Hétérogénéité non observée dans les modèles de durée

G. Horny

Discussion Paper 2007-46

Département des Sciences Économiques
de l'Université catholique de Louvain

Hétérogénéité non observée dans les modèles de durée

Guillaume Horny*

14 décembre 2007

Résumé

Cet article est une revue de la littérature où le temps passé dans un état est une variable aléatoire issue d'un mélange continu de distributions. Elle s'est constituée à partir de l'estimation de fonctions de hasards et de méthodes d'approximations d'intégrales. Nous présentons d'abord le modèle de mélange de hasards proportionnels et ses propriétés. Les conséquences des principaux résultats d'identification sont ensuite discutées. Nous présentons ensuite des méthodes d'estimations paramétriques, semi-paramétriques et bayésiennes, ainsi que les méthodes d'optimisation correspondantes.

Mots Clés : modèles de durée, hasards proportionnels, vraisemblance pénalisée, vraisemblance partielle, méthodes bayésiennes.

Classification JEL : C41, C51, C61, C11, C14.

Remerciements : Cette contribution s'inscrit dans le cadre du programme Pôles d'attraction interuniversitaires PAI pour le compte de l'Etat belge (P6/07 Economic Policy and Finance in the Global Economy : Equilibrium Analysis and Social Evaluation).

1 Introduction

Cet article est une revue de la littérature où le temps passé dans un état est une variable aléatoire issue d'un mélange continu de distributions. Cette modélisation permet d'analyser des questions variées, que ce soit l'accès à l'emploi, la durée de grèves, la décision d'investir ou les comportements de consommation.

Les premiers travaux de cette littérature datent du début des années 80, et elle s'est constituée à partir de la notion de fonction de hasards et la disponibilité de méthodes d'approximations d'intégrales.

*Département d'Economie, Université Catholique de Louvain, Belgique. guillaume.horny@uclouvain.be

Les fonctions de hasard expriment la probabilité de transition d'une unité statistique entre deux états dans un petit intervalle de temps, en terme d'hétérogénéité observée et non observée. L'hétérogénéité renvoie aux différences de facteurs pertinents dans les choix des individus, et peut être observée ou non par l'économètre, du fait d'omission lors de la collecte des données, de problèmes de mesure ou de codification. Les méthodes d'approximations d'intégrales d'évaluer les paramètres du mélange.

L'habitude de travailler sur des agents moyens ou des comportements agrégés fait perdre de vue l'absence d'homogénéité des individus, dont la découverte de l'importance dans les années 90 est évoquée par Heckman (2001). Des individus hétérogènes peuvent de plus se trouver au sein de groupes eux même hétérogènes. Une large part de l'analyse statistique est dédiée à l'évaluation de quelques quantités qui représentent autant que possible les informations contenues dans des données trop nombreuses pour être directement appréhendées (Fisher, 1922). Dès lors, comment prendre en compte des hétérogénéités non observées, potentiellement situées à plusieurs niveaux, sans faillir à l'obligation de fournir des résultats synthétiques ?

Cet article débute par une présentation générale de la modélisation des transitions et de l'hétérogénéité non observée. Dans le prolongement des travaux sur les fonctions de hasards proportionnels, nous discutons le modèle de hasards proportionnels de mélange (abrégé MPH dans la littérature anglophone). Nous montrons comment il peut prendre en compte des structures très diverses d'hétérogénéité non observée, et présentons les principaux résultats d'identifications. Nous discutons ensuite les méthodes d'inférence paramétriques, semi-paramétriques et bayésiennes. Pour toutes ces approches, nous nous attachons à expliquer la construction de la fonction objectif et à présenter les méthodes d'optimisation les plus couramment utilisées et les plus prometteuses.

Les structures d'hétérogénéité et les phénomènes qui les produisent varient grandement d'une application à l'autre. Toutefois, cette revue de littérature insiste sur les point suivants :

1. de nombreuses applications utilisent la même structure élémentaire. Les différences entre les applications relèvent principalement du champs d'étude,
2. l'identification est obtenue sous des hypothèses relativement peu contraignantes lorsque plusieurs observations ont en commun la même hétérogénéité non observée,
3. le choix de la distribution mélangeante influence les résultats et il n'y a généralement que peu d'arguments en faveur d'une distribution plutôt qu'une autre,

4. les même idées sous-tendent les méthodes d'optimisations utilisées dans un cadre paramétrique, semi-paramétrique ou bayésien.

Plusieurs revues de littérature sont consacrées aux modèles de durées. Kiefer (1988) présente les principales fonctions de hasard, et Florens *et alii* (2007) discutent précisément la prise en compte de durées discrètes et censurées. Lancaster (1990) fournit un exposé détaillé et complet de la modélisation, et van den Berg (2001) discute le modèle MPH et les principaux résultats d'identification. Par comparaison, nous nous focalisons ici sur le traitement de l'hétérogénéité non observée.

L'article est organisé de la manière suivante. Nous présentons dans la Section 2 les principaux concepts de l'étude des durées. La Section 3 présente le modèle MPH et ses propriétés. Les principaux résultats d'identification sont discutés dans la Section 4. Le choix de la distribution mélangeante est présenté dans la Section 5. Les Sections 6, 7 et 8 sont consacrées aux méthodes d'estimation paramétriques, semi-paramétriques et bayésiennes.

2 Fonctions de survie, de hasard et censure

Nous présentons d'abord le processus le plus simple générant des durées et des dates de censure. Nous définissons ensuite la distribution des durées et en déduisons les fonctions de survie et de hasard.

Considérons le processus $\{Z_t, t \in \mathbf{R}_+\}$ où Z_t est à valeur dans l'espace des états $E = \{E_0, E_1\}$. Il est initialisé par $Z_0 = E_0$ et il n'y a de transition que vers l'état E_1. La date $t = 0$ correspond à la date d'entrée d'un individu dans l'état d'intérêt E_0. On parle alors d'échantillonnage de flux, et des modalités alternatives de collectes de données sont présentées dans Lancaster (1990, p. 182-193). Soit T la variable aléatoire positive indiquant le temps qui sépare la date $t = 0$ de la date où s'est produite la transition vers E_1. Dans ce contexte, T n'est pas le temps calendaire mais le temps passé par une unité statistique en E_0.

Il est possible d'observer une suite de durées lorsqu'un passage du processus en E_1 est suivie par un retour en E_0. Ce processus génère des données contenant des épisodes multiples. Lorsque l'espace des états s'étend à $\{E_0, \ldots, E_K\}$, il est possible de quitter E_0 pour K destinations et on parle alors de modèle à risques concurrents.

Pour tenir compte d'un mécanisme de censure, étendons l'espace des états de Z_t à $\{E_0, E_1, E_C\}$, où E_C correspond à un état hors du champs de l'enquête. On note C la durée séparant la date $t = 0$

d'une transition vers E_C, et l'enquêteur n'observe que $\min(T, C)$. La censure permet ainsi d'observer partiellement les processus dont les transitions vers E_1 ne se sont pas produites avant la fin de l'enquête. Une censure est exogène lorsque T et C sont indépendantes, et résulte entre autres des mécanismes de collecte de données discutés dans Rodriguez (2001).

La fonction de répartition de T, notée F, satisfait les conditions habituelles : elle est à valeur dans $[0, 1]$, monotone non décroissante, continue à droite et vaut 1 lorsque $t \to \infty$.

Définition 2.1. ***Fonction de survie***

La fonction de survie S de la variable aléatoire T est définie comme :

$$S(t) = P(T \geq t) = 1 - F(t). \tag{1}$$

Ces propriétés se déduisent directement de celles de la fonction de répartition.

Définition 2.2. ***Fonction de hasard***

La fonction de hasard λ de la variable aléatoire T est définie comme :

$$\lambda(t) = \lim_{dt \to 0} \frac{1}{dt} P(T \in [t, t + dt[| T \geq t). \tag{2}$$

En temps continu, la fonction de hasard est la probabilité limite qu'une unité effectue une transition sachant qu'elle n'en a pas effectué au préalable. On peut montrer (voir par exemple Gourieroux et Jasiak, 2007, p. 87) la relation suivante :

$$\lambda(t) = \frac{f(t)}{S(t)}. \tag{3}$$

D'où :

$$\int_0^t \lambda(u) du = \int_0^t \frac{f(u)}{S(u)} du = - \int_0^t \frac{1}{S(u)} dS(u) = - \ln S(t). \tag{4}$$

Les écritures des fonctions de survie et de hasard en temps discret sont présentées dans Florens *et alii* (2007).

