

Bayesian Estimation of Cox Models with Random Effects: An Application to the Ratification of ILO Conventions

Guillaume Horny*, Bernhard Boockmann[†],
Dragana Djurdjevic[‡] and François Laisney[§]

February 2004

Abstract

This empirical study investigates the ratification behaviour of ILO member states. We use a multivariate hazard model for the analysis of data on the timing of ratifications of different conventions. The model accounts for two random effects, one at the country level and the other at the convention level. We use of a semi-parametric Bayesian approach, based on the partial likelihood. Our findings confirm the results of preceding studies that ratification behaviour varies substantially across member states and conventions. Furthermore, the results yield insights on the impact of unobserved heterogeneity on the ratification process.

Keywords: Gibbs sampling, partial likelihood, frailties, duration analysis

JEL Classification: C11, C14, C15, C41, D78, J80, O19

*BETA-Thème, University Louis Pasteur, 61 avenue de la Forêt Noire, 67085 Strasbourg Cedex, France. horny@cournot.u-strasbg.fr

[†]Centre for European Economic Research (ZEW), PO Box 10 3443. D-68034 Mannheim, Germany. boockmann@zew.de

[‡]SIAW, University of St. Gallen, Dufourstr 48. CH-9000 St.Gallen, Switzerland. dragana.djurdjevic@unisg.ch

[§]BETA, University Louis Pasteur, 61 avenue de la Forêt Noire, 67085 Strasbourg Cedex, France. fla@cournot.u-strasbg.fr and Centre for European Economic Research (ZEW), PO Box 10 3443. D-68034 Mannheim, Germany. laisney@zew.de

1 Introduction

The International Labour Organization (ILO) has been created at the end of World War I with the idea that a lasting peace could only be built on the basis of social justice. Thus the ILO, a specialised agency of the UN, endeavours to set since 1919 minimum standards of basic labour rights. For that purpose, it adopts conventions that each member state is free to ratify or not. There are 184 ILO conventions that have been adopted since 1919 in many subject areas. However, (see Boockmann, 2001) most conventions are ratified only by a minority of member states. Thus, it appears interesting to evaluate the determinants of the ratification behaviour of ILO members, and so doing to establish a link between economic performance, political factors and the acceptance of international labour standards.

Boockmann (2001) specifies a ratification model that takes account of the role of costs incurred by the ratifying countries, as well as the external and internal pressures, in explaining the ratification behaviour of a given country. Membership in the ILO is almost universal, and this complicates the problem addressed above: substantial differences exist between member states and some are difficult to control for. Typically, we are not able to observe all explanatory variables that affect the ratification process. As a consequence, any model which attempts to investigate the factors governing ratification has to consider the problem of unobserved heterogeneity.

Typically, two approaches have been used: fixed effects and random effects (sometime referred to as frailty effects in the survival literature). The advantage of the fixed effects approach is that it is easy to implement and that it does not require assumptions on the unobserved heterogeneity.

A more general way to deal with fixed effects is stratified partial likelihood as discussed in Kalbfleisch and Prentice (1980), Yamaguchi (1986) and Ridder and Tunali (1999). Observations are assumed to be grouped, and each group is called a stratum. Partial likelihood is widely used in duration analysis, because it is a semi-parametric approach which avoids the specification of a functional form for the baseline hazard, and thus the problems linked to a potential misspecification. Ridder and Tunali (1999) discuss the merits of fixed and random effects in their application to child mortality rates. They also propose a Hausman test to test for fixed effects (stratified estimation) versus random effects (unstratified). 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 in presence of a large number of strata. Boockmann (2001) restricts attention to a country fixed effect, and applies their estimator. However, the Hausman test did not reject the unstratified model

for the sample of developing countries (for industrial countries, the unstratified model was rejected). This information is important, because it provides some justification for the choice of a random effects approach for the dataset of developing countries used in this paper.

