DOCUMENT

DE TRAVAIL

N° 248

INFERENCE IN MIXED PROPORTIONAL HAZARD MODELS WITH K RANDOM EFFECTS

Guillaume Horny

August 2009

BANQUE DE FRANCE

EUROSYSTÈME

DIRECTION GÉNÉRALE DES ÉTUDES ET DES RELATIONS INTERNATIONALES

DIRECTION GÉNÉRALE DES ÉTUDES ET DES RELATIONS INTERNATIONALES

INFERENCE IN MIXED PROPORTIONAL HAZARD MODELS WITH K RANDOM EFFECTS

Guillaume Horny

August 2009

Inference in Mixed Proportional Hazard Models with K Random Effects

Guillaume Horny[1]

Forthcoming in Statistical Papers

[1]Banque de France, DGEI-DEMS-SAMIC, 31 rue Croix des Petits Champs, 75 049 Paris Cedex 01, France and UCLouvain, Belgium. E-mail: guillaume.horny@banque-france.fr.

Abstract

A general formulation of Mixed Proportional Hazard models with K random effects is provided. It enables to account for a population stratified at K different levels. We then show how to approximate the partial maximum likelihood estimator using an EM algorithm. In a Monte Carlo study, the behavior of the estimator is assessed and I provide an application to the ratification of ILO conventions. Compared to other procedures, the results indicate an important decrease in computing time, as well as improved convergence and stability.

Keywords: EM algorithm, penalized likelihood, partial likelihood, frailties, duration analysis.
JEL Classification: C13, C14, C41.

Résumé

Nous présentons une formulation générale d'un modèle de mélange de hasards proportionnels à K effets aléatoires. Elle permet la prise en compte d'une population stratifiée à K niveaux différents. Nous montrons ensuite comment approcher l'estimateur du maximum de la vraisemblance partielle par un algorithme EM. Le comportement de cet estimateur est étudié dans une étude Monte Carlo et nous fournissons également une application à la ratification des conventions de l'OIT. Comparativement aux autres procédures, nos résultats indiquent une baisse importante des temps de calcul ainsi qu'une amélioration de la convergence et de la stabilité de l'algorithme.

Mots-clés: Algorithme EM, vraisemblance pénalisée, vraisemblance partielle, effets aléatoires, modèles de durées.
Classification JEL: C13, C14, C41.

Non-technical summary

It is generally not reasonable to assume that we observe all the determinants leading to a transition from one state to another, and many studies in duration analysis take account of the heterogeneity that exist among agents but that is unobserved to the analyst. It is generally modeled using a random effect. Applications handling a single random effect, and thus an heterogeneity located at only one level, are common in econometrics, biometrics and demography. However, the possibility of omitted variables with group structure at different levels arises in several data generating processes. Ignoring some of the unobserved heterogeneity can lead to substantial biases (see Pakes, 1983, Moulton, 1986, and Gouriéroux and Peaucelle, 1990, for some case studies in linear models). Furthermore, accounting for unobserved heterogeneity at different levels can be a matter of importance in applied work, e.g. when studying child mortality or the spread of a disease. However, multiple unobserved heterogeneity raises some awkward problems, as it involves multidimensional integrals which typically do not admit analytical expressions. No more than two levels of clustering are handled in the literature (see Manda and Meyer, 2005, Yau, 2001 and Sastry, 1997, among others).

In this paper, I present a Mixed Proportional Hazard (MPH) model with K random effects.[2] A general EM algorithm is then proposed to approximate the estimator. We show how the estimation of a single model with K effects can be restated in the estimation of K models, each with a single frailty. This simplification enables us to use quick and stable numerical procedures. We also detail how to recover the variance of the estimates. Finally, we compare our algorithm with an accelerated one developed by Sastry (1997) on simulated and real data. Our procedure exhibit a strong decrease in computing time, and an improved stability.

[2]Mixed Proportional Hazard models with K random effects allow the unobserved characteristics to be located at K different levels. For example, it allows the sample to be divided in clusters and sub-clusters K times.

Résumé non-technique

Tous les déterminants des transitions entre deux états ne sont généralement pas observés, et la plupart des études utilisant des modèles de durés prennent en compte l'hétérogénéité qui n'est pas observée par l'analyste mais qui existe entre les agents. Elle est généralement modélisée au moyen d'un effet aléatoire. Les applications comprenant un effet aléatoire, et donc une hétérogénéité située à un unique niveau, sont courantes en économétrie, biostatistique et démographie. Toutefois, la possibilité de variables omises dont les valeurs diffèrent pour des groupes d'observations intervient dans de nombreux processus générant des données. Ne pas prendre en compte une part de l'hétérogénéité non observée peut amener à des biais substantiels (voir Pakes, 1983, Moulton, 1986, et Gouriéroux and Peaucelle, 1990, pour des modèles linéaires). De plus, l'évaluation de l'hétérogénéité non observée se trouvant à différents niveaux peut être pertinente dans les applications, par exemple lors de l'analyse de mortalité infantile, ou bien de la propagation d'une épidémie. Toutefois, la présence d'une hétérogénéité non observée multiple soulève plusieurs problèmes, car elle implique l'évaluation d'intégrales multiples qui n'admettent généralement pas de solution analytique. Il n'y a pas plus de deux niveaux qui soient étudiés simultanément dans la littérature (voir Manda and Meyer, 2005, Yau, 2001 et Sastry, 1997, parmi d'autres).

Dans ce papier, nous présentons un modèle de mélange de hasards proportionnels à K effets aléatoires.[3] Nous proposons ensuite un algorithme EM général pour approcher l'estimateur du maximum de vraisemblance. Nous montrons comment l'estimation d'un modèle à K effets peut se ramener à l'estimation de K modèles impliquant chacun un unique effet aléatoire. Cette simplification nous permet d'utiliser des procédures numériques rapides et stables. Nous montrons également comment approcher la variance de l'estimateur. Finalement, nous comparons sur des données simulées et réelle notre algorithme avec l'algorithme EM accéléré développé par Sastry (1997). Notre procédure conduit à une baisse importante des taux de calcul, et à une amélioration de la stabilité.

[3]Les modèle de mélange de hasards proportionnels à K effets aléatoires permettent à l'hétérogénéité non observée d'être située à K niveaux différents. Par exemple, elle permet à l'échantillon d'être divisé en groupes et sous groupes K fois.

1 Introduction

In this paper, I propose a general framework to estimate mixed proportional hazard models with a fixed number of risks, where unobserved heterogeneity is modeled at different levels. This class of models allows the unobserved characteristics to be located at different levels. If the random effects are nested, the sample is divided in clusters, sub-clusters and so on.

As it is not reasonable to assume that we observe all the determinants leading to a transition, many studies in duration analysis take account of unobserved heterogeneity. Applications handling a single random effect are common in econometrics, biometrics and demography. However, the possibility of omitted variables with group structure at different levels arises in several data generating processes. Ignoring some of the unobserved heterogeneity can lead to substantial biases (see Pakes, 1983, Moulton, 1986, and Gouriéroux and Peaucelle, 1990, for some case studies in linear models). Furthermore, accounting for unobserved heterogeneity at different levels can be a matter of importance in applied work, e.g. when studying child mortality or the spread of a disease. However, it raises some awkward problems for inference, as it involves multidimensional integrals which typically do not admit analytical expressions.

