cadre du modèle MPH à risques concurrents, puis leur adaptation au cas de régresseurs endogènes variant dans le temps.

### A.5.1 Preuve de l'identification du sous modèle à partir de $(Q_S^0, Q_T)$

Abbring et Van den Berg (2003a) considèrent un modèle à risques concurrents dans lequel la transition vers chacune des destinations possible est caractérisée par un modèle MPH. Les risques sont ici dépendants *via* le terme d'hétérogénéité non observé, commun à plusieurs risques. Heckman et Honoré (1989) ont déjà établi ce résultat, et Abbring et Van den Berg (2003a) l'obtiennent avec une hypothèse moins restrictive que leurs prédécesseurs.

Soit un modèle à deux risques concurrents, où  $T_A$  et  $T_B$  sont les variables aléatoires des durées précédant les transitions vers les états  $A$  et  $B$ . On note  $S(t_A, t_B|X)$  la fonction de survie jointe qui a pour expression :

$$\begin{aligned} S(t_A, t_B|X) &= \int_{\mathcal{V}} \exp[-v_A \Lambda_{0,A}(t_A) \lambda_{1,A}(X) - v_B \Lambda_{0,B}(t_B) \lambda_{1,B}(X)] dH(v_A, v_B) \\ &= \mathcal{L}[\Lambda_{0,A}(t_A) \lambda_{1,A}(X), \Lambda_{0,B}(t_B) \lambda_{1,B}(X)], \end{aligned} \quad (\text{A.35})$$

où  $\Lambda_{0,A}(t) = \int_0^t \lambda_{0,A}(u) du$ . On note  $S_A(t|X) = P(T_A > t, T_B > T_A|X)$  et  $S_B(t|X) = P(T_B > t, T_A > T_B|X)$ . Pour  $X \neq X_0$  et  $i = A, B$ , on a :

$$\frac{s_i(t|X)}{s_i(t|X_0)} = \frac{\lambda_{0,i}(t) \lambda_{1,i}(X) \mathcal{L}'_i[\Lambda_{0,A}(t) \lambda_{1,A}(X), \Lambda_{0,B}(t) \lambda_{1,B}(X)]}{\lambda_{0,i}(t) \lambda_{1,i}(X_0) \mathcal{L}'_i[\Lambda_{0,A}(t) \lambda_{1,A}(X_0), \Lambda_{0,B}(t) \lambda_{1,B}(X_0)]}, \quad (\text{A.36})$$

où  $\mathcal{L}'_i$  désigne la dérivée à l'ordre 1 de la transformée de Laplace par rapport à la variable d'indice  $i$ . Comme  $V$  est indépendante de  $X$  et a une espérance finie, la limite lorsque  $t \rightarrow 0$  et avec la normalisation  $\lambda_{1,i}(X_0) = 1, \forall i$  est :

$$\lim_{t \rightarrow 0} \frac{s_i(t|X)}{s_i(t|X_0)} = \lambda_{1,i}(X). \quad (\text{A.37})$$

Abbring et Van den Berg (2003a) supposent que  $\lambda_{1,i}$  est à valeur dans un ouvert non vide  $\Phi \subset (0, \infty)$ . Ainsi, l'observation de  $S(t_0, t_0|X)$  à une date  $t_0$  quelconque et la monotonie de  $\mathcal{L}$  permet de retrouver  $\mathcal{L}$  sur  $(0, \infty)$  à partir de (A.35). Comme le hasard est la dérivée du hasard intégré, il est possible d'obtenir l'équation différentielle suivante à partir de  $s_i(t|X)$  :

$$\lambda_{0,i}(t) = \frac{s_i(t|X)}{\lambda_{1,i}(X_i) \mathcal{L}'_i[\Lambda_{0,A}(t) \lambda_{1,A}(X), \Lambda_{0,B}(t) \lambda_{1,B}(X)]}. \quad (\text{A.38})$$

Les conditions initiales de cette équation sont fournies par la normalisation  $\lambda_{0,i}(t_0) = 1, \forall i$  où  $t_0$  est une date quelconque. Soit  $f$  la fonction déterminant le

couple  $[\lambda_{0,A}(t), \lambda_{0,B}(t)]$  à partir de  $[t, \lambda_{0,A}(t), \lambda_{0,B}(t)]$ . Par construction, cette fonction admet une solution unique pour les conditions initiales, et comme  $\lambda_{0,A}(0) = \lambda_{0,B}(0) = 0$ , le couple  $(\lambda_{0,A}(t), \lambda_{0,B}(t))$  est identifié.

Ainsi,  $\lambda_{0,i}, \lambda_{1,i}$  et  $\mathcal{L}$  sont donc identifiées à partir de  $(Q_S^0, Q_T), \forall i = A, B$ . Abbring et Van den Berg (2003a) en déduisent l'identification du modèle à régresseur endogène défini par (1.110) et (1.111).

### A.5.2 Preuve de l'identification à partir de $(Q_S, Q_T)$

On note  $\xi(t|s, X) = \int_s^t \lambda_0^T(u) \delta(u|s, X) du$ . D'où :

$$\begin{aligned} Q_S(t, s|X) &= \exp \left[ - \int_0^s v_T \lambda_0^T(u) \lambda_1^T(X) du - \int_0^s v_S \lambda_0^S(u) \lambda_1^S(X) du \right. \\ &\quad \left. - \int_s^t v_T \delta(u|s, X) \lambda_0^T(u) \lambda_1^T(X) du \right] \quad (\text{A.39}) \\ &= \exp \left[ - v_T \lambda_1^T(X) \left( \int_0^s \lambda_0^T(u) du + \int_s^t \delta(u|s, X) \lambda_0^T(u) du \right) \right. \\ &\quad \left. - v_S \lambda_1^S(X) \int_0^s \lambda_0^S(u) du \right] \\ &= \mathcal{L} [\lambda_1^T(X) (\Lambda_0^T(s) + \xi(t|s, X)), \lambda_1^S(X) \Lambda_0^S(s)], \forall s < t. \end{aligned}$$

On en déduit,  $\forall s < t$  :

$$\frac{\partial Q_S(t, s|X)}{\partial s} = \lambda_1^S(X) \lambda_0^S(s) \mathcal{L}'_{(S)} [\lambda_1^T(X) (\Lambda_0^T(s) + \xi(t|s, X)), \lambda_1^S(X) \Lambda_0^S(s)], \quad (\text{A.40})$$

où  $\mathcal{L}'_{(S)}$  désigne la dérivée de la fonction de Laplace par rapport à l'argument d'exposant  $S$ . Elle est monotone croissante et identifiée pour  $X$  et  $s$  quelconques, ce qui nous assure l'identification de  $\xi(t|s, X)$ . Or :

$$\int_s^t \delta(u|s, X) du = \int_s^t \frac{1}{\lambda_0^T(u)} \frac{\partial \xi(u|s, X)}{\partial u} du. \quad (\text{A.41})$$

Il vient que  $\int_s^t \delta(u|s, X)$  est identifié. On en déduit l'identification complète du modèle à régresseur endogène défini par (1.110) et (1.111), à partir de  $(Q_S, Q_T)$ .

## A.6. Identification du modèle de Weibull sans supposer $E(V) < \infty$

Le hasard du modèle de Weibull est  $\lambda(t|X_i, v) = v\alpha t^{\alpha-1} \exp(X_i\beta)$  où  $\alpha > 0$ . Comme le hasard intégré suit une loi exponentielle unitaire, on en déduit que  $\Lambda_0(t, \alpha) \exp(X_i\beta)$  a la même distribution que  $W/V$ , où  $W$  suit une loi exponentielle unitaire et est indépendante de  $X_i$  et  $V$ . Où encore,  $\Lambda_0(t, \alpha) \exp(X_i\beta)$  suit une loi exponentielle de paramètre  $v$  (Lancaster 1990, p.75). On peut écrire :

$$P(T > t|X_i) = \exp \left[ - \int_0^t \Lambda_0(u, \alpha) \exp(X_i\beta) du \right] = S_U [\Lambda_0(t, \alpha) \exp(X_i\beta)], \quad (\text{A.42})$$

où  $U = W/V$  suit une loi exponentielle. On rappelle que la fonction de survie du modèle exponentiel a pour expression  $S(t) = \exp(-vt)$  et est toujours décroissante. Supposons qu'il existe un triplet  $(\alpha_1, \beta_1, H_1)$  différent du vrai triplet  $(\alpha_0, \beta_0, H_0)$  et engendrant la même distribution des durées. Pour alléger les notations, on omet l'indice  $i$  des variables explicatives et on note  $X_1$  et  $X_2$  deux valeurs des variables explicatives telles que  $\exp(X_1\beta_0) > \exp(X_2\beta_0)$  et  $\exp(X_1\beta_1) > \exp(X_2\beta_1)$ , et on normalise  $\exp(X_1\beta_0) = \exp(X_1\beta_1) = 1$ . On peut écrire :

$$S_U(t) = S_U \left[ \frac{1}{\exp(X_2\beta_0)} \Lambda_0^0 (\Lambda_0^{1,-1} [t \exp(X_2\beta_1)]) \right], \quad (\text{A.43})$$

où  $\Lambda_0^i(t) = \Lambda_0(t, \alpha_i)$  avec  $i = 1, 2$  et  $\Lambda_0^{i,-1}$  désigne la fonction réciproque de  $\Lambda_0^i$ . Soit  $K(t) = \Lambda_0^0[\Lambda_0^{1,-1}(t)]$ . On a  $\Lambda_0^i(0) = 0$  et on suppose sans perte de généralité que le hasard de base est strictement croissant. On en déduit que  $K(t)$  est strictement croissant et  $K(0) = 0$ . Il vient de la relation (A.43) que  $K[t \exp(X_2\beta_1)] = \exp(X_2\beta_0)K(t)$ . Après dérivation, elle devient :

$$\frac{\exp(X_2\beta_0)}{\exp(X_2\beta_1)} K'(t) = K'(t \exp(X_2\beta_1)). \quad (\text{A.44})$$

D'où en généralisant  $K[t \exp(X_2\beta_1)^q] = \exp(X_2\beta_0)^q K(t)$ ,  $\forall q \geq 1$ , ce qui nous donne après dérivation :

$$\left[ \frac{\exp(X_2\beta_0)}{\exp(X_2\beta_1)} \right]^q K'(t) = K'[t \exp(X_2\beta_1)^q]. \quad (\text{A.45})$$



En faisant le rapport des deux dernières équations, on obtient :

$$\begin{aligned} \frac{K'(t)}{K(t)} &= \lim_{q \rightarrow \infty} \exp(X_2 \beta_0)^q \left[ \frac{\exp(X_2 \beta_1)}{\exp(X_2 \beta_0)} \right]^q \frac{K' [t \exp(X_2 \beta_1)^q]}{K [t \exp(X_2 \beta_1)^q]} \\ &= \frac{1}{t} \lim_{q \rightarrow \infty} t \exp(X_2 \beta_1)^q \frac{K'(t \exp(X_2 \beta_1)^q)}{K(t \exp(X_2 \beta_1)^q)} \end{aligned} \quad (\text{A.46})$$

Comme  $K(t) = \Lambda_0^0(\Lambda_0^{1,-1}(t))$ , on a :

$$K'(t) = \frac{1}{\lambda_0^1(t)} \lambda_0^0(\Lambda_0^{1,-1}(t)) \quad (\text{A.47})$$

Que l'on peut diviser par :

$$1 = \frac{d}{dt} \Lambda_0^1(\Lambda_0^{1,-1}(t)) = \frac{1}{\lambda_0^1(t)} \lambda_0^1(\Lambda_0^{1,-1}(t)) \quad (\text{A.48})$$

D'où :

$$K'(t) = \frac{\lambda_0^0(\Lambda_0^{1,-1}(t))}{\lambda_0^1(\Lambda_0^{1,-1}(t))} \quad (\text{A.49})$$

Que l'on peut substituer dans la relation (A.46) :

$$\frac{K'(t)}{K(t)} = \frac{1}{t} \lim_{q \rightarrow \infty} \frac{\lambda_0^0(\Lambda_0^{1,-1}[t \exp(X_2 \beta_1)^q])}{\lambda_0^1(\Lambda_0^{1,-1}[t \exp(X_2 \beta_1)^q])} \frac{t \exp(X_2 \beta_1)^q}{K [t \exp(X_2 \beta_1)^q]} = \frac{1}{t}, \quad (\text{A.50})$$

car  $\lambda_0^1(0) = \lambda_0^0(0)$ . Le taux d'accroissement de  $K(t)$  est constant et  $K$  est donc affine. Comme  $K(0) = 0$ , par définition de  $K(t)$ , on en déduit que  $K(t) = t$ . Ainsi,  $\Lambda_0^0(t) = \Lambda_0^1(t)$  et  $\alpha_0 = \alpha_1$ . Comme  $K(t) = t$ , on a  $K'(t) = 1$  et par (A.45),  $\beta_0 = \beta_1$ . Par l'unicité de la transformée de Laplace, on obtient l'identification de la loi de  $U$ . Ainsi, les seules valeurs des paramètres engendrant la même distribution des durées sont les vraies valeurs elles mêmes.

## A.7. Épisodes multiples et identification

Honoré (1993) considère l'identification sans variable explicative d'un modèle où la même réalisation  $v$  caractérise plusieurs épisodes.

Soit deux variables aléatoires  $T_1$  et  $T_2$  caractérisées respectivement par  $\lambda_1(t) = v \lambda_{0,1}(t)$  et  $\lambda_2(t) = v \lambda_{0,2}(t)$ . La fonction de survie jointe aux dates  $t_1$  et  $t_2$  a pour expression :

$$S(t_1, t_2) = \int_{\mathcal{V}} \exp[-v \Lambda_{0,1}(t_1) - v \Lambda_{0,2}(t_2)] dH(v) = \mathcal{L}[\Lambda_{0,1}(t_1) + \Lambda_{0,2}(t_2)], \quad (\text{A.51})$$

où  $\Lambda_0(t) = \int_0^t \lambda_0(u)du$  et  $\mathcal{L}$  désigne la transformée de Laplace. Honoré (1993) montre que si  $\Lambda_{0,1}$  et  $\Lambda_{0,2}$  sont différentiables et monotones croissantes, alors  $\Lambda_{0,1}, \Lambda_{0,2}$  et  $H$  sont identifiées à une normalisation près.

La preuve commence de la même manière que celle de Elbers et Ridder (1982), c'est-à-dire par la relation issue de (A.51) :

$$\frac{\partial S(t_1, t_2)/\partial t_2}{\partial S(t_1, t_2)/\partial t_1} = \frac{\mathcal{L}' [\Lambda_{0,1}(t_1) + \Lambda_{0,2}(t_2)] \lambda_{0,2}(t_2)}{\mathcal{L}' [\Lambda_{0,1}(t_1) + \Lambda_{0,2}(t_2)] \lambda_{0,1}(t_1)} = \frac{\lambda_{0,2}(t_2)}{\lambda_{0,1}(t_1)}. \quad (\text{A.52})$$

Pour deux valeurs quelconques  $t$  et  $t_0$ , le rapport entre les valeurs de l'expression (A.52) aux points  $(t_0, t_2)$  et  $(t, t_2)$  donne :

$$\frac{\lambda_{0,2}(t_2)}{\lambda_{0,1}(t_0)} \bigg/ \frac{\lambda_{0,2}(t_2)}{\lambda_{0,1}(t)} = \frac{\lambda_{0,1}(t)}{\lambda_{0,1}(t_0)}. \quad (\text{A.53})$$

On normalise  $\lambda_{0,1}(t_0) = 1/k$  et en intégrant la relation précédente de 0 à  $t$ , on obtient  $k\Lambda_{0,1}(t) + c_1$ . En reprenant le même raisonnement aux points  $(t_2, t_0)$  et  $(t_2, t)$ , on trouve  $k\Lambda_{0,2}(t) + c_2$ . Les constantes d'intégration sont déterminées à partir de  $\Lambda_{0,1}(0) = \Lambda_{0,2}(0) = 0$ , et on obtient que l'intégrale du rapports des valeurs de (A.52) en  $(t_0, t_2)$  et en  $(t, t_2)$  est déterminé de manière unique. Ainsi,  $\Lambda_{0,1}$  et  $\Lambda_{0,2}$  sont identifiés. Par l'injectivité de la transformée de Laplace, l'identification de  $\Lambda_{0,1}$  et de  $\Lambda_{0,2}$  assure celle de  $H(v)$ .

## A.8. Identification du modèle MMPH à deux effets aléatoires

Considérons le modèle :

$$\lambda_{i,1}(t|X_i, v_i, w_1) = v_i w_1 \lambda_{0,1}(t|X_i), \quad (\text{A.54})$$

$$\lambda_{i,2}(t|X_i, v_i, w_2) = v_i w_2 \lambda_{0,2}(t|X_i), \quad (\text{A.55})$$

La fonction de survie jointe de l'individu  $i$  a pour expression :

$$\begin{aligned} S_i(t_1, t_2|X_i) &= \int_{\mathcal{V}} \exp[-v w_1 \Lambda_{0,1}(t_1|X_i) - v w_2 \Lambda_{0,2}(t_2|X_i)] dH_V(v) \quad (\text{A.56}) \\ &= \mathcal{L}_V [w_1 \Lambda_{0,1}(t_1|X_i) + w_2 \Lambda_{0,2}(t_2|X_i)], \end{aligned}$$

avec  $\mathcal{L}_V$  la transformée de Laplace. En dérivant (A.56) par rapport à  $t_1$  et

$t_2$ , puis en faisant le rapport :

$$\begin{aligned} \frac{\partial S_i(t_1, t_2 | X_i) / \partial t_2}{\partial S_i(t_1, t_2 | X_i) / \partial t_1} &= \frac{w_2 \lambda_{0,2}(t_2 | X_i) \mathcal{L}'_V [w_1 \Lambda_{0,1}(t_1 | X_i) + w_2 \Lambda_{0,2}(t_2 | X_i)]}{w_1 \lambda_{0,1}(t_1 | X_i) \mathcal{L}'_V [w_1 \Lambda_{0,1}(t_1 | X_i) + w_2 \Lambda_{0,2}(t_2 | X_i)]} \\ &= \frac{w_2 \lambda_{0,2}(t_2 | X_i)}{w_1 \lambda_{0,1}(t_1 | X_i)}. \end{aligned} \quad (\text{A.57})$$

On note  $k = 1/\lambda_{0,1}(t_0 | X_i)$  et on obtient en faisant le rapport entre les valeurs de l'expression précédente aux points  $(t_0, t_2)$  et  $(t, t_2)$

$$\frac{w_2 \lambda_{0,2}(t_2 | X_i)}{w_1 \lambda_{0,1}(t_0 | X_i)} \bigg/ \frac{w_2 \lambda_{0,2}(t_2 | X_i)}{w_1 \lambda_{0,1}(t | X_i)} = \frac{\lambda_{0,1}(t | X_i)}{\lambda_{0,1}(t_0 | X_i)} = k \lambda_{0,1}(t | X_i). \quad (\text{A.58})$$

Cette dernière expression est également obtenue en reprenant les calculs précédents pour un modèle sans hétérogénéité non observée. En intégrant par rapport au temps, on obtient  $k\Lambda_{0,1}(t | X_i) + c_1$ , où  $c_1$  est déterminé grâce à  $\Lambda_{0,1}(0) = 0$ . Ainsi  $\Lambda_{0,1}(t | X_i)$  est identifié. De même, on montre que  $\Lambda_{0,2}(t | X_i)$  est identifié en faisant le rapport au point  $(t_1, t_0)$  et en  $(t_1, t)$ .