2.1 Mélanges continus

Nous montrons ici comment retrouver la fonction de survie et la densité de probabilité lorsqu'elle dépendent d'une variable latente continue. Le lecteur intéressé par les mélanges discrets peut se référer, entre autres, à Marin *et alii* (2005).

Supposons que la population soit divisée en sous-populations, chacune caractérisée par sa propre fonction de hasard. Supposons de plus que le hasard dépende d'une variable aléatoire latente V, de support $\mathcal{V}$ et de fonction de répartition H. Le hasard conditionnel $\lambda(t|v)$ s'interprète comme la probabilité de transition instantanée d'un individu dans une population homogène par rapport à v. Les fonctions de survie et de densité peuvent s'écrire :

$$S(t) = \int_{\mathcal{V}} S(t|u)dH(u) = \mathrm{E}_V[S(t|v)], \tag{5}$$

$$f(t) = \int_{\mathcal{V}} f(t|u)dH(u) = \mathrm{E}_V[f(t|v)], \tag{6}$$

$$\begin{aligned}\lambda(t) &= \frac{\int_{\mathcal{V}} f(t|u)dH(u)}{\int_{\mathcal{V}} S(t|u)dH(u)} = \int_{\mathcal{V}} \lambda(t|u)\frac{S(t|u)}{\int_{\mathcal{V}} S(t|u)dH(u)}dH(u) \\ &= \int_{\mathcal{V}} \lambda(t|u)dH(u|T \geq t) = \mathrm{E}_V[\lambda(t|v)|T \geq t]. \end{aligned} \tag{7}$$

Le hasard caractérisé par l'équation (7) est appelé hasard moyen, car obtenu en faisant une moyenne pondérée sur des sous-populations hétérogènes. Florens *et alii* (2007) présentent un exemple de calcul de hasard moyen, et Lancaster (1990, p. 59-61) montre que des erreurs de mesure sur la variable endogène ou les variables exogènes, ainsi que l'omission de variables explicatives, produisent des mélanges.

3 Modèles de mélange de hasards proportionnels

Une manière de spécifier la distribution des durées est de supposer une forme fonctionnelle pour le hasard, et nous nous intéressons à des intensités de transition issues de mélanges de hasards proportionnels.

Notons x le vecteur des variables explicatives constantes dans le temps de l'individu courant. Le modèle MPH est présent pour la première fois dans Lancaster (1979) et van den Berg (2001) en propose la définition suivante :

Définition 3.1. ***Modèle MPH***
Il existe des fonctions λ_0 et λ_1 telles que pour tout (t, x, v), la hasard conditionnel a pour expression :

$$\lambda(t|x, v) = v\lambda_0(t)\lambda_1(x). \tag{8}$$

La fonctions λ_0 est le hasard de base, commune à tous les individus et qui ne dépend que du temps. Les variables explicatives interviennent au travers de la partie systématique du hasard λ_1. Le terme d'hétérogénéité v peut être spécifié à différents niveaux, et nous en présentons plusieurs modélisations dans la sous-section 3.2. Lorsque v est égal à 1, il n'y a pas d'hétérogénéité non observée et le modèle MPH devient un modèle de hasards proportionnels comprenant le modèle de Cox (Cox, 1972) comme cas particulier.

Les conditions de régularité usuelles sont discutées dans van den Berg (2001). Elles assurent que λ_0 et λ_1 soient positives, le hasard n'est pas infini sur un intervalle de temps et tous les épisodes se terminent au plus tard à une date finie. La distribution de l'hétérogénéité a son support dans $[0, \infty[$ et v ne dépend pas du temps. Il convient de noter que ces hypothèses sont plus fortes que ce qui est requis pour la plupart des résultats présentés dans cette section. L'identification, discutée dans la section 4, requiert toutefois des hypothèses additionnelles.

3.1 Propriétés

Nous présentons tout d'abord la propriété de décroissance au fil du temps de la valeur moyenne de v parmi les survivants, qui influence le taux de variation du hasard moyen. Nous le comparons ensuite au taux de variation du hasard conditionnel.

3.1.1 Hasards proportionnels

Cox (1972) présente le modèle de hasards proportionnels. Le modèle MPH en est une extension prenant en compte l'hétérogénéité non observée.

Le rapport des fonctions de hasards des individus i et j a pour expression :

$$\frac{\lambda(t|x_i, v_i)}{\lambda(t|x_j, v_j)} = \frac{v_i \lambda_1(x_j)}{v_j \lambda_1(x_j)}. \tag{9}$$

Le ratio diffère d'un couple d'individus à l'autre selon les hétérogénéités observées et inobservées. A toute date t, la probabilité de transition de l'individu i est proportionnelle à la probabilité de transition de l'individu j. Le rapport est constant pour deux individus donnés, hormis lorsque les régresseurs varient dans le temps.

3.1.2 Décroissance de la moyenne de V parmi les survivants

La propriété que nous rappelons ici est présenté en détails dans Lancaster (1990, p. 62-65) et van den Berg (2001). Des relations (7) et (4), on déduit la valeur moyenne de l'effet aléatoire parmi les survivants à la date t :

$$\begin{aligned}
\mathrm{E}(V|T \geq t) &= \int_{\mathcal{V}} u dH(u|T \geq t) \\
&= \int_{\mathcal{V}} u \frac{S(t|u)}{\int_{\mathcal{V}} S(t|u) dH(u)} dH(u) \\
&= \frac{\int_{\mathcal{V}} u \exp[-u\Lambda(t)] dH(u)}{\int_{\mathcal{V}} \exp[-u\Lambda(t)] dH(u)},
\end{aligned} \tag{10}$$

où $\Lambda(t) = \lambda_1(x) \int_0^t \lambda_0(u) du$. D'où :

$$\begin{aligned}
\frac{d\mathrm{E}(V|T \geq t)}{dt} &= \lambda_0(t)\lambda_1(x) \left[-\frac{\int_{\mathcal{V}} u^2 S(t|u) dH(u)}{\int_{\mathcal{V}} S(t|u) dH(u)} + \left(\frac{\int_{\mathcal{V}} u S(t|u) dH(u)}{\int_{\mathcal{V}} S(t|u)) dH(u)} \right)^2 \right] \\
&= -\lambda_0(t)\lambda_1(x) \left[\mathrm{E}(V^2|T \geq t) - [\mathrm{E}(V|T \geq t)]^2 \right] \\
&= -\lambda_0(t)\lambda_1(x) \mathrm{Var}[V|T \geq t] \leq 0.
\end{aligned} \tag{11}$$

Ce résultat, appelé "weeding out" dans la littérature anglophone, indique que les survivants ont en moyenne des v de plus en plus petits lorsque le temps passe. Intuitivement, les unités caractérisées par des v élevés effectuent leurs transitions rapidement et l'échantillon des survivants est sélectionné.

3.1.3 Taux d'accroissements du hasard

Il vient de (7), (8) et (11) :

$$\frac{d\lambda(t)}{dt} = \lambda_1(X) \frac{d\lambda_0(t)}{dt} \mathrm{E}[V|T \geq t] - [\lambda_0(t)\lambda_1(X)]^2 \, \mathrm{Var}[V|T \geq t]. \tag{12}$$

Le taux de croissance du hasard moyen par rapport au temps peut s'écrire :

$$\begin{aligned}
\frac{d\ln\lambda(t)}{dt} &= \frac{1}{\lambda_0(t)}\frac{d\lambda_0(t)}{dt} - \lambda_0(t)\lambda_1(X)\frac{\mathrm{Var}[V|T\geq t]}{\mathrm{E}[V|T\geq t]} \\
&= \frac{d}{dt}\ln\lambda(t|v) - \lambda_0(t)\lambda_1(X)\frac{\mathrm{Var}[V|T\geq t]}{\mathrm{E}[V|T\geq t]} \\
&< \frac{d\ln\lambda(t|v)}{dt}.
\end{aligned} \tag{13}$$

Le hasard moyen décroît plus vite dans le temps que le hasard conditionnel. Ce résultat est un corollaire du "weeding out" : le hasard moyen est plus faible car évalué sur une sous-population hétérogène comprenant des éléments peu susceptibles d'effectuer une transition. Omettre à tort la présence d'hétérogénéité non observée revient à amplifier la décroissance du hasard.

