DOCUMENT
DE TRAVAIL
N° 249

BAYESIAN ESTIMATION OF COX MODELS WITH NON-NESTED RANDOM EFFECTS: AN APPLICATION TO THE RATIFICATION OF ILO CONVENTIONS BY DEVELOPING COUNTRIES

Bernhard Boockmann, Dragana Djurdjevic, Guillaume Horny
and François Laisney

August 2009

BANQUE DE FRANCE
EUROSYSTÈME

DIRECTION GÉNÉRALE DES ÉTUDES ET DES RELATIONS INTERNATIONALES

DIRECTION GÉNÉRALE DES ÉTUDES ET DES RELATIONS INTERNATIONALES

BAYESIAN ESTIMATION OF COX MODELS WITH NON-NESTED RANDOM EFFECTS: AN APPLICATION TO THE RATIFICATION OF ILO CONVENTIONS BY DEVELOPING COUNTRIES

Bernhard Boockmann, Dragana Djurdjevic, Guillaume Horny
and François Laisney

August 2009

Bayesian estimation of Cox models with non-nested random effects: an application to the ratification of ILO conventions by developing countries

Guillaume Horny[1] , Bernhard Boockmann[2], Dragana Djurdjevic[3] and François Laisney[4]

Forthcoming in the Annales d'Economie et de Statistique

[1]Banque de France, DGEI-DEMS-SAMIC, 31 rue Croix des Petits Champs, 75 049 Paris Cedex 01, France and Université Catholique de Louvain, Belgique. guillaume.horny@banque-france.fr

[2]Institute for Applied Economic Research (IAW), Ob dem Himmelreich 1. D-72074 Tübingen, Germany. bernhard.boockmann@iaw.edu

[3]Wüest & Partner, rue de la rôtisserie 1, CH-1204 Genève, Switzerland. djurdjevic@wuestundpartner.com

[4]BETA, University Louis Pasteur (Strasbourg I), 61 avenue de la Forêt Noire, 67085 Strasbourg Cedex, France, and Centre for European Economic Research (ZEW), PO Box 10 34 43. D-68034 Mannheim, Germany. fla@cournot.u-strasbg.fr, laisney@zew.de

Abstract

We use a multivariate hazard model to analyse the ratification behaviour of ILO conventions by developing countries. The model accounts for two random effects: one at the country level, the other at the convention level. After investigating identification, we use a semi-parametric Bayesian approach based on the partial likelihood. We find diverging results between Bayesian and frequentist estimates concerning the importance of the two unobserved heterogeneities.
Keywords: Gibbs sampling, partial likelihood, frailties, duration analysis.
JEL Classification: C11, C14, C15, C41, D78, J80, O19

Résumé

Nous utilisons un modèle multivarié de hasards pour analyser le comportement de ratification des pays en voie de développement. Le modèle comprend deux effets aléatoires: l'un au niveau du pays, l'autre au niveau de la convention. Après avoir établi l'identification, nous utilisons une approche bayésienne semi-paramétrique reposant sur la vraisemblance partielle. Les estimations bayésiennes et fréquentistes fournissent des résultats différents quant à l'importance des deux hétérogénéités non observées.
Mots-clés: échantillonnage de gibbs, vraisemblance partielle, effets aléatoires, modèles de durées.
Classification JEL: C11, C14, C15, C41, D78, J80, O19

Non-technical summary

In this paper, we focus on models describing the transitions from one state to another over time. The individuals and destination states are assumed to have unobserved characteristics. Fixed and random effects models have been used in similar contexts. The advantage of the fixed effects approach is that it is easy to implement and require weaker assumptions than the random effects approach. A general way to deal with fixed effects is provided in KALBFLEISCH and PRENTICE [1980]; YAMAGUCHI [1986]; RIDDER and TUNALI [1999]. As explained in THERNEAU and GRAMBSCH [2000], the disadvantages of their approach are that the magnitude of the frailties cannot be estimated directly and the precision of the estimates decreases with the number of strata.

In our application, we are dealing with a case where the number of grouped observations in the data is relatively large and, therefore, consider modelling by random effects only. We study the ratification of International Labour Organization (ILO) conventions by developing countries. There are currently 187 conventions in many different subject areas which may be voluntarily ratified by member states. We assume that ratification behaviour is influenced by unobserved characteristics both of countries and conventions. The presence of the convention effects stems from the fact that some conventions may be more easily ratified than others. For instance, conventions grant countries different degrees of flexibility or differ in complexity.

Random effects allow deeper understanding of the unobserved heterogeneity, as the estimation of the parameters of the frailty distributions leads to the possibility of comparing the importance of the various group effects. This comes at a price: the need to specify distributions for the frailties.

In previous work on the ILO case, BOOCKMANN [2001] restricts attention to a country fixed effect, and applies RIDDER AND TUNALI 's [1999] estimator. He provides some justification for the choice of random effects for the sample of developing countries. HORNY [2001] extends the approach to two random effects. Using the same dataset as BOOCKMANN [2001] and the EM algorithm, HORNY [2001] meet serious convergence problems. They conclude that the EM algorithm, while working satisfactorily on simulated data, is not well suited for this dataset of ILO conventions. HORNY [2009] presents a version of the EM algorithm that is both faster and more stable than previous ones, and achieves convergence with this dataset.

However, THERNEAU et al. [2003] produce evidence pointing to underestimation of the importance of unobserved heterogeneity in such EM estimation. RODRIGUEZ and GOLDMAN [2001] compare several estimation procedures, and report the underestimation of the importance of both fixed and random effects except for maximum likelihood and Bayesian estimators.

In this paper, we apply a semi-parametric Bayesian approach to a Cox model with two random effects. We also show that the model with two frailties is identified if each realization of the random effects is shared by at least two observations. To achieve this result, we require weaker assumptions than the bulk of the literature.

Résumé non-technique

Nous nous intéressons dans ce papier aux modèles décrivant les transitions entre deux états survenant au fil du temps. Les individus et destinations sont supposés dotés de caractéristiques inobservées. Des modèles à effets fixes et aléatoires ont été utilisés dans des contextes similaires. Les effets fixes ont pour avantage de supposer des hypothèses moins restrictives que les effets aléatoires. Des manières générales de spécifier des effets fixes sont fournies dans Kalbfleisch et Pentice [1980], Yamaguchi [1986], Ridder et Tunali[1999]. Comme expliqué dans Therneau et Grambsch [2000], l'inconvénient de l'approche par effets fixes est que l'ampleur de l'hétérogénéité non observée ne peut pas être estimée directement et que la précision des estimations décroît avec le nombre de strates.

Dans notre application, nous nous intéressons à un cas où les observations sont réparties en différents groupes, et où le nombre d'observations au sein de chaque groupe est relativement important. C'est pourquoi nous considérons une modélisation comprenant des effets aléatoires. Nous étudions la ratification des conventions de l'Organisation Internationale du Travail (OIT) par les pays en voie de développement. Il existe actuellement 187 conventions sur différent domaines, qui peuvent être volontairement ratifiées par les membres de l'OIT. Nous supposons que le comportement de ratification est influencé par des caractéristiques inobservées des pays et des conventions. La présence d'un effet aléatoire au niveau des conventions correspond à l'idée que certaines conventions sont plus faciles à ratifier que d'autres. Par exemple, les conventions diffèrent dans leur complexité, ainsi que dans leur flexibilité.

Les effets aléatoires permettent une compréhension plus profonde de l'hétérogénéité non observée, car l'estimation des paramètres de la distribution des effets aléatoires permet de comparer l'importance des caractéristiques inobservées propres aux différents groupes. Ceci à toutefois une contrepartie: le besoin de spécifier des distributions pour les effets aléatoires.