En considérant une population homogène par rapport à  $w_1$  et  $w_2$  (même si la réalisation de l'effet aléatoire est inobservée, la manière dont la sont constitués les groupes est connue),  $S(t_1, t_2 | X_i)$  peut être observée. Comme  $w_1$  et  $w_2$  sont constants, on retrouve  $\mathcal{L}_V$  en faisant varier  $(t_1, t_2)$  sur  $[0, \infty]^2$  et on a l'identification de  $H_V$ .

Posons  $z_{ij} = v_i w_j$ . On peut réécrire :

$$S_i(t_1, t_2 | X_i) = \mathcal{L}_Z [\Lambda_{0,1}(t_1 | X_i) + \Lambda_{0,2}(t_2 | X_i)], \quad (\text{A.59})$$

où  $\mathcal{L}_Z$  désigne la transformée de Laplace par rapport à la variable aléatoire  $Z_{ij}$ . Comme  $\Lambda_{0,1}$  et  $\Lambda_{0,2}$  sont identifiées, on en déduit que la distribution du produit  $Z$  est identifiée. L'identification de  $H_Z$  et de  $H_V$  nous assure l'identification de  $H_W$ .

## A.9. Espérances conditionnelles dans un modèle à deux termes d'hétérogénéité gamma

Pour procéder à l'étape E dans le modèle de Sastry (1997), nous avons besoin de la densité conditionnelle aux observations de chaque terme d'hétérogénéité non observée ainsi que de la densité conditionnelle de leur produit. Nous détaillons ici ces calculs, en nous appuyant sur Sastry (1995).

Considérons l'hétérogénéité résidant au niveau des communautés. On a :

$$f(v_i|t_i, \delta_i) = \frac{f(v_i, t_i, \delta_i)}{\int_{\mathcal{V}} f(v, t_i, \delta_i) dv} \quad (\text{A.60})$$

Or :

$$\begin{aligned} f(t_i, \delta_i|v_i, w_{i1}, \dots, w_{iJ_i}) &= \prod_{j=1}^{J_i} \prod_{k=1}^{K_{ij}} f(t_{ijk}, \delta_{ijk}|v_i, w_{ij}) \\ &= \prod_{j=1}^{J_i} \prod_{k=1}^{K_{ij}} [\lambda_{ijk}(t_{ijk})]^{\delta_{ijk}} \exp \left[ -v_i w_{ij} \sum_k \Lambda_{ijk} \right] \\ &= v_i^{\sum_{j=1}^{J_i} \sum_{k=1}^{K_{ij}} \delta_{ijk}} \prod_{j=1}^{J_i} w_{ij}^{\sum_{k=1}^{K_{ij}} \delta_{ijk}} \exp \left[ -v_i w_{ij} \sum_k \Lambda_{ijk} \right] \\ &\quad \prod_{k=1}^{K_{ij}} [\lambda_0(t_{ijk}) \exp(X_{ijk}(t_{ijk})\beta)]^{\delta_{ijk}}, \end{aligned} \quad (\text{A.61})$$

où  $\Lambda_{ijk} = \int_0^{t_{ijk}} \lambda_0(u) \exp(X_{ijk}(u)\beta) du$ .

Comme les effets aléatoires sont indépendants, on a :

$$\begin{aligned} f(v_i, w_{i1}, \dots, w_{iJ_i}) &= f_V(v_i) \prod_{j=1}^{J_i} f_W(w_{ij}) \\ &= \frac{\alpha^\alpha}{\Gamma(\alpha)} v_i^{\alpha-1} \exp(-\alpha v_i) \prod_{j=1}^{J_i} \frac{\eta^\eta}{\Gamma(\eta)} w_{ij}^{\eta-1} \exp(-\eta w_{ij}) \\ &= \frac{\alpha^\alpha}{\Gamma(\alpha)} \left[ \frac{\eta^\eta}{\Gamma(\eta)} \right]^{J_i} v_i^{\alpha-1} \exp(-\alpha v_i) \prod_{j=1}^{J_i} w_{ij}^{\eta-1} \exp(-\eta w_{ij}). \end{aligned} \quad (\text{A.62})$$

Et on obtient la densité jointe des observations et des effets aléatoires en multipliant (A.61) et (A.62) :

$$\begin{aligned} f(t_i, \delta_i, v_i, w_{i1}, \dots, w_{iJ_i}) &= \frac{\alpha^\alpha}{\Gamma(\alpha)} \left[ \frac{\eta^\eta}{\Gamma(\eta)} \right]^{J_i} v_i^{\alpha-1 + \sum_{j=1}^{J_i} \sum_{k=1}^{K_{ij}} \delta_{ijk}} \exp(-\alpha v_i) \\ &\quad \prod_{j=1}^{J_i} w_{ij}^{\eta-1 + \sum_{k=1}^{K_{ij}} \delta_{ijk}} \exp \left[ -\eta w_{ij} - v_i w_{ij} \sum_k \Lambda_{ijk} \right] \\ &\quad \prod_{k=1}^{K_{ij}} [\lambda_0(t_{ijk}) \exp(X_{ijk}(t_{ijk})\beta)]^{\delta_{ijk}}. \end{aligned} \quad (\text{A.63})$$

En posant :

$$c_i = \frac{\alpha^\alpha}{\Gamma(\alpha)} \left[ \frac{\eta^\eta}{\Gamma(\eta)} \right]^{J_i} \left[ \prod_{j=1}^{J_i} \prod_{k=1}^{K_{ij}} \lambda_0(t_{ijk}) \exp(X_{ijk}(t_{ijk})\beta) \right]^{\delta_{ijk}}, \quad (\text{A.64})$$

on a :

$$\begin{aligned} f(t_i, \delta_i, v_i) &= \int_{\mathcal{W}} \dots \int_{\mathcal{W}} f(v_i, w_{i1}, \dots, w_{iJ_i}, t_i, \delta_i) dw_{i1} \dots dw_{iJ_i} \\ &= c_i v_i^{\alpha-1+\sum_{j=1}^{J_i} \sum_{k=1}^{K_{ij}} \delta_{ijk}} \exp(-\alpha v_i) \int_{\mathcal{W}} \dots \int_{\mathcal{W}} \prod_{j=1}^{J_i} w_{ij}^{\eta-1+\sum_{k=1}^{K_{ij}} \delta_{ijk}} \\ &\quad \exp \left[ -\eta w_{ij} - v_i w_{ij} \sum_k \Lambda_{ijk} \right] dw_{i1} \dots dw_{iJ_i} \\ &= c_i v_i^{\alpha-1+\sum_{j=1}^{J_i} \sum_{k=1}^{K_{ij}} \delta_{ijk}} \exp(-\alpha v_i) \prod_{j=1}^{J_i} \int_{\mathcal{W}} w_{ij}^{\eta-1+\sum_{k=1}^{K_{ij}} \delta_{ijk}} \\ &\quad \exp \left[ -\eta w_{ij} - v_i w_{ij} \sum_k \Lambda_{ijk} \right] dw_{ij} \\ &= c_i v_i^{\alpha-1+\sum_{j=1}^{J_i} \sum_{k=1}^{K_{ij}} \delta_{ijk}} \exp(-\alpha v_i) \prod_{j=1}^{J_i} \left[ \eta + v_i \sum_k \Lambda_{ijk} \right]^{-\eta-\sum_{k=1}^{K_{ij}} \delta_{ijk}} \\ &\quad \Gamma \left( \eta + \sum_{k=1}^{K_{ij}} \delta_{ijk} \right). \end{aligned} \quad (\text{A.65})$$

En substituant (A.65) dans (A.60), on obtient :

$$f(v_i|t_i, \delta_i) = \frac{v_i^{\alpha-1+\sum_j \sum_k \delta_{ijk}} \exp(-\alpha v_i) \prod_j [\eta + v_i \sum_k \Lambda_{ijk}]^{-\eta-\sum_k \delta_{ijk}}}{\int_{\mathcal{V}} v^{\alpha-1+\sum_j \sum_k \delta_{ijk}} \exp(-\alpha v) \prod_j [\eta + v \sum_k \Lambda_{ijk}]^{-\eta-\sum_k \delta_{ijk}} dv}. \quad (\text{A.66})$$

De même, on montre que :

$$f(w_{ij}|t_{ij}, \delta_{ij}) = \frac{w_{ij}^{\eta-1+\sum_k \delta_{ijk}} \exp(-\eta w_{ij}) [\alpha + w_{ij} \sum_k \Lambda_{ijk}]^{-\alpha-\sum_k \delta_{ijk}}}{\int_{\mathcal{W}} w^{\eta-1+\sum_k \delta_{ijk}} \exp(-\eta w) [\alpha + w \sum_k \Lambda_{ijk}]^{-\alpha-\sum_k \delta_{ijk}} dw}. \quad (\text{A.67})$$

Et pour un sous-groupe  $j$  fixé, on a :

$$f(v_i, w_{ij}|t_i, \delta_i) = v_i^{\alpha-1+\sum_l \sum_k \delta_{ilk}} w_{ij}^{\eta-1+\sum_k \delta_{ijk}} \frac{\exp(-v_i w_{ij} \sum_k \Lambda_{ijk})}{\Gamma(\eta + \sum_k \delta_{ijk})} \frac{\exp(-\alpha v_i - \eta w_{ij}) \prod_{l \neq j} [\eta + v_i \sum_k \Lambda_{ilk}]^{-\eta - \sum_k \delta_{ilk}}}{\int_{\mathcal{V}} v^{\alpha-1+\sum_l \sum_k \delta_{ilk}} \exp(-\alpha v) \prod_{l=1}^{J_i} [\eta + v \sum_k \Lambda_{ilk}]^{-\eta - \sum_k \delta_{ilk}} dv}. \quad (\text{A.68})$$

Les différentes espérances devant être évaluées lors de l'étape E n'admettent pas de solution analytique et doivent être simulées au moyen de méthodes de Monte Carlo dans les densités (A.66), (A.67) et (A.68). Cependant, Sastry (1997) utilise les approximations suivantes :

$$E_{\beta^{(q)}, \alpha^{(q)}, \eta^{(q)}}(w_{ij}|T = t, \delta) = \frac{\eta^{(q)} + \sum_{k=1}^{K_{ij}} \delta_{ijk}}{\eta^{(q)} + v_i \sum_{k=1}^{K_{ij}} \Lambda_{ijk}^{(q)}}, \quad (\text{A.69})$$

$$E_{\beta^{(q)}, \alpha^{(q)}, \eta^{(q)}}(\ln w_{ij}|T = t, \delta) = \psi \left( \eta^{(q)} + \sum_{k=1}^{K_{ij}} \delta_{ijk} \right) - \ln \left( \eta^{(q)} + v_i \sum_{k=1}^{K_{ij}} \Lambda_{ijk}^{(q)} \right). \quad (\text{A.70})$$

La justification invoquée en faveur de ces approximations est que comme le nombre familles au sein d'une communauté est élevé, chaque  $w_{ij}$  n'a que peu d'influence dans l'estimation du  $v_i$  de la communauté correspondante, et celui-ci peut donc être considéré comme exogène dans l'estimation de chaque  $w_{ij}$ .

## A.10. Vraisemblance partielle pénalisée et hétérogénéité gamma

Lorsque plusieurs épisodes partagent la même réalisation du terme d'hétérogénéité, la vraisemblance marginale a comme expression :

$$L_V = \prod_{i=1}^I (-1)^{d_i} \mathcal{L}_V^{(d_i)} \left( \sum_{j=1}^{J_i} \int_0^{t_{ij}} \lambda_0(u) \lambda_1(X_{ij}(u)) du \right) \prod_{j=1}^{J_i} [\lambda_0(t_{ij}) \lambda_1(X_{ij}(t_{ij}))]^{\delta_{ij}}. \quad (\text{A.71})$$

L'expression (A.71) peut se réécrire :

$$\begin{aligned} \ln L_V = \sum_{i=1}^I \ln \left[ (-1)^{d_i} \mathcal{L}_V^{(d_i)} \left( \sum_{j=1}^{J_i} \int_0^{t_{ij}} \lambda_1(X_{ij}(s)) d\Lambda_0(s) \right) \right] \\ + \sum_{j=1}^{J_i} \delta_{ij} \ln \left[ \int_0^{t_{ij}} \lambda_1(X_{ij}(s)) d\Lambda_0(s) \right]. \end{aligned} \quad (\text{A.72})$$

La loi  $\gamma(\eta, \eta)$ , admet comme transformée de Laplace :

$$\begin{aligned} \mathcal{L}_v(s) &= \int_{\mathcal{V}} \exp(-sv) \frac{\eta^\eta}{\Gamma(\eta)} v^{\eta-1} \exp(-\eta v) dv & (\text{A.73}) \\ &= \frac{\eta^\eta}{\Gamma(\eta)} \int_{\mathcal{V}} v^{\eta-1} \exp(-v(s+\eta)) dv \\ &= \frac{\eta^\eta}{\Gamma(\eta)} \frac{\Gamma(\eta)}{(s+\eta)^\eta} \\ &= \left( 1 + \frac{s}{\eta} \right)^{-\eta}. \end{aligned}$$

Sa dérivée à l'ordre  $d_i$  est donc :

$$\mathcal{L}_V^{(d_i)} = \left( -\frac{1}{\eta} \right)^{d_i} \left( 1 + \frac{s}{\eta} \right)^{-(\eta+d_i)} \prod_{l=0}^{d_i-1} (\eta + l). \quad (\text{A.74})$$

En substituant dans (A.72) et en se rappelant que  $\widehat{\beta}_{EM}$  et  $\widehat{v}_{EM}$  peuvent être exprimés comme des fonctions de  $\eta$ , on a :

$$\begin{aligned} \ln L_V(\eta) &= \sum_{j=1}^{J_i} \delta_{ij} \ln \left[ \int_0^{t_{ij}} \lambda_1(X_{ij}(s)) d\Lambda_0(s) \right] + \sum_{i=1}^I \left[ -d_i \ln \eta \right. \\ &\quad \left. - (\eta + d_i) \ln \left( 1 + \frac{\sum_{j=1}^{J_i} \int_0^{t_{ij}} \lambda_1(X_{ij}(s)) d\Lambda_0(s)}{\eta} \right) \right. \\ &\quad \left. + \ln \Gamma(\eta + d_i) - \ln \Gamma(\eta) \right]. \end{aligned} \quad (\text{A.75})$$

En substituant (1.144) dans l'équation précédente, on a :

$$\begin{aligned}
 \ln L_V(\eta) &= \sum_{j=1}^{J_i} \delta_{ij} \ln \left[ \int_0^{t_{ij}} \lambda_1(X_{ij}(s)) d\Lambda_0(s) \right] + \sum_{i=1}^I \left[ -d_i \ln \eta \right. \\
 &\quad \left. - (\eta + d_i) \ln \left( 1 + \frac{\eta + d_i}{\eta \widehat{v}_{EM}^i} - 1 \right) + \ln \Gamma(\eta + d_i) - \ln \Gamma(\eta) \right] \quad (\text{A.76}) \\
 &= \sum_{j=1}^{J_i} \delta_{ij} \ln \left[ \int_0^{t_{ij}} \lambda_1(X_{ij}) d\Lambda_0(s) \right] + \sum_{i=1}^I \left[ -d_i \ln \eta - \ln \Gamma(\eta) \right. \\
 &\quad \left. - (\eta + d_i) \ln (\eta + d_i) + (\eta + d_i) \ln (\eta \widehat{v}_{EM}^i) + \ln \Gamma(\eta + d_i) \right] \\
 &= \sum_{j=1}^{J_i} \delta_{ij} \ln \left[ \int_0^{t_{ij}} \lambda_1(X_{ij}(s)) d\Lambda_0(s) \right] + \sum_{i=1}^I \left[ -d_i \ln \eta - \ln \Gamma(\eta) \right. \\
 &\quad \left. - (\eta + d_i) \ln (\eta + d_i) + \eta \ln (\eta \widehat{v}_{EM}^i) + d_i \ln (\eta \widehat{v}_{EM}^i) + \ln \Gamma(\eta + d_i) \right] \\
 &= \sum_{j=1}^{J_i} \delta_{ij} \ln \left[ \int_0^{t_{ij}} \lambda_1(X_{ij}(s)) d\Lambda_0(s) \right] + \sum_{i=1}^I \left[ d_i \ln (\widehat{v}_{EM}^i) \right. \\
 &\quad \left. - (\eta + d_i) \ln (\eta + d_i) + \eta \ln (\eta \widehat{v}_{EM}^i) + \ln \Gamma(\eta + d_i) - \ln \Gamma(\eta) \right].
 \end{aligned}$$

En ajoutant  $g(\widehat{v}_{EM}^i, \eta) - g(\widehat{v}_{EM}^i, \eta)$  à l'expression précédente, on obtient :

$$\begin{aligned}
 \ln L_V(\eta) &= \sum_{j=1}^{J_i} \delta_{ij} \ln \left[ \int_0^{t_{ij}} \lambda_1(X_{ij}(s)) d\Lambda_0(s) \right] - g(\widehat{v}_{EM}^i, \eta) \quad (\text{A.77}) \\
 &\quad + \sum_{i=1}^I \left[ -\eta \ln \widehat{v}_{EM}^i + \eta \widehat{v}_{EM}^i - (\eta + d_i) \ln (\eta + d_i) + \eta \ln (\eta \widehat{v}_{EM}^i) \right. \\
 &\quad \left. + \ln \Gamma(\eta + d_i) - \ln \Gamma(\eta) \right] \\
 &= PPL(\eta) + \sum_{i=1}^I \left[ \eta \widehat{v}_{EM}^i - (\eta + d_i) \ln (\eta + d_i) + \eta \ln (\eta) - \ln \Gamma(\eta) \right. \\
 &\quad \left. + \ln \Gamma(\eta + d_i) \right].
 \end{aligned}$$