De même, on peut montrer la relation suivante :

$$\frac{d\ln\lambda(t)}{dx_p} < \frac{d\ln\lambda(t|v)}{dx_p}, \tag{14}$$

où p ($p = 1,\ldots,P$) est l'indice des variables explicatives. L'omission d'hétérogénéité non observée dans un modèle de hasards proportionnels amène à sous-estimer l'influence des régresseurs.

L'élasticité croisée du hasard moyen par rapport au temps et aux variables explicatives dépend des trois premiers moments de la distribution mélangeante (voir Horny, 2006a, pour le détail des calculs). Lancaster (1990, p. 65-67) montre que la présence d'hétérogénéité gamma atténue les variations dues aux variables explicatives et au temps du hasard moyen par rapport au hasard conditionnel. Ce résultat ne peut toutefois pas être généralisé à d'autres distribution mélangeante et van den Berg (2001) montre que le signe est indéterminé pour une hétérogénéité discrète quelconque.

3.2 Structure de l'hétérogénéité non observée

La structure très souple de l'hétérogénéité non observée dans les modèles MPH permet de spécifier v à différents niveaux, ou de l'écrire comme produit de plusieurs effets aléatoires.

Par simplicité, nous n'avons considéré jusqu'ici que des situations où une durée est observée pour chaque individu. Toutefois, un même individu peut connaître plusieurs durées consécutives, et des individus physiquement différents peuvent partager la même hétérogénéité non observée. Un ensemble d'individus ayant en commun une même hétérogénéité non observée est appelé "strate" .

3.2.1 Formulations à un effet aléatoire

Soit i ($i = 1, \ldots, N$) l'indice de durée et j ($j = 1, \ldots, J$) l'indice de strate. La spécification la plus courante est :

$$\lambda_i(t|x_i, v_i) = v_i \lambda_0(t) \lambda_1(x_i). \tag{15}$$

A chaque épisode correspond une unique réalisation de l'effet aléatoire. Les v_i peuvent être indépendants (Clayton, 1978; McGilchrist et Aisbett, 1991), ou bien issus d'une loi multivariée.[1] Les durées sont indépendantes conditionnellement aux variables explicatives et à V. Les applications comprennent des études du chômage (Lancaster, 1979), de la mortalité infantile (Guo et Rodriguez, 1992), etc. Lorsqu'il y n'y a pas d'hétérogénéité non observée mais uniquement des erreurs de mesure, v_i capte la surdispersion.

L'hétérogénéité peut être multiple et située à la fois au niveau des durées et des strates, d'où l'écriture du hasard :

$$\lambda(t|x_{ij}, v_{ij}) = v_{ij} \lambda_0(t) \lambda_1(x_{ij}). \tag{16}$$

McGilchrist (1993) l'utilise pour l'étude de durées, où la survie d'un individu est corrélée avec la survie des autres membres familles. La matrices de variance de V est bloc-diagonale.

3.2.2 Formulations à deux effets aléatoires

Parner (1997) considère la structure d'hétérogénéité suivante :

$$\lambda(t|x_i, v_j, w_i) = (v_j + w_i) \lambda_0(t) \lambda_1(x_i). \tag{17}$$

Les deux effets sont supposés indépendants et distribués selon des lois gamma de même paramètre d'échelle, amenant la somme $v_j + w_i$ à suivre elle aussi une loi gamma. Le modèle est similaire à celui défini par (16), et comprend le cas particulier du modèle à épisodes uniques lorsque Var(v_j)=0.

Sastry (1997) spécifie la fonction de hasard :

$$\lambda(t|x_{ijk}, v_k, w_j) = v_k w_j \lambda_0(t) \lambda_1(x_{ijk}), \tag{18}$$

[1] Bonnal *et alii* (1997) et Fougère et Kamionka (2005) introduisent une relation de dépendance en posant $v_i = \exp(a_i u_1 + b_i u_2)$, où u_1 et u_2 sont des variables latentes indépendantes.

où k $(k = 1, \ldots, K)$ est un indice de groupes d'unités. Son application porte sur la mortalité infantile, où l'enfant i est issu de la famille j qui réside dans le village k. Les effets sont ici emboîtés car une famille appartient à un unique village. Une structure d'hétérogénéité non-emboîtée seraient adaptée si de nombreuses familles changeaient de lieu de résidence.

4 Identification

Une abondante littérature s'intéresse à l'identification non paramétrique du modèle MPH, c'est-à-dire sans spécification de λ_0, λ_1 et H. Le modèle est identifié lorsqu'il existe un unique $(\lambda_0, \lambda_1, H)$ pouvant générer les données observées, ce qui n'est pas trivial car $\lambda(t|x, v) = vk\lambda_0(t)k^{-1}\lambda_1(x)$, $\forall k \in \mathbf{R}_*^+$. En plus des hypothèses habituelles ci dessous, la solution consiste généralement à normaliser $(\lambda_0, \lambda_1, H)$.

4.0.3 Cas général

L'identification non paramétrique est établie pour la première fois par Elbers et Ridder (1982), en posant les conditions de régularité évoquées dans la Section 2 et les hypothèses suivantes :

Hypothèse 4.1. *V est indépendant de X.*

Hypothèse 4.2. *L'ensemble $\mathcal{X}$ contient aux moins deux valeurs différentes, et $\lambda_1(x)$ n'est pas constante sur $\mathcal{X}$.*

Hypothèse 4.3. *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 4.4. $\mathrm{E}(V) < \infty$.

L'hypothèse H4.1 est relativement forte car elle suppose que les déterminants inobservés des transitions ne sont pas liés aux déterminants observés. L'hypothèse H4.2 signifie qu'une variabilité est requise dans la partie déterministe du hasard et qu'un modèle comprenant seulement une constante n'est pas identifié. Les choix de la durée et de l'individu de référence sont faits dans l'hypothèse H4.3. La normalisation de H est effectuée dans la dernière hypothèse, qui est relativement lourde dans la mesure où des spécifications comme une loi positive stable Hougaard (1984) sont tout à fait plausibles mais n'admettent pas une espérance de V finie. Heckman et Singer (1984b) lui proposent une alternative portant uniquement sur l'épaisseur des queues de distribution. Ce choix inclut alors la loi positive stable dans les spécifications identifiables, mais exclue les lois dégénérées (van den Berg, 2001). Choisir entre H4.4 et l'alternative de Heckman et Singer (1984b) revient à choisir entre une inférence basée sur une spécification peu flexible de H, ou exclure a priori les modèles MPH ne contenant pas d'hétérogénéité non observée.

L'identification de modèles de risques concurrents est caractérisée par le résultat de Cox (1962) et Tsiatis (1975). Ils montrent que tout modèle de risques concurrents dont les durées latentes sont indépendantes est observationnellement équivalent à un modèle de risques concurrents dont les durées latentes ne sont pas indépendantes. Supposer l'indépendance pour obtenir l'identification n'est toutefois pas indispensable, comme le montrent Heckman et Honoré (1989) et Abbring et Van den Berg (2003a) avec un modèle MPH contenant des variables explicatives.

4.0.4 Rôle des variables explicatives et de la structure d'hétérogénéité

La famille des modèles de Box-Cox, qui comprend les modèles de Weibull et de Gompertz, est identifiée en l'absence de variable explicative et les hypothèses afférentes dans Heckman et Taber (1994). Ce résultat n'a pas été étendu à d'autres modèles paramétriques. Ils montrent également que la présence de régresseurs variant dans le temps permet d'identifier un modèle MPH sans supposer H4.4. Ce dernier résultat laisse penser que l'hypothèse H4.4, et les considérations relativement pessimistes qu'elle suscite, n'est pas incontournable dès lors que l'on dispose de variation supplémentaire dans une composante du modèle. Le cas des régresseurs dichotomiques endogènes variant dans le temps est discuté dans Abbring et Van den Berg (2003b). Ils établissent l'identification sous les hypothèses de Elbers et Ridder (1982) auxquelles s'ajoute une hypothèse garantissant que l'endogénéité des régresseurs ne vient que des informations passées et non pas d'anticipations sur le futur.