Dans une étude antérieure sur l'OIT, Boockmann [2001] se concentre sur un effet fixe au niveau des pays et applique l'estimateur de Ridder et Tunali [1999]. Il justifie également le choix d'effets aléatoires pour l'échantillon de pays en voie de développement. Horny [2001] étend ce travail en utilisant deux effets aléatoires. En utilisant les mêmes données que Boockmann [2001] et un algorithme EM, Horny [2001] rencontre de sérieux problèmes de convergence. La conclusion en est que l'algorithme EM, bien que fonctionnant de manière satisfaisante sur données simulées, n'est pas approprié pour ces données de ratification des conventions de l'OIT. Horny [2009] présente une version de l'algorithme EM qui est à la fois plus rapide et plus stable que les versions antérieures, et obtient la convergence

des procédures numériques sur ces données.

Toutefois, Therneau et al. [2003] fournissent des éléments indiquant que l'algorithme EM tend à sous-estimer l'importance de l'hétérogénéité non observée. Rodriguez et Goldman [2001] comparent plusieurs méthodes d'estimations et concluent à une sous-estimation des effets fixes et aléatoires, hormis pour l'estimateur du maximum de vraisemblance et l'estimateur bayésien.

Dans ce papier, nous utilisons une approche bayésienne semi-paramétrique pour estimer un modèle de Cox à deux effets aléatoires. Nous montrons également que le modèle à deux effets aléatoires est identifié si les réalisations de chacun des effets aléatoires sont partagées par au moins deux observations. Ce résultat est obtenu avec des hypothèses moins contraignantes qu'une large partie de la littérature sur les modèles de durée.

1 Introduction

In this paper, we develop an approach for modelling unobserved heterogeneity in a duration model for the case where each cross-sectional unit (individual) faces a number of distinct but comparable risks. Both individuals and risks are assumed to have unobserved characteristics (sometimes referred to as frailties in the survival literature). Fixed and random effects models have been used in similar contexts. The advantage of the fixed effects approach is that it is easy to implement and that it does not require assumptions on the joint distribution of observed covariates and unobserved heterogeneity. A general way to deal with fixed effects is stratified partial likelihood (Kalbfleisch and Prentice, 1980; Yamaguchi, 1986; Ridder and Tunali, 1999). As explained in Therneau and Grambsch [2000], the disadvantages of stratified partial likelihood are that the magnitude of the frailties cannot be estimated directly and the precision of the estimates decreases with the number of strata.

In our application, we are dealing with a case where the number of strata is relatively large and, therefore, consider modelling by random effects only. We study the ratification of International Labour Organization (ILO) conventions by developing countries. There are currently 187 conventions in many different subject areas which may be voluntarily ratified by member states. We assume that ratification behaviour is influenced by unobserved characteristics both of countries and conventions. The presence of the convention effects stems from the fact that some conventions may be more easily ratified than others. For instance, conventions grant countries different degrees of flexibility or differ in complexity. As a consequence, clustering occurs at two different hierarchical levels.

Random effects allow deeper understanding of the unobserved heterogeneity, as the estimation of the parameters of the frailty distributions leads to the possibility of comparing the importance of the various group effects. This comes at a price: the need to specify distributions for the frailties, most often assumed to be independent of the covariates. In this vein, Guo and Rodriguez [1992] consider the proportional hazard model with one frailty term and, assuming a piecewise constant baseline hazard, focus on efficient parametric estimation of observed covariate and frailty effect parameters. Following Louis [1982], they use an accelerated Expectation-Maximization (EM) algorithm (Dempster et al., 1977). The case of two frailties is presented in the survey of Liang et al. [1995]. These can be nested, as in Gustafson [1997]; Sastry [1997] or Milcent [2003] where several individuals share a common frailty. Sastry [1997] extends this approach to the case of a Cox model (Cox, 1972) with two nested random effects. Bol-

STAD and MANDA [2001] propose a Bayesian approach to estimate SASTRY's [1997] model.

In previous work on the ILO case, BOOCKMANN [2001] restricts attention to a country fixed effect, and applies RIDDER AND TUNALI 's [1999] estimator. He provides some justification for the choice of random effects for the sample of developing countries. DJURDJEVIC [2000] extends the approach of GUO and RODRIGUEZ [1992] to the partial likelihood framework, relaxing the assumption of a piecewise constant baseline hazard. HORNY [2001] generalizes this to a Cox model with two nested random effects estimated through partial likelihood, extending also SASTRY [1997]. Using the same dataset as BOOCKMANN [2001] and the EM algorithm, both DJURDJEVIC [2000] and HORNY [2001] meet serious convergence problems. They conclude that the EM algorithm, while working satisfactorily on simulated data, is not well suited for this dataset of ILO conventions.[5] HORNY [2009] presents a version of the EM algorithm that is both faster and more stable than previous ones, and achieves convergence with this dataset.

However, THERNEAU et al. [2003] produce Monte Carlo evidence pointing to underestimation of the importance of unobserved heterogeneity in such EM estimation. Considering logistic regression, RODRIGUEZ and GOLDMAN [2001] compare the maximum likelihood, marginal quasi likelihood, penalized quasi likelihood and Bayesian approaches using clustered data. Except for maximum likelihood and Bayesian estimators, they report underestimation of the importance of both fixed and random effects. Although bias can be removed by bootstrapping, the procedure proves to be computationally more intensive than Markov chain Monte Carlo estimation and sometimes fails to converge.

In this paper, we investigate Bayesian estimation of this type of model. We apply a semi-parametric Bayesian approach, based on partial likelihood, to a Cox model with two non-nested random effects. We base inference on partial likelihood for two reasons. First, this flexible semi-parametric approach does not require specifying the baseline hazard. Second, for a model with time-varying covariates as in our application, LANCASTER [1990, pp. 272-274] shows that the joint distribution of durations and explanatory variables can be written as the product of the partial likelihood and a factor which is independent of the survival indicators under the exogeneity condition. "This is a most important conclusion because it shows that econometric

[5] LANCASTER [1990] explains why such problems can occur when the likelihood is bounded and when the variance of the random effect tends to zero. But here the problem goes in the opposite direction: the variance becomes large enough to create numerical difficulties. BOLSTAD and MANDA [2001] trace these back to inaccurate rounding in case of high variance.

inferences in models with time-varying covariates must generally be partial likelihood inferences, even when the hazard is specified fully parametrically" (Lancaster, 1990, p. 274).

We show that the model with two frailties is identified if each realization of the random effects is shared by at least two observations. To achieve this result, we do not need to assume that the frailties are independent of the covariates or that they have finite mean, in contrast with the bulk of the literature on duration models with random effects. To estimate the model, we use the approach described by Kalbfleisch [1978] justifying the partial likelihood approach in the Bayesian paradigm. After the full specification of the model, we estimate the parameters using Gibbs sampling.

The paper is organized in 5 sections. Section 2 describes the data. Section 3 presents the Cox model with two random effects (see Sastry, 1997) and two special cases: the Cox model with one random effect (see, for example, Lancaster, 1979) and the Cox model without random effect, thereafter referred as the standard Cox model. In Section 4, we discuss the choice of the prior distributions, the implementation of the partial likelihood, and estimation using Gibbs sampling. The results are presented in Section 5.

2 Data

The dataset used for the empirical analyses provides information on the ratification behaviour of ILO conventions. We focus on a group of 80 non-industrialized (or developing) countries (no OECD members). Our observation period is restricted to conventions that have been adopted over the period 1975-1995. This covers 29 conventions for a total of 228 ratifications. A description of this data can be found in Boockmann [2001].

Since we observe the date of adoption of any convention over the time interval 1975-1995, we can use a flow sampling scheme to define the number of days between adoption and ratification as the time unit of analysis. For ease of computation in the practical part of this study, we set up the data as if they were recorded every 15 days. However, the following descriptive statistics show that this "aggregation" is unlikely to affect the results in any significant way.

Given ratification, a convention is ratified 8 years after its adoption on average. Substantial heterogeneity characterizes the timing of ratification: nearly 20% of ratifications occur within 3 years after adoption while 20% of all uncensored spells last over 13 years. These differences in ratification behaviour also exist by countries. Table 1 shows the distribution of the number of ratifications across countries. Three member states have ratified

Table 1: Number of ratifications

Number of ratifications	Number of countries
More than 12	3
9-10	3
7-8	1
5-6	10
3-4	17
1-2	20
0	26
228	80

more than a dozen conventions (Brazil, Mexico and Uruguay) and 6 countries more than 8, while 26 members did not ratify any.