Notons  $\widehat{\xi}_{EM}^i = \ln \widehat{v}_{EM}^i$ . On remarque que la fonction  $g(\widehat{\xi}_{EM}^i, \eta)$  atteint son minimum par rapport à  $\widehat{\xi}_{EM}^i$  pour  $I = \sum_{i=1}^I \exp(\widehat{\xi}_{EM}^i) = \sum_{i=1}^I \widehat{v}_{EM}^i$ , et donc que cette relation est vérifiée au maximum de la vraisemblance partielle pé-



nalisee. On peut ainsi reecrire l'equation ci-dessus :

$$\ln L_V(\eta) = PPL(\eta) + \sum_{i=1}^I \left[ \eta - (\eta + d_i) \ln(\eta + d_i) + \eta \ln(\eta) - \ln \Gamma(\eta) + \ln \Gamma(\eta + d_i) \right]. \quad (\text{A.78})$$

ANNEXE B

## Annexes du Chapitre 2

## B.1. Calcul des espérances conditionnelles

Nous détaillons ici le calcul des espérances conditionnelles des  $v_i^j$  pour  $(i, j)$  donné. Notons par  $\mathcal{C}_{ij}$  l'ensemble des unités au sein du groupe  $j$ , défini au niveau de stratification  $i$ . On a :

$$f(v_1^j, \dots, v_i^j, T, d) = \prod_{k \in \mathcal{C}_{ij}} \left[ \left( \prod_{i=1}^I v_i^j \right) \lambda_0(t_k) \lambda_1 [X_k(t_k)] \right]^{\delta_k} \exp \left[ - \left( \prod_{i=1}^I v_i^j \right) \int_0^{t_k} \lambda_0(u) \lambda_1 [X_k(u)] du \right] \prod_{i=1}^I h_i(v_i^j; \alpha_i). \quad (\text{B.1})$$

En intégrant par rapport à  $v_{1j}$ , on obtient :

$$f(v_2^j, \dots, v_i^j, T, d) = \left( \prod_{k \in \mathcal{C}_{ij}} \left[ \left( \prod_{i=2}^I v_i^j \right) \lambda_0(t_k) \lambda_1 [X_k(t_k)] \right]^{\delta_k} \prod_{i=2}^I H_i(v_i^j; \alpha_i) \right) \int_{\mathcal{V}_1} v^{l_{1ik}} \exp \left[ - \left( \prod_{i=1}^I v_i^j \right) \sum_{k \in \mathcal{C}_{ij}} \int_0^{t_k} \lambda_0(u) \lambda_1 [X_k(u)] du \right] h_1(v; \alpha_1) dv, \quad (\text{B.2})$$

où  $l_{1ik}$  est le nombre de transitions observées dans l'ensemble construit comme l'intersection des groupes définis aux niveaux 1 et  $i$  contenant l'élément  $k$ . Cette densité peut être réécrite :

$$f(v_2^j, \dots, v_i^j, T, d) = \left( \prod_{k \in \mathcal{C}_{ij}} \left[ \left( \prod_{i=2}^I v_i^j \right) \lambda_0(t_k) \lambda_1 [X_k(t_k)] \right]^{\delta_k} \prod_{i=2}^I h_i(v_i^j; \alpha_i) \right) (-1)^{l_{1ik}} \mathcal{L}_1^{(l_{1ik})} \left[ \left( \prod_{i=2}^I v_i^j \right) \sum_{k \in \mathcal{C}_{ij}} \int_0^{t_k} \lambda_0(u) \lambda_1 [X_k(u)] du \right], \quad (\text{B.3})$$

où  $\mathcal{L}_1^{(l_{1ik})}$  est la dérivée à l'ordre  $l_{1ik}$  de la transformée de Laplace d'une variable aléatoire positive ou nulle  $v_1$ , définie comme :

$$\mathcal{L}_1(s) = \int_{\mathcal{V}} \exp(-sv) dH(v), \quad (\text{B.4})$$

où  $s \geq 0$ . En intégrant par rapport à  $v_{2j}$  et en omettant l'argument de la

## B.1 Calcul des espérances conditionnelles

---

transformée de Laplace pour alléger les notations, on a :

$$\begin{aligned}
 f(v_3^j, \dots, v_i^j, T, d) &= \left( \prod_{k \in \mathcal{C}_{ij}} \left[ \left( \prod_{i=3}^I v_i^j \right) \lambda_0(t_k) \lambda_1 [X_k(t_k)] \right]^{\delta_k} \prod_{i=3}^I h_i(v_i^j; \alpha_i) \right) \\
 &\quad (-1)^{l_{1ik}} \int_{\mathcal{V}_2} v^{l_{2ij}} \mathcal{L}_1^{(l_{1ik})} h_2(v, \alpha_2) dv \\
 &= \left( \prod_{k \in \mathcal{C}_{ij}} \left[ \left( \prod_{i=3}^I v_i^j \right) \lambda_0(t_k) \lambda_1 [X_k(t_k)] \right]^{\delta_k} \prod_{i=3}^I h_i(v_i^j; \alpha_i) \right) \\
 &\quad (-1)^{l_{1ik}} \mathbb{E} \left[ v_2^{j, l_{2ik}} \mathcal{L}_1^{(l_{1ik})} \right].
 \end{aligned} \tag{B.5}$$

En intégrant successivement, on montre que :

$$\begin{aligned}
 f(v_i^j, T, d) &= \left[ \prod_{k \in \mathcal{C}_{ij}} (v_i^j \lambda_0(t_k) \lambda_1 [X_k(t_k)])^{\delta_k} h_i(v_i^j; \alpha_i) \right] (-1)^{l_{1ik}} \\
 &\quad \mathbb{E} \left[ (v_I^j)^{l_{Iik}} \dots \mathbb{E} \left[ (v_{(i+1)}^j)^{l_{(i+1)ik}} \mathbb{E} \left[ (v_{(i-1)}^j)^{l_{(i-1)ik}} \dots \mathbb{E} \left[ (v_2^j)^{l_{2ik}} \mathcal{L}_1^{(l_{1ik})} \right] \right] \right] \right] \tag{B.6}
 \end{aligned}$$

Pour alléger les expressions, notons :

$$\xi_{(-i)} = \mathbb{E} \left[ (v_i^j)^{l_{Iik}} \dots \mathbb{E} \left[ (v_{(i+1)}^j)^{l_{(i+1)ik}} \mathbb{E} \left[ (v_{(i-1)}^j)^{l_{(i-1)ik}} \dots \mathbb{E} \left[ (v_2^j)^{l_{2ik}} \mathcal{L}_1^{(l_{1ik})} \right] \right] \right] \right]. \tag{B.7}$$

Ainsi :

$$\begin{aligned}
 \mathbb{E} [v_i^j | T, d] &= \frac{\int_{\mathcal{V}_i} v f(v, T, d) dv}{\int_{\mathcal{V}_i} f(v, T, d) dv} \\
 &= \frac{\int_{\mathcal{V}_i} v^{1+l_{iik}} \xi_{(-i)} h_i(v, \alpha_i) dv}{\int_{\mathcal{V}_i} v^{l_{iik}} \xi_{(-i)} h_i(v, \alpha_i) dv} \\
 &= \frac{\mathbb{E} [(v_i^j)^{1+l_{iik}} \xi_{(-i)}]}{\mathbb{E} [(v_i^j)^{l_{iik}} \xi_{(-i)}]}.
 \end{aligned} \tag{B.8}$$

## B.2. Évaluation de la matrice d'information

Le gradient se déduit des équations de (2.6) et (2.7) :

$$\frac{\partial \ln L}{\partial \alpha_i} = \sum_{k=1}^N \frac{\partial \ln h_i}{\partial \alpha_i}(v_i^j; \alpha_i), \quad (\text{B.9})$$

$$\begin{aligned} \frac{\partial \ln L}{\partial \beta} = \sum_{k=1}^N \left[ \delta_k \frac{\partial \ln \lambda_1}{\partial \beta} [X_k(u); \beta] - \left( \prod_{i=1}^I v_i^j \right) \int_0^{t_k} \lambda_0(u) \frac{\partial \lambda_1}{\partial \beta} [X_k(u); \beta] \right. \\ \left. \lambda_1 [X_k(u); \beta] du \right]. \end{aligned} \quad (\text{B.10})$$

Le dernier terme de (2.16) est la variance du score conditionnel aux observations. Pour l'évaluer, nous avons besoin de spécifier les distributions mélangeantes et le cas de la loi gamma est présenté en détails dans l'Annexe B.3.

Le premier terme de (2.16) est la matrice d'information dans le modèle complet. Le Hessien est :

$$H(\alpha, \beta) = \begin{pmatrix} H_{11}(\alpha) & 0 \\ 0 & H_{22}(\beta) \end{pmatrix}, \quad (\text{B.11})$$

où :

$$H_{11}(\alpha) = \sum_{k=1}^N \frac{\partial^2 \ln h_i}{\partial \alpha_i \partial \alpha_i'}(v_i^j, \alpha_i), \quad (\text{B.12})$$

$$H_{22}(\beta) = \sum_{k=1}^N \left( \prod_{i=1}^I v_i^j \right) \int_0^{t_k} \lambda_0(u) \frac{\partial^2 \lambda_1}{\partial \beta \partial \beta'} [X_k(u); \beta] \lambda_1 [X_k(u); \beta] du. \quad (\text{B.13})$$

L'espérance de la matrice d'information, conditionnellement aux données complètes, est évaluée une fois que l'algorithme EM a convergé vers les vraies valeurs. Nous avons donc :

$$E[-H(\alpha, \beta) | T, d, v_1, \dots, v_I] = -H(\alpha^{(*)}, \beta^{(*)}), \quad (\text{B.14})$$

où  $\alpha^{(*)}$  et  $\beta^{(*)}$  sont les estimations à la dernière étape de l'algorithme EM.

### B.3. Calcul de la matrice d'information en présence d'hétérogénéité gamma

L'évaluation de la matrice d'information d'équation (2.16), p. 103, fait intervenir l'espérance de la matrice hessienne et la variance du score. Le hessien est une matrice bloc, dont le bloc  $H_{11}(\alpha)$ , présenté dans l'équation B.12, est une matrice diagonale. Lorsque la distribution mélangeante est une loi gamma, elle a pour élément  $(i, i)$  :

$$h_{11}(i, i) = J \left( \frac{1}{\alpha_i} - \psi'(\alpha_i) \right). \quad (\text{B.15})$$

Le bloc  $H_{22}(\beta)$ , caractérisé dans l'équation (B.13), n'est pas influencé par le choix de la distribution mélangeante.

L'évaluation de la matrice d'information fait intervenir le gradient, dont les équations (B.9) et (B.10) ont pour expression :

$$\frac{\partial \ln L}{\partial \alpha_i} = \sum_{i=1}^I [\ln \alpha_i + 1 + \psi(\alpha_i) - v_i^j + \ln v_i^j], \quad (\text{B.16})$$

$$\frac{\partial \ln L}{\partial \beta} = \sum_{k=1}^N \left[ \delta_k X_k(t_k) - \left( \prod_{i=1}^I v_i^j \right) \int_0^{t_k} \lambda_0(u) X_k(t_k) \exp [X_k(u)\beta] du \right], \quad (\text{B.17})$$

où  $\psi$  est la fonction digamma, définie comme la dérivée de :

$$\ln \Gamma(\alpha_i) = \ln \int_0^{\infty} u^{\alpha_i-1} \exp(-u) du. \quad (\text{B.18})$$

La variance, conditionnelle aux observations, du score requiert  $\text{Var}(v_i^j | T, d)$ ,  $\text{Var}(\ln v_i^j | T, d)$ ,  $\text{Var} \left[ \left( \prod_{i=1}^I v_i^j \right) | T, d \right]$  et  $\text{Cov} \left( \frac{\partial \ln L}{\partial \beta}, \frac{\partial \ln L}{\partial \alpha_i} | T, d \right)$ . Elles sont obtenues à partir de la densité des termes d'hétérogénéité, conditionnelle aux observations :

$$\begin{aligned} f(v_i^j | T, d) &= \frac{f(v_i^j, T, d)}{f(T, d)} \\ &= \frac{f(v_i^j, T, d)}{\int_{\mathcal{V}_i} f(v, T, d), dv}. \end{aligned} \quad (\text{B.19})$$

Le numérateur correspond à l'équation (B.1), qui en présence d'hétérogénéité gamma admet comme expression :

$$\begin{aligned}
 f(v_1^j, \dots, v_I^j, T, d) &= \left[ \prod_{k \in \mathcal{C}_{ij}} \left( \prod_{m=1}^I v_m^j \right) \lambda_0(t_k) \lambda_1[X_k(t_k)]^{\delta_k} \right] \\
 &\quad \exp \left[ - \left( \prod_{m=1}^I v_m^j \right) \sum_{k \in \mathcal{C}_{ij}} \int_0^{t_k} \lambda_0(u) \lambda_1[X_k(u)] du \right] \\
 &\quad \prod_{m=1}^I \frac{\alpha_m^{\alpha_m}}{\Gamma(\alpha_m)} (v_m^j)^{\alpha_m-1} \exp(-\alpha_m v_m^j) \\
 &= K_i^j \left( \prod_{m=1}^I (v_m^j)^{l_{mjk} + \alpha_m - 1} \right) \exp \left[ - \sum_{m=1}^I \alpha_m v_m^j \right. \\
 &\quad \left. - \left( \prod_{m=1}^I v_m^j \right) \sum_{k \in \mathcal{C}_{ij}} \int_0^{t_k} \lambda_0(u) \lambda_1[X_k(u)] du \right], \tag{B.20}
 \end{aligned}$$

où :

$$K_i^j = \prod_{m=1}^I \frac{\alpha_m^{\alpha_m}}{\Gamma(\alpha_m)} \prod_{k \in \mathcal{C}_{ij}} \lambda_0(t_k) \lambda_1[X_k(t_k)]^{\delta_k}. \tag{B.21}$$

On en déduit :

$$\begin{aligned}
 f(v_i^j, T, d) &= \int_{\mathcal{V}_1} \dots \int_{\mathcal{V}_{i-1}} \int_{\mathcal{V}_{i+1}} \dots \int_{\mathcal{V}_I} f(v_1^j, \dots, v_I^j, T, d) dv_1^j \dots dv_I^j \\
 &= K_i^j (v_i^j)^{l_{ijk} + \alpha_i - 1} \exp(-\alpha_i v_i^j) \prod_{n \neq i} \int_{\mathcal{V}_n} (v_n^j)^{l_{njc} + \alpha_n - 1} \\
 &\quad \exp \left[ -\alpha_n v_n^j - \left( \prod_{m=1}^I v_m^j \right) \sum_{k \in \mathcal{C}_{ij}} \int_0^{t_k} \lambda_0(u) \lambda_1[X_k(u)] du \right] dv_n^j. \\
 &= K_i^j (v_i^j)^{l_{ijk} + \alpha_i - 1} \exp(-\alpha_i v_i^j) \prod_{n \neq i} \left[ \alpha_n + \left( \prod_{m=1}^I v_m^j \right) \right. \\
 &\quad \left. \sum_{k \in \mathcal{C}_{ij}} \int_0^{t_k} \lambda_0(u) \lambda_1[X_k(u)] du \right]^{-l_{njc} - \alpha_n} \Gamma(l_{njc} + \alpha_n - 1). \tag{B.22}
 \end{aligned}$$

### B.3 Calcul de la matrice d'information en présence d'hétérogénéité gamma

D'où :

$$f(v_i^j|T, d) = \frac{(v_i^j)^{p_{ijk}} \exp(-\alpha_i v_i^j) \prod_{n \neq i} \left[ \alpha_n + \left( \prod_{m=1}^I v_m^j \right) \sum_{k \in \mathcal{C}_{ij}} \Lambda_k \right]^{-1-p_{njk}}}{\int_{\mathcal{V}_i} v^{p_{ijk}} \exp(-\alpha_i v) \prod_{n \neq i} \left[ \alpha_n + \left( \prod_{m=1}^I v_m^j \right) \sum_{k \in \mathcal{C}_{ij}} \Lambda_k \right]^{-1-p_{njk}} dv}, \quad (\text{B.23})$$

où  $p_{ijk} = l_{ijk} + \alpha_i - 1$  et  $\Lambda_k = \int_0^{t_k} \lambda_0(u) \lambda_1[X_k(u)] du$ . Les quantités  $\text{Var}(v_i^j|T, d)$  et  $\text{Var}(\ln v_i^j|T, d)$  peuvent être évaluées avec les moments d'ordres 1 et 2, approchés par méthodes de Monte Carlo à partir de (B.23). On en déduit  $\text{Var}\left[\left(\prod_{i=1}^I v_i^j\right) | T, d\right]$  et  $\text{Cov}\left(\frac{\partial \ln L}{\partial \beta}, \frac{\partial \ln L}{\partial \alpha_i} | T, d\right)$  :

$$\begin{aligned} \text{Var}\left[\left(\prod_{i=1}^I v_i^j\right) | T, d\right] &= \text{E}\left[\left(\prod_{i=1}^I (v_i^j)^2\right) | T, d\right] - \text{E}\left[\left(\prod_{i=1}^I (v_i^j)\right) | T, d\right]^2 \\ &= \left\{ \prod_{i=1}^I \text{E}\left[(v_i^j)^2 | T, d\right] \right\} - \prod_{i=1}^I \left\{ [\text{E}(v_i^j | T, d)]^2 \right\} \end{aligned} \quad (\text{B.24})$$

L'évaluation de  $\text{Cov}\left(\frac{\partial \ln L}{\partial \beta}, \frac{\partial \ln L}{\partial \alpha_i} | T, d\right)$  demande  $\text{Cov}\left(\prod_{i=1}^I v_i^j, v_i^j - \ln v_i^j | T, d\right)$ , qui est :

$$\begin{aligned} &\text{Cov}\left(\prod_{i=1}^I v_i^j, v_i^j - \ln v_i^j | T, d\right) \\ &= \text{E}\left[\left(\prod_{i=1}^I v_i^j\right) (v_i^j - \ln v_i^j) | T, d\right] - \text{E}\left[\prod_{i=1}^I v_i^j | T, d\right] \text{E}\left[v_i^j - \ln v_i^j | T, d\right] \\ &= \text{E}\left[\prod_{i' \neq i} v_{i'}^j | T, d\right] \text{E}\left[(v_i^j)^2 - v_i^j \ln v_i^j | T, d\right] \\ &\quad - \text{E}\left[\prod_{i' \neq i} v_{i'}^j | T, d\right] (\text{E}\left[v_i^j | T, d\right] \text{E}\left[v_i^j - \ln v_i^j | T, d\right]) \\ &= \left(\prod_{i' \neq i} \text{E}\left[v_{i'}^j | T, d\right]\right) [\text{Var}(v_i^j | T, d) - \text{Cov}(v_i^j, \ln v_i^j | T, d)]. \end{aligned} \quad (\text{B.25})$$