No more than two levels of clustering are handled in the literature. Manda and Meyer (2005) consider a model in discrete time, Yau (2001) and Sastry (1997) study two nested random effects with log-normal and gamma mixing distributions, respectively. In this last study, inference is based on the Expectation Maximization (EM) algorithm (see Dempster *et al*, 1977, for the first general formulation). This algorithm is ideally suited for mixture models, due to their missing data structure, and is used in numerous studies involving the Cox model with one frailty, such as Clayton and Cuzick (1985), Gill (1985) and Parner (1997).[4]

However, this theoretical attractivity it balanced by numerical drawbacks. First, there are several cases of non-convergence in the literature. Bolstad and Manda (2001) pointed out one case where the variance of the random effect becomes large enough to raise numerical issues. Conversely, Lancaster (1990, p. 267) described the case of an unbounded likelihood with a variance of the unobserved heterogeneity tending to zero. Second, step E can have no analytical solutions and the expressions thus require an evaluation using numerical integration or Monte Carlo methods.[5] The EM algorithm typically asks for a large number of iterations and thus involves a huge computational cost, even for relatively simple models.

In this paper, I present a Mixed Proportional Hazard (MPH) model with K random effects. We then use an EM algorithm for the inference, general to all mixing distributions admitting a Laplace transform. We show furthermore how to transform the estimation of a single model with K effects in the estimation of K MPH models, each with a single frailty. This simplification enables us

[4]The term "frailty" is used as a synonym for random effect. It comes from the first studies in this literature, which were on demography.

[5]The Monte Carlo EM (MCEM) algorithm is described in Wei and Tanner (1990).

to use quick and stable procedures during the estimation. The EM algorithm does not provide an estimate of the Information matrix, and we detail how to recover it using Louis' (1982) methodology. Finally, we compare the behavior of our algorithm with the accelerated EM algorithm developed by Sastry (1997) on simulated and real data.

Some papers have been published on related topics. Hybrid ML-EM algorithm are provided in the one gamma frailty setting by Vu *et al.* (2001) and by Vu and Knuiman (2002a) for a log-normal frailty. An extension to one shared gamma random effect is provided in Vu and Knuiman (2002b) and the events before entry are included in Vu (2003) and Vu (2004). In comparison, we focus on the multiple frailties setting and devote a particular attention to gamma random effects. Indeed, Abbring and Van den Berg (2006) show that the mixing distribution among survivors converges to a gamma distribution and the model described in this paper nest many common models as special cases.[6]

The remainder of the paper is organized as follows. Section 2 presents the statistical model. Section 3 proposes the devoted estimation method. We compare our estimator with the one detailed in Sastry (1997) in a Monte Carlo study in Section 4, and in an illustrative application on the timing of ratification of ILO conventions in Section 5.

2 Mixed Proportional Hazard model with random effects

Individuals can move among a set of states, and a transition is a movement from one state to another. We are focusing here on the state, or the sequence of states, that are occupied and the times at which movements between them occurred. When there is only a single state, econometricians usually call this single spell data, whereas it is referred to as survival data in biometrics. When we observe the duration of stay in a sequence of states, these are multiple spell data. Individual histories can be observed through panel or retrospective surveys. Let T^* denote the time elapsed in a given state. The investigator does not observe T^* for every individual: some of them did not experienced any transition before the study ends and all one knows is that T^* is at least as great as the observed duration. Let T denote the vector of the followup time and δ a vector of indicators which are 0 if a transition is observed and 1 if the observation is censored. We observe the pair (T, δ).

Duration data can be studied using the widespread Cox proportional hazard model and its extension, the MPH model. Van den Berg (2001) provides a detailed survey where the MPH model is defined by the hazard function:

$$\lambda_i(t|X_i, v) = v\lambda_0(t)\lambda_1[X_i(t), \beta], \tag{1}$$

[6] The result of Abbring and Van den Berg, 2006, requires the heterogeneity distribution to be regularly varying at 0, as defined in Feller (1971). This requirement is weak and distributions such as the exponential, uniform, beta and all distributions with a mass point at 0 fulfill it.

where i is the individual index. The random effect v usually captures unobserved characteristics specific to individual i, making him more or less prone to experiment a transition. These unobserved characteristics may also be shared by all the individuals within a group, and thus one realization v can be common to several agents. The random effect is drawn from a probability distribution such that $Pr(0 < v < \infty) = 1$. In the following, we assume that the mixing distribution $h(v;\alpha)$ admits a known Laplace transform. A large variance of a random effect means a tighter positive association among units of the same group and greater differences between the groups defined at this level. The baseline hazard λ_0 is positive, common to all the observations and depends only on the elapsed time. The function λ_1 is also positive and called the "systematic part of the hazard". It depends on the explanatory variables which can be time varying. We make the standard regularity assumption that the process $X_i(t)$ is absolutely continuous with respect to the Lebesgue measure.

Random effects are used to model unobserved characteristics common to several individuals. The related observations are then sharing the same frailty variable. Dependence among individuals can be settled at different levels, requiring a model with several frailties. For example, a model involving 2 random effects is well suited when the population is divided into clusters and each cluster is divided into subclusters. We specify the following general hazard function involving K frailties:

$$\lambda_{ik}^{j}(t) = \left(\prod_{k=1}^{K} v_k^j\right) \lambda_0(t)\lambda_1\left[X_i(t),\beta\right], \tag{2}$$

where i is the cross-sectional unit index ($i = 1,\ldots,N$), j ($j = 1,\ldots,J_k$) is the index of the group containing unit i, and k ($k = 1,\ldots,K$) is the level index.[7]

This flexible unobserved heterogeneity structure covers many uses and special cases:

- Single random effect

  $$\lambda_{i1}^{j}(t) = v_1^j \lambda_0(t)\lambda_1\left[X_i(t),\beta\right]. \tag{3}$$

  Units are correlated through v_1^j. There is only one level and the model is relevant in multiple spells settings, where each agent j experiences several durations. The limiting case is reached when $j = i$, when each group is made of one observation (single spell setting). In this case, the random effect does not depict dependence, but overdispersion instead. Model (3) contains the Cox model (Cox, 1972), as well as the shared frailty model (see e.g. Vu and Knuiman, 2002a) as special cases. Identification is discussed in Van den Berg (2001).

- Two nested random effects

  $$\lambda_{ik}^{j}(t) = v_1^j v_2^{j'} \lambda_0(t)\lambda_1\left[X_i(t),\beta\right]. \tag{4}$$

[7] Since each unit belongs to a group, a complete notation would require the use of $j(i)$. Still, the simplified notation should not lead to any confusion.

We extend model (3) to allow for 2 dependence levels. Model (4) is used in demography to allow the survival to depends on the family the individual belongs to, and on the community where the family is settled. In this case, the random effects are nested. They are non-nested in studies where each individual belongs to several groups. This involves studies on purchasing behavior where a consumer can visit different shopping centers. As before, one random effect can be defined to handle overdispersion. Identification is shown in Horny *et al.* (2005).