Random effects permit us to go deeper in the analysis of unobserved heterogeneity, by estimating the parameters of the frailty distribution and the importance of the group effect. The disadvantages of this approach are that we need to suppose a distribution for the frailties and that they are generally assumed to be independent of the covariates. 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. If we suspect that dependence among observations is due to more than one frailty, one way to proceed is to specify a model with several random effects. The case of two frailties is presented in the survey of Liang et al. (1995). They can be independent, or nested as in Gustafson (1997), Sastry (1997) or Milcent (2003) where several individuals share a common frailty. Following Louis (1982), Guo and Rodriguez (1992) used a speeded up Expectation-Maximization algorithm (Dempster, Laird, and Rubin, 1977). Sastry (1997) extends this approach to the case of a Cox model with two nested random effects. Bolstad and O. Manda (2001) propose a Bayesian approach to estimate Sastry's (1997) model.

Djurdjevic (2000) extends the Guo and Rodriguez (1992) methodology to the partial likelihood approach, relaxing the assumption of a piecewise constant baseline hazard. Horny (2001) generalizes this work 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, they fail to provide any sensible result because of convergence problems. They conclude that the EM algorithm, while working satisfactorily on simulated data, is not well suited for this dataset of ILO conventions.¹

Considering a logistic regression, Rodriguez and Goldman (2001) compare the maximum likelihood, marginal quasi likelihood, penalized quasi likelihood and Bayesian approaches using clustered data. Except in the cases of maximum likelihood and Bayesian estimator, results underestimate both fixed and random effects. Although bias can be removed by bootstrapping, the procedure proved to be computationally more intensive than Markov chain Monte Carlo estimation and sometime failed to converge.

¹Lancaster (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 other direction: the variance becomes large enough to create numerical problems. Bolstad and O. Manda (2001) explains it by inaccurate rounding off in case of high variance.

The main contribution of this paper is to apply a semi-parametric Bayesian approach, based on partial likelihood, to a Cox model with two non-nested random effects. In line with Boockmann (2001), we use a reduced form approach in order to analyse the observed and unobserved determinants affecting the duration between adoption and ratification of a convention. Following Boockmann (2001), we formulate three assumptions on ratification behaviour. The first one, concerning economic and administrative costs, posits that political actors are more inclined to ratify conventions inducing low economic and administrative costs. The second hypothesis is about internal pressure: implementing an international standard *via* the ratification of an ILO convention may be easier from the point of view of domestic acceptance than implementing an equivalent domestic standard. The third hypothesis deals with external pressure. Despite the voluntary character of ratification, some countries can be influenced by other countries, or international organizations such as the World Bank or the IMF, in their ratification decision. We will use approximatively the same explanatory variables as Boockmann (2001). Readers interested in the details of the choice of these variables or their expected effect on the hazard function are referred to this source. Explanatory variables may take different values depending on time, country and convention, and can logically be separated in three groups corresponding to the three hypotheses above.

The first group is composed of explanatory variables that can measure the economic and administrative costs of ratification. Real per capita GDP, in dollars and varying over time and countries, enters both linearly and quadratically, which allows a non-monotonic impact on the hazard. Two other variables in this group are dummy variables. The first one indicates whether or not the convention under consideration is an explicit update of an existing convention. The second one indicates whether or not a member state has ratified in the past a convention whose explicit update is now considered. Both variables are working together. While the second variable seems to be the most meaningful, the first one introduces a difference between two states: if the second variable is equal to zero, it may be due to the non-existence of a “preceding” convention or to the non-ratification of an existing convention. In the first case, ratification is more likely than in the second case. The last variable is the population, used as a proxy of the administrative capacity of a country.

The second group of explanatory variables is composed of dummies which measure the impact of internal pressures. The first one, deduced from Alvarez et al. (1996), indicates whether the country is a democracy or not. Another dummy is equal to one for a left wing parliamentary majority. Two other variables refer to the vote of the national delegates (representing government

and employers) during the adoption of the convention under consideration at the International Labour Conference. The vote of the delegate of the unions is not taken into account, because it is equal to one for nearly 99% of the observations.

The third and last group of variables refers to potential external pressures. The development aid received each year by a country, in percent of GDP, is the first explanatory variable of this set. We also consider the IMF lending and World Bank credits, in percent of GDP. Total exports and exports towards industrialized democracies, both in percent of GDP, are two other explanatory variables. These variables can suffer from a potential endogeneity problem if development aids, trade policies or the instruments of the IMF and the Worldbank depend on past ratification decisions. We use three years moving averages and lagged values to deal with this problem, as in Boockmann (2001). We also differentiate oil-exporting countries from the others, and consider as explanatory variables an oil-exporting dummy and the interaction of its complement to one with the variable “exports towards industrialized democracies”.² We also create a set of dummies to take account of differences between subjects of conventions.