As regards conventions, some of them are ratified by a large number of countries. For instance, Convention No. 144 on the role of tripartite consultation in promoting international labour standards has been ratified 38 times, and Convention No. 159 on vocational rehabilitation received 25 ratifications. At the other extreme, Convention No. 143 on migrant workers and Convention No. 157 on social security rights have been ratified by only 6 countries and 1 country, respectively. To assess the heterogeneity across conventions, we performed a log rank test over the survival functions which we estimated using the Kaplan-Meier nonparametric estimator. The test statistic rejects the null hypothesis of equality across conventions. In view of these results, we consider two types of random effects: a country-specific effect and a convention effect, on top of the observed heterogeneity in the regressors described below.

As regards observed heterogeneity, we consider three categories of explanatory variables: variables relating to the economic and administrative costs and benefits of ratification, as well as to domestic political circumstances and to external pressure for ratification. These variables vary over time and may take different values depending on country and convention.

Economic costs are captured by the level of GDP per capita and the ratio of exports to GDP (taken from the Penn Tables and at 1985 international prices in US-$). We expect that workers from richer countries have a higher demand for labour standards which increases the ratification propensity. In addition, we allow for a flexible specification by considering that GDP per capita enters the hazard function quadratically. Higher-order polynomial terms were found to be insignificant. The next two variables in this group are dummy variables which indicate whether or not the convention under consideration is an explicit update of an existing convention and whether the country had ratified this "preceding" convention. We expect that previous ratifications make current ratification less costly. Symmetrically, non-ratification of a previous convention makes ratification less likely. The last variable in this category is total population, used as an inverse proxy for per-capita administrative costs of ratification.

The second group of explanatory variables is composed of indicators for internal political circumstances favouring ratification. The first one, taken from ALVAREZ et al. [1996] and updated by BOOCKMANN [2001], is a dummy for democracies. A second indicator, derived from various internet sources, equals one if the main legislative body is dominated by the political left (BOOCKMANN, 2001). Two further variables refer to the vote of the national delegates representing government and employers at the *adoption* of the convention in the International Labour Conference. They are set as one if the delegate voted against the convention or abstained. The vote of the unions' delegate is not taken into account, because it is in favour of adoption for nearly 99% of the observations.

Variables capturing external pressure derive their justification from the fact that, despite the voluntary character of ratification, some countries may be influenced in their ratification decision by other countries or by international organizations such as the World Bank or the IMF. The higher the dependence on these organizations or on other countries, the higher the external pressure that can potentially be applied. The amount of development aid received each year by a country is the first explanatory variable of this group. IMF lending and World Bank credits are additional variables capturing the external pressure effect. Lastly, we control for exports towards industrialized democracies. This variable captures countries' vulnerability with respect to trade sanctions. The rationale is that trade sanctions would most likely be applied by industrialized countries and not by other developing countries.[6] To control for the fact that pressure is less likely to occur for an oil exporter country, we interact exports towards industrialized countries with an indicator variable for OPEC membership. All these variables are measured in per cent of GDP.

An endogeneity problem arises if trade policies, or credits of the IMF and the WorldBank, are used as rewards or sanctions regarding the decision to ratify. To avoid such problem, we use a three year moving average lagged one period for these variables.

[6] Over the period considered, industrialized democracies are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Luxembourg, Netherlands, New-Zealand, Norway, Portugal, Spain, Sweden, Switzerland, UK and the USA.

We also created a set of indicators to take account of differences between the *subjects* of conventions. Table 2 shows the number of convention within each subject group.

Table 2: Number of conventions per subject

Subject	Number of conventions
Basic human rights	4
Employment	3
Labour administration	3
Conditions of work	5
Migrant workers	2
Special workers	3
Total	29

Readers interested in more details on the choice of these variables or their expected effect on the hazard function are referred to BOOCKMANN [2001], who also presents descriptive statistics and Kaplan-Meier estimates of the ratification probabilities.

3 Mixed Proportional Hazard Models

In the following, we consider three different nested models belonging to the Mixed Proportional Hazard family (hereafter referred to as MPH models), which can deal with time varying covariates. We will first introduce the Cox model with two random effects, and deduce the other two models, the Cox model with one random effect and the standard Cox model, by imposing restrictions. As a random effect represents a source of clustering between the observations, the Cox model with two frailties takes account of unobserved heterogeneity in the finest manner among the models we consider, while the standard Cox model does not deal with it at all.

Before going into the details, it is useful to discuss the indexing and the concept of time used here. Let i=1,..., I denote the country index, and j=1,...,J_i the convention index.[7] We denote by t_{ij} the time between the adoption of convention j during the International Labour Conference and its

[7] Due to the sampling scheme, the relevant conventions differ from one country to the other. A complete notation would require introducing a function j_i indexing the conventions relevant for coutry i as $j_i(1), \ldots, j_i(J_i)$. Still, the simplified notation introduced above should not lead to any confusion.

ratification by country i. In other words, t_{ij} does not denote calendar time but the duration from the start of the ratification spell.

Van den Berg [2001] provides a complete discussion of the MPH model and its properties. In this general framework, the hazard function is supposed to be the product of three terms: a term of unobserved heterogeneity, a function of possibly time varying observed explanatory variables and a function of time common to all individuals. The idea underlying the use of random effects is that data are clustered in some way and a realization of a random effect is common to all observations in the same cluster. The MPH model allows us to represent unobserved heterogeneity in several ways.

We begin with the model where the unobserved heterogeneity term is modelled as the product of two random effects: a country effect and a convention effect.[8] There is no hierarchy in the clustering here. We assume the two effects to be independent. In our case, this may be justified with respect to the universality of ILO standards. At least officially, there are no ILO conventions intended for certain regions in particular. Observed durations are assumed independent conditional on the covariates and both frailty terms, and the hazard can be expressed as:

$$\lambda_{ij}(t_{ij}|x_{ij}, \xi_i, \psi_j) = \xi_i \psi_j \, \lambda_0(t_{ij}) \, \exp\left[\beta' x_{ij}(t_{ij})\right], \tag{1}$$

where ξ_i denotes a country effect, ψ_j a convention effect, $\lambda_0(t_{ij})$ the baseline hazard which depends only on the time elapsed since the adoption of the convention under study, and β is a parameter vector common to all observations. The hazard function can also be written:

$$\lambda_{ij}(t_{ij}|x_{ij}, v_i, w_j) = \lambda_0(t_{ij}) \, \exp\left[\beta' x_{ij}(t_{ij}) + v_i + w_j\right], \tag{2}$$

where v_i denotes the log country effect and w_j the log convention effect.

Elbers and Ridder [1982] show that identification of the MPH model in a single spell setting requires independence between covariates and random effect, as well as a finite mean for the latter. Honoré [1993] proves that both assumptions can be relaxed in a multivariate setting if the realization of the random effect is shared among at least two observations. In our setting, we show in Appendix A that the model is identified without assuming $E(A) < \infty$, where $A = \xi_i, \psi_j$, or that A and x_{ij} are independent. The underlying idea is that each random effect can be held constant depending on whether the

[8] Considering the limited number of conventions, one could think of treating the convention-specific effect as a fixed effect. However, previous estimations on this dataset suggest that the empirical identification of the convention specific fixed effects is extremely difficult.

model is formulated for a given country or a given convention. We can thus switch between these two viewpoints and use HONORÉ's [1993] approach.