Conditionally on the random effects and explanatory variables, the observations are assumed independent. If we assume the random effects to be independent, we obtain the likelihood by taking the product of the frailty densities $h_i(v_i^j;\alpha_i)$ with the likelihood function conditional on the random effects, which is the conditional hazard times the conditional survivor function:

$$L(\alpha,\beta;T,d,X,v) = \prod_{i=1}^{N} \left\{ \left[\left(\prod_{k=1}^{K} v_k^j \right) \lambda_0(t_i)\lambda_1\left[X_i(t_i),\beta\right] \right]^{1-\delta_i} \right.$$
$$\left. \exp\left[-\left(\prod_{k=1}^{K} v_k^j \right) \int_0^{t_i} \lambda_0(u)\lambda_1\left[X_i(k),\beta\right] du \right] \prod_{k=1}^{K} h_k(v_k^j;\alpha_k) \right\}. \quad (5)$$

The log-likelihood is:

$$\ln L(\alpha,\beta;T,d,X,v) = \ln L_1(\alpha;v,d) + \ln L_2(\beta;T,d,X,v), \quad (6)$$

where

$$\ln L_1(\alpha;v) = \sum_{i=1}^{N} \left[\sum_{k=1}^{K} \ln h_k(v_k^j,\alpha_k) + (1-\delta_i) \sum_{k=1}^{K} \ln v_k^j \right], \quad (7)$$
$$\ln L_2(\beta;T,d,X,v) = \sum_{i=1}^{N} \left[(1-\delta_i)\ln\left\{\lambda_0(t_i)\lambda_1\left[X_i(t_i),\beta\right]\right\} \right.$$
$$\left. - \left(\prod_{k=1}^{K} v_k^j \right) \int_0^{t_i} \lambda_0(u)\lambda_1\left[X_i(u),\beta\right] du \right]. \quad (8)$$

The maximum likelihood estimator does not have an analytical expression, but it can be approximated using an EM algorithm.

3 Inference using an EM algorithm

Had the v_k^j been observed, the evaluation of the maximum likelihood estimator would have been straightforward. The EM algorithm proceeds in two steps. In a first step, we compute the expectation of the likelihood conditional on the data and the current estimates. This is maximized with respect to the parameters in a second step, and the procedure iterates between steps 1 and 2 until convergence is achieved.

Durations specific to group j at stratification level k are denoted by T_{jk}, and censoring indicators are denoted by δ_{jk}. If we assume all the random effects to be independent, the E step at iteration (q) is the evaluation of:

$$Q\left(\alpha,\beta;\alpha^{(q)},\beta^{(q)}\right)=Q_1\left(\alpha;\alpha^{(q)},\beta^{(q)}\right)+Q_2\left(\beta;\alpha^{(q)},\beta^{(q)}\right), \tag{9}$$

where

$$\begin{aligned} Q_1\left(\alpha;\alpha^{(q)},\beta^{(q)}\right) =& \sum_{i=1}^{N}\Bigg(\sum_{k=1}^{K}\mathrm{E}^{(q)}\left[\ln h_k(v_k^j,\alpha_k)|T_{jk},\delta_{jk}\right] \\ & + (1-\delta_i)\sum_{k=1}^{K}\mathrm{E}^{(q)}\left[\ln v_k^j|T_{jk},\delta_{jk}\right]\Bigg), \qquad (10) \\ Q_2\left(\beta;\alpha^{(q)},\beta^{(q)}\right) =& \sum_{i=1}^{N}\Bigg((1-\delta_i)\ln\lambda_0(t_i)\lambda_1\left[X_i(t_i,\beta\right] \\ & - \prod_{k=1}^{K}\mathrm{E}^{(q)}\left[v_k^j|T_{jk},\delta_{jk}\right]\int_0^{t_i}\lambda_0(u)\lambda_1\left[X_i(u),\beta\right]du\Bigg), \qquad (11) \end{aligned}$$

were $\mathrm{E}^{(q)}$ denotes the expectation with respect to the structure describe by $(\alpha^{(q)},\beta^{(q)})$ These equations admit an analytical solution only in the case of a single gamma random effect (see Clayton and Cuzick, 1985). Otherwise, the expectations in (11) have to computed using numerical procedures.[8]

3.1 Inference with one gamma frailty

Define the risk set as the set of spells still not completed at any instant before t_i, denoted by R_i. Therneau *et al.* (2003) show that the estimates of a model with one gamma frailty can be exactly obtained by maximizing the following penalized partial likelihood:

$$\ln L_{PPL}(\beta,v,\alpha_1)=\ln L_{PL}(\beta,v)-\frac{1}{\alpha_1}\sum_{j=1}^{J_1}(\ln v^j-v^j), \tag{12}$$

where $L_{PL}(\beta,v)$ is the partial likelihood:

$$L_{PL}(\beta,v)=\prod_{i=1}^{N}\left[\frac{v^j\lambda_1\left[X_i(u),\beta\right]}{\sum_{m\in R_i}v^l\lambda_1\left[X_i(u),\beta\right]}\right]^{1-\delta_i}. \tag{13}$$

In a general penalized likelihood setting, $1/\alpha$ is a smoothing parameter indicating the tradeoff between the fit to the data and smoothness of the fitted curve

[8]See Vu and Knuiman (2002b) for one log-normal frailty and Sastry (1997) for two gamma frailties

in the penalized likelihood. The solution maximizing the penalized partial likelihood above is equivalent to the EM solution for an MPH model with a gamma frailty. As shown in de Montricher *et al.* (1975), this result relies on the choice of the penalty function and does not hold if one uses as a penalty function the quantity $\int_0^\infty \left(\lambda_0^{(2)}(u)\right)^2 du$, where $\lambda_0^{(2)}(u)$ stands for the second derivative of the baseline hazard, as done for example in Rondeau *et al.* (2003).

3.2 Inference with two or more frailties

Sastry (1997) specifies two gamma frailties and organizes the algorithm as follows. After the E step at iteration (q), function Q_1 is optimized separately using a sub-EM algorithm. Step M of iteration (q) is carried out afterward for all the parameters. Using these sub-routines enables to achieve efficiency and speed gains, and Sastry (1997) does not need to implement further acceleration techniques such as the ones described for example in Meilijson (1989). However, his approach is specific to two gamma frailties. Here, we propose a general approach for models with K frailties whose distribution admits a Laplace transform.

We show in Appendix A that for all mixing distribution with a Laplace transform:

$$\mathrm{E}\left[v_k^j|T,d\right] = \frac{\mathrm{E}\left[(v_k^j)^{1+l_{kkj}}\xi_{(-k)}\right]}{\mathrm{E}\left[(v_k^j)^{1+l_{kkj}}\xi_{(-k)}\right]}, \tag{14}$$

where:

$$\xi_{(-k)} = \mathrm{E}\left[(v_k^j)^{1+l_{kkj}}\dots\mathrm{E}\left[(v_{(k+1)}^j)^{l_{(k+1)kj}}\mathrm{E}\left[(v_{(k-1)}^j)^{l_{(k-1)kj}}\dots\mathrm{E}\left[(v_2^j)^{l_{2kj}}\mathcal{L}_1^{(l_{1kj})}\right]\right]\right]\right], \tag{15}$$

and $l_{kk'j}$ is the number of transitions observed in the subcluster defined as the intersection of the clusters at levels k and k' containing unit i .

The M step asks for expression (9) to be maximised. An estimate of the cumulative baseline hazard for fixed β is:

$$\widehat{\Lambda}_0^{(q)}(t) = \sum_{t_i<t} \frac{1-\delta_i}{\sum_{m\in R_i}\left(\prod_{k=1}^K v_k^{l,(q)}\right)\lambda_1\left[X_i(t_i),\beta^{(q)}\right]}. \tag{16}$$

3.3 Organisation of the EM algorithm

Consider the following algorithm. Let us denote by $\beta^{(0)}$ the estimates of a model without random effects as starting values, and by $\beta^{(q,k)}$ the estimates of the coefficient β at iteration q obtained by treating only the effect at level k as random. Set the initial values $\alpha_k = 1$ and $\beta^{(0)} = \beta^{(0,k)}$, $\forall k$. At iteration (q), proceed as follows:

Step 1 Set k=1 and $\beta^{(q)} = \beta^{(q,1)}$. Compute $Q_1\left(\alpha_1;\alpha^{(q)},\beta^{(q,1)}\right)$ and maximise it to obtain $\alpha_1^{(q+1)}$.

