Discussion Paper No. 05-23

Bayesian Estimation of Cox Models With Non-Nested Random Effects: An Application to the Ratification of ILO Conventions by Developing Countries

Guillaume Horny, Bernhard Boockmann,
Dragana Djurdjevic and François Laisney

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 05-23

Bayesian Estimation of Cox Models With Non-Nested Random Effects: An Application to the Ratification of ILO Conventions by Developing Countries

Guillaume Horny, Bernhard Boockmann,
Dragana Djurdjevic and François Laisney

Non-technical summary

In this paper, we investigate the determinants of ratification of ILO conventions in a sample of 80 developing countries over the period from 1975 to 1995. Since ratifying ILO standards may raise labour costs, we expect that countries' economic conditions are crucially important in determining ratification behaviour. On the other hand, the interests of domestic political actors also have to be taken into account. Finally, pressure from foreign actors, including pressure from other international organisations, may be mobilised to speed up the ratification process.

A hazard rate model is introduced to explain the duration up to ratification. Pooling over the 29 conventions adopted during the observation period, we obtain a large dataset with three dimensions: country, convention, and time. An important issue in any cross-country study is to control for unobserved heterogeneity. Under certain assumption, the random effects model provides a consistent estimator. In our data, we are faced with a multi-level structure of unobserved effects. While time effects are estimated by a trend variable, we allow for both country-specific and convention-specific time-constant random effects. These effects are non-nested and assumed to operate independently from one another.

Since estimation by classical methods proved to be infeasible, we follow the Bayesian paradigm and use Markov Chain Monte Carlo (MCMC) methods. The Gibbs sampler is used to draw samples from the simulated posterior density of the model. Posterior means derived from this procedure confirm most of the results from an earlier study (Boockmann, 2001). In particular, GDP per capita and previous ratification of similar conventions raises the ratification probability. As opposed to the earlier study, democracy also influences ratification positively. The results suggest that external pressure is nonexistent in the ratification decision. Finally, the results confirm that it is very important to control for unobserved heterogeneity in this framework.

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

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

October 2004

Abstract

We use a multivariate hazard model for the analysis of data on the timing of ratifications of different ILO conventions by developing countries. The model accounts for two random effects, one at the country level and the other at the convention level. After investigating identification, we use a semi-parametric Bayesian approach based on the partial likelihood for the inference. Our findings confirm the results of preceding studies that ratification depends both on economic and political factors. Furthermore, the results yield insights on the impact of unobserved heterogeneity across member states and conventions on the ratification process.

Keywords: Gibbs sampling, partial likelihood, frailties, duration analysis
JEL Classification: C11, C14, C15, C41, D78, J80, O19

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

[†]Centre for European Economic Research (ZEW), PO Box 10 34 43. D-68034 Mannheim, Germany. boockmann@zew.de

[‡]Institut für Volkswirtschaftslehre, Empirische Wirtschaftsforschung und Angewandte Ökonometrie, Technische Universität Darmstadt, Residenzschloss, Marktplatz 15, 64283 Darmstadt, Germany. djurdjevic@vwl.tu-darmstadt.de

[§]BETA-*Theme*, University Louis Pasteur, Strasbourg I, 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 34 43. D-68034 Mannheim, Germany. laisney@zew.de

Acknowledgements: We thank Mathieu Brezovski and Viktor Winschel for detailed comments on previous versions of this paper, as well as participants at the Econometric Society European Meeting, August 2004, Madrid.

1 Introduction

In duration models, one rarely observes all determinants leading to an event. In this paper, we develop an approach for modelling unobserved heterogeneity for the case where each cross-sectional unit (individual) faces a fixed number of different risks. Both individuals and risks are assumed to have unobserved characteristics. As an application, we consider the ratification of ILO conventions by ILO member states (see Boockmann, 2001). There are 184 conventions in many different subject areas which member states are free to ratify. It seems reasonable to assume that there are differences in the frequency of ratification produced both by country-specific as well as convention-specific unobserved effects.

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 general way to deal with fixed effects is stratified partial likelihood as discussed in Kalbfleisch and Prentice (1980), Yamaguchi (1986) and Ridder and Tunalı (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 Tunalı discuss the merits of fixed and random effects in their application to child mortality rates. They also propose a Hausman 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 Ridder and Tunalı's (1999) 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 parameters. 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 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 an accelerated 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 Manda (2001) propose a Bayesian approach to estimate Sastry's (1997) model.

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 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 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.[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 for maximum likelihood and Bayesian estimators, the results underestimate the importance of both fixed and random effects. Although bias can be removed by boostrapping, the procedure proved to be computationally more intensive than Markov chain Monte Carlo estimation and sometimes failed to converge.

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. Furthermore, we show that the model is identified if each

[1]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 problems. Bolstad and Manda (2001) trace these difficulties to inaccurate rounding off in case of high variance.

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 a finite mean. 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 the conclusions of Boockmann (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). We also discuss the identification of the model with two frailties in our setting. 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 survival data analyzed in this study come from a database described in Boockmann (2001). The spells 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 (no OECD members), and we call them 'developing countries' for simplicity. It covers 29 conventions and a total of 228 ratifications. Considering spells ended by a ratification, the mean length of a spell is 8 years, but durations differ widely: nearly 20% of ratifications occur within 3 years after adoption while 20% of all spells last over 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 conventions (Brazil, Mexico and Uruguay) and 6 countries more than 8, while 26 members did not ratify any.