L'identification en présence d'épisodes multiples et sans variable explicative est établie dans Honoré (1993) pour une famille de modèles incluant le cas MPH, sans supposer H4.1 et H4.4. L'intuition est que les variations de λ_0 peuvent être isolées à partir du moment où l'on dispose d'un même v pour plusieurs durées, entraînant l'identification de H. Ce résultat est important car les hypothèses d'indépendance de l'hétérogénéité observée et non observée, ainsi que la finitude de $\mathrm{E}(V)$, ne peuvent pas être testées dans les modèles à épisodes uniques. Horny *et alii* (2005) montrent l'identification de modèles MPH où le terme d'hétérogénéité non observée s'écrit comme le produit de deux effets aléatoires, chacun ayant toutes ses réalisations communes à plusieurs durées, sans supposer H4.1 et H4.4.

L'ajout de variables explicative est nécessaire pour identifier un modèle MPH incluant une fonction du temps passé dans l'état qui précède l'état actuel (Honoré, 1993; Frijters, 2002).

5 Choix d'une distribution mélangeante

L'inférence dans les modèles MPH requiert généralement la spécification d'une forme paramétrique pour la distribution mélangeante. Malheureusement, les estimations sont biaisées en cas de mauvaise spécification (voir par exemple Baker et Melino, 2000 et Gaure *et alii*, 2005 pour des étude sur données simulées), et de nombreuses études pointent une sensibilité des résultats au choix de la distribution mélangeante (Heckman et Singer, 1984a; Hougaard *et alii*, 1994; Keiding *et alii*, 1997).

5.1 Spécifications paramétriques

La loi gamma, introduite comme distribution mélangeante dans Lancaster (1979) et Vaupel *et alii* (1979), est de loin la plus utilisée dans les applications. L'idée fait vraisemblablement écho au modèle Negbin de Greenwood et Yule (1920). Abbring et van den Berg (2007) donnent une justification a posteriori intéressante à ce choix. Ils montrent que pour une large classe de modèles de hasards proportionnels, la distribution mélangeante de l'hétérogénéité non observée parmi les survivants converge vers une loi gamma. Ce résultat s'applique entre autres lorsque la loi de V dans la population d'origine est uniforme, normale tronquée, beta ou encore discrète avec un point de masse en 0. Il n'est par contre pas valide pour une distribution ne comprenant pas 0 dans son support.

On peut également spécifier une loi générale qui admet comme cas particuliers plusieurs distributions usuelles. Hougaard (2000) propose ainsi la loi stable positive, qui équivaut entre autres à une loi gaussienne ou gamma selon les paramètres estimés.

5.2 Spécifications non-paramétriques

La structure de l'hétérogénéité non observée dans une spécification non-paramétrique a ceci de spécifique que l'appartenance d'une durée à un groupe n'est plus spécifiée ex ante dans le modèle mais estimée. Heckman et Singer (1984a) proposent un estimateur non paramétrique du maximum de vraisemblance (NPMLE), qui suppose une distribution mélangeante discrète avec des points de masse à estimer. L'estimation retourne généralement 3 ou 4 points de masse différents. Baker et Melino (2000) et Gaure *et alii* (2005) étudient le comportement de NPMLE dans de larges études de Monte Carlo.

Horowitz (1999) montre comment estimer le hasard de base et la distribution mélangeante au sein d'une approche non-paramétrique, lorsque les régresseurs sont constants dans le temps. Hausman et Woutersen (2005) proposent une approche non-paramétrique pour des régresseurs variant dans le temps.

Le cadre bayésien se prête également à la modélisation d'une distribution mélangeante non paramétrique, notamment lorsque les probabilités de transitions sont tirées dans des lois de Dirichlet. Florens *et alii* (1999) passent en revue la littérature bayésienne des modèles de durée où la distribution de Dirichlet est utilisée comme a priori. Lau (2006) montre comment procéder à l'inférence par échantillonnage de Gibbs.

Les résultats méthodologiques de Heckman et Taber (1994) et Horowitz (1999) indiquent que les durées sont généralement peu informatives sur la forme de la distribution mélangeante. Il n'y a que peu d'arguments en faveur d'une distribution plutôt qu'une autre, qu'elle soit discrète ou continue. Des changements minimes de spécification peuvent amener à des ajustements très proches et des estimations très différentes, particulièrement en présence d'épisodes uniques. Ce constat doit être nuancé en présence d'épisodes multiples, car conformément à ce que laissent penser les résultats d'identification, plusieurs études empiriques retournent des résultats plus stables pour différents choix de H (Guo et Rodriguez, 1992; Gönül et Srinivasan, 1993; Bonnal *et alii*, 1997).

6 Inférence paramétrique

L'inférence est paramétrique lorsque la distribution jointe des données est complètement spécifiée par l'économètre, hormis pour un nombre fini de paramètres. L'estimation procède généralement par maximisation de la vraisemblance, qui a l'avantage de fournir des estimateurs invariant à une transformation fonctionnelle, efficaces et convergeant à la vitesse $\sqrt{N}$.

Le processus générant les données ne correspond pas nécessairement à un mélange, et les estimations obtenues par maximum de vraisemblance peuvent se trouver sur la frontière de l'espace paramétrique où le gradient de la log-vraisemblance n'est pas nul. Un tel exemple et présenté dans Lancaster (1990, p. 194-197). Une méthode d'optimisation exempte de cette difficulté et adaptée aux vraisemblances issues de mélanges et l'algorithme EM de Dempster *et alii* (1977), que nous présentons.

A chaque date t, l'unité peut rester dans l'état courant, effectuer une transition vers un autre état ou bien connaître une censure exogène. Une observation censurée contribue à la vraisemblance à hauteur de la fonction de survie évaluée à la date de censure, tandis qu'une transition observée contribue à hauteur de la probabilité jointe de la survie jusqu'à t et d'une transition en t. Soit d_i une variable indiquant une transition observée pour la durée i. La contribution à la vraisemblance marginale de l'épisode i est :

$$L^i(\beta) = \int_{\mathcal{V}} [v_i\lambda_0(t_i)\lambda_1(x_i;\beta)]^{d_i} \exp\left(-\int_0^{t_i} v_i\lambda_0(u)\lambda_1(x_i;\beta)du\right) dH(v;\alpha). \tag{19}$$

Si les v_i étaient observés, on pourrait optimiser :

$$\ln L_V(\beta|v) = \sum_{i=1}^{N}\left[d_i \ln\left[v_i\lambda_0(t_i)\lambda_1(x_i;\beta)\right] - \left(\int_0^{t_i} v_i\lambda_0(u)\lambda_1(x_i;\beta)du\right) + \ln h(v_i;\alpha)\right]. \tag{20}$$

Une méthode d'approximation du maximum de vraisemblance en présence de variables latentes est l'algorithme EM (Dempster *et alii*, 1977). L'intuition sous-jacente est d'approcher les v_i par leur espérance conditionnelle (étape E), puis d'optimiser la vraisemblance ainsi obtenue (étape M). Comme les estimations des v_i dépendent des paramètres estimés, les deux étapes s'inscrivent au sein d'une procédure itérative.

L'espérance conditionnelle aux données de la log-vraisemblance à l'itération $(q+1)$ peut s'écrire :

$$\begin{aligned} Q(\beta,\beta^{(q)}) = \sum_{i=1}^{N}\Bigg[& d_i \mathrm{E}_{\beta^{(q)},\alpha^{(q)}}(\ln V|T=t,d) + d_i \ln\left[\lambda_0(t_i)\lambda_1(x_i;\beta)\right] \\ & - \mathrm{E}_{\beta^{(q)},\alpha^{(q)}}(V|T=t,d)\int_0^{t_i}\lambda_0(u)\lambda_1(x_i;\beta)du + \mathrm{E}_{\beta^{(q)},\alpha^{(q)}}\left[\ln h(V)|T=t,d\right]\Bigg], \end{aligned} \tag{21}$$

où $\mathrm{E}_{\beta^{(q)},\alpha^{(q)}}$ indique que les espérances de fonctions de V dépendent des β estimés lors de l'itération précédente. On peut donc organiser l'algorithme EM de la manière suivante :

Étape A Évaluer la relation (21)

Étape B Optimiser le résultat de l'étape A, récupérer $(\beta^{(q)},\alpha^{(q)})$.