Step 2 Compute $Q_2\left(\beta;\alpha_1^{(q+1)},\alpha_2^{(q)},\ldots,\alpha_K^{(q)},\beta^{(q,1)}\right)$ and maximise it to recover $\beta^{(q+1,1)}$.

Step 3 Iterate between steps 1 and 2 until convergence.

Step 4 For $k=2,\ldots,K$, iterate between steps 1 and 3 and recover $\alpha_2^{(q+1)},\ldots,\alpha_K^{(q+1)}$.

Step 5 Compute $Q_2\left(\beta;\alpha_1^{(q+1)},\ldots,\alpha_K^{(q+1)},\beta^{(q)}\right)$ and maximise it to recover $\beta^{(q+1)}$.

Step 6 Iterate between steps 1 and 5 until convergence.

The $\beta^{(q,k)}$ are involved in Steps 1 and 2 to help the estimator of α_k to converge. One could think of going directly from Step 1 to Step 4. However, Q_1 is a function of $\beta^{(q,k)}$, iterating between steps 1 and 2 provides more stable estimates $\alpha_k^{(q+1)}$ and reduces the number of iterations of the whole algorithm.

By profiling with respect to $\alpha_{k'}^{(q)}$, $\forall k' \neq k$, this approach is equivalent to alternately estimating K models involving each one random effect, the other random effects being treated as offsets. One can thus rewrite the algorithm as follows:

Step A Set k=1. Estimate model:

$$\lambda_{i1}^j(t) = v_1^j\left(\prod_{k=2}^{K} v_k^{j,(q)}\right)\lambda_0(t)\lambda_1\left[X_i(t),\beta^{(q,1)}\right]. \qquad (17)$$

Recover $v_1^{j,(q+1)}$ and $\beta^{(q+1,1)}$.

Step B Perform Step 1 for $k=2,\ldots,K$

Step C Estimate model:

$$\lambda_{ik}^j(t) = \left(\prod_{k=1}^{K} v_k^{j,(q+1)}\right)\lambda_0(t)\lambda_1\left[X_i(t),\beta^{(q)}\right]. \qquad (18)$$

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

Step D Iterate between steps A and C until convergence.

The gains of this approach are twofold: the E step is easier because it involves fewer integrals and the M step is quicker as the optimisations are carried over spaces with smaller dimensions.

The algorithms described above only require the mixing distribution to admit a Laplace transform, so that the expectations of the random effects can be computed. In studies involving a single frailty setting, the gamma and log-normal distributions are commonly used, as well as finite discrete distributions. For all mixing distribution with a mass point at 0 or a finite positive limit at 0, Abbring and Van den Berg (2006) show convergence to a gamma distribution. Approaches designed in a one gamma frailty setting can thus be implemented within our EM algorithm, and we use the penalized partial likelihood maximization described in Therneau *et al.* (2003) to perform steps 1 to 3, or equivalently steps A and B.

Note that since the α_i are treated as constants while estimating the coefficients, the standard errors of the β recovered in Step 5 or C are underestimated. We use Louis' (1982) approach, described in Appendix C, to evaluate the information matrix on the last iteration of the EM algorithm. This can be written:

$$I = \mathrm{E}\left[-H(\alpha,\beta)|T,\delta,v_1,\dots,v_I\right] - \mathrm{Var}\left[S(\alpha,\beta)|T,\delta\right], \tag{19}$$

where I is the observed information matrix, H the Hessian and S the score. Details are given in Appendix B.

4 Monte Carlo Experiments

The aim of this Monte Carlo study is to compare the behavior of the algorithm we propose with the behavior of the accelerated EM algorithm described in Sastry (1997), for different sizes of groups and subgroups in presence of two frailties. Our algorithm is hereafter referred to as Expectation Maximization algorithm based on Penalized Likelihood (EMPL).

4.1 Sample design and starting values

For each setting, 200 samples were simulated. We consider samples of size 2000 with two levels of clustering, and we vary the number of spells per group and subgroup accordingly. The smallest groups we design contain 10 observations and the largest 100. Under 10 spells, it is unlikely that the other frailty is defined at a finer level, and over 100, stratifying the baseline hazard is often preferred in applications. Subgroup sizes go from 1 to 5, the largest value compatible with groups of size 10.

We set $\mathrm{E}(v_1^j)$=$\mathrm{E}(v_2^j)$=1, $\mathrm{Var}(v_1^j)$=$\mathrm{Var}(v_2^j)$=0.5, consider a constant baseline hazard and no censoring. Two standard gaussian covariates receive the coefficients $\beta_a = 1$ and $\beta_b = -1$, and there is no constant as it is not identified in a partial likelihood setting.

As pointed out in Ng *et al.* (2004), convergence and thus computing time of the EM algorithm is sensitive to the choice of starting values. We set them at 1 for v_1^j and v_2^j, 1 for α_1 and α_2, and the estimates of a standard Cox model for the coefficients. All the fitting programs are designed under the R 2.0.1 software (Team, 2005), the inference using penalized partial likelihood calling the function 'coxph' of the package 'survival'. The source codes are available upon request.

4.2 Results

A preliminary comment is that the EMPL and accelerated EM does converge for all samples. Furthermore, both provide the same estimated $(\alpha_1, \alpha_2, \beta)$ up to a thousandth. Indeed, the starting values are close enough to the true value and both procedures are approximating the maximum likelihood. Both algorithm have the same accuracy and we compare them on the basis of the computational cost.

Computing time differ widely even when the estimators are applied to samples with the same clustering. In most cases, the ratio of the longest computing time over the shortest for a given sample design is equal to 100 for the EMPL and 500 for the accelerated EM. Table 1 reports the 25th, 50th and 75th centiles of the computing times. Values are printed in italic when the accelerated EM is quicker than the EMPL.

The EMPL algorithm is fast, with computing time quantiles generally far under their counterparts for the EM algorithm. We notice it especially when there are 20 durations or less per group: in this case, the third quartile for the EMPL computing time is under the first quartile for the EM computing time. This result of the EMPL algorithm being quicker than the EM does not hold in presence of a single spell per subgroup with 20 spells or more per group.

Group and subgroup sizes have a mixed impact on the computing times of the EMPL algorithm. First, these decrease monotonically with subgroup size. However, computing times are not much influenced by the number of spells per group. Due to this, a model with groups containing 10 spells and subgroups of 5 observations will be estimated much more quickly than a model where groups contain 100 spells and subgroups 2 observations. By contrast, computing times for the accelerated EM algorithm have an inverted $\cup$ profile with a maximum around 4 spells per subgroup. The more subgroups there are, the more v_2^j have to be evaluated at each iteration, which raises computing times until a threshold. But more v_2^j also imply more information, which speeds up convergence after the threshold.

5 Ratification of the International Labour Organization (ILO) conventions

As an example with real, we apply the EMPL algorithm on real data reporting the timing of the ratification of ILO conventions. The dataset is presented and analysed in Boockmann (2001). The survival time represents the time between the adoption of ILO conventions and their ratification by developing countries over the period 1975-1995. They comprise 80 countries and 29 conventions for a total of 228 ratifications. Horny *et al.* (2005) provide Bayesian estimates of a Cox model with 2 non-nested random effects. The hazard function is written as: $\lambda_{i2}^j(t) = v_1^j v_2^j \lambda_0(t) \exp[X_i(t)\beta]$, where v_1^j is a convention effect and v_2^j a country effect. They estimate the model using a Bayesian approach based on partial likelihood.