## B.4. Programme de l'algorithme EMPL

Le programme ci dessous correspond à un algorithme EMPL dont les écarts types estimés sont évalués par l'approche de Louis (1982), codé pour le logiciel R (Team, 2005).

```
rm(list=ls())

library(survival) #pour la fonction coxph
library(MASS) #pour la fonction area

Nech <- 1 #nombre d'échantillons à estimer
Nreg <- 2 #nombre de régresseurs
Nef <- 2 #nombre d'effets
#J <-#nombre

nbiter1v <- numeric(Nech)
nbiter1w <- numeric(Nech)
varfrailtiesv <- numeric(Nech)
varfrailtiesw <- numeric(Nech)

breslowk <- numeric(1)
partdetk <- numeric(1)
nelson.aalenwk <- numeric(1)
ntrans <- numeric(1)
numg <- numeric(1)
numd <- numeric(1)
numera <- numeric(1)

Ev <- numeric(1)
Ev2 <- numeric(1)
Vv <- numeric(1)
Elnv <- numeric(1)
Elnv2 <- numeric(1)
Vlnv <- numeric(1)
Evlv <- numeric(1)
Covvlnv <- numeric(1)

Ew <- numeric(1)
Ew2 <- numeric(1)
Vw <- numeric(1)
```

## B.4 Programme de l'algorithme EMPL

---

```
Elnw <- numeric(1)
Elnw2 <- numeric(1)
Vlnw <- numeric(1)
Ewlnw <- numeric(1)
Covwlnw <- numeric(1)

#emplacements et noms des fichiers de donnees
toto <- paste("d:/Econométrie/2f-var-2/cens-unif-0/ech-1000-\\
100-",c(seq(2, 2, length=Nech)), "/2-frailty-trie-1000-", 1:\\
Nech, ".raw", sep = "")

#emplacements des fichiers de sauvegarde
#titi <- paste("d:/Program Files/R/rw2001/EMPL-Louis-2-0-1000\\
-100-", c(2) , ".RData", sep = "")

simuldata <- read.table(toto[1])

#####lecture des données#####

com <- simuldata[,1] #conventions
fam <- simuldata[,2] #pays
#v <- simuldata[,3]
#vfam <- simuldata[,4]
x <- simuldata[,5]
y <- simuldata[,7]
d <- simuldata[,13]
tobs <- simuldata[,15]
id <- simuldata[,16]
#stock la valeur true dans la variable "event" si une transition \\
est observée, false sinon
event <- d == 1
N <- length(tobs)
Nfam <- max(fam)
Ncom <- max(com)

#mise en forme matricielle des régresseurs
X <- cbind(x,y)

#####Valeurs initiales#####
```

```
# Frailty Model, gamma distribution, sur l'effet famille
mod.vfam <-coxph(Surv(tobs,event) ~ x + y + frailty(fam, \\
distribution="gamma"))

# Frailty Model, gamma distribution, sur l'effet communauté
mod.vcom <-coxph(Surv(tobs,event) ~ x + y + frailty(com, \\
distribution="gamma"))

#récupération des résultats de la regression avec hno au \\
niveau des fam
lnwchap <- numeric(1)
for(i02 in 1:Nfam) {lnwchap[which(fam==i02)] <- mod.vfam$\\
frail[i02]}

#récupération des résultats de la regression avec hno au \\
niveau des com
lnvchap <- numeric(1)
for(i01 in 1:Ncom) {lnvchap[which(com==i01)] <- mod.vcom$\\
frail[i01]}

##### EMPL #####

#initialisation du critere de convergence
#eps <- .001
eps <- .0001
crit <- 1
niter <- 0
b1q <- 2
b2q <- 2

#enregistrement du temps
ptm <- proc.time()

while (crit > eps) {
#vrais pénalisée sur effet fam + offset com
mod.w <-coxph(Surv(tobs,event) ~ x + y + offset(lnvchap) + \\
frailty.gamma(fam, sparse=TRUE))

#opération de sauvegarde
lnwchapold <- lnwchap
```

## B.4 Programme de l'algorithme EMPL

---

```
#récupération des w estimés
for(i01 in 1:Nfam) {lnwchap[which(fam==i01)] <- \\
mod.w$frail[i01]}

#vrais pénalisée sur effet com + offset fam
mod.v <-coxph(Surv(tobs,event) ~ x + y + offset(lnwchap) + \\
frailty.gamma(com, sparse=TRUE))

#opération de sauvegarde
lnvchapold <- lnvchap

#récupération des v estimés
for(i02 in 1:Ncom) {lnvchap[which(com==i02)] <- \\
mod.v$frail[i02]}

#PPL avec frailty estimé en offset
mod.b <-coxph(Surv(tobs,event) ~ x + y + offset(lnvchap)+ \\
offset(lnwchap))

#récupération des coefs
b1vieux <- b1q
b2vieux <- b2q
b1q <- coef(mod.b)[1]
b2q <- coef(mod.b)[2]

#calcul et affichage critere de convergence
crit <- sqrt((b1q - b1vieux)^2 + (b2q - b2vieux)^2)
niter <- niter + 1

#pour passer a l'ech suivant si ca itere a l'infini
if (niter > 50) crit <- 0

} #fin des iterations sur un echantillon donne

#temps d'estimation sur l'échantillon
timepersample <- proc.time() - ptm

##### Louis #####
```

```

#récupération des paramètres des distro mélangeantes
nbiter1v <- mod.v[[5]][1]
nbiter1w <- mod.w[[5]][1]
varfrailtiesv <- mod.v$history$"frailty.gamma(com, \\  

sparse = TRUE)"$history[nbiter1v,1] #isgt de la \\  

variance de v
varfrailtiesw <- mod.w$history$"frailty.gamma(fam, \\  

sparse = TRUE)"$history[nbiter1w,1] #isgt de la \\  

variance de w

alpha <- 1/(varfrailtiesv) #paramètre effet \\  

convention (com)
eta <- 1/(varfrailtiesw) #paramètre effet \\  

pays (fam)

#calcul hasard intégré
#évaluation de la partie déterministe du hasard en \\  

vue de Breslow
partdet <- exp(x * b1q + y * b2q + lnvchap + lnwchap)

#estimateur de breslow du hasard de base intégré
breslow <- numeric(1)
breslow <- cumsum(d/ (sum(partdet) - cumsum(partdet) \\  

+ partdet))

#calcul de l'estimateur de Nelson-Aalen du hasard intégré\ \  

(régresseurs constants dans le temps)
nelson.aalen <- breslow * exp( x * b1q + y * b2q)
nelson.aalenvw <- breslow * exp(lnvchap + lnwchap + \\  

x * b1q + y * b2q)

#####\blocs de (- Hessien) #####
h1 <- - Ncom * (1/alpha - trigamma(alpha))
h2 <- - Nfam * (1/eta - trigamma(eta))
#h3 <- t(X) %*% diag(nelson.aalenvw, N, N) %*% X

#lorsque bcp obsř:
b <- matrix(0, Nreg, N)
h3 <- matrix(0, Nreg, Nreg)
for (i0 in 1:Nreg) {
for (i1 in 1:N) {

```

## B.4 Programme de l'algorithme EMPL

---

```
b[i0, i1] <- sum( t(X)[i0, i1] * nelson.aalenvw[i1] )
}
}
for (i0 in 1:Nreg) {
for (i1 in 1:Nreg) {
h3[i0, i1] <- sum(b[i0, 1:N] * X[1:N,i1])
}
}

#assemblage de - la matrice Hessiene
H <- matrix(0,Nreg+2,Nreg+2)
H[1,1] <- h1
H[2,2] <- h2
H[3:(Nreg+2), 3:(Nreg+2)] <- h3

##### Variance du score #####

#calcul du numérateur de f(v|T,d)

for(i03 in 1:Ncom) {
#évaluation de la partie déterministe du hasard en vue \\
de Breslow (contient hno)
partdetk[which(com==i03)] <- exp(x * b1q + y * b2q + \\
lnvchap + lnwchap)

#estimateur de breslow du hasard de base intégré
breslowk[which(com==i03)] <- cumsum(d[which(com==i03)]/ \\
(sum(partdetk[which(com==i03)]) - cumsum(partdetk[which\\
(com==i03)]) + partdetk[which(com==i03)]))

#calcul de l'estimateur de Nelson-Aalen du hasard intégré \\
(régresseurs constants dans le temps)
nelson.aalenwk[which(com==i03)] <- breslowk[which(com==i03)] \\
* exp( lnwchap + x * b1q + y * b2q)

ntrans[which(com==i03)] <- sum(d[which(com==i03)])
numg[which(com==i03)] <- exp(- alpha * exp(lnvchap\\
[which(com==i03)])) * exp(lnvchap[which(com==i03)]) ^ \\
(alpha + ntrans[which(com==i03)] - 1)
numd[which(com==i03)] <- (eta + nelson.aalenwk\\
```

```

[which(com==i03)]^( -alpha - ntrans[which(com==i03)] )
numera[which(com==i03)] <- numg[which(com==i03)] * \
numd[which(com==i03)]
}

#calcul du dénominateur de f(v|T,d)
numerav <- fonction(v) {exp(- alpha * v) * v ^ \
(alpha + ntrans - 1) * (eta + v * breslowk * \
exp(x * b1q + y * b2q) )^( -alpha - ntrans)}
denov <- area(numerav, 0, 100)
denov <- replace(denov, which(denov==NA), 0.00001^6)

#densité f(v|T,d)
fv <- numera / denov

#tirage dans une loi exponentielle unitaire
u <- rexp(100)

#calcul de E(v|T,d), E(lnv|T,d), V(v|T,d), \
V(lnv|T,d) et cov(v,lnv|T,d)
for(i in 1:N) {
Evs <- u * fv[i] / exp(-u)
Ev2s <- u^2 * fv[i] / exp(-u)
Vvs <- Ev2 - Ev^2

Elnvs <- log(u) * fv[i] / exp(-u)
Elnv2s <- log(u)^2 * fv[i] / exp(-u)
Vlnvs <- Elnv2 - Elnv^2

Evlnvs <- u * log(u) * fv[i] / exp(-u)
Covvlnvs <- Evlnv - Ev * Elnv

Ev[i] <- sum(Evs)
Ev2[i] <- sum(Ev2s)
Vv[i] <- sum(Vvs)

Elnv[i] <- sum(Elnvs)
Elnv2[i] <- sum(Elnv2s)
Vlnv[i] <- sum(Vlnvs)

Evlnv[i] <- sum(Evlnvs)

```

## B.4 Programme de l'algorithme EMPL

---

```
Covvlnv[i] <- sum(Covvlnvs)

}

#calcul du numérateur de f(w|T,d)

for(i04 in 1:Nfam) {
#évaluation de la partie déterministe du \
hasard en vue de Breslow (contient hno)
partdetk[which(fam==i04)] <- exp(x * b1q + \
y * b2q + lnvchap + lnwchap)

#estimateur de breslow du hasard de base \
intégré
breslowk[which(fam==i04)] <- cumsum(\
d[which(fam==i04)]/ (sum(partdetk[which(fam==i04)]) \
- cumsum(partdetk[which(fam==i04)])) + \
partdetk[which(fam==i04)]))

#calcul de l'estimateur de Nelson-Aalen du hasard \
intégré (régresseurs constants dans le temps)
nelson.aalenwk[which(fam==i04)] <- breslowk\
[which(fam==i04)] * exp( lnwchap + x * b1q + y * b2q)

ntrans[which(fam==i04)] <- sum(d[which(fam==i04)])
numg[which(fam==i04)] <- exp(- eta * exp(lnvchap\
[which(fam==i04)])) * exp(lnvchap[which(fam==i04)])\
^ (eta + ntrans[which(fam==i04)] - 1)
numd[which(fam==i04)] <- (alpha + nelson.aalenwk\
[which(fam==i04)])^( -alpha - ntrans[which(fam==i04)] )
numera[which(fam==i04)] <- numg[which(fam==i04)] * \
numd[which(fam==i04)]
}

#calcul du dénominateur de f(v|T,d)
numeraw <- fonction(w) {exp(- eta * w) * w ^ \
(eta + ntrans - 1) * (alpha + w * breslowk * \
exp(x * b1q + y * b2q) )^( -alpha - ntrans)}
denow <- area(numeraw, 0, 100)
denow <- replace(denow, which(denow==NA), 0.00001^6)
```



```

#densité f(v|T,d)
fw <- numera / denow

#calcul de E(w|T,d), E(lnw|T,d), w(w|T,d),\
w(lnw|T,d) et cov(w,lnw|T,d)
for(i in 1:N) {
Ews <- u * fw[i] / exp(-u)
Ew2s <- u^2 * fw[i] / exp(-u)
Vws <- Ew2 - Ew^2

Elnws <- log(u) * fw[i] / exp(-u)
Elnw2s <- log(u)^2 * fw[i] / exp(-u)
Vlnws <- Elnw2 - Elnw^2

Ewlnws <- u * log(u) * fw[i] / exp(-u)
Covwlnws <- Ewlnw - Ew * Elnw

Ew[i] <- sum(Ews)
Ew2[i] <- sum(Ew2s)
Vw[i] <- sum(Vws)

Elnw[i] <- sum(Elnws)
Elnw2[i] <- sum(Elnw2s)
Vlnw[i] <- sum(Vlnws)

Ewlnw[i] <- sum(Ewlnws)
Covwlnw[i] <- sum(Covwlnws)

}

Vvw <- Ev2 * Ew2 - Ev^2 * Ew^2

#calcul de la variance du score
v11 <- Ncom/N * sum( Vlnv + Vv - 2 * Covvlnv )
v22 <- Nfam/N * sum(Vlnw + Vw - 2 * Covwlnw )
v33 <- t(X) %*% diag(Vvw * nelson.aalen^2, N, N) %*% X
v13 <- t(nelson.aalen * (Vv - Covvlnv)) %*% X
v23 <- t(nelson.aalen * (Vw - Covwlnw)) %*% X

#assemblage de la matrice de variance du score
VS <- matrix(0,Nreg+2,Nreg+2)

```

## B.5 Programme de l'algorithme EM accéléré

---

```
VS[1,1] <- v11
VS[2,2] <- v22
VS[3:(Nreg+2), 3:(Nreg+2)] <- v33
VS[1, 3:(Nreg+2)] <- v13
VS[3:(Nreg+2), 1] <- t(v13)
VS[2, 3:(Nreg+2)] <- v23
VS[3:(Nreg+2), 2] <- t(v23)

#matrice d'information
I <- H - VS
sd <- chol(diag(diag(solve(I)),4,4))

#sauvegarde resultats
save.image(titi)
```

## B.5. Programme de l'algorithme EM accéléré

Le code ci dessous est une extension de l'algorithme EM accéléré de Sastry (1997) à la vraisemblance partielle, au moyen d'un argument similaire de Johansen (1983). Le programme est conçu pour le logiciel R (Team, 2005).

```
rm(list=ls())

ptm <- proc.time()

library(survival) #pour la fonction coxph
library(stats) #pour les fonctions optim et offset

simuldata <- read.table("c:/Econométrie/data/simul/2_nested\\
_frailties_tries26.raw")

#####
#lecture des données #
#####

com      <-simuldata[,1]
v        <-simuldata[,2]
fam      <-simuldata[,3]
vfam     <-simuldata[,4]
```

```

x          <-simuldata[,5]
beta1     <-simuldata[,6]
y         <-simuldata[,7]
beta2     <-simuldata[,8]
varexplib <-simuldata[,9]
L0        <-simuldata[,10]
theta     <-simuldata[,11]
torig     <-simuldata[,12]
tcens     <-simuldata[,13]
tobs      <-simuldata[,14]
d         <-simuldata[,15]
id <- simuldata[,16]
#stock la valeur true dans la variable "event" si une \
transition est observée, false sinon
event     <- d == 1
n <- length(tobs)
Nfam <- max(fam)
Ncom <- max(com)

#####
# Frailty Model, gamma distribution, sur l'effet famille
#mod.1 <-coxph(Surv(tobs,event) ~ x + y )

# Frailty Model, gamma distribution, sur l'effet famille
mod.2 <-coxph(Surv(tobs,event) ~ x + y + frailty(fam))
#summary(mod.2)

# Frailty Model, gamma distribution, sur l'effet communauté
mod.3 <-coxph(Surv(tobs,event) ~ x + y + frailty(com))
#summary(mod.3)
#####

#récupération des résus de cox ne considérant que des \
effets fam
b1 <- coef(mod.3)[1] #utilise les résus de l'effet \
famille pour réduire le "downward bias"
b2 <- coef(mod.3)[2]

#affectation au bon id des effets famille estimés

```

## B.5 Programme de l'algorithme EM accéléré

---

```
wchap <- numeric(n)
for(i0 in 1:Nfam) {wchap[which(fam==i0)] <- \\  
exp(mod.2[[10]][i0])}
#wchap <- rep(1,n)

#récupération des résultats de la regression avec hno \\  
au niveau des com
vchap <- numeric(n)
for(i01 in 1:Ncom) {vchap[which(com==i01)] <- \\  
exp(mod.3[[10]][i01])}
#vchap <- rep(1,n)

#valeurs initiales des paramètres
alpha <- 5
eta <- 5

lnvchap <- numeric(n)
lnwchap <- numeric(n)

#initialisation du critere de convergence pour les \\  
boucles globales
epstot <- .0001
crittot <- 1
niter.tot <- 1

while (crittot > epstot) {

#####
### Algorithme EM sur l'effet communauté ###
#####

#####Algorithme EM sur Qalpha#####

#initialisation du critere de convergence pour les \\  
boucles sur Qalpha
epsEM <- .001
critEM <- 1
niter.EM <- 1

while (critEM > epsEM) {
```

```

#Initialisation critere de convergence et nb \\
iteration pour boucles Lancaster
epsL <- .001
critL <- 1
niter.L <- 1

#####Etape E sur V #####
#####Boucles de Lancaster sur v #####

while (critL > epsL) {

#évaluation de la partie déterministe du hasard en\\
vue de Breslow (contient hno) (OK)
partdet <- vchap * wchap * exp(x * b1 + y * b2 )

#estimateur de breslow du hasard de base intégré
#survfit calcul la fonction de survie et basehaz le \\
hasard de base
#(cf le fichier breslow 3 working)
breslow <- numeric(n)
breslow <- cumsum(d/ (sum(partdet) - cumsum(partdet)\\
+ partdet))

#calcul de l'estimateur de Nelson-Aalen du hasard \\
intégré (sans hno)
nelson.aalen <- breslow * exp(x * b1 + y * b2)

#calcul de vchap et lnvchap, en utilisant equation \\
(A5) de l'annexe A
vchaps <- numeric(Ncom)
lnvchaps <- numeric(Ncom)

for (i1 in 1:Ncom) {
#calcul de vchap
numerav <- alpha + sum(d[which(com==i1)])
denomiv <- alpha + sum((wchap * nelson.aalen)\\
[which(com==i1)])
vchaps[i1] <- numerav/denomiv
#calcul de lnvchap
lnvchaps[i1] <- digamma(numerav) - log(denomiv)
}