The identification result is obtained in a much more general setting than the one considered in this application. First, we do not need to assume that the mixing distributions have a finite mean. This allows us to specify improper priors. Improper priors are frequently used in Bayesian inference, as the posterior distribution can be proper even though the prior is not. However, the finite mean assumption will still be maintained in this paper in order to simplify the computation to some extent. Second, the identification result is obtained without covariates and allows the part of the hazard not depending on the unobserved heterogeneity, i.e. the baseline hazard, to vary for each observation. As pointed out by VAN DEN BERG [2001], covariates, and especially time varying ones, ease identification in duration analysis. Therefore, once identification is achieved without them, we do not need to impose more structure on the model if we want to consider time varying covariates later on. Third, the model is identified without assuming independence between the random effects and the covariates. In the paper, we still assume independence between frailties and covariates. It is possible to be more general and specify a model with an effect correlated with the covariates, and an independent effect. However, our identification proof is non-constructive and does not yield to an estimator of the potential association.

By constraining this model, we obtain the Cox model with one random effect (used for example by CLAYTON, 1978, 1991, GUO and RODRIGUEZ, 1992, and many others). If we set one random effect to 1 in equation (1), the remaining level of unobserved heterogeneity can be used to represent alternatively the convention effect or the country effect. As clustering at the country level is finer than at the convention level, we specify the unique random effect as a country frailty.[9] The hazard can be written as follow:

$$\lambda_{ij}(t_{ij}|x_{ij}, v_i) = \lambda_0(t_{ij}) \exp\left[\beta' x_{ij}(t_{ij}) + v_i\right]. \tag{3}$$

The standard Cox model can be deduced by assuming that both random effects are set to 1. The hazard function is:

$$\lambda_{ij}(t_{ij}|x_{ij}) = \lambda_0(t_{ij}) \exp\left[\beta' x_{ij}(t_{ij})\right]. \tag{4}$$

As the Cox model with one random effect and the standard Cox model are submodels of the one with two frailties, they are also identified.

[9]Estimators deduced from models involving random effects converge for a number of strata infinitely large (RIDDER and TUNALI, 1999).

Convention j ratified by country i at time t_{ij} contributes to the likelihood through:

$$L_{ij}(t_{ij}|\beta, \lambda_0, v_i, w_j) = \lambda_{ij}(t_{ij}) \exp\left(-\int_0^{t_{ij}} \lambda_{ij}(u)du\right). \tag{5}$$

A commonly used approach to estimate the Cox model and its refinements is the partial likelihood approach.[10] To proceed, we have to define the risk set as the set of spells still not completed at the instant just before t_{ij}. This set is denoted by R_{ij}. Convention j ratified by country i at time t_{ij} contributes to the partial likelihood through:

$$L_{ij}^P(t_{ij}|\beta, v_i, w_j) = \frac{\exp\left[\beta' x_{ij}(t_{ij}) + v_i + w_j\right]}{\sum_{(k,l)\in R_{ij}} \exp\left[\beta' x_{kl}(t_{ij}) + v_k + w_l\right]}. \tag{6}$$

Note that the baseline hazard cancels out, but the random effects do not, as long as they take at least two different values in R_{ij}. The complete and partial likelihoods of nested models can be deduced from equations (5) and (6) by ignoring the different random effects. The whole likelihood is obtained in each case by taking the product of L_{ij} or L_{ij}^P over i and j.

Equations (5) and (6) are conditional on unobserved heterogeneity terms, and we proceed by making assumptions on their distribution. A wide choice is available for ξ_i and ψ_j: gamma distribution, inverse-gaussian, log-normal, positive stable (see for example Clayton, 1991; Milcent, 2003; Gustafson, 1997 and Hougaard, 2000, respectively).[11] We assume that the log-frailties v_i and w_j follow a Gaussian distribution, as McGilchrist [1993]; Sargent [1998]; Yau [2001]; Ripatti and Palmgren [2000]; Vaida and Xu [2000] and many others. There are two motivations for this choice. First, we have no reason to think that the log-frailties induce positive or negative deviations on the hazard. Thus, we assume a symmetric distribution. Furthermore, the choice of a Gaussian distribution matches the view that unobserved heterogeneity is due to a large number of unobserved country and convention specific covariates.[12] We suppose that $\{v_i\}_{i=1}^{I}$ and $\{w_j\}_{j=1}^{J_i}$

[10] See Lancaster [1990] for a detailed discussion.

[11] Alternative specifications include a discrete mixture model, implying a defective duration distribution, (a Heckman-Singer type approach with a mass point at an infinite duration) to capture the unobserved heterogeneity of the 26 countries that did not ratify any convention. This would also be feasible in a Bayesian framework. But here 20% of the ratifications occur after 13 years and the high number of non-ratifying countries results from the time interval constraint: the rates of transition are low but strictly positive, leading to more ratifications occurring as time passes. On a finite time interval, the truncation of the duration distribution is modelled by the censoring mechanism.

[12] We also tried to estimate a model with gamma distributions for the frailties but did not obtain convergence.

are independent and normally distributed:

$$v_i \sim N(0, \tau^2), w_j \sim N(0, \alpha^2). \tag{7}$$

We make the zero mean assumption so that the effects represent deviation from the mean.

4 Bayesian Inference

The parameters of the models presented are estimated using a Bayesian approach. We first need to combine the information carried by the data with prior beliefs, in order to simulate draws from the posterior distribution using MCMC methods. In this section, we explain the choice of the priors, the form of the posterior, and the estimation procedure using Gibbs sampling.

4.1 Prior and Posterior Distributions

Inference is based on the partial likelihood, justified from the Bayesian viewpoint by Kalbfleisch [1978], and we recall briefly his approach in Appendix B. It enables to estimate the baseline hazard, while remaining in a semi-parametric setting. Assume that β has an improper prior and that the baseline hazard is distributed according to a gamma prior. This approach is based on a limit argument in which both parameters of the gamma prior tend to zero.

Following the requirements of the software we use, we specify proper priors but fairly uninformative ones. We assume independent priors for the regression coefficients. We assign each of them a normal univariate distribution with mean 0 and variance $\sigma_\beta^2 = 10^6$. Alternatively, one could think of a multivariate normal distribution as prior but it would induce much more complexity in the model, given that we would have to specify a variance matrix for 22 parameters. We specify for the baseline hazard a gamma prior with both parameters set to $a = 0.001$.[13] This assumption implies an expectation equal to 1, which does not restrict the hazard because the term $\beta' x_{ik}(t_{ik})$ is not constrained, and a variance equal to 1000.

The precisions of both log-frailties (i.e. τ^{-2} and α^{-2}) are assumed to follow a gamma distribution with expectation one and variance equal to $c = 10^3$. The gamma distribution is a conjugate prior for the precision of the Gaussian distribution (see for example Gouriéroux and Monfort, 1990, pp. 381-382) and we choose it to speed up computation.

[13] We refer here to the parameterization $\gamma(\alpha, \lambda)$ leading to an expectation equal to α/λ and a variance equal to α/λ^2.

Denote by T the vector of the durations and by M the number of observed explanatory variables. The posterior density of the Cox model with two frailties is:

$$\pi(\beta, \lambda_0, \alpha, \tau|T) \propto \left[\prod_{i=1}^{I}\prod_{j=1}^{J_i} L_{ij}(t_{ij}|\beta, \lambda_0(t_{ij}), v_i, w_j) f(\lambda_0(t_{ij})|a)\right] \prod_{i=1}^{I}\left[f(v_i|\tau)\right.$$
$$\left.\prod_{j=1}^{J_i} f(w_j|\alpha)\right] \prod_{m=1}^{M} f(\beta_m) f(\alpha) f(\tau). \quad (8)$$

On the basis of this posterior, we can deduce the special cases when we consider the Cox model with one frailty and the standard Cox model. The marginal posterior for β contains a factor equal to the partial likelihood, as shown in Appendix B.

4.2 Parameter Estimation

An analytical solution for the posterior distribution is not easily available. However, this distribution can be approximated using Markov Chain Monte Carlo (MCMC) methods, especially Gibbs sampling. Reviews on MCMC methods include Robert [1996]; Neal [1997] and Robert and Casella [1999].