Table 1: Computing time in seconds (Var(v_1^j)=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

Note: Italic entries indicate that accelerated EM is quicker than the EMPL.

We estimate this model with gamma heterogeneity, using both the accelerated EM and the EMPL algorithm. Convergence was not achieved by the accelerated EM algorithm. Indeed, the variances of both mixing distributions increase progressively after each iteration of the sub-EM algorithms optimising function Q_1, characterized in equation (10). Finally, the algorithm collapses in the M step while returning a coefficient tending to $-\infty$. The whole process took 8 hours and a half. This convergence problem can be explained by the weak amount of information available in the data. The sample is heavily censored with only 5% of the durations ended by an observed transition. The sub-EM algorithm involves approximation of the cumulative hazard by the semi-parametric estimator of Nelson (1969), which is typically inefficient with so few transitions. Conversely, the EMPL does converge. Computation took 66 seconds and the results, reported in Appendix D, are close to the Bayesian estimates obtained in Horny *et al.* (2005).

6 Conclusion

This paper presents an MPH model with K random effects. We show how to estimate it using an EM algorithm in a general framework where we only require the mixing distributions to have a known Laplace transform. We then recover the information matrix using the approach of Louis (1982). The methodology is semi-parametric as it relies on partial likelihood. The provided EMPL algorithm is not only fast, but also simple and stable. We assess its properties in a Monte Carlo study and in an illustrative application on the timing of ratification of ILO conventions.

The computing times are notably reduced by using the EMPL algorithm, thus not asking for speeding-up routines as the ones described in Meilijson (1989) to be implemented in moderate size samples. The case of a random effect defining groups of one spell is an exception and speed depends on the size of the groups defined by the other frailties. Furthermore, the EMPL does converge in some settings where the EM does not.

We assume that all the random effects are continuous. This assumption can be questionable when the population at hand is divided only in a few groups at the more aggregated levels. With 2 levels of heterogeneity, one possibility is to switch to a fixed effects approach, stratifying the baseline hasard at the broader levels and defining a frailty at the finest one, as proposed by Xue and Brookmeyer (1996).

Acknowledgements: I thank Gerard van den Berg and François Laisney for detailed comments on previous versions of this paper. I am grateful to to Bernhard Boockmann for kindly providing the data on the ratification of ILO conventions. I also thank two anonymous referees for careful comments that were in helpfull in revising the paper. I would like to thank Narayan Sastry for explanations about his algorithm, as well as participants at the seminar of the

Tinbergen Institute, Amsterdam, March 2005, and at the Scottish Economic Society Annual Conference, Perth, April 2006.

A Derivation of $\mathbf{E}[v_k^j|T,d]$

In this appendix, we derive the conditional expectations of v_k^j for a given (j,k). Let us denote by $\mathcal{C}_{jk}$ the set of the observations in cluster j defined at stratification level k. We have:

$$f(v_1^j,\ldots,v_k^j,T_{jk},\delta_{jk}) = \prod_{i\in\mathcal{C}_{jk}} \left[\left(\prod_{k=1}^{K} v_k^j\right)\lambda_0(t_i)\lambda_1\left[X_i(t_i)\right]\right]^{1-\delta_i} \exp\left[-\left(\prod_{k=1}^{K} v_k^j\right)\int_0^{t_i}\lambda_0(u)\lambda_1\left[X_i(u)\right]du\right]\prod_{k=1}^{K} h_k(v_k^j;\alpha_k). \quad (20)$$

Integrating over v_1^j, we obtain:

$$f(v_2^j,\ldots,v_k^j,T_{jk},\delta_{jk}) = \left(\prod_{k\in\mathcal{C}_{jk}} \left[\left(\prod_{k=2}^{K} v_k^j\right)\lambda_0(t_i)\lambda_1\left[X_i(t_i)\right]\right]^{1-\delta_i}\prod_{k=2}^{K} h_k(v_k^j;\alpha_k)\right) \int_{\mathcal{V}_1} v^{l_{ik1}}\exp\left[-\left(\prod_{k=1}^{K} v\right)\sum_{k\in\mathcal{C}_{jk}}\int_0^{t_i}\lambda_0(u)\lambda_1\left[X_i(u)\right]du\right]h_1(v;\alpha_1)dv, \quad (21)$$

where l_{ik1} is the number of transitions observed in the group obtained as the intersection of the two clusters defined at stratification levels k and 1 and containing unit i. We can rewrite this expression as:

$$f(v_2^j,\ldots,v_k^j,T_{jk},\delta_{jk}) = \left(\prod_{k\in\mathcal{C}_{jk}} \left[\left(\prod_{k=2}^{K} v_k^j\right)\lambda_0(t_i)\lambda_1\left[X_i(t_i)\right]\right]^{1-\delta_i}\prod_{k=2}^{K} h_k(v_k^j;\alpha_k)\right) (-1)^{l_{ik1}}\mathcal{L}_1^{(l_{ik1})}\left[\left(\prod_{k=2}^{K} v_k^j\right)\sum_{k\in\mathcal{C}_{jk}}\int_0^{t_i}\lambda_0(u)\lambda_1\left[X_i(u)\right]du\right], \quad (22)$$

where $\mathcal{L}_1^{(l_{ik1})}$ is the l_{ik1}-th derivative of the Laplace transform of a non-negative random variable v_1, defined as:

$$\mathcal{L}_1(s) = \int_{\mathcal{V}}\exp(-sv)dH(v), \quad (23)$$

where $s \geq 0$. Integrating over v_2^j and omitting the argument of the Laplace transform, we obtain:

$$\begin{aligned} f(v_3^j, \ldots, v_k^j, T_{jk}, \delta_{jk}) = & \left(\prod_{k \in \mathcal{C}_{jk}} \left[\left(\prod_{k=3}^{K} v_k^j \right) \lambda_0(t_i) \lambda_1 \left[X_i(t_i) \right] \right]^{1-\delta_i} \prod_{k=3}^{K} h_k(v_k^j; \alpha_k) \right) \\ & (-1)^{l_{ik1}} \int_{\mathcal{V}_2} v^{l_{i2j}} \mathcal{L}_1^{(l_{ik1})} h_2(v, \alpha_2) dv \\ = & \left(\prod_{k \in \mathcal{C}_{jk}} \left[\left(\prod_{k=3}^{K} v_k^j \right) \lambda_0(t_i) \lambda_1 \left[X_i(t_i) \right] \right]^{1-\delta_i} \prod_{k=3}^{K} h_k(v_k^j; \alpha_k) \right) \\ & (-1)^{l_{ik1}} \mathrm{E}_2 \left[(v_2^j)^{l_{ik2}} \mathcal{L}_1^{(l_{ik1})} \right]. \end{aligned} \tag{24}$$

By further integrations, we can show that:

$$\begin{aligned} f(v_k^j, T_{jk}, \delta_{jk}) = & \left(\prod_{k \in \mathcal{C}_{jk}} \left[v_k^j \lambda_0(t_i) \lambda_1 \left[X_i(t_i) \right] \right]^{1-\delta_i} h_k(v_k^j; \alpha_k) \right) (-1)^{l_{ik1}} \\ & \mathrm{E}_K \left[(v_K^j)^{l_{ikK}} \ldots \mathrm{E}_{(k+1)} \left[(v_{(k+1)}^j)^{l_{ik(k+1)}} \mathrm{E}_{(k-1)} \left[(v_{(k-1)}^j)^{l_{ik(k-1)}} \ldots \mathrm{E}_2 \left[(v_2^j)^{l_{ik2}} \mathcal{L}_1^{(l_{ik1})} \right] \right] \right] \right]. \end{aligned} \tag{25}$$