Following the conclusions of Boockmann (2001) and Chau and Kanbur (2001), we consider two types of effects: country effects, and convention effects. The presence of the latter type of effects stems from the fact that some conventions may be more easily ratified than others. For instance, conventions allow countries different degrees of flexibility or differ in complexity. As a consequence, clustering occurs in two different ways and we are investigating the degree of association among observations from both sides.

To estimate the model, we use the approach described by Kalbfleisch (1978) which allows to use the partial likelihood approach in the Bayesian paradigm. After the full specification of the model, we estimate the parameters using Gibbs sampling (introduced in the seminal paper of Gelfand and Smith, 1990).

This 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 (see Cox, 1972). 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

²These 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.

Section 5.

2 Data

The survival data analyzed in this study come from a database described in Boockmann (2001). The duration were collected using flow sampling over the period 1975-1995. Durations are defined in the original database as the number of days between the adoption of a convention and its ratification by a particular country. Due to computational limitations, we set up the data as if they were recorded every 15 days.

Our data comprises 80 ILO member states, none of them industrialized (OECD members) and 29 conventions for a total of 228 ratifications. Considering spells ended by a ratification, the mean length of spell is 8 years but durations differ widely: nearly 20% of ratifications occur before 3 years and further 20% after 13 years. The values of the explanatory variables are unlikely to remain constant over such spells. We expect the probability of ratification to be influenced by the changes of the covariates as time passes, and as a consequence, the data were set up to account for time varying covariates. Table 1 summarises the distribution of the number of ratifications per country. Three member states have ratified more than a dozen conven-

Table 1: Number of ratifications for developing countries

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

tions (Brazil, Mexico and Uruguay) and 6 countries more than 9, while 26 members did not ratify any.

Considering conventions, we see that some of them are ratified by a huge number of countries whereas others are not. For example, conventions numbered 144 and 159 in the ILO classification have been respectively ratified 38 and 25 times. On the other side, conventions 143 and 157 have been ratified

by only 6 and 1 countries respectively. We estimate the survival function for each convention using the Kaplan-Meier non-parametric estimator, and perform a log rank test to assess their equality. Based on the null hypothesis of no difference, the test statistic yields a value of 1002.55 and we conclude that the cumulative probabilities of ratification are significantly different.³

As a consequence, we consider two types of random effects: a country specific effect and a convention effect.

3 Mixed Proportional Hazard Models

In the following, we consider three different nested models belonging to the Mixed Proportional Hazard family (hereafter referred as MPH models, see Van den Berg, 2001, for a survey), which can deal with time varying covariates. We will first introduce the Cox model with two random effects, and by imposing restrictions, deduce the other two models: Cox model with one random effect and the standard Cox model. 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 $i=1, \dots, I$ denote the country index, and $j=1, \dots, J_i$ the convention index.⁴ We denote by t_{ij} the time between the adoption of the convention j during the International Labour Conference and its ratification by state i . That is, t_{ij} does not denote calendar time but 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 the observed explanatory variables (possibly time varying) 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 (for instance because they share the same genetic factors, or the same environmental characteristics), 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.

³Critical values are 39.38 at the 5% level and 48.28 at the 1% level.

⁴Due to the sampling scheme, the number of conventions varies from one country to the other.

We begin here with the model where the unobserved heterogeneity term is modelled as the product of two random effects. There is no hierarchy in the clustering here. Observed durations are assumed independent conditional on the covariates and both frailty terms, and the hazard can be expressed as:

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

where ξ_i denote 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:

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

where v_i denote the log-country effect and w_j the log-convention effect.