Gibbs sampling is an iterative procedure requiring to simulate at each iteration draws from conditional distributions. The distribution of one random variable conditional on the other variables is proportional to the prior distribution of that random variable times the conditional distribution of any directly related variables. The conditional distributions are:[14]

1. Fixed effects

$$f(\beta|T, \xi, \psi) \propto \exp\left[-\frac{1}{2\sigma_\beta^2}\beta'\beta\right] \prod_{i=1}^{I}\prod_{j=1}^{J_i} \lambda_0(t_{ij}) \exp[\beta' x_{ij}(t_{ij})]$$
$$\exp\left(-\xi_i\psi_j \int_0^{t_{ij}} \lambda_0(u)\exp[\beta' x_{ij}(u)]du\right). \quad (9)$$

[14] Recall that $v_i = \ln \xi_i$ and $w_j = \ln \psi_j$; we still use both v and ξ (as well as w and ψ) to obtain simpler expressions.

2. Country effect

$$f(v|T,\beta,\psi) \propto \exp\left(-\frac{1}{2\tau^2}v'v\right)$$
$$\prod_{i=1}^{I}\prod_{j=1}^{J_i}\exp\left(-\xi_i\psi_j\int_0^{t_{ij}}\lambda_0(u)\exp[\beta' x_{ij}(u)]du\right). \quad (10)$$

3. Country effect standard error

$$f(\tau|v) \propto \frac{1}{\tau^{I+c+1}}\exp\left(-\frac{c}{\tau}-\frac{1}{2\tau^2}\sum_{i=1}^{I}v_i^2\right). \quad (11)$$

4. Convention effect

$$f(w|T,\beta,\xi) \propto \exp\left(-\frac{1}{2\alpha^2}w'w\right)$$
$$\prod_{i=1}^{I}\prod_{j=1}^{J_i}\exp\left(-\xi_i\psi_j\int_0^{t_{ij}}\lambda_0(u)\exp[\beta' x_{ij}(u)]du\right). \quad (12)$$

5. Convention effect standard error

$$f(\alpha|w) \propto \frac{1}{\alpha^{\sum_{i=1}^{I}J_i+c+1}}\exp\left(-\frac{c}{\alpha}-\frac{1}{2\alpha^2}\sum_{i=1}^{I}\sum_{j=1}^{J_i}w_j^2\right). \quad (13)$$

6. Baseline hazard

$$f(\lambda_0(t_{ij})|\beta,\xi,\psi) \propto \lambda_0(t_{ij})^a$$
$$\exp\left[-\lambda_0(t_{ij})\left(a+\xi_k\psi_l\exp\left[\beta' x_{kl}(t_{ij})\right]\right)\right]. \quad (14)$$

All densities from (9) to (13) are log-concave and draws are obtained with an Adaptative Rejection Sampling (ARS) algorithm (see GILKS and WILD, 1992). The conditional posterior expectation of the baseline hazard is:

$$\mathrm{E}[\lambda_0(t_{ij})|\beta,\xi,\psi] = \frac{a}{a+\sum_{(k,l)\in R_{ij}}\xi_k\psi_l\exp\left[\beta' x_{kl}(t_{ij})\right]}. \quad (15)$$

As shown in KALBFLEISCH [1978], the Bayesian estimator of λ_0 is thus a compromise between the Nelson-Aalen (maximum likelihood, see NELSON, 1969) estimator and a prior belief of strength a.

5 Results

In this section, we present the results for the three models. Two chains with different initial values were run for each model. Previous runs indicated that convergence for the variances is slower than for β. Thus we initially set the regression coefficient to 0 for both chains and chose different values for the variances: τ^2 and α^2 were set to 1 for the first chain and to 50 for the second chain. 11000 iterations were run for the standard Cox model, 12000 for the model with a country frailty and 35000 for the model with two frailties.

We assess convergence using quantile plots and the Gelman and Rubin [1992] statistic, which compares the variances between and within the Markov chains. Parallel chains are started from different initial values, which are over-dispersed with respect to the posterior distribution. The variability of the draws will initially be overstated by the variance between chains and understated by the variance within chains. Both variances become closer as the Markov chain converges. We concluded that 5000 iterations were necessary as a burn-in period for the models with no effects or one effect, and 10000 for model with two effects. Estimation was performed with a 2.5 Ghz Pentium and it took three months for the Cox model with two frailties.[15] All posterior summary statistics are based on iterations of the two chains after the burn-in step.

We also estimate the models using frequentist approaches and gamma heterogeneity. Results are obtained by maximum partial likelihood for the standard Cox model, by penalised partial likelihood for the model with one frailty, as described in Therneau et al. [2003], and by the Expectation Maximisation algorithm based on penalised likelihood (hereafter referred as the EMPL algorithm, see Horny, 2009) for the model with two frailties.[16] This approach yields a very fast algorithm, but the assumption of gamma frailties is necessary for the use of penalised likelihood. In this respect, the Bayesian approach is, in principle, more flexible, since it allows other choices for the frailty distributions. However, we have not obtained convergence of the Bayesian algorithm when specifying gamma frailties.

[15] The computation times are fairly small using WinBUGS with moderately sized datasets. However, other studies (see for example Brooks and Morgan, 2004), and also our own experiments, suggest that the computational cost increases quickly with the size of the dataset and the complexity of the model. Computation could also be speeded up by including more informative priors, but at the cost of more subjectivity. Still, it should not be expected that the Gibbs sampler is always faster than classical methods like simulated ML, and Train [2003, Chapter 12, Table 12.2] documents cases where it is indeed much slower.

[16] They have been obtained using the software R 2.0.1, and the package 'survival' is required for penalised likelihood. R is a free software available at http://www.r-project.org/.

Table 3 shows the estimates of the unobserved heterogeneity distribution parameters. The results obtained with a frequentist approach are similar

Table 3: Estimates of the standard-errors of the log-frailties distributions

Type of heterogeneity		Parameter	Median	2.5%	97.5%
Bayesian approach (gaussian frailties)					
Simple:	country effect	τ	0.49	0.42	0.57
Twofold:	country effect	τ	0.49	0.42	0.57
	convention effect	α	0.85	0.66	1.12
Frequentist approach (gamma frailties)					
Simple:	country effect	τ	0.53	0.45	0.68
Twofold:	country effect	τ	0.53	0.45	0.68
	convention effect	α	0.26	0.23	0.33

Note: frequentist results are based on the asymptotic normal approximation of the distribution of estimated parameters.

to the Bayesian estimates as far as the country effect is concerned. But we find a striking difference as regards the convention effect. Whereas the Bayesian estimate of its standard error is about twice the result for the country effect, the frequentist estimates are about half the result for the country effect. Thus, we confirm the result of underestimation reported for the country effect by Rodriguez and Goldman [2001] and Therneau et al. [2003], but only for the convention effect. This point clearly deserves further investigation.

Table 4 shows the posterior means and confidence intervals at the 5% level for the β parameters for the three models, and results obtained with penalised likelihood based methods are in Table 5. The entries "Log-likelihood with penalty" in Table 4 are ingredients for the computation of Bayesian information criteria, whereas the entries "penalised log-likelihood" in Table 5 are the optimum values of the objective functions.