To make equations shorter, let us denote:

$$\begin{aligned} \xi_{(-k)} = \mathrm{E}_K \Big[(v_k^j)^{l_{ikK}} \ldots \mathrm{E}_{(k+1)} \Big[(v_{(k+1)}^j)^{l_{ik(k+1)}} \mathrm{E}_{(k-1)} \Big[(v_{(k-1)}^j)^{l_{ik(k-1)}} \ldots \\ \mathrm{E}_2 \Big[(v_2^j)^{l_{ik2}} \mathcal{L}_1^{(l_{ik1})} \Big] \Big] \Big] \Big]. \end{aligned} \tag{26}$$

Thus:

$$\begin{aligned} \mathrm{E} \left[v_k^j | T, d \right] &= \frac{\int_{\mathcal{V}_k} v f(v, T, d) dv}{\int_{\mathcal{V}_k} f(v, T, d) dv} \\ &= \frac{\int_{\mathcal{V}_k} v^{1+l_{ikk}} \xi_{(-k)} h_k(v, \alpha_k) dv}{\int_{\mathcal{V}_k} v^{l_{ikk}} \xi_{(-k)} h_k(v, \alpha_k) dv} \\ &= \frac{\mathrm{E}_k \left[(v_k^j)^{1+l_{ikk}} \xi_{(-k)} \right]}{\mathrm{E}_k \left[(v_k^j)^{l_{ikk}} \xi_{(-k)} \right]}. \end{aligned} \tag{27}$$

B Evaluation of the Information Matrix

Using equations from (6) to (8), the gradient vector is made of:

$$\frac{\partial \ln L}{\partial \alpha_k} = \sum_{k=1}^{N} \frac{\partial \ln h_k}{\partial \alpha_k}(v_k^j; \alpha_k), \tag{28}$$

$$\frac{\partial \ln L}{\partial \beta} = \sum_{k=1}^{N} \left[\delta_i \frac{\partial \ln \lambda_1}{\partial \beta}\left[X_i(u);\beta\right] - \left(\prod_{k=1}^{K} v_k^j\right) \int_0^{t_i} \lambda_0(u) \frac{\partial \lambda_1}{\partial \beta}\left[X_i(u);\beta\right] \lambda_1\left[X_i(u);\beta\right] du \right]. \tag{29}$$

The last term in (19) is the variance of the score, conditional of the observed data. To be computed, it asks for the mixing distributions to be specified and we present in detail the case of gamma mixing distributions in Appendix C.

The first member in (19) is the expected information matrix, conditional on the full data. The Hessian is:

$$H(\alpha,\beta) = \begin{pmatrix} H_{11}(\alpha) & 0 \\ 0 & H_{22}(\beta) \end{pmatrix}, \tag{30}$$

where:

$$H_{11}(\alpha) = \sum_{i=1}^{N} \frac{\partial^2 \ln h_k}{\partial \alpha_k \partial \alpha_k'}(v_k^j, \alpha_k), \tag{31}$$

$$H_{22}(\beta) = \sum_{i=1}^{N} \left(\prod_{k=1}^{K} v_k^j\right) \int_0^{t_i} \lambda_0(u) \frac{\partial^2 \lambda_1}{\partial \beta \partial \beta'}\left[X_i(u);\beta\right] \lambda_1\left[X_i(u);\beta\right] du. \tag{32}$$

The expected information matrix, conditional on the full data, can be evaluated once EM algorithm converged. We have thus:

$$E\left[-H(\alpha,\beta)|T,d,v_1,\ldots,v_K\right] = -H\left(\alpha^{(*)},\beta^{(*)}\right), \tag{33}$$

where $\alpha^{(*)}$ and $\beta^{(*)}$ are the estimates at the last iteration of the algorithm.

C Computation of the Information Matrix with Gamma Heterogeneity

Evaluating the information matrix (19) requires the expectation of the second derivative matrix $H(\alpha,\beta)$ and the conditional variance of the gradient vector

$S(\alpha, \beta)$. The Hessian is presented in equation (30) and the block $H_{11}(\alpha)$ is a diagonal matrix whose element (i,i) is, in case of gamma heterogeneity:

$$H_{11}(i,i) = J\left(\frac{1}{\alpha_k} - \psi'(\alpha_k)\right), \tag{34}$$

where ψ is the digamma function. The bloc $H_{22}(\beta)$, characterised in (32), does not depend on the choice of the unobserved heterogeneity distribution.

The score vector in equations (28) and (29) becomes in presence of gamma mixing distribution:

$$\frac{\partial \ln L}{\partial \alpha_k} = \sum_{k=1}^{K}\left[\ln \alpha_k + 1 + \psi(\alpha_k) - v_k^j + \ln v_k^j\right], \tag{35}$$

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{I}\left[\delta_i X_i(t_i) - \left(\prod_{k=1}^{K} v_k^j\right)\int_0^{t_i} \lambda_0(u) X_i(t_i)\exp\left[X_i(u)\beta\right]du\right]. \tag{36}$$

The variance of the gradient vector conditional on the observed data requires the computation of $\mathrm{Var}\left(v_k^j|T,d\right)$, $\mathrm{Var}\left(\ln v_k^j|T,d\right)$, $\mathrm{Var}\left[\left(\prod_{k=1}^{K} v_k^j\right)|T,d\right]$ and $\mathrm{Cov}\left(\frac{\partial \ln L}{\partial \beta}, \frac{\partial \ln L}{\partial \alpha_k}|T,d\right)$ to be computed. These are evaluated from density:

$$\begin{aligned} f(v_k^j|T,d) &= \frac{f(v_k^j, T, d)}{f(T,d)} \\ &= \frac{f(v_k^j, T, d)}{\int_{\mathcal{V}_k} f(v,T,d), dv}. \end{aligned} \tag{37}$$

The numerator is equation (20), which becomes in presence of gamma heterogeneity:

$$\begin{aligned} f(v_1^j, \ldots, v_K^j, T, d) = &\left[\prod_{k\in\mathcal{C}_{jk}}\left(\prod_{m=1}^{K} v_l^j\right)\lambda_0(t_i)\lambda_1[X_i(t_i)]^{\delta_i}\right] \\ &\exp\left[-\left(\prod_{m=1}^{K} v_m^j\right)\sum_{k\in\mathcal{C}_{jk}}\int_0^{t_i}\lambda_0(u)\lambda_1[X_i(u)]du\right] \\ &\prod_{m=1}^{K}\frac{\alpha_m^{\alpha_m}}{\Gamma(\alpha_m)}(v_m^j)^{\alpha_m-1}\exp(-\alpha_m v_m^j) \\ =&K_k^j\left(\prod_{m=1}^{K}(v_l^j)^{l_{mjk}+\alpha_m-1}\right)\exp\left[-\sum_{m=1}^{K}\alpha_m v_m^j\right. \\ &\left.-\left(\prod_{m=1}^{K} v_m^j\right)\sum_{k\in\mathcal{C}_{jk}}\int_0^{t_i}\lambda_0(u)\lambda_1[X_i(u)]du\right], \end{aligned} \tag{38}$$

where:

$$K_k^j = \prod_{m=1}^{K} \frac{\alpha_m^{\alpha_m}}{\Gamma(\alpha_m)} \prod_{i \in \mathcal{C}_{jk}} \lambda_0(t_i)\lambda_1[X_i(t_i)]^{\delta_i}. \tag{39}$$