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 the clustering at the country level is finer than at the convention level, we choose to use the unique random effect as a country frailty:

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

The standard Cox model (Cox, 1972) can be deduced by assuming that both random effects are set to 1. It assumes that conditional on the observed vector of covariates, the t_{ij} are independent, hence not allowing for unobserved heterogeneity. The hazard function is:

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

A commonly used approach to estimate the Cox model and its refinements is the partial likelihood approach.⁵ This semi-parametric approach does not require the specification of a functional form for the baseline hazard, avoiding therefore all problems linked to misspecification of the latter. To proceed, we have to define the risk set as the set of spells still not completed at the instant just before t_{ij} , denoted by R_{ij} . Convention j ratified by country i at time t_{ij} contributes to the partial likelihood through:

$$L_{ij}(\beta) = \frac{\exp [\beta' x_{ij}(t_{ij}) + v_i + w_j]}{\sum_{(k,l) \in R_{ij}} \exp [\beta' x_{kl}(t_{ij}) + v_k + w_l]}. \quad (5)$$

⁵See Lancaster (1990) for a detailed discussion.

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 partial likelihood of nested models can be deduced from equation (5) by ignoring absent effects. The whole partial likelihood is obtained by taking the product of $L_{ij}(\beta)$ over i and j .

Equation (5) is 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 respectively Clayton, 1991, Milcent, 2001, Gustafson, 1997, and Hougaard, 2000, for example) ... We assume that the log-frailties v_i and w_j follow a Gaussian distribution, as McGilchrist (1993), Sargent (1998), Yau and McGilchrist (1998), Ripatti and Palmgren (2000), Vaida and Xu (2000) and many others. Our choice is motivated by two reasons. First, we have no reason to think that the log-frailties induce positive or negative deviation on the hazard. Thus, we assume a symmetric distribution. Furthermore, the choice of a Gaussian distribution matches the view that unobserved heterogeneity is here due to a large number of unobserved country and convention specific covariates. We suppose that $\{v_i\}_{i=1}^I$ and $\{w_j\}_{j=1}^{J_i}$ are independent and:

$$v_i \sim N(0, \tau^2), w_j \sim N(0, \alpha^2) \quad (6)$$

We make the zero mean assumption so that the effects represent deviation from the mean. A possible extension would be to assume variances specific to each country or convention.

The partial likelihood approach can easily be extended to deal with censored observations, assuming that censoring is non-informative (see Lancaster, 1990). The risk set at time t_{ik} now also contains spells censored at or after t_{ik} . Censored spells contribute only to the partial likelihood through their presence in the risk sets, and thus in the denominator of equation (5).

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, to proceed afterward to the estimation using MCMC methods. In this section, we explain the choice of the priors and the parameter estimation using Gibbs sampling.

4.1 Prior Distributions

The first stage of the inference is based on the partial likelihood, justified from the Bayesian viewpoint by Kalbfleisch (1978). Basing his analysis on counting processes, he showed that considering the baseline hazard values as following a gamma distribution with adequate parameters leads to a posterior density which is proportional to the partial likelihood. The direct calculation of the partial likelihood through the specification of the hazard function is equivalent to writing a counting process with a baseline hazard distributed according a gamma prior. Note that this approach is based on a limit argument in which both parameters of the gamma prior tend to zero. This is why we assume in our application that the baseline hazard follows a gamma prior with both parameters equal to 0.001. This assumption implies an expectation equal to 1 and a variance equal to 1000,⁶ which does not restrict the hazard because the term $\exp[\beta'x_{ik}(t_{ik}) + v_i + w_j]$ is not constrained.

The counting process-based formulation of the model assumes that time is continuous, in the sense that two ratifications cannot occur simultaneously. We have only a few tied observations in the sample, handled with the approximation of Breslow (1972).

The second stage of the Bayesian approach is to assign prior densities to the parameters of the model. In our study, we used only proper but uninformative priors, to make computation easier.

The last parameters to be estimated are the regression coefficients. We specify independent priors to simplify computation. 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 covariance matrix for 22 parameters. We assign each of them a normal univariate distribution with 0 mean and variance equal to 10^6 . Two reasons motivate this choice: scale invariance and symmetry. Scale invariance is important, because it ensures that the measuring unit for the covariates does not influence the inference. For example, if an explanatory variable is changed from dollars to thousands of dollars, this has the effect of dividing the coefficient by a factor 10^3 . The property of symmetry of the normal distribution is useful because, as for the log-frailties, positive and negative deviations seem equally likely.