```

## B.5 Programme de l'algorithme EM accéléré

---

```
#calcul et affichage critere de convergence
vcrit <- numeric(Ncom)
for(i2 in 1:Ncom) {vcrit[i2] <- vchap[which(com==i2)][1]}
critL <- sqrt(sum((vchaps - vcrit)^2))
#cat("Lancaster:", niter.L, "CritL:", critL, "Vchaps:"\\
, vchaps, "\\n")
niter.L <- niter.L + 1

#Sauvegarde des resultats et mise en forme
for (i3 in 1:Ncom) {
vchap[which(com==i3)] <- vchaps[i3]
lnvchap[which(com==i3)] <- lnvchaps[i3]
}
}

##### Fin boucles de Lancaster sur v #####

##### Etape M sur v #####
Qalpha <- fonction(alpha) {Ncom * alpha * log(alpha) + \\
(alpha-1) * sum(lnvchaps) - alpha * sum(vchaps) - Ncom * \\
  lgamma(alpha)}
alpha.max <- optimize(Qalpha, l=0, u=1000 , maximum = TRUE, \\
tol = 0.0001)

#calcul et affichage critere convergence de l'algo EM sur V
critEM <- sqrt((alpha.max[[1]] - alpha)^2)
cat("EM sur V:", niter.EM, "CritEM:", critEM, "Alpha:", \\
alpha.max[[1]], "\\n")
niter.EM <- niter.EM + 1

#récupération des résultats
alpha <- alpha.max[[1]]

}

#####Algorithme EM sur Qeta#####

#initialisation du critere de convergence pour les boucles \\
sur Qeta
epsEMw <- .001
```

```

critEMw <- 1
niter.EMw <- 1

while (critEMw > epsEMw) {

#Initialisation critere de convergence et nb iteration pour \\  

boucles Lancaster
epsLw <- .001
critLw <- 1
niter.Lw <- 1

#####Etape E sur W #####

#####Boucles de Lancaster sur W #####

while (critLw > epsLw) {

#évaluation de la partie déterministe du hasard en vue de \\  

Breslow (contient hno) (OK)
partdetw <- vchap * wchap * exp(x * b1 + y * b2 )

#estimateur de breslow du hasard de base intégré
#survfit calcul la fonction de survie et basehaz le hasard \\  

de base
#(cf le fichier breslow 3 working)
bresloww <- numeric(n)
bresloww <- cumsum(d/ (sum(partdetw) - cumsum(partdetw) + \\  

partdetw))

#calcul de l'estimateur de Nelson-Aalen du hasard intégré \\  

(sans hno)
nelson.aalenw <- bresloww * exp(x * b1 + y * b2)

#calcul de wchap et lnwchap, en utilisant equation (A3) et \\  

(A4) de l'annexe A
wchaps <- numeric(Nfam)
lnwchaps <- numeric(Nfam)

for (i1 in 1:Nfam) {
#calcul de wchap
numeraw <- eta + sum(d[which(fam==i1)])

```

## B.5 Programme de l'algorithme EM accéléré

---

```
denomiw <- eta + sum((vchap * nelson.aalenw)\
[which(fam==i1)])
wchaps[i1] <- numeraw/denomiw
#calcul de lnwchap
lnwchaps[i1] <- digamma(numeraw) - log(denomiw)
}

#calcul et affichage critere de convergence
wcrit <- numeric(Nfam)
for(i2 in 1:Nfam) {wcrit[i2] <- wchap[which(fam==i2)][1]}
critLw <- sqrt(sum((wchaps - wcrit)^2))
#cat("Lancaster W:", niter.Lw, "CritLw:", critLw, "wchaps:"\
, wchaps, "\n")
niter.Lw <- niter.Lw + 1

#Sauvegarde des resultats et mise en forme
for (i3 in 1:Nfam) {
wchap[which(fam==i3)] <- wchaps[i3]
lnwchap[which(fam==i3)] <- lnwchaps[i3]
}
}

##### Fin boucles de Lancaster sur W #####

##### Etape M sur W #####
Qeta <- fonction(eta) {Nfam * eta * log(eta) + (eta-1) * \
sum(lnwchaps) - eta * sum(wchaps) - Nfam * lgamma(eta)}
eta.max <- optimize(Qeta, l=0, u=10000 , maximum = TRUE,\
tol = 0.0001)

#calcul et affichage critere convergence de l'algo EM sur w
critEMw <- sqrt((eta.max[[1]] - eta)^2)
cat("EM sur W:", niter.EMw, "CritEMw:", critEMw, "Eta:",\
eta.max[[1]], "\n")
niter.EMw <- niter.EMw + 1

#récupération des résultats
eta <- eta.max[[1]]
}
```



```
#####
# Frailty Model, gamma distribution sur l'effet famille et\\
  récupération des v estimés #####
#####
logvchap <- log(vchap)
logwchap <- log(wchap)
mod.4 <- coxph(Surv(tobs,event) ~ x + y + offset(logvchap) \\
+ offset(logwchap))
#summary(mod.4)

#affectation au bon id des effets famille estimés
#for(i4 in 1:Nfam) {wchap[which(fam==i4)] <- \\
exp(mod.4[[10]][i4])}

#récupération des coefs
b1vieux <- b1
b2vieux <- b2
b1 <- coef(mod.4)[1]
b2 <- coef(mod.4)[2]

#calcul et affichage critere convergence de l'algo EM total
crittot <- sqrt((b1 - b1vieux)^2 + (b2 - b2vieux)^2)
varV <- 1/ alpha.max[[1]]
varVfam <- 1/ eta.max[[1]]
cat("EM:", niter.tot, "Alpha:", alpha.max[[1]], "Eta:", \\
eta.max[[1]], "b1:", b1, "b2:", b2, "V(v):", varV, \\
"V(vfam):", varVfam, "\\n")
niter.tot <- niter.tot + 1

}

totaltime <- proc.time() - ptm
cat(totaltime, "\\n")

##### Louis #####

alpha <- alpha.max[[1]] #paramètre effet \\
convention (com)
eta <- eta.max[[1]] #paramètre effet \\
pays (fam)
```

## B.5 Programme de l'algorithme EM accéléré

---

```
#calcul hasard intégré
#évaluation de la partie déterministe du hasard en \\  
vue de Breslow
partdet <- exp(x * b1 + y * b2 + lnvchap + lnwchap)

#estimateur de breslow du hasard de base intégré
breslow <- numeric(1)
breslow <- cumsum(d/ (sum(partdet) - cumsum(partdet) \\  
+ partdet))

#calcul de l'estimateur de Nelson-Aalen du hasard intégré\<\  
(régresseurs constants dans le temps)
nelson.aalen <- breslow * exp( x * b1 + y * b2)
nelson.aalenvw <- breslow * exp(lnvchap + lnwchap + \\  
x * b1 + y * b2)

#####\blocs de (- Hessien) #####
h1 <- - Ncom * (1/alpha - trigamma(alpha))
h2 <- - Nfam * (1/eta - trigamma(eta))
#h3 <- t(X) %%% diag(nelson.aalenvw, N, N) %%% X

#lorsque bcp obsř:
b <- matrix(0, Nreg, N)
h3 <- matrix(0, Nreg, Nreg)
for (i0 in 1:Nreg) {
  for (i1 in 1:N) {
    b[i0, i1] <- sum( t(X)[i0, i1] * nelson.aalenvw[i1] )
  }
}
for (i0 in 1:Nreg) {
  for (i1 in 1:Nreg) {
    h3[i0, i1] <- sum(b[i0, 1:N] * X[1:N,i1])
  }
}

#assemblage de - la matrice Hessiene
H <- matrix(0,Nreg+2,Nreg+2)
H[1,1] <- h1
H[2,2] <- h2
H[3:(Nreg+2), 3:(Nreg+2)] <- h3
```

```
##### Variance du score #####

#calcul du numérateur de f(v|T,d)

for(i03 in 1:Ncom) {
#évaluation de la partie déterministe du hasard en vue \
de Breslow (contient hno)
partdetc[which(com==i03)] <- exp(x * b1 + y * b2 + \
lnvchap + lnwchap)

#estimateur de breslow du hasard de base intégré
breslowk[which(com==i03)] <- cumsum(d[which(com==i03)]/ \
(sum(partdetc[which(com==i03)]) - cumsum(partdetc[which\
(com==i03)]) + partdetc[which(com==i03)]))

#calcul de l'estimateur de Nelson-Aalen du hasard intégré \
(régresseurs constants dans le temps)
nelson.aalenwk[which(com==i03)] <- breslowk[which(com==i03)] \
* exp( lnwchap + x * b1 + y * b2)

ntrans[which(com==i03)] <- sum(d[which(com==i03)])
numg[which(com==i03)] <- exp(- alpha * exp(lnvchap\
[which(com==i03)])) * exp(lnvchap[which(com==i03)]) ^ \
(alpha + ntrans[which(com==i03)] - 1)
numd[which(com==i03)] <- (eta + nelson.aalenwk\
[which(com==i03)])^( -alpha - ntrans[which(com==i03)] )
numera[which(com==i03)] <- numg[which(com==i03)] * \
numd[which(com==i03)]
}

#calcul du dénominateur de f(v|T,d)
numerav <- function(v) {exp(- alpha * v) * v ^ \
(alpha + ntrans - 1) * (eta + v * breslowk * \
exp(x * b1 + y * b2) )^( -alpha - ntrans)}
denov <- area(numerav, 0, 100)
denov <- replace(denov, which(denov==NA), 0.00001^6)

#densité f(v|T,d)
fv <- numera / denov
```

## B.5 Programme de l'algorithme EM accéléré

---

```
#tirage dans une loi exponentielle unitaire
u <- rexp(100)

#calcul de E(v|T,d), E(lnv|T,d), V(v|T,d), \\  
V(lnv|T,d) et cov(v,lnv|T,d)
for(i in 1:N) {
  Evs <- u * fv[i] / exp(-u)
  Ev2s <- u^2 * fv[i] / exp(-u)
  Vvs <- Ev2 - Ev^2

  Elnvs <- log(u) * fv[i] / exp(-u)
  Elnv2s <- log(u)^2 * fv[i] / exp(-u)
  Vlnvs <- Elnv2 - Elnv^2

  Evlnvs <- u * log(u) * fv[i] / exp(-u)
  Covvlnvs <- Evlnv - Ev * Elnv

  Ev[i] <- sum(Evs)
  Ev2[i] <- sum(Ev2s)
  Vv[i] <- sum(Vvs)

  Elnv[i] <- sum(Elnvs)
  Elnv2[i] <- sum(Elnv2s)
  Vlnv[i] <- sum(Vlnvs)

  Evlnv[i] <- sum(Evlnvs)
  Covvlnv[i] <- sum(Covvlnvs)
}

#calcul du numérateur de f(w|T,d)

for(i04 in 1:Nfam) {
  #évaluation de la partie déterministe du \\  
hasard en vue de Breslow (contient hno)
  partdetk[which(fam==i04)] <- exp(x * b1 + \\  
y * b2 + lnvchap + lnwchap)

  #estimateur de breslow du hasard de base \\  
intégré
  breslowk[which(fam==i04)] <- cumsum(\
```

```

d[which(fam==i04)] / (sum(partdetk[which(fam==i04)]) \\  

- cumsum(partdetk[which(fam==i04)]) + \\  

partdetk[which(fam==i04)])

#calcul de l'estimateur de Nelson-Aalen du hasard \\  

intégré (régresseurs constants dans le temps)
nelson.aalenwk[which(fam==i04)] <- breslowk\<\  

[which(fam==i04)] * exp( lnwchap + x * b1 + y * b2)

ntrans[which(fam==i04)] <- sum(d[which(fam==i04)])
numg[which(fam==i04)] <- exp(- eta * exp(lnvchap\<\  

[which(fam==i04)])) * exp(lnvchap[which(fam==i04)])\\  

^ (eta + ntrans[which(fam==i04)] - 1)
numd[which(fam==i04)] <- (alpha + nelson.aalenwk\<\  

[which(fam==i04)])^( -alpha - ntrans[which(fam==i04)] )
numera[which(fam==i04)] <- numg[which(fam==i04)] * \<\  

numd[which(fam==i04)]
}

#calcul du dénominateur de f(v|T,d)
numeraw <- fonction(w) {exp(- eta * w) * w ^ \<\  

(eta + ntrans - 1) * (alpha + w * breslowk * \<\  

exp(x * b1 + y * b2) )^( -alpha - ntrans)}
denow <- area(numeraw, 0, 100)
denow <- replace(denow, which(denow==NA), 0.00001^6)

#densité f(v|T,d)
fw <- numera / denow

#calcul de E(w|T,d), E(lnw|T,d), w(w|T,d),\<\  

w(lnw|T,d) et cow(w,lnw|T,d)
for(i in 1:N) {
Ews <- u * fw[i] / exp(-u)
Ew2s <- u^2 * fw[i] / exp(-u)
Vws <- Ew2 - Ew^2

Elnws <- log(u) * fw[i] / exp(-u)
Elnw2s <- log(u)^2 * fw[i] / exp(-u)
Vlnws <- Elnw2 - Elnw^2

Ewlnws <- u * log(u) * fw[i] / exp(-u)

```

## B.5 Programme de l'algorithme EM accéléré

---

```
Covwlnws <- Ewlnw - Ew * Elnw

Ew[i] <- sum(Ews)
Ew2[i] <- sum(Ew2s)
Vw[i] <- sum(Vws)

Elnw[i] <- sum(Elnws)
Elnw2[i] <- sum(Elnw2s)
Vlnw[i] <- sum(Vlnws)

Ewlnw[i] <- sum(Ewlnws)
Covwlnw[i] <- sum(Covwlnws)

}

Vvw <- Ev2 * Ew2 - Ev^2 * Ew^2

#calcul de la variance du score
v11 <- Ncom/N * sum( Vlnv + Vv - 2 * Covvlnv )
v22 <- Nfam/N * sum(Vlnw + Vw - 2 * Covwlnw )
v33 <- t(X) %*% diag(Vvw * nelson.aalen^2, N, N) %*% X
v13 <- t(nelson.aalen * (Vv - Covvlnv)) %*% X
v23 <- t(nelson.aalen * (Vw - Covwlnw)) %*% X

#assemblage de la matrice de variance du score
VS <- matrix(0,Nreg+2,Nreg+2)
VS[1,1] <- v11
VS[2,2] <- v22
VS[3:(Nreg+2), 3:(Nreg+2)] <- v33
VS[1, 3:(Nreg+2)] <- v13
VS[3:(Nreg+2), 1] <- t(v13)
VS[2, 3:(Nreg+2)] <- v23
VS[3:(Nreg+2), 2] <- t(v23)

#matrice d'information
I <- H - VS
sd <- chol(diag(diag(solve(I)),4,4))
```

## B.6. Résultats concernant la ratification des conventions du BIT

TAB. B.1 – Estimations des  $\beta$ 

Variable	Bayes		EMPL	
	Coef.	S.d	Coef.	S.d
<b>Coûts</b>				
PIB réel par tête <sup>a</sup>	<b>3.81</b>	1.40	<b>3.03</b>	1.39
PIB réel par tête, au carré	<b>-3.19</b>	1.51	-2.41	1.51
Pas de mise à jour	<b>1.39</b>	0.27	<b>1.26</b>	0.37
Ratification antérieure si mise à jour	<b>1.62</b>	0.36	<b>1.52</b>	0.38
Population <sup>b</sup>	-0.02	0.05	-0.03	0.05
<b>Pressions Intérieures</b>				
Démocratie	<b>0.34</b>	0.15	0.29	0.15
Majorité de gauche	<b>-0.69</b>	0.31	<b>-0.62</b>	0.31
Vote contre :				
Gouvernement	-0.22	0.23	-0.08	0.24
Employeurs	<b>0.38</b>	0.20	0.28	0.21
<b>Pressions Extérieurs</b>				
Aides au développement <sup>c</sup>	<b>-7.65</b>	2.05	<b>-8.56</b>	2.16
Prêts de la Banque Mondiale <sup>c</sup>	2.00	1.55	<b>3.15</b>	1.57
Crédits du FMI <sup>c</sup>	3.96	1.95	3.68	1.98
Exportations <sup>c</sup>	-0.79	1.30	0.27	1.06
Exportations vers les pays industrialisés <sup>c</sup>	-0.18	3.66	-2.54	3.80
Exportations vers les pays industrialisés <sup>c</sup> (hors OPEP)	-0.77	3.48	0.52	3.71
Pays de l'OPEP	0.25	0.68	-0.01	0.71

Note : Les entrées en gras sont significatives au seuil de 5%.

a. Prix internationaux de 1985, en 10 000\$. b. Centaines de millions. c. Pourcents du PIB.