In the following, we concentrate on the Bayesian estimates. Considering cost variables, we see that the second and third models show many more significant parameters than the first one. In particular, the dummy variable indicating a new convention is significantly positive, while past ratification of the convention now updated positively influences the hazard. This points to the presence of dynamics in the ratification process. The models with frailties indicate a non-monotonic impact of real GDP per capita.[17] Comparing

[17]For both, the profile is an inverted $\cup$ with a maximum at a GDP per capita value of about \$6000, so that the relationship is essentially increasing and concave (the mean GDP

Table 4: Bayesian estimates of the β parameters

Variable	Standard Cox			Cox: one frailty			Cox: two frailties		
	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%
Cost									
Real GDP per capita[a]	2.01	-0.76	4.92	**4.00**	1.49	6.71	**3.81**	1.12	6.64
Real GDP per capita, squared	-2.24	-5.46	0.77	**-3.32**	-6.25	-0.69	**-3.19**	-0.33	-6.29
No explicit update	0.53	-0.05	1.15	**0.98**	0.45	1.58	**1.39**	0.88	1.95
Own past ratification if explicit update	**1.50**	0.77	2.23	**1.37**	0.64	2.10	**1.62**	0.91	2.34
Population[b]	-0.05	-0.14	0.03	-0.01	-0.09	0.08	-0.02	-0.11	0.07
Internal pressure									
Democracy	0.12	-0.18	0.43	**0.36**	0.06	0.67	**0.34**	0.04	0.64
Left majority	-0.39	-1.04	0.19	**-0.72**	-1.37	-0.14	**-0.69**	-0.10	-1.33
Vote against convention:									
Government	-0.15	-0.59	0.33	-0.18	-0.62	0.28	-0.22	-0.66	0.24
Employers	-0.15	-0.56	0.28	0.25	-0.14	0.68	**0.38**	0.01	0.79
External pressure									
Development aid[c]	**-0.06**	-0.11	-0.03	**-7.43**	-11.69	-3.55	**-7.65**	-3.81	-11.85
Worldbank loans[c]	**-4.20**	-7.63	-0.82	2.55	-0.59	5.71	2.00	-1.11	4.97
IMF credits[c]	**6.23**	2.03	10.10	3.61	-0.24	7.29	3.96	-0.09	7.62
Exports[c]	-0.76	-3.64	1.45	-0.78	-3.54	1.12	-0.79	-3.76	1.21
Exports into industrialized countries[c]	-0.07	-6.37	6.29	0.22	-6.20	6.61	-0.18	-7.43	6.96
Exports into industrialized countries (non oil exporting countries)[c]	-0.55	-6.13	5.28	-1.10	-7.14	5.00	-0.77	-7.51	6.07
Non oil exporting country	0.19	-0.91	1.30	0.26	-0.87	1.50	0.25	-1.04	1.65
Log-likelihood with penalty[d]		-1728			-1630			-1536	
Subject of convention		Yes			Yes			No	
Number of spells		2349			2349			2349	
Number of ratifications		228			228			228	

Note: Bold entries are significant at the 5% level. *a.* 1985 international prices, in \$10 000. *b.* hundred milions. *c.* percent of GDP. The other variables are indicators, and convention subject indicators are included for the first two models. *d.* The penalty is $\frac{d}{2}\ln n$, where d is the number of parameters and n the sample size.

with Boockmann [2001], we remark that we have the same significant cost

per capita in the sample is \$2280).

Table 5: Frequentists estimates of the β parameters

Variable	**Standard Cox**			**Cox: one frailty**			**Cox: two frailties**		
	Coef.	2.5%	97.5%	Coef.	2.5%	97.5%	Coef.	2.5%	97.5%
Cost									
Real GDP per capita[a]	**3.84**	1.10	6.58	**3.94**	1.20	6.67	**3.06**	0.55	5.57
Real GDP per capita, squared	**-3.14**	-6.08	-0.20	**-3.19**	-0.27	-6.12	-2.47	-5.29	0.35
No explicit update	**0.96**	0.40	1.52	**0.90**	0.33	1.46	**0.90**	0.37	1.43
Own past ratification if explicit update	**1.37**	0.66	2.08	**1.38**	0.67	2.10	**1.38**	0.68	2.09
Population[b]	0.01	-0.08	0.09	0.01	-0.08	0.10	-0.03	-0.13	0.07
Internal pressure									
Democracy	**0.37**	0.06	0.67	**0.39**	0.09	0.69	0.27	-0.03	0.57
Left majority	**-0.68**	- 1.28	-0.07	**-0.69**	-1.29	-0.08	**-0.62**	-1.23	-0.01
Vote against convention:									
Government	-0.19	-0.66	0.27	-0.13	-0.60	0.33	-0.09	-0.54	0.36
Employers	0.23	-0.17	0.64	0.26	-0.15	0.67	0.24	-0.15	0.63
External pressure									
Development aid[c]	**-7.22**	-11.25	-3.19	**-7.22**	-11.26	-3.18	**-8.55**	-4.41	-12.69
Worldbank loans[c]	2.62	-0.41	5.65	2.74	-0.30	5.78	**3.27**	0.21	6.33
IMF credits[c]	3.65	-0.14	7.44	3.49	-0.29	7.27	3.61	-0.25	7.47
Exports[c]	-0.16	-2.87	2.56	-0.09	-2.72	2.54	0.23	-1.89	2.35
Exports into industrialized countries[c]	-0.84	-8.10	6.43	-0.85	-8.10	6.39	-2.50	-9.93	4.93
Exports into industrialized countries (non oil exporting countries) [c]	-0.67	-7.43	6.10	-0.78	-7.57	6.01	0.56	-4.08	5.20
Non oil exporting country	0.13	-1.19	1.44	0.16	-1.16	1.48	-0.02	-0.82	0.86
Penalized log-likelihood		-2203			- 2156			-2143	
Subject of convention		Yes			Yes			No	
Number of spells		2349			2349			2349	
Number of ratifications		228			228			228	

Note: Bold entries are significant at the 5% level. *a*. 1985 international prices, in $10 000. *b*. hundred milions. *c*. percent of GDP. The other variables are indicators, and convention subject indicators are included for the first two models.

variables in both studies.[18]

Turning to internal pressure variables, none of them is significant in the

[18]Results for the Cox model differ between the two studies because Boockmann [2001] included a time trend. We omit it here because simulated samples for this variable were highly autocorrelated and this slowed down convergence.

specification without frailty terms. Considering at least one frailty greatly alters this conclusion. The democracy indicator has a significant positive impact. This result is plausible because individuals whose working conditions are improved by ILO conventions, such as farmers and industrial workers, are more likely to be politically represented in a democracy than in authoritarian regimes. Surprisingly, the left majority indicator has a negative coefficient. This is probably due to the fact that the left-right distinction is often inadequate to capture the domestic politics of developing countries. Government voting has no effect on ratification. By contrast, a negative vote of the employer delegate increases the probability of ratification. These conventions may be those that formulate the most advanced labor standards. Therefore, trade unions may mobilise most political power for the ratification of these conventions (note that the corresponding variable for worker delegate had to be omitted, as discussed above). By contrast, BOOCKMANN [2001] finds no significant parameter in this category when controlling for unobserved heterogeneity with fixed effects or stratification.

The next group of variables captures external pressure. Development aid has a negative influence on the hazard, in particular if unobserved heterogeneity is controlled for. An explanation may be that countries receiving a large amount of aid also have to cope with temporary economic problems not accounted for by the other variables. As in Boockmann's study, World Bank loans seem to discourage ratification in the first model, but the effect becomes insignificant in the models with frailties.[19] IMF credits are also insignificant in these models. Finally, there is no impact of potential exposure to trade sanctions measured as exports into industrialised countries. In sum, the Bayesian results suggest that external pressure is nonexistent in the ratification decision.

Surprisingly, the Bayesian estimates of the standard Cox model are rather different from those obtained by maximum likelihood. A reason may be that we have tied observations, handled with the Breslow approximation (BRESLOW, 1974) in the Bayesian approach and with the Efron approximation (EFRON, 1977) in the frequentist one. A detailed discussion on handling ties in a partial likelihood framework is provided in THERNEAU and GRAMBSCH [2000, pp. 48-53]. HERTZ-PICCIOTTO and ROCKHILL [1997] show that the Breslow approximation leads to underestimating β in the standard Cox model, whereas the Efron approximation performs better but asks for more computing time.[20] We do not observe such differences for the other mod-

[19] Here we note a divergence with the frequentist results, since in Table 5 the coefficient of World Bank loans is insignificant except in the model with two frailties, where it is positive.

[20] In both cases, the ease of computation motivated the choice of the approximation.

els, because the Bayesian results are not underestimated with respect to the frequentists ones when unobserved heterogeneity is accounted for.