The last stage concerns the choice of priors for the parameter of the unobserved heterogeneity distribution. Standard errors of both log-frailties are assumed to follow an inverse gamma distribution, which is a conjugate prior for the Gaussian distribution, with expectation one and variance equal to

⁶We refer here to the parameterization $\gamma(\alpha, \lambda)$ leading to an expectation equal to α/λ and a variance equal to α/λ^2 .

10^3 .

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

$$\pi(\beta, \alpha, \tau | t_{ij}) \propto \left[\prod_{i=1}^I \prod_{j=1}^{J_i} L_{ij}(t_{ij}) f(w_j | \alpha) \right] \prod_{i=1}^I f(v_i | \tau) \prod_{m=1}^M f(\beta_m) f(\alpha) f(\tau). \quad (7)$$

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. We just need to omit unobserved heterogeneity distribution and to be careful when taking products.

4.2 Parameter Estimation

An analytical solution for the posterior distribution is not easily available because it involves the evaluation of complex integrals. However, it can be approximated using Markov Chain Monte Carlo (MCMC) methods, especially Gibbs sampling. Recent reviews on MCMC methods include Robert (1996), Neal (1997) and Robert and Casella (1999).

Gibbs sampling is implemented in the WinBUGS software⁷ and works as follows. Suppose we consider a vector X , composed of n ($n > 1$) random variables. The algorithm samples each x_i from its conditional distribution $f_i(x_i | \{x_j\}_{j \neq i})$, given the current values of x_j , for $i = 1 \dots n$. The sequence x_i obtained by repeating the procedure is a Markov chain. Once convergence is achieved, the Gibbs sampler produces a sample close to one sampled directly from the posterior distribution, from which the expectations of quantities of interest can be estimated.

The regression coefficients β are simulated from their conditional distribution using the Acceptation-Rejection sampling (ARS) (see Robert, 1996, or Robert and Casella, 1999, for detailed discussion on this sampler). The variances τ and α are simulated directly from a gamma density.

Results obtained are presented in the next section.

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 run indicate that variance estimator converge slower than β estimator. Thus we initially set

⁷Freely available at <http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml>, see Spiegelhalter et al. (2000) for the manual and examples.

the regression coefficient to 0 for both chains and differentiate the values for the parameter of the frailties distributions. Variances τ^2 and α^2 were set to 1 for the first chain and to 50 for the second chain. Results for the Cox model with one or two frailties presented in the following are only preliminary because convergence is not totally achieved. 11000 iterations were run for the standard Cox model, 5000 for the model with a country frailty and 12000 for the model with two frailties. Using the convergence diagnostic tool of Gelman and Rubin (1992) and quantile plots, we concluded that 4000 iterations were sufficient in every cases for the burn-in period. Estimation was performed with a 2.5 Ghz Pentium and it took 610 hours for the standard Cox model, 265 hours for the Cox model with on frailty and 660 hours for the Cox model with two frailties. All posterior summary statistics are based on iterations of the two chains after the burn-in step.

Table 2 shows the estimates of the unobserved heterogeneity distribution parameters.

Table 2: Estimates of the standard-errors of the log-frailties distributions

Type of heterogeneity		Parameter	Mean	2.5%	97.5%
Simple:	country effect	τ	0.56	0.34	0.96
Twofold:	country effect	τ	0.49	0.42	0.57
	convention effect	α	0.85	0.66	1.11

Considering the confidence interval at the 5% level, we remark that the convention effect is significantly more important than the country effect. That is, observations are more strongly correlated among conventions than among countries.

Table 3 shows the posterior means and confidence intervals at the 5% level for the β parameters for the three models.

We see that the third model is characterized by many more significant parameters than the two first and an important increase of the log-likelihood. Comparing with Boockmann (2001), we remark that we have the same significant cost variables in both studies.