ANNEXE C

# Annexes du Chapitre 3



## C.1. Identification du MPH à deux effets aléatoires en présence d'épisodes multiples

La preuve ci dessous est une extension de la première preuve de Honoré (1993) à une fonction de hasard comprenant deux effets aléatoires. Elle est plus générale que ce qui est requis dans ce Chapitre car les deux effets ne sont pas supposés emboîtés. Les fonctions de hasard pour les conventions 1 et 2 soumises au pays  $i$  s'écrivent :

$$\lambda_{i,1}(t_{i1}|\xi_i, \psi_1) = \xi_i \psi_1 \lambda_{0,1}(t_{i1}), \quad (\text{C.1})$$

$$\lambda_{i,2}(t_{i2}|\xi_i, \psi_2) = \xi_i \psi_2 \lambda_{0,2}(t_{i2}). \quad (\text{C.2})$$

Elles comprennent le cas particulier  $\lambda_{i,1}(t) = \lambda_0(t) \exp(\beta' x_i)$ . La fonction de survie jointe pour le pays  $i$  a pour expression :

$$\begin{aligned} S(t_{i1}, t_{i2}) &= \int_{\xi} \exp[-\xi \psi_1 \Lambda_{0,1}(t_{i1}) - \xi \psi_2 \Lambda_{0,2}(t_{i2})] dH_{\xi}(\xi) \quad (\text{C.3}) \\ &= \mathcal{L}_{\xi} [\psi_1 \Lambda_{0,1}(t_{i1}) + \psi_2 \Lambda_{0,2}(t_{i2})], \end{aligned}$$

où  $\mathcal{L}_V$  est la transformée de Laplace de la variable  $\xi$  et  $\Lambda_{0,j}(t_{ij}) = \int_0^{t_{ij}} \lambda_{0,j}(u) du$ . On observe  $S(t_{i1}, t_{i2})$  avec un grand nombre de durées issues d'une population homogène par rapport à  $\psi_1$  et  $\psi_2$ . La structure d'hétérogénéité est connue et on observe  $S(t_{i1}, t_{i2})$  pour  $\psi_1$  et  $\psi_2$  fixés, c'est-à-dire pour l'ensemble défini comme l'intersection des groupes caractérisés par  $\psi_1$  et  $\psi_2$ . En dérivant (C.3) par rapport à  $t_1$  et  $t_2$ , puis en faisant le rapport :

$$\begin{aligned} \frac{\partial S_i(t_{i1}, t_{i2})/\partial t_{i2}}{\partial S_i(t_{i1}, t_{i2})/\partial t_{i1}} &= \frac{\psi_2 \lambda_{0,2}(t_{i2}) \mathcal{L}'_{\xi} [\psi_1 \Lambda_{0,1}(t_{i1}) + \psi_2 \Lambda_{0,1}(t_{i1})]}{\psi_1 \lambda_{0,1}(t_{i1}) \mathcal{L}'_{\xi} [\psi_1 \Lambda_{0,1}(t_{i1}) + \psi_2 \Lambda_{0,2}(t_{i2})]} \quad (\text{C.4}) \\ &= \frac{\psi_2 \lambda_{0,2}(t_{i2})}{\psi_1 \lambda_{0,1}(t_{i1})}. \end{aligned}$$

On note  $k = 1/\lambda_0(t_1|X_i)$ , où  $t_0$  est une date de référence. En faisant le rapport entre les valeurs de (A.57) aux dates  $(t_{i0}, t_{i2})$  et  $(t, t_{i2})$ , on a :

$$\frac{\psi_2 \lambda_{0,2}(t_{i2})}{\psi_1 \lambda_{0,1}(t_{i0})} \bigg/ \frac{\psi_2 \lambda_{0,2}(t_{i2})}{\psi_1 \lambda_{0,1}(t)} = \frac{\lambda_{0,1}(t)}{\lambda_{0,1}(t_{i0})} = k \lambda_{0,1}(t). \quad (\text{C.5})$$

En intégrant l'équation précédente par rapport aux temps, on obtient  $k \Lambda_{0,1}(t) + c_1$ , où  $c_1$  est déterminé grâce à la condition initiale  $\Lambda_0(0) = 0$ . Ainsi,  $\Lambda_0(t|X_i)$  est identifiée à la constante de normalisation  $k$  près. De même, on montre que  $\Lambda_{0,2}(t|X_i)$  est identifiée en partant du rapport de (A.57) aux

## C.2 Construction de la Vraisemblance Partielle

---

dates  $(t_{i1}, t_{i0})$  et  $(t_{i1}, t)$ . La fonction de survie  $S(t_{i1}, t_{i2})$  dépend du temps de manière connue, et on retrouve  $\mathcal{L}_V$  en faisant varier  $(t_1, t_2)$  sur  $[0, \infty[^2$  et on a l'identification de  $H_V$ .

Changeons le points de vue et considérons la convention  $j$  soumise aux pays 1 et 2. Le modèle est :

$$\lambda_{1,j}(t_{1j}|\xi_1, \psi_j) = \xi_1 \psi_j \lambda_{0,1}(t_{1j}), \quad (\text{C.6})$$

$$\lambda_{2,j}(t_{1j}|\xi_2, \psi_j) = \xi_2 \psi_j \lambda_{0,2}(t_{2j}). \quad (\text{C.7})$$

Il s'agit du même cas de figure que pour le système (C.1), où  $\xi$  et  $\psi$  ont été permutés. On a donc l'identification de  $H_\psi$ .

## C.2. Construction de la Vraisemblance Partielle

Nous adaptons ici la justification de la vraisemblance partielle fournie par Kalbfleisch (1978) à un modèle de Cox à deux effets aléatoires. On note  $\delta_{ij}$  une indicatrice prenant la valeur 1 si une ratification est observée à la date  $t_{ij}$ . Supposons que  $\delta_{ij}$  suive une loi de Poisson de paramètres  $\lambda_{ij}(t_{ij}|x_{ij}, v_i, w_j)$ . La contribution à la vraisemblance de la ratification de la convention  $j$  par le pays  $i$  à la date  $t_{ij}$  est :

$$L^{Poisson}(t_{ij}, \delta_{ij}|v_i, w_j, x_{ij}, \lambda_0) = \lambda_{ij}(t_{ij}|x_{ij}, v_i, w_j)^{\delta_{ij}} \exp \left[ - \sum_{(k,l) \in R_{ij}} \lambda_{kl}(t_{ij}|x_{kl}, v_k, w_l) \right]. \quad (\text{C.8})$$

L'équation C.8 correspond à la vraisemblance d'un modèle de durée de hasard  $\lambda_{ij}(t_{ij}|x_{ij}, v_i, w_j)$ . Supposons que le hasard de base suit comme a priori une loi gamma de paramètres  $c\lambda_0^*(t)$  et  $c$ , où  $\lambda_0^*(t)$  est une croyance a priori sur

$\lambda_0(t)$ , et intégrons par rapport au hasard de base :

$$\begin{aligned}
 L(t_{ij}, \delta_{ij} | v_i, w_j, x_{ij}) &= \prod_{i=1}^I \prod_{j=1}^{J_i} \exp[\beta' x_{ij}(t_{ij}) + v_i + w_j] \int_0^\infty \lambda_0(t_{ij}) \\
 &\quad \exp \left[ - \sum_{(k,l) \in R_{ij}} \lambda_{kl}(t_{ij} | x_{kl}, v_k, w_l) \right] \lambda_0(t_{ij})^{c\lambda_0^*(t)-1} \\
 &\quad \exp[-c\lambda_0(t_{ij})] d\lambda_0(t_{ij}) \\
 &= \prod_{i=1}^I \prod_{j=1}^{J_i} \exp[\beta' x_{ij}(t_{ij}) + v_i + w_j] \int_0^\infty \lambda_0(t_{ij})^{c\lambda_0^*(t)} \\
 &\quad \exp \left[ - \lambda_0(t_{ij}) \left( c + \sum_{(k,l) \in R_{ij}} \exp[\beta' x_{ij}(t_{ij}) + v_i + w_j] \right) \right] \\
 &\quad d\lambda_0(t_{ij}) \\
 &\propto \prod_{i=1}^I \prod_{j=1}^{J_i} \frac{\exp[\beta' x_{ij}(t_{ij}) + v_i + w_j]}{\left( c + \sum_{(k,l) \in R_{ij}} \exp[\beta' x_{ij}(t_{ij}) + v_i + w_j] \right)^{c\lambda_0^*(t)}}.
 \end{aligned} \tag{C.9}$$

On obtient la vraisemblance partielle (3.4) avec  $(c\lambda_0^*(t), c) \rightarrow 0$ , c'est-à-dire en supposant que le hasard de base suit un a priori gamma non-informatif.

### C.3. Programme du modèle à deux effets aléatoires

Le programme ci dessous procède à l'estimation du modèle à deux effets aléatoires par l'intermédiaire du logiciel WinBUGS 1.3. Le code des sous-modèles s'en déduit directement.

```

model
{

# Set up data
for(i in 1 : N) {
  for(j in 1 : T) {
    # risk set = 1 if obs.t >= t
    Y[i, j] <- step(obst[i] - t[j] + eps)
    # counting process jump = 1 if obst in [ t[j], t[j+1] [

```

### C.3 Programme du modèle à deux effets aléatoires

---

```
#           i.e. if t[j] <= obst < t[j+1]
dN[i, j] <- Y[i, j ] *step(t[j+1] - obst[i] - eps)*fail[i]
}
}

for(i in 1:N) {
  Idt1[i] <- exp( bexports * exports[i] + bexpnonopep \\
* expnonopep[i] + baid2 * aid2[i] + bexpind * expind[i]\\
+ bcredits * credits[i] + bwrldbank * wrldbank[i] \\
+ balvarez * alvarez[i] + bprev_r * prev_r[i]\\
+ bpop * pop[i] + bleft_maj * left_maj[i] + bgov * gov[i]\\
+ bempl * empl[i] + brgdpc * rgdp_pc[i] + brgdpc_sq \\
* rgdp_sq[i] + bnopredec * nopredec[i] + bnon_oil \\
* non_oil[i] + brelrat * relrat[i] \\
+ bpays[ pays[i] ] + bconv[conv[i] ] )
}
#model
for(j in 1:T) {
  for(i in 1:N) {
    dN[i, j] ~ dpois(Idt[i, j])           # Likelihood
    Idt[i, j] <- Y[i, j] * Idt1[i] * dL0[j] #Intensity
  }
}

# prior distributions of the betas
bexports ~dnorm(0, 1.0E-6)
brgdpc ~dnorm(0, 1.0E-6)
brgdpc_sq~dnorm(0, 1.0E-6)
bnopredec~dnorm(0, 1.0E-6)
bprev_r~dnorm(0, 1.0E-6)
bpop~dnorm(0, 1.0E-6)
bleft_maj~dnorm(0, 1.0E-6)
bgov~dnorm(0, 1.0E-6)
bempl~dnorm(0, 1.0E-6)
bnon_oil~dnorm(0, 1.0E-6)
bexpnonopep~dnorm(0, 1.0E-6)
baid2~dnorm(0, 1.0E-6)
bexpind~dnorm(0, 1.0E-6)
bcredits~dnorm(0, 1.0E-6)
bwrldbank~dnorm(0, 1.0E-6)
balvarez~dnorm(0, 1.0E-6)
```

```
breirat~dnorm(0, 1.0E-6)

#prior distribution of the country effect
for(k in 1 : Ncountries) { bpays[k] ~dnorm(0, tau) }
tau ~ dgamma(0.001, 0.001)
sigmatau <- sqrt(1/tau)

#prior distribution of the convention effect
for(n in 1:Nconv) {bconv[n] ~dnorm(0, alphac)}
alphac ~ dgamma(0.001, 0.001)
sigmalphac <- sqrt(1/alphac)

#prior distribution on baseline hazard (T = number of \\
subspells)
for(j in 1:T) { dL0[j] ~ dgamma(c , c) }

}
```

ANNEXE D

# Annexes du Chapitre 4

## D.1. Calcul des espérances et variances conditionnelles

On note  $V$  le vecteur  $(v_1, \dots, v_I)'$ .  $V$  et les  $w_j$  ont pour loi jointe une distribution gaussienne. Ainsi,  $E(w_j|V)$  est l'espérance linéaire et  $V(w_j|V)$  la variance partielle (voir Gouriéroux et Monfort, 1990). Comme  $E(w_{ij}) = E(v_i) = 0$ , on a :

$$\begin{aligned} E(w_j|V) &= \text{cov}(w_j, V)\text{var}(V)^{-1}V \\ &= \frac{1}{\sigma_f^2}E(w_j V')V \\ &= \frac{\rho\sigma_w}{\sigma_f} \left( \sum_{l=1}^I \delta_{lj}v_l \right). \end{aligned} \quad (\text{D.1})$$

La variance partielle s'écrit :

$$\begin{aligned} \text{var}(w_j|V) &= \text{var}(w_j) - \text{cov}(w_j, V)\text{var}(V)^{-1}\text{cov}(V, w_j) \\ &= \sigma_w^2 - \frac{1}{\sigma_f^2}\text{cov}(w_j, V)\text{cov}(V, w_j) \\ &= \sigma_w^2 \left( 1 - \rho^2 \sum_{l=1}^I \delta_{lj} \right). \end{aligned} \quad (\text{D.2})$$

## D.2. Statistiques descriptives des durées

TAB. D.1 – Transitions observées

Durée de l'épisode	Échantillon complet	Sous-échantillon	Sous-échantillon contraint
4 et plus	7	5	6
3	9	9	9
2	20	20	22
1	64	66	63
Total	100	100	100

Note : les durées sont en années.

### D.3 Statistiques descriptives des variables explicatives

TAB. D.2 – Nombre d'épisodes par individu

Nombre d'épisodes	Échantillon complet		Sous-échantillon	Sous-échantillon contraint	
3 et plus	2		2		2
2	12		13		14
1	86		85		84
Total	100		100		100

Note : les durées sont en années.

## D.3. Statistiques descriptives des variables explicatives

TAB. D.3 – caractéristiques des entreprises

Variable	Échantillon complet		Sous-échantillon		Sous-échantillon contraint	
	Moy.	S.d	Moy.	S.d	Moy.	S.d
Plusieurs sites	0.39	0.49	0.37	0.48	0.29	0.45
Secteur :						
industrie	0.34	0.47	0.36	0.48	0.39	0.49
construction	0.16	0.37	0.17	0.37	0.16	0.37
commerce	0.31	0.46	0.33	0.46	0.32	0.47
finance	0.13	0.34	0.14	0.33	0.09	0.29
Région :						
Nord	0.33	0.47	0.34	0.47	0.36	0.48
Centre	0.14	0.35	0.13	0.34	0.15	0.36
Lisbonne et Vallée du Tage	0.42	0.49	0.44	0.50	0.40	0.49
Alentejo, Algarve et Isles	0.05	0.23	0.08	0.22	0.06	0.23
Nombre d'entreprises	77 603		7 307		7325	



TAB. D.4 – Caractéristiques des individus

Variable	Échantillon complet		Sous-échantillon		Sous-échantillon contraint	
	Moy.	S.d	Moy.	S.d	Moy.	S.d
Femme	0.39		0.38		0.37	
Age :						
16 - 25	0.34	0.48	0.34	0.47	0.32	0.46
26 - 35	0.38	0.48	0.38	0.49	0.38	0.49
36 - 55	0.28	0.45	0.28	0.45	0.30	0.46
Niveau d'études :						
primaire	0.32	0.47	0.32	0.47	0.34	0.47
secondaire	0.43	0.50	0.44	0.50	0.44	0.50
post-baccalauréat	0.25	0.43	0.24	0.43	0.22	0.42
Temps-partiel	0.11		0.10		0.08	
Salaires	696.05		690.29	441.08	670.60	432.63
Nombre d'employés	756	120	10	139	8468	

## D.4. Estimation Bayésienne sur l'échantillon non contraint

TAB. D.5 – Estimations des écarts-types des distributions mélangeantes sur l'échantillon non contraint

Type d'hétérogénéité	Paramètre	Moy.	2.5%	97.5%
<b>Effets corrélés</b>				
effet entreprise	$\sigma_f$	0.37	0.27	0.50
effet employé	$\sigma_w$	0.47	0.35	0.59
corrélation	$\rho$	0.54	0.44	0.58
<b>Effets indépendants</b>				
effet entreprise	$\sigma_f$	0.47	0.34	0.61
effet employé	$\sigma_w$	0.55	0.34	0.96
<b>Effet unique</b>				
effet employé	$\sigma_w$	0.42	0.32	0.54

### D.3 Statistiques descriptives des variables explicatives

TAB. D.6 – Estimations Bayésiennes sur l'échantillon non contraint

Variable	None	Employé	Effet(s) aléatoire(s)	
			Indépendants	Corrélés
<b>Années en poste</b>				
2	<b>-0.28</b>	<b>-0.27</b>	<b>-0.23</b>	<b>-0.21</b>
3	<b>-0.54</b>	<b>-0.51</b>	<b>-0.44</b>	<b>-0.42</b>
4	<b>-0.93</b>	<b>-0.90</b>	<b>-0.82</b>	<b>-0.79</b>
5 et plus	<b>-1.21</b>	<b>-1.17</b>	<b>-1.08</b>	<b>-1.04</b>
<b>Caractéristiques des individus</b>				
Femme	<b>-0.30</b>	<b>-0.30</b>	<b>-0.31</b>	<b>-0.32</b>
Age :				
16 - 25	<b>0.47</b>	<b>0.48</b>	<b>0.49</b>	<b>0.49</b>
26 - 35	<b>0.26</b>	<b>0.27</b>	<b>0.27</b>	<b>0.28</b>
Niveau d'études :				
primaires	0.01	0.01	-0.02	-0.03
secondaires	0.05	0.05	0.02	0.00
Temps partiel	0.14	0.15	0.16	0.15
<b>Caractéristiques des entreprises</b>				
Plusieurs sites	<b>0.19</b>	<b>0.20</b>	<b>0.21</b>	<b>0.16</b>
Région :				
Centre	0.12	0.12	0.13	0.13
Lisbonne et Vallée du Tage	<b>0.25</b>	<b>0.25</b>	<b>0.24</b>	<b>0.22</b>
Alentejo, Algarve et Isles	<b>0.24</b>	<b>0.24</b>	<b>0.23</b>	<b>0.25</b>
Secteur :				
Construction	<b>0.20</b>	<b>0.20</b>	0.19	0.16
Commerce	<b>0.19</b>	<b>0.19</b>	<b>0.18</b>	0.17
Finance	<b>0.58</b>	<b>0.59</b>	<b>0.53</b>	<b>0.45</b>
Constante	<b>-3.06</b>	<b>-3.16</b>	<b>-3.25</b>	<b>-3.29</b>
Log-vraisemblance	- 4598	-4493	-4318	-4278
DIC	9677	9461	9088	9046
Nombre d'employés	10 139	10 139	10 139	10 139
Nombre d'entreprises	7 307	7307	7307	7307

Note : les coefficients en caractères gras sont significatifs au seuil de 5%.