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

Le choix de la distribution mélangeante détermine les expressions des espérances conditionnelles. Guo et Rodriguez (1992) utilisent l'algorithme EM pour un modèle à hétérogénéité gamma, où les espérances admettent une expression analytique. Sastry (1997) étend leur approche à deux effets aléatoires gamma emboîtés et approche les espérances par des méthodes d'intégration numérique. En présence d'hétérogénéité gaussienne, elles peuvent être approchées par quadrature de Gauss-Hermite (Friedl et Kauermann, 2000; Maples *et alii*, 2002), méthode de Monte Carlo (Ripatti et Palmgren, 2002) ou échantillonnage de Gibbs (Vaida et Xu, 2000).

7 Inférence semi-paramétrique

Les approches paramétriques imposent au hasard de base une forme déterminée par un nombre fini de paramètres. Dans cette section, nous nous intéressons à l'approche semi-paramétrique de la vraisemblance partielle (Cox, 1975), qui ne contraint pas les variations du hasard de base. En contrepartie, la variance des estimateurs est plus importante. La meilleure précision de l'estimateur du maximum de vraisemblance n'est toutefois d'aucune utilité s'il repose sur une spécification erronée.

L'optimisation de la vraisemblance partielle se fait généralement par l'algorithme EM. Nous attachons également une attention toute particulière aux méthodes récentes de vraisemblances partielles pénalisées, dont nous montrons qu'elles peuvent s'insérer dans une structure EM pour l'améliorer.

7.1 La vraisemblance partielle

Cox (1972) propose, en parallèle au modèle qui prendra son nom, de baser l'inférence sur une autre fonction que la vraisemblance et fournit dans Cox (1975) l'approche semi-paramétrique de la vraisemblance partielle. Cameron et Trivedi (2005, p. 594-596) la présentent en détails et elle admet comme expression :

$$L_p(\beta) = \prod_{i=1}^{N} \left[\frac{\lambda_1(x_i;\beta)}{\sum_{k \in R(t_i)} \lambda_1(x_k;\beta)} \right]^{d_i}, \tag{22}$$

où $R(t_i)$ est l'ensemble des indices des unités susceptibles d'effectuer une transition à la date t_i. La vraisemblance partielle est le ratio de la probabilité instantanée de transition de l'individu i en t_i sur la probabilité instantanée de transition d'un survivant en t_i. Si une constante est incluse dans x_i, elle se simplifie et n'est pas identifiée dans cette approche. Les durées censurées contribuent à la vraisemblance partielle par le dénominateur de la relation (22).

Dans une approche paramétrique en présence de régresseurs variant dans le temps, Lancaster (1990, p. 272-274) montre que la distribution jointe des durées et des variables explicative peut s'écrire comme le produit de la vraisemblance partielle et d'un facteur ne dépendant pas des durées. L'inférence devrait alors être basée sur la vraisemblance partielle, quand bien même l'approche est complètement paramétrique.

La contribution de l'individu i à la vraisemblance partielle en présence d'hétérogénéité non observée peut s'écrire :

$$L_p^i(\beta) = \int_{\mathcal{V}} \cdots \int_{\mathcal{V}} \frac{v_i \lambda_1(x_i;\beta)}{\sum_{k\in R(t_i)} v_k \lambda_1(x_k;\beta)} dH(v_1,\ldots,v_I). \tag{23}$$

L'expression (23) peut être optimisée avec l'algorithme EM (Nielsen *et alii*, 1992), en présence d'un effet aléatoire (Parner, 1997; Vu et Knuiman, 2002; Hsu *et alii*, 2004) ou de plusieurs (Horny, 2006b). Bien que l'algorithme soit conçu pour mener l'inférence en présence de variables latentes, les difficultés numériques qui lui sont inhérentes (Ng *et alii*, 2004) deviennent problématique en présence d'hétérogénéité non observée et plusieurs cas de non convergence sont pointés dans la littérature (Lancaster, 1990, p. 267, et Bolstad et Manda, 2001).

7.2 La vraisemblance partielle pénalisée

L'économètre peut disposer d'une information sur la forme générale d'une composante du modèle, même s'il retient une spécification non-paramétrique pour conserver un maximum de flexibilité. Les méthodes de vraisemblance pénalisée sont une manière de prendre en compte cette information (voir Green, 1999, pour une présentation générale). Elles proposent d'optimiser une fonction objectif comprenant la vraisemblance et une fonction de pénalité. La vraisemblance traduit les hypothèses portant sur le processus générant les données, tandis que la fonction de pénalité attribue des pondérations aux observations et assouplit la vraisemblance. L'économètre choisit la fonction de pénalité, qui permet d'introduire une information dont il dispose et qui n'est pas contenue dans les données.[2]

Notons β le vecteur des paramètres de λ_1 et α un vecteur de paramètres spécifiques à la fonction de pénalité. On peut écrire la log-vraisemblance partielle pénalisée suivante :

$$\ln L_{PPL}(\alpha,\beta,v) = \ln L_{PL}(\beta|v) - g(v,\alpha), \tag{24}$$

où L_{PL} est la vraisemblance partielle et $g(v,\alpha)$ une fonction positive quelconque. Cette représentation indique que les observables sont générés par leur probabilité jointe conditionnelle aux inobservables, tandis qu'une information sur la distribution des inobservables est véhiculée par g.

[2] L'approche bayésienne reconnaît explicitement l'idée de croyance a priori sur le processus générant les données (Lancaster, 2004, p. 2-8).

La pénalité choisie ici ne dépend pas des paramètres de L_{PL} et l'estimateur du maximum de la vraisemblance partielle pénalisée (PPL) est identique à l'estimateur du maximum de la vraisemblance partielle (PL) pour les paramètres de λ_1. L'estimateur des v_i est obtenu par les conditions de premier ordre et dépend du choix de g.

Therneau *et alii* (2003) montrent que l'argument de la maximisation d'une vraisemblance partielle pénalisée où la fonction de pénalité est le logarithme du noyau d'une loi gamma est égal à l'argument de la maximisation de la vraisemblance partielle en présence d'hétérogénéité gamma par l'algorithme EM. La simplicité des conditions de premier ordre permet d'optimiser la fonction objectif au moyen de méthodes de Newton-Raphson, plus rapides en convergence ainsi qu'en temps de calcul, et plus stables qu'un algorithme EM. Leur résultat ne se généralise toutefois pas à d'autres distributions mélangeantes (Therneau et Grambsch, 2000, p. 255-256).

7.3 Un algorithme EM pour modèles à plusieurs effets aléatoires

Les modèles faisant intervenir deux effets aléatoires, présentés dans la sous-section 3.2.2, sont adaptées à des échantillons comprenant des durées situées à deux niveaux : il existe des groupes d'observations qui contiennent des sous-groupes, chaque sous-groupe contenant lui même plusieurs observations. Sastry (1997) en présente une application à l'étude de la mortalité d'individus vivant dans des communautés, chaque communauté comprenant les familles de plusieurs individus comme sous-groupes.

Afin de prendre en compte une population stratifiée à K niveaux différents, on peut spécifier un modèle MPH dont le terme d'hétérogénéité non observé s'écrit comme le produit de K effets aléatoires.[3] Un algorithme EM général pour ce type de modèles est présenté dans Horny (2006b). Chacune des K distributions mélangeantes est supposée admettre une transformée de Laplace.

Considérons la fonction de hasard :

$$\lambda_i^j(t) = \left(\prod_{k=1}^{K} v_k^j\right) \lambda_0(t)\lambda_1\left(x_i;\beta\right), \qquad (25)$$

où i est l'indice des durées ($i = 1, \ldots, N$), j ($j = 1, \ldots, J_k$) l'indice du groupe contenant la durée i et situé au niveau k, et k ($k = 1, \ldots, K$) l'indice de niveau.[4] On note T le vecteur des durées et d le vecteur des

[3]L'identification de modèles MPH à K effets aléatoires s'obtient directement en généralisant la preuve fournie par Horny *et alii* (2005) dans le cas de deux effets aléatoires.