Turning to internal pressure variables, no variable is significant in the specification without frailty terms. Considering two frailties greatly alters this conclusion. The democracy dummy has a significant positive impact, which is not surprising: groups whose working conditions are concerned by a convention are more likely represented in a democracy than in other regimes. Surprisingly, the left majority dummy has a negative coefficient. This is probably due to the fact that the left-right distinction is often inadequate

Table 3: 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	2.01	-0.76	4.92	0.21	-2.83	3.34	3.84	1.24	6.53
Real GDP per capita, squared	-2.24	-5.46	0.77	-0.88	-4.38	2.42	-3.23	-0.49	-6.17
No explicit update	0.53	-0.05	1.15	0.53	-0.01	1.11	1.38	0.86	1.95
Own past ratification if explicit update	1.50	0.77	2.23	1.55	0.87	2.24	1.61	0.89	2.33
Population	-0.05	-0.14	0.03	-0.12	-0.23	-0.02	-0.02	-0.11	0.07
Internal pressure									
Democracy	0.12	-0.18	0.43	-0.14	-0.50	0.21	0.34	0.04	0.64
Left majority	-0.39	-1.04	0.19	-0.24	-0.87	0.33	-0.69	-0.09	-1.32
Vote against convention:									
Government	-0.15	-0.59	0.33	-0.10	-0.54	0.38	-0.22	-0.66	0.24
Employers	-0.15	-0.56	0.28	-0.14	-0.51	0.26	0.39	0.01	0.79
External pressure									
Development aid	-0.06	-0.11	-0.03	-0.09	-0.13	-0.05	-7.64	-3.85	-11.77
Worldbank loans	-4.20	-7.63	-0.82	-4.25	-7.76	-1.68	2.00	-1.10	4.93
IMF credits	6.23	2.03	10.10	6.64	2.41	10.68	3.92	-0.09	7.67
Exports	-0.76	-3.64	1.45	0.19	-2.27	1.86	-0.77	-3.58	1.19
Exports into industrialized countries	-0.07	-6.37	6.29	-4.11	-11.15	3.38	-0.12	-7.35	6.70
Exports into industrialized countries (non oil exporting countries)	-0.55	-6.13	5.28	2.25	-5.00	9.15	-0.84	-7.27	5.93
Non oil exporting country	0.19	-0.91	1.30	-0.30	-1.65	1.23	0.26	-1.02	1.59
BIC				47.38			120.88		

Note: Bold entries are significant at the 5% level.

to capture the domestic politics of developing countries. Votes of the employers against a convention enhance the probability of ratification. They do not seem to mobilize enough political power to prevent the ratification of a convention against which they have voted. Comparing these results to Boockmann's earlier study, the coefficients are similar to the ones he obtained for the Cox model without heterogeneity. They are, however, different from those obtained from stratified estimation.

The three models give closer results about external pressure variables. Development aid has a negative influence on the hazard. One reason for this is that countries receiving huge amount of this aid also have to cope

with acute development problems, and so cannot bear the economic cost induced by ratification. Worldbank loans seems to discourage ratification in the two first models. Maybe the conditions a country has to fulfill to obtain loans are antagonist with some conventions. Another reason is that like the development aid, the countries receiving loans cannot face the economic costs raised by the convention. IMF credits have the opposite effect, but is not significant in the model with two frailties.

We compare the models using the Bayesian Information Criterium (BIC) (Schwarz, 1978). It has been designed to find the most probably model given the data, and takes account of Occam's Razor. In other words, the more parsimonious model is chosen when two models fit the data equally well. Wasserman (2000) reports that under mild regularity conditions, BIC approximates the log Bayes factor. If one set the priors odds of each model to be equal, the Bayes factor is the posterior odds ratio off a model versus the other one. It indicates how many time a model is more likely than the other. The BIC is equal to 47.38 when comparing the standard Cox model to the model with one frailty, and 120.88 when comparing the model with two frailties to the one with one frailty.⁸ The Bayes factor between the last two models is equal to $\exp(120.88)$ giving thus strong evidence for the model with two random effects.

These results confirm those obtained by Boockmann (2001) about the cost variables, and yield more insights on the internal and external pressure variables.