In order to compare the models on the basis of the Bayesian Information Criterium (BIC, Schwarz, 1978), we computed log-likelihoods with penalty equal to the log-likelihood minus half the number of parameters of the model times the log of the number of observations. The BIC has been designed to find the most probable model given the data, and it takes account of Occam's razor, i.e. the more parsimonious model is chosen when two models fit the data comparably well. Wasserman [2000] reports that under mild regularity conditions, the BIC approximates the log Bayes factor. If one sets the prior odds of each model to be equal, the Bayes factor is the posterior odds ratio of one model versus the other one. The BIC, obtained by taking differences of the log-likelihoods with penalty, is equal to 98 when comparing the standard Cox model to the model with one frailty, and 94 when comparing the model with two frailties to the one with one frailty.[21] The Bayes factor between the last two models is thus exp(94), giving strong evidence in favour of the model with two random effects. This result was not obvious ex ante, because the model with two frailties does not include the convention subject variables used in the other two models.

6 Conclusion

Our study uses a Bayesian approach to estimate MPH models with different specifications of unobserved heterogeneity, the most detailed one using two random effects. Rather than assuming a parametric form for the baseline hazard, we use Cox's partial likelihood semi-parametric approach to avoid misspecification problems. This approach has been justified from a Bayesian viewpoint by Kalbfleisch [1978]. After having established the identification of the models, we estimate them using Gibbs sampling. In order to simulate from posterior marginal densities, we also use other simulation-based computational algorithms such as the acceptation-rejection sampling.

Our results confirm that it is important to control for the country under study. Some unobserved explanatory country-specific variables seem to have

It is possible to use to Efron approximation in a Bayesian setting, as well as the Breslow approximation in a frequentist approach, and to compare the results of the four models. However, the expected results of such a comparison are likely to be outweighted by the cost of the investment.

[21]The formula is: $BIC_{ij} = \ln L_i - \ln L_j + \frac{d_j - d_i}{2} \ln n$, where i and j are the models indexes, L the likelihood, d the number of parameters of the model and n the size of the sample.

a large influence on the ratification behaviour. The Bayesian results also show the presence of an even larger amount of heterogeneity among conventions. They differ in this from the frequentist results, for which the heterogeneity between conventions is less important than the unobserved heterogeneity between countries. Some of the conventions are very consensual among member states of the ILO or do not induce important economic or political costs. Other conventions can be less easily ratified because they are more ambitious and imply too important costs. Switching from the frequentist paradigm to a Bayesian approach does not seem to influence other estimation results in a major way, because the findings confirm most results from previous studies on ILO ratification behaviour, especially BOOCKMANN [2001]. However, frequentist approaches seem to understate the amount of unobserved heterogeneity, in contrast to the Bayesian approach. Further research efforts are needed to understand this phenomenon.[22]

Acknowledgements: We thank Gerard Van den Berg, Mathieu Brezovski, Jean-Pierre Florens, Denis Fougère and Viktor Winschel for comments on previous versions of this paper. We also thank an anonymous referee and an Editor for careful comments that helped us revise the paper. We also thank participants in the Econometric Society European Meeting, August 2004, Madrid.

A Identification: proof

The proof is an extension of HONORÉ's [1993] first proof to the case of 2 frailties. As we do not restrict the random effects to be nested, it is more general than what is required for the models at hand. Consider the case of two conventions submitted to country i. The model is:

$$\lambda_{i,1}(t_{i1}|\xi_i, \psi_1) = \xi_i \psi_1 \, \lambda_{0,1}(t_{i1}), \tag{A.1}$$

$$\lambda_{i,2}(t_{i2}|\xi_i, \psi_2) = \xi_i \psi_2 \, \lambda_{0,2}(t_{i2}). \tag{A.2}$$

This includes of course the case $\lambda_{i,1}(t) = \lambda_0(t)\exp(\beta' x_i)$. The joint survivor function for country i is:

$$\begin{aligned} S(t_{i1}, t_{i2}) &= \int_\xi \exp\left[-\xi\psi_1 \, \Lambda_{0,1}(t_{i1}) - \xi\psi_2 \, \Lambda_{0,2}(t_{i2})\right] dH_\xi(\xi) \qquad \text{(A.3)} \\ &= \mathcal{L}_\xi \left[\psi_1 \, \Lambda_{0,1}(t_{i1}) + \psi_2 \, \Lambda_{0,2}(t_{i2})\right], \end{aligned}$$

[22] A thorough comparison between Bayesian and frequentist approaches would ask for the specification of identical mixing distributions in both frameworks. This is not feasible with the data studied here, as the Bayesian estimator does not converge, for numerical reasons, on this dataset in presence of gamma heterogeneity, and the penalised likelihood estimator requires gamma heterogeneity.

where $\Lambda_{0,j}(t_{ij}) = \int_0^{t_{ij}} \lambda_{0,j}(u)du$, H_ξ is the cumulative distribution function of ξ and $\mathcal{L}_\xi$ is the Laplace transform for ξ. Notice that $S(t_{i1}, t_{i2})$ is observable by taking a large number of observations for a population homogeneous with respect to ψ_1 and ψ_2. The way the population is clustered is known and we can thus observe $S(t_{i1}, t_{i2})$ for invariant ψ_1 and ψ_2, that is for a group defined by the intersection of the group characterized by ψ_1 and the group characterized by ψ_2. By differentiating (A.3) over t_{i1} and t_{i2} and taking the ratio, we obtain:

$$\begin{aligned}\frac{\partial S_i(t_{i1}, t_{i2})/\partial t_{i2}}{\partial S_i(t_{i1}, t_{i2})/\partial t_{i1}} &= \frac{\psi_2\lambda_{0,2}(t_{i2})\mathcal{L}'_\xi\left[\psi_1\Lambda_{0,1}(t_{i1}) + \psi_2\Lambda_{0,1}(t_{i1})\right]}{\psi_1\lambda_{0,1}(t_{i1})\mathcal{L}'_\xi\left[\psi_1\Lambda_{0,1}(t_{i1}) + \psi_2\Lambda_{0,2}(t_{i2})\right]} \\ &= \frac{\psi_2\lambda_{0,2}(t_{i2})}{\psi_1\lambda_{0,1}(t_{i1})}.\end{aligned} \tag{A.4}$$

Let us denote by k the quantity $1/\lambda_{0,1}(t_{i0})$, where t_0 is a reference time. Taking the ratio of the value of (A.4) at (t_{i0}, t_{i2}) to the value at (t, t_{i2}), we obtain:

$$\frac{\psi_2\lambda_{0,2}(t_{i2})}{\psi_1\lambda_{0,1}(t_{i0})}\bigg/\frac{\psi_2\lambda_{0,2}(t_{i2})}{\psi_1\lambda_{0,1}(t)} = \frac{\lambda_{0,1}(t)}{\lambda_{0,1}(t_{i0})} = k\lambda_{0,1}(t). \tag{A.5}$$

By integrating over time, we obtain $k\Lambda_{0,1}(t) + c_1$, where c_1 is obtained by the initial condition $\Lambda_{0,1}(0) = 0$. Then, $\Lambda_{0,1}(t|X_i)$ is identified up to normalization k. Similarly, one can prove that $\Lambda_{0,2}(t|X_i)$ is identified by taking the ratio of (A.4) at (t_{i1}, t_{i0}) over its value at (t_{i1}, t). As ψ_1 and ψ_2 are fixed, the joint survivor function $S(t_{i1}, t_{i2}|\psi_1, \psi_2)$ depends only on time in a known way. We can thus trace out $\mathcal{L}_\xi$ by letting (t_1, t_2) vary over $[0, \infty[^2$ and H_ξ is identified.

Reverse now the viewpoint and consider the case of convention j submitted to two differents countries, indexed by 1 and 2. The model is:

$$\lambda_{1,j}(t_{1j}|\xi_1, \psi_j) = \xi_1\psi_j\,\lambda_{0,1}(t_{1j}), \tag{A.6}$$
$$\lambda_{2,j}(t_{1j}|\xi_2, \psi_j) = \xi_2\psi_j\,\lambda_{0,2}(t_{2j}). \tag{A.7}$$

The model is exactly the preceding one where ξ and ψ have been reverted, and we have thus the identification of the distribution of ψ.