## D.5. Estimations fréquentistes

TAB. D.7 – Estimations fréquentistes sur l'échantillon complet

Variable	Effet(s) aléatoire(s)		
	Aucun	Employé	Employeur et employé
<b>Années en poste</b>			
2	<b>-0.31</b>	<b>-0.29</b>	<b>-0.26</b>
3	<b>-0.50</b>	<b>-0.46</b>	<b>-0.41</b>
4	<b>-0.81</b>	<b>-0.76</b>	<b>-0.71</b>
5 et plus	<b>-1.22</b>	<b>-1.15</b>	<b>-1.08</b>
<b>Caractéristiques de l'employé</b>			
Femme	<b>-0.28</b>	<b>-0.29</b>	<b>-0.29</b>
Age :			
16-25	<b>0.47</b>	<b>0.48</b>	<b>0.49</b>
26-35	<b>0.29</b>	<b>0.29</b>	<b>0.30</b>
Niveau d'études :			
primaires	<b>0.09</b>	<b>0.09</b>	<b>0.09</b>
secondaires	<b>0.08</b>	<b>0.08</b>	<b>0.09</b>
Temps partiel	<b>0.19</b>	<b>0.20</b>	<b>0.20</b>
<b>Caractéristiques de l'entreprise</b>			
Plusieurs sites	<b>0.14</b>	<b>0.15</b>	<b>0.15</b>
Région :			
Centre	<b>0.11</b>	<b>0.11</b>	<b>0.12</b>
Lisbonne et Vallée du Tage	<b>0.22</b>	<b>0.23</b>	<b>0.23</b>
Alentejo et Algarve	<b>0.27</b>	<b>0.28</b>	<b>0.28</b>
Isles	0.03	0.02	0.02
Secteur :			
Construction	<b>0.17</b>	<b>0.18</b>	<b>0.18</b>
Commerce	<b>0.21</b>	<b>0.21</b>	<b>0.22</b>
Transport	-0.01	-0.01	-0.01
Finance	<b>0.41</b>	<b>0.41</b>	<b>0.43</b>
Constante	<b>-2.93</b>	<b>-3.06</b>	<b>-3.19</b>
Log-vraisemblance	-388034	-387719	-387701
$\sigma_w$	No	0.48	0.51
$\sigma_f$	No	No	0.48

## D.6. Programme du modèle à deux effets aléatoires corrélés

Le programme ci dessous procède à l'estimation du modèle à deux effets aléatoires corrélés, le code des sous-modèles s'en déduisant directement. Nous utilisons le logiciel OpenBUGS 2.2.0, un successeur du projet WinBUGS à l'heure actuelle arrêté. Comme son prédécesseur, OpenBUGS est un logiciel libre, disponible à l'adresse <http://mathstat.helsinki.fi/openbugs/>.

```
model {

#Set up data

for(i in 1:N) {
for(j in 1:T) {
# risk set = 1 if obs_t>=t
Y[i, j] <- step(obs_t[i] - t[j] + eps)
#counting process jump = 1 if obs_t in [ t[j], t[j+1] [
# ie if t[j] <= obs_t < t[j+1]
dN[i, j] <- Y[i, j] * step(t[j+1] - obs_t[i] - eps) \\
* fail[i]

dN[i, j] ~ dbern(hazard[i, j]) #Likelihood
cloglog(hazard[i, j]) <- bcst + b0[1] * tD2[i] + b0[2] \\
* tD3[i] + b0[3] * tD4[i] + bsx * sexo[i] + bage1 * ageD1[i] \\
+ bage2 * ageD2[i] + bedc1 * educD1[i] + bedc2 * educD2[i] \\
+ bpart * part[i] + bw * hwstd[i] + bsct2 * sectorD2[i] \\
+ bsct4 * sectorD4[i] + bsct5 * sectorD5[i] + brg2 * regD2[i] \\
+ brg3 * regD3[i] + brg4 * regD4[i] + bmltpl * multipl[i] \\
+ bfsz * f_sizestd[i] + v[firm[i]] + w[worker[i]] \\
#hazard
}
}

#mixing distribution at the firm level
for(k in 1:Nfirm) {v[k] ~ dnorm(0, tauv)}
tauv ~ dgamma(9,1)
sigmav <- 1 / sqrt(tauv)

#mixing distribution at the worker level
```

```

for(l in 1:Nworker) {
mean[l] <- (sigmaw / sigmav) * rho * sum( toto[l, ] )
prec[l] <- tauw / (1-rho * rho * nb_firms[l])
for(m in 1:N){toto[l, m] <- equals( worker[m], l ) \\
* v[firm[m]] }
w[l] ~ dnorm(mean[l], prec[l])
}
tauw ~ dgamma(4,1)
sigmaw <- 1 / sqrt(tauw)

#Prior distributions of the coefficients
bcst ~dnorm(0, 1.0E-3)
bsx ~dnorm(0, 1.0E-3)
bage1 ~dnorm(0, 1.0E-3)
bage2 ~dnorm(0, 1.0E-3)
bedc1 ~dnorm(0, 1.0E-3)
bedc2 ~dnorm(0, 1.0E-3)
bpart ~dnorm(0, 1.0E-3)
bw ~dnorm(0, 1.0E-3)
bsct2 ~dnorm(0, 1.0E-3)
bsct4 ~dnorm(0, 1.0E-3)
bsct5 ~dnorm(0, 1.0E-3)
brg2 ~dnorm(0, 1.0E-3)
brg3 ~dnorm(0, 1.0E-3)
brg4 ~dnorm(0, 1.0E-3)
bmltpl ~dnorm(0, 1.0E-3)
bfsz ~dnorm(0, 1.0E-3)

#Prior distributions of the baseline hazard
for(n in 1:Npieces) {b0[n] ~ dnorm(0, 1.0E-3)}

#Prior distribution of the correlation
rho ~dunif(-1,1)

}

```

# Bibliographie

- AALEN, O. (1978) : “Non-parametric inference for a family of counting processes,” *Annals of Statistics*, 6, 701–726.
- (1988) : “Heterogeneity in survival analysis,” *Statistics in Medicine*, 7, 1121–1137.
- ABBRING, J., ET G. VAN DEN BERG (2001) : “The unobserved heterogeneity distribution in duration analysis,” Document de Travail, Free University Amsterdam, Amsterdam.
- (2003a) : “The identifiability of the mixed proportional hazards competing risks model,” *Journal of the Royal Statistical Society, Series B*, 65, 701–711.
- (2003b) : “The nonparametric identification of treatment effects in duration models,” *Econometrica*, 71(5), 1491–1518.
- ABOWD, J., F. KRAMARZ, ET S. ROUX (2006) : “Wages, mobility and firm performance : Advantages and insights from using matched worker-firm data,” *The Economic Journal*, 116(512), F245–F285.
- ALVAREZ, M., J. CHEIBUB, F. LIMONGI, ET A. PRZEWORSKI (1996) : “Classifying political regimes,” *Studies in Comparative International Development*, 31(2), 2–36.
- ANDERSEN, P., ET R. GILL (1982) : “Cox’s regression model for counting processes : A large sample study,” *Annals of Statistics*, 10, 1100–1120.
- ANDERSEN, P., J. KLEIN, ET M.-J. ZHANG (1999) : “Testing for centre effects in multicenter survival studies : a Monte Carlo comparison of fixed and random effects tests,” *Statistics in Medicine*, 18, 1489–1500.
- ANDERSON, J., ET A. SENTHILSELVAN (1980) : “Smooth estimates for the hazard function,” *Journal of the Royal Statistical Society, Series B*, 42, 322–327.
- BAKER, M., ET A. MELINO (1999) : “Duration dependence and nonparametric heterogeneity : a Monte Carlo study,” Document de Travail M5S 3G7, University of Toronto, Toronto.
- BEFFY, M., M. BUCHINSKY, D. FOUGÈRE, T. KAMIONKA, ET F. KRAMARZ (2006) : “The returns to seniority in France (and why are they lower

- than in the United States ?),” Discussion Paper Series 5486, CEPR, London.
- BELLMANN, L., S. BENDER, ET U. HORNSTEINER (2000) : “Job tenure of two cohorts of young German men 1979 - 1990 : An analysis of the (West-) German Employment Statistic Register Sample concerning multivariate failure times and unobserved heterogeneity,” Discussion Paper Series 106, IZA.
- BIGGERI, L., M. BINI, ET L. GRILLI (2001) : “The transition from university to work : a multilevel approach to the analysis of the time to obtain the first job,” *Journal of the Royal Statistical Society. Series A*, 164, 293–305.
- BIJWAARD, G., P. FRANSES, ET R. PAAP (2003) : “Modeling purchases as repeated events,” .
- BIJWAARD, G., ET G. RIDDER (2005) : “Correcting for selective compliance in a re-employment bonus experiment,” *Journal of Econometrics*, 125, 77–111.
- BLUMEN, I., M. KOGEN, ET P. MCCARTHY (1955) : *The industrial mobility of workers as a probability process*, vol. 6 of *Cornell studies in industrial and labor relations*. Cornell University, Ithaca, NY.
- BOLSTAD, W., ET S. MANDA (2001) : “Investigating child mortality in Malawi using family and community random effects : a Bayesian analysis,” *Journal of the American Statistical Association*, 96, 12–19.
- BONNAL, L., D. FOUGÈRE, ET A. SÉRANDON (1997) : “Evaluating the impact of French employment policies on individual labour market histories,” *Review of Economic Studies*, 64, 683–713.
- BOOCKMANN, B. (2001) : “The ratification of ILO conventions : a hazard rate analysis,” *Economics and Politics*, 13, 281–309.
- BRESLOW, N. (1972) : “Discussion of Professor Cox’s paper,” *Journal of the Royal Statistical Society, Series B*, 34, 216–217.
- (1974) : “Covariance analysis of censored survival data,” *Biometrics*, 30, 89–99.
- BROWNING, M., ET J. CARRO (2005) : “Heterogeneity and microeconomics modelling,” dans *Advances in Economics and Econometrics, Theory and Applications : Ninth World Congress of the Econometric Society*, vol. 3.
- BUCHINSKY, M., D. FOUGÈRE, F. KRAMARZ, ET R. TCHERNIS (2005) : “Interfirm Mobility, Wages, and the Returns to Seniority and Experience in the U.S.,” Discussion Paper Series 1521, IZA.

## BIBLIOGRAPHIE

---

- BURNHAM, K., ET D. ANDERSON (1998) : *Model Selection and Inference*. Springer, New York.
- CAMPOLIETI, M. (2001) : “Bayesian semiparametric estimation of discrete duration models : an application of the Dirichlet process prior,” *Journal of Applied Econometrics*, 16, 1–22.
- CHAMBERLAIN, G. (1986) : “Asymptotic efficiency in estimation with conditional moment restriction,” *Journal of Econometrics*, 34, 305–334.
- CLAYTON, D. (1978) : “A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence,” *Biometrika*, 65, 141–151.
- (1991) : “A Monte Carlo method for Bayesian inference in frailty models,” *Biometrics*, 47, 467–485.
- CLAYTON, D., ET J. CUZICK (1985) : “Multivariate generalizations of the proportional hazards model,” *Journal of the Royal Statistical Society, Series A*, 148, 82–108.
- CONAWAY, M. (1990) : “A random effect model for binary data,” *Biometrics*, 46, 317–328.
- CORTINAS ABRAHANTES, J., ET T. BURZYKOWSKI (2004) : “A version of the EM algorithm for proportional hazard model with random effects,” Document de Travail, IAP Statistics Network.
- COX, D. (1959) : “The analysis of exponentially distributed lifetimes with two types of failure,” *Journal of the Royal Statistical Society, Series B*, 21, 411–421.
- (1972) : “Regression models and life tables,” *Journal of the Royal Statistical Society, Series B*, 34, 187–220.
- COX, D. (1975) : “Partial likelihood,” *Biometrika*, 62(2), 269–276.
- CUNHA, F., J. HECKMAN, ET N. SALVADOR (2005) : “Separating uncertainty from heterogeneity in life cycle earnings,” *Oxford Economic Papers*, 57, 191–261.
- DE MONTRICHER, G., R. TAPIA, ET J. THOMPSON (1975) : “Nonparametric maximum likelihood estimation of probability densities by penalty function methods,” *The Annals of Statistics*, 3, 1329–1348.
- DEL BOCA, D., ET R. SAUER (2006) : “Life cycle employment and fertility across institutional environments,” Mimeo.
- DEMPSTER, A., N. LAIRD, ET D. RUBIN (1977) : “Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion),” *Journal of the Royal Statistical Society, Series B*, 39, 1–38.



- DJURDJEVIC, D. (2000) : “Estimation de modèles de durée multivariés : application à la ratification des conventions par les pays membres du BIT,” Mémoire de DEA, Université Louis Pasteur, Strasbourg.
- DOLTON, P., ET W. VAN DER KLAUW (1995) : “Leaving teaching in the U.K. : a duration analysis,” *Economic Journal*, 105, 431–444.
- DORMONT, B., D. FOUGÈRE, ET A. PRIETO (2001) : “L’effet de l’allocation unique dégressive sur la reprise d’emploi,” *Economie et Statistiques*, 343, 3–28.
- DOSTIE, B. (2005) : “Job turnover and the returns to seniority,” *Journal of Business and Economic Statistics*, 23(2), 192–199.
- ECKSTEIN, Z., ET G. VAN DEN BERG (2003) : “Empirical Labor Search : A Survey,” Discussion Paper 929, IZA.
- EFRON, B. (1977) : “The Efficiency of Cox’s Likelihood Function for Censored Data,” *Journal of the American Statistical Association*, 72, 557–565.
- ELBERS, C., ET G. RIDDER (1982) : “True and spurious duration dependence : The identifiability of the proportional hazard model,” *Review of Economic Studies*, 49, 403–410.
- FARBER, H. (1999) : “Mobility and stability : The dynamics of job change in labor markets,” dans *Handbook of Labor Economics*, ed. O. Ashenfelter, et D. Card, vol. 3, chap. 37, pp. 2439–2483. Elsevier, Amsterdam.
- FELLER, W. (1971) : *An introduction to probability theory and its applications*, vol. 2. Wiley, New-York, 2<sup>e</sup> edn.
- FERGUSON, T. (1973) : “A Bayesian analysis of some nonparametric problems,” *The Annals of Statistics*, 1, 209–230.
- FERREIRA, A., ET N. GARCIA (2002) : “Simulation study for misspecifications on a frailty model,” *Brazilian Journal of Probabilities and Statistics*, 15(2), 121–134.
- FIRTH, D., C. PAYNE, ET J. PAYNE (1999) : “Efficacy of programmes for the unemployed : discrete time modelling of duration data from a matched-comparison study,” *Journal of the Royal Statistical Society, Series A*, 162, 11–120.
- FLEMING, T., ET D. HARRINGTON (1991) : *Counting processes and survival analysis*. Wiley, New-York.
- FLORENS, J., ET D. FOUGÈRE (1996) : *The Econometrics of Panel Data*, vol. 33, chap. Point Processes, pp. 537–572. Kluwer Academic Publishers, second revised edition edn.

## BIBLIOGRAPHIE

---

- FLORENS, J., D. FOUGÈRE, ET M. MOUCHART (1996) : *The Econometrics of Panel Data*, chap. Duration Models, pp. 491–536. Kluwer Academic Press.
- FLORENS, J.-P., M. MOUCHART, ET J.-M. ROLIN (1999) : “Semi- and non-parametric Bayesian analysis of duration models with Dirichlet priors : a survey,” *International Statistical Review*, 67, 187–210.
- FORTIN, B., D. FOUGÈRE, ET G. LACROIX (2001) : *Institutional and Financial Incentives for Social Insurance*, chap. The Effects of Welfare Benefits on the Duration of Welfare Spells : Evidence from a Natural Experiment in Canada, pp. 1–23. Kluwer Academic.
- FOUGÈRE, D., ET T. KAMIONKA (2003) : “Bayesian inference for the mover-stayer model in continuous time with an application labour market transition data,” *Journal of Applied Econometrics*, 18, 697–723.
- (2005) : “Econometrics of Individual Labor Market Transitions,” Document de Travail 1850, IZA.
- FRIEDL, H., ET G. KAUERMANN (2000) : “Standard errors for EM estimates in generalized linear models with random effects,” *Biometrics*, 56(3), 761–767.
- FRIJTERS, P. (2002) : “The non-parametric identification of lagged duration dependence,” *Economic letters*, 75, 289–292.
- GAUDERMAN, W., ET D. THOMAS (1994) : “Censored survival models for genetic epidemiology : a Gibbs sampling approach,” *Genetic Epidemiology*, 11, 171–188.
- GELFAND, A., ET A. SMITH (1990) : “Sampling based approaches to calculating marginal densities,” *Journal of the American Statistical Association*, 85, 398–409.
- GELMAN, A., ET D. RUBIN (1992) : “Inference from iterative simulation using multiple sequences,” *Statistical Science*, 7, 457–472.
- GEMAN, S., ET D. GEMAN (1984) : “Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.
- GILKS, W., ET P. WILD (1992) : “Adaptative rejection sampling for Gibbs sampling,” *Journal of the Royal Statistical Society, Series C*, 41, 337–348.
- GILL, R. (1985) : “Discussion of the paper by D. Clayton and J. Cuzick,” *Journal of the Royal Statistical Society, Series A*, 148, 108–109.
- GÖNÜL, F., ET K. SRINIVASAN (1993) : “Consumer purchase behavior in a frequently bought product category : estimation issues and managerial

- insights from a hazard function model with heterogeneity,” *Journal of the American Statistical Association*, 88, 1219–1227.
- GOOD, I., ET R. GASKINS (1971) : “Nonparametric roughness penalties for probability densities,” *Biometrika*, 58, 255–277.
- GOODMAN, L. (1961) : “Statistical methods for the mover-stayer model,” *Journal of the American Statistical Association*, 56, 841–868.
- GOURIÉROUX, C. (1989) : *Econométrie des variables qualitatives*. Economica, 2d edn.
- (1990) : “Hétérogénéité II. Étude des biais de représentativité (sous l’hypothèse d’exogénéité faible),” *Annales d’Économie et de Statistique*, 17, 185–204.
- GOURIÉROUX, C., ET J. JASIAK (2001) : “Duration models,” dans *Companion in Theoretical Econometrics*, ed. B. Baltagi, chap. 21, pp. 444–465. Blackwell.
- GOURIÉROUX, C., ET A. MONFORT (1990) : *Statistiques et modèles économétriques*, vol. 1-2. Economica, Paris, 2 edn.
- GOURIÉROUX, C., ET I. PEAUCELLE (1990) : “Hétérogénéité I. Étude des biais d’estimation dans le cas linéaire,” *Annales d’Économie et de Statistique*, 17, 163–184.
- GREEN, P. (1999) : “Penalized likelihood,” dans *Encyclopaedia of Statistical Sciences*, ed. S. Kotz, C. Read, et D. Banks, vol. 3 update, pp. 578–586. Wiley.
- GREENWOOD, M., ET G. YULE (1920) : “An Inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents,” *Journal of the Royal Statistical Society*, 83, 255–279.
- GRILLI, L. (2005) : “The random-effects proportional hazards model with grouped survival data : a comparison between the grouped continuous and continuation ration versions,” *Journal of the Royal Statistical Society, Series A*, 168, 83–94.
- GUO, G., ET G. RODRIGUEZ (1992) : “Estimating a multivariate proportional hazards model for clustered data using the EM algorithm, with an application to child survival in Guatemala,” *Journal of the American Statistical Association*, 87, 969–976.
- GUSTAFSON, P. (1997) : “Large hierarchical Bayesian analysis of multivariate survival data,” *Biometrics*, 53, 230–242.
- HAHN, J. (1994) : “The efficiency bound of the Mixed Proportional Hazard model,” *Review of Economic Studies*, 61, 607–629.