6 Conclusion

Our study uses a Bayesian approach to estimate MPH models with different specifications of unobserved heterogeneity, the more detailed using two random effects. Rather than assuming a parametric form for the baseline hazard, we use Cox's partial likelihood semi-parametric approach, justified from a Bayesian viewpoint by Kalbfleisch (1978), to avoid misspecification problems. The models are estimated using Gibbs sampling. In order to simulate from posterior marginal densities, we also use other simulation-based computational algorithms such as the acceptance-rejection sampling.

Switching from the frequentist paradigm to a Bayesian approach does not seem to influence a lot estimation results a lot, because the findings confirm results of previous studies on ILO ratification behaviour, especially

⁸The formula used 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.

Boockmann (2001). That is, we can also conclude that it is important to control for the country under study. Some unobserved explanatory country-specific variables seem to have a huge influence on the ratification behaviour. Furthermore, there appears to be an even larger amount of heterogeneity among conventions. Each one is specific and cannot be treated as a perfect substitute for another, some of them being very consensual among member states of the ILO or not inducing important economic or political costs. At the same time, other conventions can be less easily ratified because they are more ambitious and raise too important costs for the ratifying countries. It is thus important to investigate states specific determinants of ratification, and go deeper in the understanding of the specificity of each convention.

References

- ALVAREZ, M., ET AL. (1996): “Classifying political regimes,” *Studies in Comparative International Development*, 31(2), 2–36.
- BOLSTAD, W., AND S. O. 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. (1972): “Discussion of Professor Cox’s paper,” *Journal of the Royal Statistical Society B*, 34, 216–217.
- CHAU, N., AND R. KANBUR (2001): “The adoption of international labor standards convention: who, when, why?,” in *Brookings trade forum: 2001*. Brookings Institution Press, Washington.
- 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 (with discussion),” *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*, 39, 1–38.
- DJURDJEVIC, D. (2000): “Estimation de modèles de durée multivariés: application à la ratification des conventions par les pays membres du BIT,” Mémoire de DEA, Université Louis Pasteur, Strasbourg.
- GELFAND, A., AND A. SMITH (1990): “Sampling based approaches to calculating marginal densities,” *Journal of the American Statistical Association*, 85, 398–409.
- GELMAN, A., AND D. RUBIN (1992): “Inference from alternative simulation using multiple sequences,” *Statistical Science*, pp. 457–472.
- 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.
- HORNY, G. (2001): “Estimation de modèles de durée multivariés au moyen de l’algorithme EM,” Mémoire de DEA, Université Louis Pasteur, Strasbourg.
- HOUGAARD, P. (2000): *Analysis of multivariate survival data*. Springer-Verlag, New York.
- KALBFLEISCH, J. (1978): “Non-parametric Bayesian analysis of survival time data,” *Journal of the Royal Statistical Society*, 40, 214–221.
- KALBFLEISCH, J., AND R. PRENTICE (1980): *The Statistical analysis of failure time data*. John Wiley, New York.
- LANCASTER, T. (1979): “Econometric methods for the duration of unemployment,” *Econometrica*, 47, 939–956.
- (1990): *The econometric analysis of transition data*. Cambridge University Press, Cambridge.
- LIANG, K.-Y., ET AL. (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*, Ser.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,” Discussion Paper No. 9722, Department of Statistics, University of Toronto.
- 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*. 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 the study 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.
- SPIEGELHALTER, D., A. THOMAS, AND N. BEST (2000): *Winbugs user manual and examples: version 1.3* Medical research Council of Biostatistics unit, Cambridge.
- THERNEAU, M., AND P. GRAMBSCH (2000): *Modeling Survival Data: Extending the Cox model*, Statistics for Biology and Health. Springer, New-York.

- 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*, vol. 5, chap. 55. James J. Heckman and Edward Leamer, Amsterdam.
- WASSERMAN, L. (2000): “Bayesian model selection and model averaging,” *Journal of Mathematical Psychology*, 44(1), 92–107.
- YAMAGUCHI, K. (1986): “Alternative approaches to unobserved heterogeneity in the analysis of repeatable events,” in *Sociological Methodology 1986*, ed. by N. Tuma. Jossey-Bass, Washington D.C.
- YAU, K., AND A. MCGILCHRIST (1998): “ML and REML estimation in survival analysis with time dependent correlated frailty,” *Statistics in Medicine*, 17, 1201–1213.