B Implementing the partial likelihood

We recall here the justification of the partial likelihood in a Bayesian setting described in Kalbfleisch [1978]. Let us denote by δ_{ij} an indicator equal to one if a ratification is observed at time t_{ij}. Assume that δ_{ij} follows a

Poisson distribution with parameter $\lambda_{ij}(t_{ij}|x_{ij},\xi_i,\psi_j)$. The contribution to the likelihood of convention j ratified by country i at time t_{ij} is:

$$L^{Poisson}(\delta_{ij}|\xi_i,\psi_j) = \lambda_{ij}(t_{ij}|x_{ij},\xi_i,\psi_j)^{\delta_{ij}} \exp\left[-\sum_{(k,l)\in R_{ij}} \lambda_{kl}(t_{ij}|x_{kl},\xi_k,\psi_l)\right]. \tag{A.8}$$

This last equation is exactly the likelihood of a duration model with hazard $\lambda_{ij}(t_{ij}|x_{ij},\xi_i,\psi_j)$. Assume that the baseline hazard has a gamma prior with parameters $c\lambda_0^*(t)$ and c, where $\lambda_0^*(t)$ is a prior guess on $\lambda_0(t)$, and integrate it out:

$$\begin{aligned}\pi(\beta|\lambda_0^*,v,w,c) = \prod_{i=1}^{I}\prod_{j=1}^{J_i} \xi_i\psi_j \exp\left[\beta' x_{ij}(t_{ij})\right] \int_0^\infty \lambda_0(t_{ij})^{c\lambda_0^*(t)} \exp\Bigg[&- \lambda_0(t_{ij}) \\ \Big(c + \sum_{(k,l)\in R_{ij}} \xi_i\psi_j \exp\left[\beta' x_{ij}(t_{ij})\right]\Big)&\Bigg] d\lambda_0(t_{ij}) \\ \propto \prod_{i=1}^{I}\prod_{j=1}^{J_i} \frac{\xi_i\psi_j \exp\left[\beta' x_{ij}(t_{ij})\right]}{\Big(c + \sum_{(k,l)\in R_{ij}} \xi_i\psi_j \exp\left[\beta' x_{ij}(t_{ij})\right]\Big)^{c\lambda_0^*(t)+1}}&,\end{aligned} \tag{A.9}$$

where R_{ij} denote the set of spells still not completed just before t_{ij}. By letting $(c\lambda_0^*(t), c) \to 0$, that is assuming the baseline hazard follows a gamma non-informative prior, the marginal posterior distribution for β is proportional to the partial likelihood.

References

ALVAREZ, M., J. CHEIBUB, F. LIMONGI, and A. PRZEWORSKI (1996).– «Classifying political regimes», *Studies in Comparative International Development*, 31(2), 2–36.

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.

BRESLOW, N. (1974).– «Covariance analysis of censored survival data», *Biometrics*, 30, 89–99.

BROOKS, S., K. R. and B. MORGAN (2004).– «A Bayesian approach to combining animal abundance and demographic data», To appear in *Animal Biodiversity and Conservation.*

CLAYTON, D. (1978).– «A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence», *Biometrika*, 65, 141–151.

——— (1991).– «A Monte Carlo method for Bayesian inference in frailty models», *Biometrics*, 47, 467–485.

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

DEMPSTER, A., N. LAIRD, and D. RUBIN (1977).– «Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion)», *Journal of the Royal Statistical Society. Series B*, 39, 1–38.

DJURDJEVIC, D. (2000).– «Estimation de modèles de durée multivariés: application à la ratification des conventions par les pays membres du BIT», Master's thesis, Université Louis Pasteur, Strasbourg.

EFRON, B. (1977).– «The Efficiency of Cox's Likelihood Function for Censored Data», *Journal of the American Statistical Association*, 72, 557–565.

ELBERS, C. and G. RIDDER (1982).– «True and spurious duration dependence: The identifiability of the proportional hazard model», *Review of Economic Studies*, 49, 403–410.

Gelman, A. and D. Rubin (1992).– «Inference from iterative simulation using multiple sequences», *Statistical Science*, 7, 457–472.

Gilks, W. and P. Wild (1992).– «Adaptative rejection sampling for Gibbs sampling», *Journal of the Royal Statistical Society. Series C*, 41, 337–348.

Gouriéroux, C. and A. Monfort (1990).– *Statistiques et modèles économétriques*, vol. 1-2, Paris: Economica, 2 ed.

Guo, G. and G. Rodriguez (1992).– «Estimating a multivariate proportional hazards model for clustered data using the EM algorithm, with an application to child survival in Guatemala», *Journal of the American Statistical Association*, 87, 969–976.

Gustafson, P. (1997).– «Large hierarchical Bayesian analysis of multivariate survival data», *Biometrics*, 53, 230–242.

Hertz-Picciotto, I. and B. Rockhill (1997).– «Validity and Efficiency of Approximation Methods for Tied Survival Times in Cox Regression», *Biometrics*, 53, 1151–1156.

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

Horny, G. (2001).– «Estimation de modèles de durée multivariés au moyen de l'algorihme EM», Master's thesis, Université Louis Pasteur, Strasbourg.

——— (2009).– «Inference in Mixed Proportional Hazard Models with K Random Effects», *Statistical Papers*, 50 (3), 481–499.

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

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

Kalbfleisch, J. and R. Prentice (1980).– *The Statistical analysis of failure time data*, New York: John Wiley.

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

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

LIANG, K.-Y., S. SELF, K. BANDEEN-ROCHE, and S. ZEGER (1995).– «Some recent developments for regression analysis of multivariate failure time data», *Lifetime data analysis*, 1, 403–415.

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

MCGILCHRIST, C. (1993).– «REML estimation for survival model with frailty», *Biometrics*, 49, 221–225.

MILCENT, C. (2003).– «Ownership, system of reimbursement and mortality rate relationships», *mimeo*.

NEAL, R. (1997).– «Markov Chain Monte Carlo methods based on 'slicing' the density function», Tech. Rep. 9722, Department of Statistics, University of Toronto.

NELSON, W. (1969).– «Hazard plotting for incomplete failure time data», *Journal of Quality Technology*, 1, 27–52.

RIDDER, G. and I. TUNALI (1999).– «Stratified partial likelihood estimation», *Journal of Econometrics*, 92, 193–232.

RIPATTI, S. and J. PALMGREN (2000).– «Estimation of multivariate frailty models using penalized partial likelihood», *Biometrics*, 56, 1016–1022.

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

ROBERT, C. and G. CASELLA (1999).– *Monte Carlo statistical methods*, Springer.

RODRIGUEZ, G. and N. GOLDMAN (2001).– «Improved estimation procedures for multilevel models with binary responses: a case study», *Journal of the Royal Statistical Society*, 164, 339–355.

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

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.

SCHWARZ, G. (1978).– «Estimating the dimension of a model», *The Annals of Statistics*, 6, 461–464.

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

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

TRAIN, K. (2003).– *Discrete Choice Methods with Simulation*, Cambridge University Press.

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

VAN DEN BERG, G. (2001).– «Duration models: specification, identification and multiple durations», in *Handbook of Econometrics*, Elsevier, Amsterdam: J. J. Heckman and E. Leamer (eds.), vol. 5, chap. 55, 3381–3463.

WASSERMAN, L. (2000).– «Bayesian model selection and model averaging», *Journal of Mathematical Psychology*, 44, 92–107.

YAMAGUCHI, K. (1986).– «Alternative approaches to unobserved heterogeneity in the analysis of repeatable events», in *Sociological Methodology 1986*, ed. by N. Tuma, Washington D.C.: Jossey-Bass.

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

249. Bernhard Boockmann, Dragana Djurdjevic, G. Horny and François Laisney, " Bayesian estimation of Cox models with non-nested random effects: an application to the ratification of ILO conventions by developing countries ", 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