## BIBLIOGRAPHIE

---

- HAUSMAN, J., ET T. WOUTERSEN (2004) : “A semi-parametric duration model with heterogeneity that does not need to be estimated,” Presented at the Econometric Society World Congress 2005.
- HECKMAN, J. (1991) : “Identifying the hand of past : distinguishing state dependence from heterogeneity,” *American Economic Review*, 81 (supplement), 71–79.
- (2001) : “Micro Data, Heterogeneity, and the Evaluation of Public Policy : Nobel Lecture,” *Journal of Political Economy*, 109, 673–748.
- HECKMAN, J., ET B. HONORÉ (1989) : “The identifiability of the competing risks model,” *Biometrika*, 76, 325–330.
- HECKMAN, J., ET B. SINGER (1984a) : “Econometric duration analysis,” *Journal of Econometrics*, 24, 62–132.
- (1984b) : “The identifiability of the proportional hazards model,” *Review of Economic Studies*, 60, 231–243.
- (1984c) : “A method for minimizing the impact of distributional assumptions in econometric models for duration data,” *Econometrica*, 52, 271–320.
- HECKMAN, J., ET C. TABER (1994) : “Econometric mixture models and more general models for unobservables in duration analysis,” *Statistical Methods in Medical Research*, 3, 279–302.
- HEITMÜLLER, A. (2005) : “A note on decompositions in fixed effects models in the presence of time-invariant characteristics,” IZA Discussion Paper Series 1886, IZA.
- HENDERSON, C. (1975) : “Best Linear Unbiased Estimation and Prediction under a Selection Model,” *Biometrics*, 31 (2), 423–447.
- HERTZ-PICCIOTTO, I., ET B. ROCKHILL (1997) : “Validity and Efficiency of Approximation Methods for Tied Survival Times in Cox Regression,” *Biometrics*, 53, 1151–1156.
- HONORÉ, B. (1990) : “Identification of duration models with unobserved heterogeneity,” Unpublished manuscript, Northwestern University.
- (1993) : “Identification results for duration models with multiple spells,” *The Review of Economic Studies*, 60, 241–246.
- HORNY, G. (2001) : “Estimation de modèles de durée multivariés au moyen de l’algorithme EM,” Mémoire de DEA, Université Louis Pasteur, Strasbourg.
- HOROWITZ, J. (1996) : “Semiparametric estimation of a regression model with an unknown transformation of the dependent variable,” *Econometrica*, 64, 103–137.

- (1999) : “Semiparametric estimation of a proportional hazard model with unobserved heterogeneity,” *Econometrica*, 67, 1001–1028.
- HOROWITZ, J., ET S. LEE (2004) : “Semiparametric estimation of a panel data proportional hazards model with fixed effects,” *Journal of Econometrics*, 119, 155–198.
- HOSMER, D., ET S. LEMESHOW (1999) : *Applied Survival Analysis*. John Wiley, New York.
- HOUGAARD, P. (1984) : “Life table methods for heterogeneous populations : distributions describing the heterogeneity,” *Biometrika*, 71, 75–83.
- (2000) : *Analysis of multivariate survival data*. Springer-Verlag, New York.
- ISHWARAN, H. (1996) : “Identifiability and rates of estimation for scale parameters in location mixture models,” *The Annals of Statistics*, 24, 1560–1571.
- JOHANSEN, S. (1983) : “An extension of Cox’s regression model,” *International Statistical Review*, 51, 258–262.
- JOHNSON, W. (1978) : “A theory of job shopping,” *Quarterly Journal of Economics*, 92(2), 261–278.
- KALBFLEISCH, J. (1978) : “Non-parametric Bayesian analysis of survival time data,” *Journal of the Royal Statistical Society, Series B*, 40, 214–221.
- KALBFLEISCH, J., ET R. PRENTICE (1980) : *The Statistical analysis of failure time data*. John Wiley, New York.
- KAPLAN, E., ET P. MEIER (1958) : “Nonparametric estimation from incomplete observations,” *Journal of the American Statistical Association*, 53, 457–481.
- KIEFER, N. (1988) : “Economic duration data and hazard function,” *Journal of Economic Literature*, 26, 646–679.
- KLEIN, J. (1992) : “Semiparametric estimation of random effects using the Cox model based on the EM algorithm,” *Biometrics*, 48, 795–806.
- KORSGAARD, I., P. MADSEN, ET J. JENSEN (1998) : “Bayesian inference in the semiparametric log normal frailty model using Gibbs sampling,” *Genetics Selection Evolution*, 30, 241–256.
- LANCASTER, T. (1979) : “Econometric methods for the duration of unemployment,” *Econometrica*, 47, 939–956.
- (1990) : *The econometric analysis of transition data*. Cambridge University Press, Cambridge.

## BIBLIOGRAPHIE

---

- LAU, J. (2006) : “Bayesian semi-parametric modeling for mixed proportional hazard models with right censoring,” *Statistics and Probability Letters*, 76, 719–728.
- LEE, Y., ET J. NELDER (2003) : “Extended-REML estimators,” *Journal of Applied Statistics*, 30 (8), 845–856.
- LIGHT, A., ET M. URETA (1992) : “Panel estimates of male and female job turnover behavior : can female nonquitters be identified?,” *Journal of Labor Economics*, 10, 156–181.
- LIMA, F. (2004) : “How much mobility? Careers, promotions, and wages,” Document de Travail 3/2004, CEG-IST, Lisbon.
- LINDEBOOM, M., ET M. KERKHOFS (2000) : “Multistate models for clustered duration-An application to workplace effects on individual sickness absenteeism,” *The Review of Economics and Statistics*, 82, 668–684.
- LINDEBOOM, M., ET G. VAN DEN BERG (1994) : “Heterogeneity in models for bivariate survival : The importance of the mixing distribution,” *Journal of the Royal Statistical Society. Series B*, 56, 49–60.
- LINDSAY, B. (1983a) : “The geometry of mixture likelihoods : A general theory,” *The Annals of Statistics*, 11, 86–94.
- (1983b) : “The geometry of mixture likelihoods, Part II : The exponential family,” *The Annals of Statistics*, 11, 783–792.
- LOUIS, T. (1982) : “Finding the observed Information matrix when using the EM algorithm,” *Journal of the Royal Statistical Society, Series B*, 44, 226–233.
- MACCULLAGH, P., ET J. NELDER (1996) : *Generalized Linear Models*. Chapman Hall, London, UK, 2nd edn.
- MANDA, S., ET R. MEYER (2005) : “Age at first marriage in Malawi : a Bayesian multilevel analysis using a discrete time model,” *Journal of the Royal Statistical Society, Series A*, 168, 439–455.
- MAPLES, J., S. MURPHY, ET W. AXINN (2002) : “Two level proportional hazards model,” *Biometrics*, 58(4), 754–763.
- MARIN, J., K. MENGERSEN, ET C. ROBERT (2005) : *Handbook of Statistics* vol. 25, chap. Bayesian modelling and inference on mixtures of distributions, pp. 1–56. Dey, D. Dey and Rao, C.
- MCGILCHRIST, C. (1993) : “REML estimation for survival model with frailty,” *Biometrics*, 49, 221–225.
- MCGILCHRIST, C., ET C. AISBETT (1991) : “Regression with frailty in survival analysis,” *Biometrics*, 47(2), 461–466.

- MEILIJSON, I. (1989) : “A fast improvement to the EM algorithm on its own terms,” *Journal of the Royal Statistical Society, Series B*, 51, 127–138.
- MENDES, R., G. VAN DEN BERG, ET M. LINDEBOOM (2005) : “An empirical assessment of assortative matching in the labor market,” dans *Econometric Society World Congress*.
- MEYER, B. (1990) : “Unemployment insurance and unemployment spells,” *Econometrica*, 58, 757–782.
- MILCENT, C. (2003) : “Ownership, system of reimbursement and mortality rate relationships,” Mimeo.
- MITTELHAMMER, R., G. JUDGE, ET D. MILLER (2000) : *Econometric foundations*. Cambridge University Press, Cambridge.
- MOULTON, B. (1986) : “Random group effects and the precision of regression estimates,” *Journal of Econometrics*, 32, 385–397.
- NEAL, R. (1993) : “Probabilistic inference using Markov Chain Monte Carlo methods,” Document de Travail CRG-TR-93-1, Department of Computer Science, University of Toronto.
- (1994) : “Bayesian learning for neural networks,” Thèse de Doctorat, University of Toronto.
- (1997) : “Markov Chain Monte Carlo methods based on ‘slicing’ the density function,” Document de Travail 9722, Department of Statistics, University of Toronto.
- NELSON, W. (1969) : “Hazard plotting for incomplete failure time data,” *Journal of Quality Technology*, 1, 27–52.
- NG, S., T. KRISHNAN, ET G. MCLACHLAN (2004) : *Handbook of computational statistics*, chap. The EM algorithm. Springer.
- NIELSEN, G., R. GILL, P. ANDERSEN, ET T. SORENSEN (1992) : “A counting process approach to maximum likelihood estimation in frailty models,” *Scandinavian Journal of Statistics*, 19, 25–44.
- ONDRICH, J., ET K. PRASAD (1997) : “A duration model with unobserved heterogeneity as a mixture of Dirichlet processes,” *Economic Letters*, 55, 19–25.
- PARNER, E. (1997) : “Inference in semiparametric frailty models,” Thèse de Doctorat, University of Aarhus.
- RIDDER, G. (1990) : “The non-parametric identification of generalized hazard models,” *Review of Economic Studies*, 57, 167–182.
- RIDDER, G., ET I. TUNALI (1999) : “Stratified partial likelihood estimation,” *Journal of Econometrics*, 92, 193–232.

## BIBLIOGRAPHIE

---

- RIDDER, G., ET T. WOUTERSEN (2003) : “The singularity of the information matrix of the mixed proportional hazard model,” *Econometrica*, 71(5), 1579–1589.
- RIPATTI, S. LARSEN, K., ET J. PALMGREN (2002) : “Maximum likelihood inference for multivariate frailty models using an automated Monte Carlo EM algorithm,” *Lifetime Data Analysis*, 8, 349–360.
- RIPATTI, S., ET J. PALMGREN (2000) : “Estimation of multivariate frailty models using penalized partial likelihood,” *Biometrics*, 56, 1016–1022.
- ROBERT, C. (1996) : *Méthodes de Monte Carlo par chaînes de Markov*. Economica, Paris.
- ROBERT, C., ET G. CASELLA (1999) : *Monte Carlo statistical methods*. Springer.
- RODRIGUEZ, G. (2001) : “Survival models,” Lecture Notes, Chapter 7.
- RODRIGUEZ, G., ET N. GOLDMAN (2001) : “Improved estimation procedures for multilevel models with binary responses : a case study,” *Journal of the Royal Statistical Society, Series A*, 164, 339–355.
- RONDEAU, V., D. COMMENGES, ET P. JOLY (2003) : “Maximum penalized likelihood estimation in a gamma-frailty model,” *Lifetime Data Analysis*, 9, 139–153.
- SAINT AUGUSTIN, . (vers 400) : “Les Confessions,” livre XI, chap. 14 et 20, traduction L. Moreau (1879).
- SARGENT, D. (1998) : “A general framework for random effects survival analysis in the Cox proportional hazards setting,” *Biometrics*, 54, 1486–1497.
- SASTRY, N. (1995) : “Community characteristics, individual attributes and child survival in Brazil : a multilevel hazard model analysis,” Thèse de Doctorat, Princeton University.
- (1997) : “A nested frailty model for survival data, with an application to the study of child survival in northeast Brazil,” *Journal of the American Statistical Association*, 92, 426–435.
- SCHMIDT, P. (1989) : “Predicting recidivism using ‘split population’ survival time models,” *Journal of Econometrics*, 40(1), 141–159.
- SCHWARZ, G. (1978) : “Estimating the dimension of a model,” *The Annals of Statistics*, 6, 461–464.
- SO, Y., ET W. KUHFIELD (1995) : “Multinomial logit models,” SAS Technical Note, TS-694G.



- SPIEGELHALTER, D., N. BEST, B. CARLIN, ET A. VAN DER LINDE (2002) : “Bayesian Measures of model complexity and fit,” *Journal of the Royal Statistical Society, Series B*, 64, 583–639.
- SPIEGELHALTER, D., A. THOMAS, ET N. BEST (2000) : *Winbugs user manual and examples : version 1.3*, Medical research Council of Biostatistics unit, Cambridge.
- SPIELERMAN, S. (1972) : “Extensions of the mover-stayer model,” *American Journal of Sociology*, 78, 599–626.
- TANNER, M., ET W. WONG (1987) : “The calculation of posterior distributions by data augmentation,” *Journal of the American Statistical Association*, 82, 528–540.
- TEAM, R. D. C. (2005) : *R : A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
- THERNEAU, M., ET P. GRAMBSCH (2000) : *Modeling survival data : extending the Cox model*, Statistics for Biology and Health. Springer, New-York.
- THERNEAU, T., P. GRAMBSCH, ET V. PANKRATZ (2003) : “Penalized survival models and frailty,” *Journal of Computational and Graphical Statistics*, 12, 156–175.
- TSIATIS, A. (1975) : “A nonidentifiability aspect of the problem of competing risks,” *Proceedings of the National Academy of Sciences*, 72, 20–22.
- VAIDA, F., ET R. XU (2000) : “Proportional hazards model with random effects,” *Statistics in Medicine*, 19, 3309–3324.
- VAN DEN BERG, G. (2001) : “Duration models : specification, identification and multiple durations,” dans *Handbook of Econometrics*, vol. 5, chap. 55, pp. 3381–3463. J. J. Heckman and E. Leamer (eds.), Elsevier, Amsterdam.
- VAN DEN BERG, G., A. HOLM, ET J. VAN OURS (2002) : “Do stepping-stone job exist ? Early career paths in the medical profession,” *Journal of Population Economics*, 15, 647–666.
- VAUPEL, J., K. MANTON, ET E. STALLARD (1979) : “The impact of heterogeneity in individual frailty on the dynamics of mortality,” *Demography*, 16(3), 439–454.
- VU, H., ET M. KNUIMAN (2002) : “A hybrid ML-EM algorithm for calculation of maximum likelihood estimates in semiparametric shared frailty model,” *Computational Statistics and Data Analysis*, 40, 173–187.
- WASSERMAN, L. (2000) : “Bayesian model selection and model averaging,” *Journal of Mathematical Psychology*, 44(1), 92–107.

## BIBLIOGRAPHIE

---

- WEI, G., ET M. TANNER (1990) : “Monte Carlo implementation of the EM algorithm and the poor man’s data augmentation algorithm,” *Journal of the American Statistical Association*, 85, 699–704.
- XUE, X. (1998) : “Multivariate survival data under bivariate frailty : An estimating equation approach,” *Biometrics*, 54(4), 1631–1637.
- XUE, X., ET R. BROOKMEYER (1996) : “Bivariate frailty model for the analysis of multivariate survival time,” *Lifetime Data Analysis*, 2, 277–289.
- XUE, X., ET Y. DING (1999) : “Assessing heterogeneity and correlation of paired failure times with the bivariate frailty model,” *Statistics in Medicine*, 18, 907–918.
- YAMAGUCHI, K. (1986) : “Alternative approaches to unobserved heterogeneity in the analysis of repeatable events,” dans *Sociological Methodology 1986*, ed. N. Tuma. Jossey-Bass, Washington D.C.
- YASHIN, A., I. IACHINE, A. BEGUN, ET J. VAUPEL (2001) : “Hidden frailty : myths and reality,” Document de Travail 34, University of Southern Denmark, Department of Statistics.
- YASHIN, A., J. VAUPEL, ET I. IACHINE (1995) : “Correlated individual frailty : an advantageous approach to survival analysis of bivariate data,” *Mathematical Population Studies*, 5, 145–59.
- YAU, K. (2001) : “Multilevel models for survival analysis with random effects,” *Biometrics*, 57, 96–102.
- YAU, K., ET A. MCGILCHRIST (1998) : “ML and REML estimation in survival analysis with time dependent correlated frailty,” *Statistics in Medicine*, 17, 1201–1213.
- ZHOU, M. (2002) : Handout 3 for the book *Survival Analysis Using the SAS System* from Allison.
- ZIJL, M., G. VAN DEN BERG, ET A. HEYMA (2004) : “Stepping-stones for the unemployed : the effect of temporary jobs on the duration until regular works,” Presented at the Econometric Society World Congress 2005.

## Résumé

LE temps passé dans un état est au coeur de nombreuses questions économiques, concernant notamment le marché du travail. Les comportements individuels sont divers, et peuvent correspondre à des interactions complexes que l'économètre n'observe qu'en partie. Cette thèse porte ainsi sur la prise en compte d'une hétérogénéité non observée multiple au sein des modèles de durée multivariés.

Les outils existants pour estimer les modèles de durée comprenant plusieurs termes d'hétérogénéité non observée sont souvent spécifiques et dérivés d'approches paramétriques. Cette thèse apporte une contribution **méthodologique** en fournissant des approches semi-paramétriques plus souples. Nous proposons ainsi une méthode d'estimation générale, plus rapide que l'algorithme de référence et exempte de ses difficultés numériques. Nous nous attachons également à mettre en perspective les approches fréquentistes et Bayésiennes, qui nous permettent de modéliser des structures d'interactions complexes.

Cette thèse apporte également une contribution **empirique** à l'étude du marché du travail. Nous nous intéressons à sa régulation par l'Organisation Internationale du Travail (OIT), et nos résultats indiquent que les conventions ne demandant qu'un accord de principe sont ratifiées facilement, tandis que les textes plus précis sont difficilement signés. L'adéquation des systèmes juridiques et économiques en place avec le texte de la convention est un facteur déterminant, bien plus que l'orientation politique des dirigeants ou d'éventuelles sanctions commerciales. Nous étudions également la mobilité professionnelle avec des données d'appariements employeurs-employés. La plupart des travaux ne prennent pas en compte les politiques de rétention du personnel car elles sont difficilement observables, et nous montrons qu'elles sont très diverses d'une firme à l'autre. De plus, les résultats laissent penser que l'adéquation entre les caractéristiques des entreprises et des employés, non observées par les statisticiens, influencent la mobilité.

**Mots-Clés :** effets aléatoires, approches semi-paramétriques, Organisation Internationale du Travail, mobilité professionnelle, modèle MPH, données d'appariements, vraisemblance partielle, algorithme EM.