[4]Comme chaque groupe est spécifique à un niveau, une notation complète demanderait d'utiliser $j(k)$. Les écritures

indicatrices de transition. Lorsque les effets aléatoires sont indépendants, la vraisemblance conditionnelle s'écrit alors :

$$L(\alpha,\beta;T,d,x,v)=\prod_{i=1}^{N}\left\{\left[\left(\prod_{k=1}^{K}v_k^j\right)\lambda_0(t_i)\lambda_1\left(x_i,\beta\right)\right]^{d_i}\exp\left[-\left(\prod_{k=1}^{K}v_k^j\right)\int_0^{t_i}\lambda_0(u)\lambda_1\left(x_i,\beta\right)du\right]\right.$$
$$\left.\prod_{k=1}^{K}h_k(v_k^j;\alpha_k)\right\}, \quad (26)$$

où $h_k(v_k^j;\alpha_k)$ est la distribution mélangeante au niveau k. L'optimisation par l'algorithme EM est équivalente à l'algorithme dont l'itération (q) est la suivante :

Étape A Pour un k' quelconque, estimer le modèle :

$$\lambda_i^j(t)=v_{k'}^j\left(\prod_{k\neq k'}^{K}v_k^{j,(q)}\right)\lambda_0(t)\lambda_1\left(x_i,\beta^{(q,k')}\right). \quad (27)$$

On obtient $v_1^{j,(q+1)}$ et $\beta^{(q+1,k')}$.

Étape B Répéter l'étape A pour tout $k\in\{1,\ldots,K\}-\{k'\}$

Étape C Estimer le modèle :

$$\lambda_{ik}^j(t)=\left(\prod_{k=1}^{K}v_k^{j,(q+1)}\right)\lambda_0(t)\lambda_1\left(x_i,\beta^{(q)}\right). \quad (28)$$

On obtient $\beta^{(q+1)}$.

Étape D Itérer entre les étapes A et C jusqu'à la convergence.[5]

Chaque itération de l'étape A peut être effectuée avec un algorithme EM. Les gains de cette reformulation sont déjà doubles : l'évaluation des espérances implique moins d'intégrales et les maximisations s'effectuent sur des espaces paramétriques de dimensions plus faibles.

Lorsque le théorème de convergence vers une loi gamma de la distribution mélangeante parmi les survivants d'Abbring et van den Berg (2007) s'applique, on peut procéder aux étapes A, B et C en maximisant des vraisemblances partielles pénalisées. L'algorithme ainsi obtenu converge là où un algorithme

simplifiée utilisées ici ne devraient toutefois pas prêter à confusion.

[5]Les valeurs intermédiaires $\beta^{(q,k)}$ servent à étudier la convergence de l'algorithme.

EM standard ne converge pas, et les temps calcul peuvent être réduits de 66% sur certaines structures d'hétérogénéitée par rapport à un algorithme EM standard (Horny, 2006b).

8 L'approche bayésienne

L'approche bayésienne est l'application de la formule de Bayes à l'analyse économétrique.[6] La conception des probabilité qui en résulte est fondamentalement différente de la vision fréquentiste habituelle, et on ne peut comparer en profondeur les deux approches que sur la base de critères philosophiques. De façon plus pragmatique, l'approche bayésienne permet d'analyser des modèles dont la vraisemblance n'admet pas de forme analytique, tout comme les méthodes de maximum de vraisemblance simulées présentées entre autres dans Gouriéroux et Monfort (1995). L'approche bayésienne ne requiert toutefois pas l'augmentation du nombre de simulations à un taux approprié avec la taille de l'échantillon. De plus, les propriétés de l'estimateur bayésien sont à distance finie, et l'optique bayésienne se prête naturellement à la comparaison de modèles. L'inconvénient de cette approche, outre la subjectivité, est généralement le temps de calcul.

Nous présentons dans un premier temps des éléments de formulation et d'interprétations de modèles bayésiens. Le traitement des mélanges est discuté, ainsi qu'un algorithme général d'évaluation de l'estimateur bayésien dans un modèle MPH. La méthode semi-paramétrique de la vraisemblance partielle peut être justifiée dans un cadre bayésien, et nous passons en revue les principales applications.

8.1 Formulation et interprétation d'un modèle bayésien

La théorie des probabilités décrit les propriétés de l'application P() mais ne nous apprend rien sur son interprétation. Dans une conception fréquentiste, elle s'interprète comme la fréquence d'apparition d'un événement lors d'un nombre élevé de répétitions. Dans une conception subjectiviste, les probabilités s'interprètent comme la croyance en la réalisation d'un évènement.[7] Une interprétation suggestive prévaut dans l'optique bayésienne, où les paramètres sont issus de distributions de probabilités. Une démarche bayésienne peut être vue comme un processus de mise à jour de croyances sur les paramètres du modèle résultant de la collecte de nouvelles données (Lancaster, 2004, p. 2-8).

[6] L'article de Thomas Bayes est republié dans Barnard et Bayes (1958).

[7] L'interprétation suggestive des probabilités est proche de l'idée que nous avons dans la vie de tous les jours lorsque nous disons d'un événement qu'il est "très probable" ou "peu probable".

L'approche bayésienne est séquentielle. Dans un premier temps, l'économètre formule le processus générant les données en un modèle probabiliste. Ensuite, les croyances sur les paramètres sont traduites en distributions de probabilités a priori. On a déduit la loi a posteriori, qui est le produit de la probabilité jointe d'un échantillon avec les densités a priori. Les données sont collectées dans une étape ultérieure, et l'estimateur bayésien est obtenu comme l'espérance de la loi a posteriori.

8.2 Loi a posteriori et mélanges

La loi a posteriori a pour expression :

$$\pi(\theta|T, d) \propto f(T, d|\theta)p(\theta), \tag{29}$$

où θ est le vecteur des paramètres, $f(T, d|\theta)$ la densité jointe des données et p la densité a priori.

En présence d'inobservables, la loi a posteriori peut s'écrire :

$$\pi(\theta|T, d) = \int_{\mathcal{V}} \pi(\theta|T, d, v)dP(v). \tag{30}$$

Tanner et Wong (1987) introduisent cette décomposition reprenant l'idée de l'algorithme EM de Dempster *et alii* (1977), qu'ils citent, de traiter les variables latentes comme des données manquantes dont on connaît toutefois la distribution. L'estimateur bayésien est l'espérance des paramètres sous la loi a posteriori, et a pour expression lorsque l'on considère un modèle MPH à un effet aléatoire :

$$\mathrm{E}_{\pi}(\theta|T, d) \propto \int_{\Theta} \theta \int_{\mathcal{V}} \prod_{i=1}^{N} [v_i\lambda_0(t_i)\lambda_1(x_i)]^{d_i} \exp\left(-\int_0^{t_i} v_i\lambda_0(u)\lambda_1(x_i)du\right) dP(v|\theta)p(\theta)d\theta. \tag{31}$$

Tandis que les difficultés d'utilisation du maximum de vraisemblance sont majoritairement des problèmes d'optimisation, l'approche bayésienne est confrontée à des problèmes d'intégrations. Elle se prête donc naturellement à la modélisation et l'estimation de mélanges.

8.3 Approximation de l'estimateur Bayésien

La relation (30) suggère d'approcher de l'estimateur bayésien par des simulations issues de l'algorithme suivant :

Étape A Tirer θ^{sim} dans la loi a priori

Étape B Utiliser θ^{sim} pour tirer des v^{sim} dans la distribution mélangeante

Étape C Evaluer la vraisemblance au point $(T, d, v^{\mathrm{sim}}, \theta^{\mathrm{sim}})$.

Étape D Procéder aux étapes B et C pour un nombre Nrep_1 de réplications.

Étape E Approcher $\mathrm{E}_{V|\theta}(L)$, où L est la vraisemblance conditionnelle aux v, par la moyenne sur les réplications.

Étape F Procéder aux étapes A et E pour un nombre Nrep_2 de réplications.

Étape G Approcher $\mathrm{E}_\pi(\theta|T, d)$ par la moyenne sur les réplications.

La loi des grands nombres s'applique et pour un nombre suffisamment élevé de réplications, les moments d'ordre 1 calculés sur les simulations convergent en probabilité vers leurs espérances.