We deduce:

$$\begin{aligned}
f(v_k^j, T, d) =& \int_{\mathcal{V}_1} \dots \int_{\mathcal{V}_{k-1}} \int_{\mathcal{V}_{k+1}} \dots \int_{\mathcal{V}_K} f(v_1^j, \dots, v_K^j, T, d) dv_1^j \dots dv_K^j \\
=& K_k^j (v_k^j)^{l_{ijk}+\alpha_k-1} \exp(-\alpha_k v_k^j) \prod_{n \neq k} \int_{\mathcal{V}_n} (v_n^j)^{l_{njk}+\alpha_n-1} \\
& \exp\left[-\alpha_n v_n^j - \left(\prod_{m=1}^{K} v_m^j\right) \sum_{k \in \mathcal{C}_{jk}} \int_0^{t_i} \lambda_0(u)\lambda_1[X_i(u)]du\right] dv_n^j. \\
=& K_k^j (v_k^j)^{l_{ijk}+\alpha_k-1} \exp(-\alpha_k v_k^j) \prod_{n \neq k} \left[\alpha_n + \left(\prod_{m=1}^{K} v_m^j\right)\right. \\
& \left.\sum_{k \in \mathcal{C}_{jk}} \int_0^{t_i} \lambda_0(u)\lambda_1[X_i(u)]du\right]^{-l_{njk}-\alpha_n} \Gamma(l_{njk}+\alpha_n-1). \qquad (40)
\end{aligned}$$

Hence:

$$f(v_k^j | T, d) = \frac{(v_k^j)^{p_{ijk}} \exp(-\alpha_k v_k^j) \prod_{n \neq k} \left[\alpha_n + \left(\prod_{m=1}^K v_m^j\right) \sum_{i \in \mathcal{C}_{jk}} \Lambda_i\right]^{-1-p_{ijn}}}{\int_{\mathcal{V}_k} v^{p_{ijk}} \exp(-\alpha_k v) \prod_{n \neq k} \left[\alpha_n + \left(\prod_{m=1}^K v_m^j\right) \sum_{i \in \mathcal{C}_{jk}} \Lambda_i\right]^{-1-p_{ijn}} dv}, \tag{41}$$

where $p_{ijk} = l_{ijk} + \alpha_k - 1$ and $\Lambda_i = \int_0^{t_i} \lambda_0(u)\lambda_1[X_i(u)]du$. The quantities $\mathrm{Var}\Big(v_k^j | T, d\Big)$ and $\mathrm{Var}\Big(\ln v_k^j | T, d\Big)$ are computed with the first and second order moments, approximated by Monte Carlo simulations from (41). We deduce $\mathrm{Var}\Big[\Big(\prod_{k=1}^K v_k^j\Big) | T, d\Big]$ and $\mathrm{Cov}\Big(\frac{\partial \ln L}{\partial \beta}, \frac{\partial \ln L}{\partial \alpha_k} | T, d\Big)$:

$$\begin{aligned}
\mathrm{Var}\left[\left(\prod_{k=1}^K v_k^j\right) | T, d\right] =& \quad \mathrm{E}\left[\left(\prod_{k=1}^K (v_k^j)^2\right) | T, d\right] - \mathrm{E}\left[\left(\prod_{k=1}^K (v_k^j)\right) | T, d\right]^2 \\
=& \left\{\prod_{k=1}^K \mathrm{E}\left[(v_k^j)^2 | T, d\right]\right\} - \prod_{k=1}^K \left\{\left[\mathrm{E}\left(v_k^j | T, d\right)\right]^2\right\}
\end{aligned} \tag{42}$$

Evaluating $\mathrm{Cov}\left(\frac{\partial \ln L}{\partial \beta}, \frac{\partial \ln L}{\partial \alpha_k}\middle| T, d\right)$ asks for $\mathrm{Cov}\left(\prod_{k=1}^K v_k^j, v_k^j - \ln v_k^j | T, d\right)$, that is:

$$
\begin{aligned}
&\operatorname{Cov}\left(\prod_{k=1}^{K} v_k^j, v_k^j - \ln v_k^j | T, d\right) \\
&\quad = \mathrm{E}\left[\left(\prod_{k=1}^{K} v_k^j\right)\left(v_k^j - \ln v_k^j\right) | T, d\right] - \mathrm{E}\left[\prod_{k=1}^{K} v_k^j | T, d\right] \mathrm{E}\left[v_k^j - \ln v_k^j | T, d\right] \\
&\quad = \mathrm{E}\left[\prod_{k' \neq k} v_{k'}^j | T, d\right] \mathrm{E}\left[\left(v_k^j\right)^2 - v_k^j \ln v_k^j | T, d\right] \\
&\qquad - \mathrm{E}\left[\prod_{k' \neq k} v_{k'}^j | T, d\right] \left(\mathrm{E}\left[v_k^j | T, d\right] \mathrm{E}\left[v_k^j - \ln v_k^j | T, d\right]\right) \\
&\quad = \left(\prod_{k' \neq k} \mathrm{E}\left[v_{k'}^j | T, d\right]\right) \left[\operatorname{Var}(v_k^j | T, d) - \operatorname{Cov}(v_k^j, \ln v_k^j | T, d)\right].
\end{aligned}
\tag{43}
$$

D Results for the ratification of ILO conventions

Table 2 reports the Bayesian estimates provided in Horny *et al.* (2005) and the results of the EMPL.

Table 2: Estimates of the β parameters

Variable	Bayes		EMPL	
	Coef.	S.d	Coef.	S.d
Cost				
Real GDP per capita[a]	**3.81**	1.40	**3.03**	1.39
Real GDP per capita, squared	**-3.19**	1.51	-2.41	1.51
No explicit update	**1.39**	0.27	**1.26**	0.37
Own past ratification if explicit update	**1.62**	0.36	**1.52**	0.38
Population[b]	-0.02	0.05	-0.03	0.05
Internal pressure				
Democracy	**0.34**	0.15	0.29	0.15
Left majority	**-0.69**	0.31	**-0.62**	0.31
Vote against convention:				
Government	-0.22	0.23	-0.08	0.24
Employers	**0.38**	0.20	0.28	0.21
External pressure				
Development aid[c]	**-7.65**	2.05	**-8.56**	2.16
Worldbank loans[c]	2.00	1.55	**3.15**	1.57
IMF credits[c]	3.96	1.95	3.68	1.98
Exports[c]	-0.79	1.30	0.27	1.06
Exports into industrialized countries[c]	-0.18	3.66	-2.54	3.80
Exports into industrialized countries[c] (non oil exporting countries)	-0.77	3.48	0.52	3.71
Non oil exporting country	0.25	0.68	-0.01	0.71

Note: Bold entries are significant at the 5% level. *a.* 1985 international prices, in \$10 000. *b.* hundred milions. *c.* percent of GDP.

References

Abbring, J. and G. Van den Berg (2006): "The unobserved heterogeneity distribution in duration analysis," Discussion Paper 059/3, Tinbergen Institute.

Bolstad, W. and 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.

Boockmann, B. (2001): "The ratification of ILO conventions: a hazard rate analysis," *Economics and Politics*, 13, 281–309.

Clayton, D. and J. Cuzick (1985): "Multivariate generalizations of the proportional hazards model," *Journal of the Royal Statistical Society. Series A*, 148, 82–108.

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

de Montricher, G., R. Tapia and J. Thompson (1975): "Nonparametric maximum likelihood estimation of probability densities by penalty function methods," *The Annals of Statistics*, 3, 1329–1348.

Feller, W. (1971): *An introduction to probability theory and its applications*, vol. 2. Wiley, New-York, 2 edn.

Gill, R. (1985): "Discussion of the paper by D. Clayton and J. Cuzick," *Journal of the Royal Statistical Society. Series A*, 148, 108–109.

Gouriéroux, C. and I. Peaucelle (1990): "Hétérogénéité I. Étude des biais d'estimation dans le cas linéaire," *Annales d'Économie et de Statistique*, 17, 163–184.

Horny, G., B. Boockmann, D. Djurdjevic and 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.

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

Louis, T. (1982): "Finding the observed Information matrix when using the EM algorithm," *Journal of the Royal Statistical Society. Series B*, 44, 226–233.

Manda, S. and 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.

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.

Moulton, B. (1986): "Random group effects and the precision of regression estimates," *Journal of Econometrics*, 32, 385–397.

Nelson, W. (1969): "Hazard plotting for incomplete failure time data," *Journal of Quality Technology*, 1, 27–52.

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

Pakes, A. (1983): "On group effects and errors in variables in aggregation," *The Review of Economics and Statistics*, 65, 168–173.

Parner, E. (1997): "Inference in semiparametric frailty models," Ph.D. thesis, University of Aarhus.

Rondeau, V., D. Commenges and P. Joly (2003): "Maximum penalized likelihood estimation in a gamma-frailty model," *Lifetime Data Analysis*, 9, 139–153.

Sastry, N. (1997): "A nested frailty model for survival data, with an application to thestudy of child survival in northeast Brazil," *Journal of the American Statistical Association*, 92, 426–435.

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, T., P. Grambsch and V. Pankratz (2003): "Penalized survival models and frailty," *Journal of Computational and Graphical Statistics*, 12, 156–175.

Van den Berg, G. (2001): "Duration models: specification, identification and multiple durations," in *Handbook of Econometrics*, vol. 5, chap. 55, pp. 3381–3463. J. J. Heckman and E. Leamer (eds.), Elsevier, Amsterdam.

Vu, H. (2003): "Parametric and semiparametric conditional shared gamma frailty models with events before study entry," *Communications in Statistics - Simulation and Computation*, 32(4), 1223–1248.

——— (2004): "Estimation in semiparametric conditional shared frailty models with events before study entry," *Comutational Statistics and Data Analysis*, 45, 621–637.

Vu, H. and M. Knuiman (2002a): "Estimation in semiparametric marginal shared gamma frailty models," *Australian and New Zealand Journal of Statistics*, 44(4), 489–501.

——— (2002b): "A hybrid ML-EM algorithm for calculation of maximum likelihood estimates in semiparametric shared frailty model," *Computational Statistics and Data Analysis*, 40, 173–187.

Vu, H., M. Segal, M. Knuiman and I. James (2001): "Asymptotic and small sample statistical properties of random frailty variance estimates for shared gamma frailty models," *Communications in Statistics - Simulation and computation*, 30(3), 581–595.

Wei, G. and 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. and R. Brookmeyer (1996): "Bivariate frailty model for the analysis of multivariate survival time," *Lifetime Data Analysis*, 2, 277–289.

Yau, K. (2001): "Multilevel models for survival analysis with random effects," *Biometrics*, 57, 96–102.

Documents de Travail

229. Ph. Aghion, Ph. Askenazy, R. Bourlès, G. Cette and N. Dromel, "Education, Market Rigidities and Growth," January 2009

230. G. Cette and M. de Jong, "The Rocky Ride of Break-even-inflation rates," January 2009

231. E. Gautier and H. Le Bihan, "Time-varying *(S,s)* band models: empirical properties and interpretation," January 2009

232. K. Barhoumi, O. Darné and L. Ferrara, "Are disaggregate data useful for factor analysis in forecasting French GDP ? " February 2009

233. R. Cooper, H. Kempf and D. Peled, "Monetary rules and the spillover of regional fiscal policies in a federation" February 2009

234. C. Jardet, A. Monfort, and F. Pegoraro, "No-arbitrage Near-Cointegrated VAR(p) Term Structure Models, Term Premia and GDP Growth" June 2009

235. C. Jardet, A. Monfort, and F. Pegoraro, "New Information Response Functions," June 2009

236. S. Adjemian, C. Cahn, A. Devulder et N. Maggiar, « Variantes en Univers Incertain », Juin 2009

237. P-A. Chevalier, R. Lecat et N. Oulton, « Convergence de la productivité des entreprises, mondialisation, technologies de l'information et conccurence », Juin 2009

238. S. Avouyi-Dovi, M. Bardos, C. Jardet, L. Kendaoui and J. Moquet, "Macro stress testing with a macroeconomic credit risk model: Application to the French manufacturing sector," June 2009

239. O. Darné and L. Ferrara, "Identification of slowdowns and accelerations for the euro area economy," June 2009

240. H. Kempf and G. Rota Graziosi, "Leadership in Public Good Provision: a Timing Game Perspective", July 2009

241. S. Avouyi-Dovi et J.-G. Sahuc, « Comportement du banquier central en environnement incertain », Juillet 2009

242. J. Coffinet, S. Lin and C. Martin, "Stress testing French banks' income subcomponents," August 2009

243. P. Fève, J. Matheron and J.-G. Sahuc, "Inflation Target Shocks and Monetary Policy Inertia in the Euro Area", August 2009

244. P. Fève, J. Matheron et J.-G. Sahuc, « La TVA sociale : bonne ou mauvaise idée? », Août 2009

245. P. Fève, J. Matheron and J.-G. Sahuc, " Minimum Distance Estimation and Testing of DSGE Models from Structural VARs ", August 2009

246. P. Fève, J. Matheron et J.-G. Sahuc, « Une estimation de la cible implicite d'inflation dans la zone euro », Août 2009

247. P. Fève, J. Matheron et J.-G. Sahuc, « Désinflation et chômage dans la zone euro : une analyse à l'aide d'un modèle VAR structurel », Août 2009

248. G. Horny, " Inference in Mixed Proportional Hazard Models with K Random Effects ", August 2009

Pour accéder à la liste complète des Documents de Travail publiés par la Banque de France veuillez consulter le site :
http://www.banque-france.fr/fr/publications/documents_de_travail/documents_de_travail_09.htm

For a complete list of Working Papers published by the Banque de France, please visit the website:
http://www.banque-france.fr/fr/publications/documents_de_travail/documents_de_travail_09.htm

Pour tous commentaires ou demandes sur les Documents de Travail, contacter la bibliothèque de la Direction Générale des Études et des Relations Internationales à l'adresse suivante :

For any comment or enquiries on the Working Papers, contact the library of the Directorate General Economics and International Relations at the following address :

BANQUE DE FRANCE
49- 1404 Labolog
75049 Paris Cedex 01
tél : 0033 (0)1 42 92 49 55 ou 62 65 ou 48 90 ou 69 81
fax :0033 (0)1 42 92 62 92
email : thierry.demoulin@banque-france.fr
jeannine.agoutin@banque-france.fr
veronique.jan-antuoro@banque-france.fr
nathalie.bataille-salle@banque-france.f