De nombreuses méthodes d'intégrations numériques sont utilisées pour approcher l'estimateur bayésien, parmi lesquelles l'algorithme de Metropolis-Hasting (Geman et Geman, 1984), l'échantillonnage de Gibbs (Gelfand et Smith, 1990), et plus généralement les Méthodes de Monte Carlo par Chaînes de Markov (MCMC) que Robert (1996) passe en revue.

8.4 Utilisations de la vraisemblance partielle dans une approche bayésienne

Kalbfleisch (1978) propose une justification de la vraisemblance partielle dans un cadre Bayésien. Spécifions l'a priori suivant pour le hasard de base :

$$\lambda_0(t) \sim \gamma(c\lambda_0^*(t), c), \tag{32}$$

où γ désigne la loi gamma, c un paramètre et $\lambda_0(t)^*$ une fonction connue. On a : $\mathrm{E}[\lambda_0(t)] = \lambda_0^*(t)$ et $\mathrm{V}[\lambda_0(t)] = \lambda_0^*(t)/c$. La fonction $\lambda_0^*(t)$ peut s'interpréter comme une croyance a priori sur le hasard de base en t et c comme son degré de confiance. On peut montrer que lorsque $c \to 0$, l'estimateur bayésien des coefficients de λ_1 converge vers l'estimateur du maximum de vraisemblance partielle.

Clayton (1991) reprend cet argument dans l'estimation d'un modèle MPH à hétérogénéité gamma, fournissant une alternative bayésienne à l'algorithme EM de Klein (1992). Une 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 alii* (1998) et Sargent (1998). Un modèle à deux effets aléatoires gamma est estimé dans une approche paramétrique par Bolstad et Manda (2001). Dans un cadre semi-paramétrique, Horny *et alii* (2005) estiment un modèle à deux effets gamma indépendants et Horny *et alii* (2006) un modèle à deux effets log-normaux corrélés. Un modèle MPH en temps discret est étudié dans Manda et Meyer (2005).

9 Conclusion

Les modèles MPH à hétérogénéité non observée fournissent un cadre d'analyse pour de nombreuses questions auxquelles sont confrontés économètres, actuaires et démographes. La disponibilité récente de données employeurs-employés ou localisées permet de plus d'étudier les questions d'appariements ou de spatialiser les dynamiques individuelles.

Existe-t'il des régularités empiriques en faveur des modèles à hétérogénéité non observée sur les modèles sans hétérogénéité non observée ? Tout dépend de la question de recherche. De nombreux travaux empiriques concluent, dans leurs domaines respectifs, à la présence d'une influence significative de l'hétérogénéité non observée. Les modèles de hasards proportionnels (Cox, 1972), où les agents ne diffèrent que dans leurs caractéristiques observées, peut être vu comme un modèle de référence aux hypothèses minimales.

L'identification de modèle où chaque terme d'hétérogénéité est commun à plusieurs observations requiert relativement peu d'hypothèses. L'influence de la censure lorsque des durées successives sont observées est toutefois peu étudiée. En effet, si des individus sont suivis sur un intervalle de temps prédéterminé, plus les premières durées seront longues et plus les durées ultérieures seront courtes (Ridder et Tunalı, 1999). Une censure endogène est prise en compte dans Woutersen (2000) pour un modèle de hasard où l'hétérogénéité non observée est représentée par un effet fixe, éventuellement corrélé aux autres variables explicatives. A notre connaissance, il n'existe pas d'estimateur qui lui soit similaire pour une modélisation à effets aléatoires.

Les principaux résultats d'identification nous apprennent également que les données sont généralement peu informatives quant à la distribution mélangeante. Horny *et alii* (2005) fournissent l'exemple d'un même modèle estimé par maximum de vraisemblance partielle pénalisée et par une approche bayésienne, dont les résultats concernant la distribution mélangeante diffèrent.[8] Dans une application utilisant des

[8]Rodriguez et Goldman (2001) et Therneau *et alii* (2003) concluent également à la sous-estimation de l'influence de

données appariées employeurs-employés, Horny *et alii* (2006) obtiennent des résultats stables . Dans quelle mesure la disponibilité récente et croissante de ces informations permet de stabiliser les estimations est une question ouverte.

La prise en compte de l'hétérogénéité des agents dans les modèles théoriques se poursuit, notamment dans l'étude du marché du travail (Cahuc *et alii*, 2006). Le traitement de l'hétérogénéité non observée en est la contrepartie empirique, et nous avons présenté les contours actuels de cette littérature.

Références

ABBRING J. et VAN DEN BERG G. (2003a), The identifiability of the mixed proportional hazards competing risks model, *Journal of the Royal Statistical Society*, Series B, 65, pp. 701–711.

ABBRING J. et VAN DEN BERG G. (2003b), The nonparametric identification of treatment effects in duration models, *Econometrica*, 71(5), pp. 1491–1518.

ABBRING J. et VAN DEN BERG G. (2007), The unobserved heterogeneity distribution in duration analysis, *Biometrika*, 94, pp. 87–99.

BAKER M. et MELINO A. (2000), Duration dependence and nonparametric heterogeneity : a Monte Carlo study, *Journal of Econometrics*, 96 (2), pp. 357–393.

BARNARD G. et BAYES T. (1958), Studies in the history of probability and statistics : IX. Thomas Bayes's essay towards solving a problem in the doctrine of chances., *Biometrika*, 45 (3/4), pp. 293–315.

VAN DEN BERG G. (2001), Duration models : specification, identification and multiple durations, dans *Handbook of Econometrics*, J. J. Heckman and E. Leamer (eds.), Elsevier, Amsterdam, tome 5, chapitre 55, pp. 3381–3463.

BOLSTAD W. et MANDA S. (2001), Investigating child mortality in Malawi using family and community random effects : a Bayesian analysis, *Journal of the American Statistical Association*, 96, pp. 12–19.

BONNAL L., FOUGÈRE D. et SÉRANDON A. (1997), Evaluating the impact of French employment policies on individual labour market histories, *Review of Economic Studies*, 64, pp. 683–713.

l'hétérogénéité par différentes méthodes de maximisation de la vraisemblance.

Cahuc P., Postel Vinay F. et Robin J.M. (2006), Wage bargaining with on-the-job search : A structural econometric model, *Econometrica*, 74 (2), pp. 323–364.

Cameron A. et Trivedi P. (2005), *Microeconometrics*, Cambridge University Press.

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, pp. 141–151.

Clayton D. (1991), A Monte Carlo method for Bayesian inference in frailty models, *Biometrics*, 47, pp. 467–485.

Cox D. (1962), *Renewal Theory*, Methuen, London.

Cox D. (1972), Regression models and life tables, *Journal of the Royal Statistical Society*, Series B, 34, pp. 187–220.

Cox D. (1975), Partial likelihood, *Biometrika*, 62(2), pp. 269–276.

Dempster A., Laird N. et Rubin D. (1977), Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion), *Journal of the Royal Statistical Society*, Series B, 39, pp. 1–38.

Elbers C. et Ridder G. (1982), True and spurious duration dependence : The identifiability of the proportional hazard model, *Review of Economic Studies*, 49, pp. 403–410.

Fisher R. (1922), On the mathematical foundations of theoretical statistics, *Philosophical Transactions of the Royal Society of London*, 222, pp. 309–368.

Florens J.P., Fougère D. et Mouchart M. (2007), Duration models and point processes, Discussion paper 2971, IZA.

Florens J.P., Mouchart M. et Rolin J.M. (1999), Semi- and non-parametric Bayesian analysis of duration models with Dirichlet priors : a survey, *International Statistical Review*, 67, pp. 187–210.

Fougère D. et Kamionka T. (2005), Econometrics of Individual Labor Market Transitions, Discussion paper 1850, IZA.

Friedl H. et Kauermann G. (2000), Standard errors for EM estimates in generalized linear models with random effects, *Biometrics*, 56 (3), pp. 761–767.

FRIJTERS P. (2002), The non-parametric identification of lagged duration dependence, *Economic letters*, 75, pp. 289–292.

GAUDERMAN W. et THOMAS D. (1994), Censored survival models for genetic epidemiology : a Gibbs sampling approach, *Genetic Epidemiology*, 11, pp. 171–188.

GAURE S., ROED K. et ZHANG T. (2005), Time and causality : a monte carlo assessment of the timing-of-events approach, Memorandum 19, University of Oslo.

GELFAND A. et SMITH A. (1990), Sampling based approaches to calculating marginal densities, *Journal of the American Statistical Association*, 85, pp. 398–409.

GEMAN S. et GEMAN D. (1984), Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, pp. 721–741.

GÖNÜL F. et SRINIVASAN K. (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, pp. 1219–1227.

GOURIEROUX C. et JASIAK J. (2007), *The econometrics of individual risk : credit, insurance, and marketing*, Princeton university press.

GOURIÉROUX G. et MONFORT A. (1995), *Simulation Based Econometric Methods*, Oxford University Press, Oxford.

GREEN P. (1999), Penalized likelihood, dans KOTZ S., READ C. et BANKS D. (éds), *Encyclopaedia of Statistical Sciences*, Wiley, tome 3 update, pp. 578–586.

GREENWOOD M. et YULE G. (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, pp. 255–279.

GUO G. et RODRIGUEZ G. (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, pp. 969–976.

Hausman J. et Woutersen T. (2005), A semi-parametric duration model with heterogeneity that does not need to be estimated, Presented at the Econometric Society World Congress 2005.

Heckman J. (2001), Micro Data, Heterogeneity, and the Evaluation of Public Policy : Nobel Lecture, *Journal of Political Economy*, 109, pp. 673–748.

Heckman J. et Honoré B. (1989), The identifiability of the competing risks model, *Biometrika*, 76, pp. 325–330.

Heckman J. et Singer B. (1984a), Econometric duration analysis, *Journal of Econometrics*, 24, pp. 62–132.

Heckman J. et Singer B. (1984b), The identifiability of the proportional hazards model, *Review of Economic Studies*, 60, pp. 231–243.

Heckman J. et Taber C. (1994), Econometric mixture models and more general models for unobservables in duration analysis, *Satistical Methods in Medical Research*, 3, pp. 279–302.

Honoré B. (1993), Identification results for duration models with multiple spells, *The Review of Economic Studies*, 60, pp. 241–246.

Horny G. (2006a), *Modèles de durée multivariés avec hétérogénéité multiple : applications au marché du travail*, Thèse de doctorat, Strasbourg I.

Horny G. (2006b), Partial Likelihood Estimation of a Cox Model With Random Effects : an EM Algorithm Based on Penalized Likelihood, Working Paper 10, BETA.

Horny G., Boockmann B., Djurdjevic D. et Laisney F. (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.

Horny G., Mendes R. et van den Berg G.J. (2006), Job mobility in Portugal : a Bayesian study with matched worker-firm data, Working Paper 32, BETA.

Horowitz J. (1999), Semiparametric estimation of a proportional hazard model with unobserved heterogeneity, *Econometrica*, 67, pp. 1001–1028.

Hougaard P. (1984), Life table methods for heterogeneous populations : distributions describing the heterogeneity, *Biometrika*, 71, pp. 75–83.

Hougaard P. (2000), *Analysis of multivariate survival data*, Springer-Verlag, New York.

Hougaard P., Myglegaard P. et Borch-Johnsen K. (1994), Heterogeneity models of disease susceptibility, with an application to diabetic nephropathy, *Biometrics*, 50, pp. 1178–88.

Hsu L., Chen L.and Gorfine M. et Malone K. (2004), Semiparametric estimation of marginal hazard function from casecontrol family studies, *biometrics*, 60, pp. 936–944.

Kalbfleisch J. (1978), Non-parametric Bayesian analysis of survival time data, *Journal of the Royal Statistical Society*, Series B, 40, pp. 214–221.

Keiding N., Andersen P. et Klein J. (1997), The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates, *Statistics in Medicine*, 16, pp. 215–224.

Kiefer N. (1988), Economic duration data and hazard function, *Journal of Economic Literature*, 26, pp. 646–679.

Klein J. (1992), Semiparametric estimation of random effects using the Cox model based on the EM algorithm, *Biometrics*, 48, pp. 795–806.

Korsgaard I., Madsen P. et Jensen J. (1998), Bayesian inference in the semiparametric log normal frailty model using Gibbs sampling, *Genetics Selection Evolution*, 30, pp. 241–256.

Lancaster T. (1979), Econometric methods for the duration of unemployment, *Econometrica*, 47, pp. 939–956.

Lancaster T. (1990), *The econometric analysis of transition data*, Cambridge University Press, Cambridge.

Lancaster T. (2004), *An introduction to modern Bayesian econometrics*, Blackwell, Oxford.

Lau J. (2006), Bayesian semi-parametric modeling for mixed proportional hazard models with right censoring, *Statistics and Probability Letters*, 76, pp. 719–728.

Manda S. et Meyer R. (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, pp. 439–455.

Maples J., Murphy S. et Axinn W. (2002), Two level proportional hazards model, *Biometrics*, 58(4), pp. 754–763.

Marin J., Mengersen K. et Robert C. (2005), *Handbook of Statistics*, Dey, D. Dey and Rao, C., tome 25, chapitre Bayesian modelling and inference on mixtures of distributions, pp. 1–56.

McGilchrist C. (1993), REML estimation for survival model with frailty, *Biometrics*, 49, pp. 221–225.

McGilchrist C. et Aisbett C. (1991), Regression with frailty in survival analysis, *Biometrics*, 47 (2), pp. 461–466.

Ng S., Krishnan T. et McLachlan G. (2004), *Handbook of computational statistics,*, Springer, chapitre The EM algorithm.

Nielsen G., Gill R., Andersen P. et Sorensen T. (1992), A counting process approach to maximum likelihood estimation in frailty models, *Scandinavian Journal of Statistics*, 19, pp. 25–44.

Parner E. (1997), *Inference in semiparametric frailty models*, Thèse de doctorat, University of Aarhus.

Ridder G. et TunaliI. (1999), Stratified partial likelihood estimation, *Journal of Econometrics*, 92, pp. 193–232.

Ripatti S. Larsen K. et Palmgren J. (2002), Maximum likelihood inference for multivariate frailty models using an automated Monte Carlo EM algorithm, *Lifetime Data Analysis*, 8, pp. 349–360.

Robert C. (1996), *Méthodes de Monte Carlo par chaînes de Markov*, Economica, Paris.

Rodriguez G. (2001), Survival models, Lecture Notes, Chapter 7.

Rodriguez G. et Goldman N. (2001), Improved estimation procedures for multilevel models with binary responses : a case study, *Journal of the Royal Statistical Society*, Series A, 164, pp. 339–355.

Sargent D. (1998), A general framework for random effects survival analysis in the Cox proportional hazards setting, *Biometrics*, 54, pp. 1486–1497.

SASTRY N. (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, pp. 426–435.

TANNER M. et WONG W. (1987), The calculation of posterior distributions by data augmentation, *Journal of the American Statistical Association*, 82, pp. 528–540.

THERNEAU M. et GRAMBSCH P. (2000), *Modeling survival data : extending the Cox model*, Statistics for Biology and Health, Springer, New-York.

THERNEAU T., GRAMBSCH P. et PANKRATZ V. (2003), Penalized survival models and frailty, *Journal of Computational and Graphical Statistics*, 12, pp. 156–175.

TSIATIS A. (1975), A nonidentifiability aspect of the problem of competing risks, *Proceedings of the National Academy of Sciences*, 72, pp. 20–22.

VAIDA F. et XU R. (2000), Proportional hazards model with random effects, *Statistics in Medicine*, 19, pp. 3309–3324.

VAUPEL J., MANTON K. et STALLARD E. (1979), The impact of heterogeneity in individual frailty on the dynamics of mortality, *Demography*, 16(3), pp. 439–454.

VU H. et KNUIMAN M. (2002), Estimation in semiparametric marginal shared gamma frailty models, *Australian and New-Zealand Journak of Statistics*, 44, pp. 489–501.

WOUTERSEN T. (2000), Consistent Estimators for Panel Duration Data with Endogenous Censoring and Endogenous Regressors, dans *Econometric Society World Congress*.

Département des Sciences Économiques
de l'Université catholique de Louvain
Institut de Recherches Économiques et Sociales

Place Montesquieu, 3
1348 Louvain-la-Neuve, Belgique

ISSN 1379-244X D/2007/3082/046