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

As regards conventions, some of them are ratified by a large number of countries. For instance, conventions numbered 144 and 159 in the ILO classification have been ratified 38 and 25 times, respectively. On the other side, conventions 143 and 157 have been ratified by only 6 countries and 1 country, respectively. We estimated the survival function for each convention using the Kaplan-Meier non-parametric estimator, and performed a log rank test to assess their equality. Based on the null hypothesis of no difference, we concluded that the cumulative probabilities of ratification are significantly different.[2] Interpreting this result as evidence of heterogeneity across conventions, we consider two types of random effects: a country-specific effect and a convention effect.

The explanatory variables may take different values depending on country and time. They capture influences on ratification behaviour often discussed in the political economy literature (for further elaboration and for exact data sources, see Boockmann, 2001). They fall into three categories: variables relating to the economic and administrative costs and benefits of ratification,

[2]The value of the test statistic was over 1000 and critical values are 39.38 at the 5% level and 48.28 at the 1% level.

to domestic political circumstances, and to external pressure for ratification.

In the first group, the level of GDP (measured in constant US-$) and the ratio of exports to GDP are used. Workers from richer countries may have a higher demand for labour standards, which increases the likelihood of ratification. By contrast, greater openness may make ratification more difficult since countries are more concerned about their international competitiveness. To allow for a non-monotonic impact on the hazard, GDP enters both linearly and quadratically. In previous estimations, higher-order polynomial terms were never found to be significant. The next two variables in this group are indicators. The first one specifies whether or not the convention under consideration is an explicit update of an existing convention (information was taken from the wording of the convention). The second one indicates whether or not a member state has ratified in the past a convention for which the convention under consideration is an explicit update. We expect that previous ratifications make current ratification less costly. Moreover, non-ratification of a previous convention, given that it exists, makes ratification less likely as compared to the case where no previous convention exists. The last variable in this category is total population, used as a 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 resources, equals one if there is a left wing parliamentarian majority. Two further variables refer to the vote of the national delegates (representing government and employers) at the adoption of the convention at the International Labour Conference. They are set as one if the delgate 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 by applied. The amount of development aid received each year by a country is the first explanatory variable of this group. We also consider IMF lending and World Bank credits. Exports towards industrialized democracies stand for countries' vulnerability with respect to trade sanctions. The reason is that trade sanctions would most likely be applied by industrialized countries and not by other developing countries.[3] As pressure is less likely if the country is an oil exporter, we include an indicator for OPEC members and interact it with exports towards industrialized democracies. All variables are measured in per cent of GDP. To deal with a potential endogeneity problem, we use a three year moving average lagged one period for these variables. We also create a set of indicators to take account of differences between the subjects of conventions. Readers interested in more details on the choice of these variables or their expected effect on the hazard function are referred to this Boockmann (2001), who also presents descriptive statistics including Kaplan-Meier estimations.

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 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.[4] We denote by t_{ij} the time between the adoption of the convention j during the International Labour Conference and its ratification by country 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

[3] 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.

[4] 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.

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.

We begin 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. 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 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:

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

where v_i denote the log country effect and w_j the log convention effect. Due to the universality of ILO standards, we assume the two effects to be uncorrelated. At least officially, there are no ILO conventions intended for certain regions in particular.

As regards identification, 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 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. In fact, the result is obtained without covariates and letting the part of the hazard not depending of the unobserved heterogeneity vary for each observation, that is, a much more general setting than the one considered in this application. 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.

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:

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

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:

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

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.

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 juste before t_{ij}, denoted by R_{ij}. Convention j ratified by country i at time t_{ij} contributes to the partial likelihood through:

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

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

[5] See Lancaster (1990) for a detailed discussion.

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 (2001), 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 deviations 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:[6]

$$v_i \sim N(0, \tau^2), w_j \sim N(0, \alpha^2). \tag{6}$$

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, which are nearly 90% of the sample, 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, the form of the posterior, and the estimation procedure using Gibbs sampling.

4.1 Prior and Posterior Distributions

The first stage of the inference is based on the partial likelihood, justified from the Bayesian viewpoint by Kalbfleisch (1978) and we recall briefly his approach in appendix (B). Basing his analysis on counting processes, Kalbfleisch 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

[6] We also tried to estimate a model with gamma distributions for the frailties but did not obtain convergence. This confirms the problem met by Djurdjevic (2000) and by Horny (2001) with the EM algorithm for the same model.

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. This approach is based on a limit argument in which both parameters of the gamma prior tend to zero. Therefore we assume in our application that the baseline hazard follows a gamma prior with both parameters equal to 0.001.[7] This assumption implies an expectation equal to 1 and a variance equal to 1000, 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. For 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.

The last stage concerns the choice of priors for the parameters of the unobserved heterogeneity distribution. 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 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.

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

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

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.

[7] 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.

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 implemented in the WinBUGS software and can be summarily described as follows.[8] 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 $i = 1 \ldots n$. 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 Gilks and Wild, 1992, or Robert and Casella, 1999, for detailed discussion on this sampler). The variances τ^2 and α^2 are simulated directly from an inverse gamma density.

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 the 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. Using the convergence diagnostic tool of Gelman and Rubin (1992) and quantile plots, we concluded that 5000 iterations were necessary for the burn-in period for the first two models, and 10000 for the third one. Estimation was performed with a 2.5 Ghz Pentium and it took one month for the standard Cox model, two months for the Cox model with one frailty, and 3 months for the Cox model with two frailties.[9] All posterior summary

[8]WinBUGS is freely available at http://www.mrc-bsu.cam.ac.uk/bugs/. See Spiegelhalter et al. (2000) for the manual and examples.

[9]Other studies (see for example Brooks and Morgan, 2004), and also our own experiments, show that WinBUGS is very slow when dealing with large datasets. For the simpler models, faster software is available and we also performed estimation using R 1.9.1. This required less than 3 seconds for the Cox model and less than 5 seconds for the model with

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	0.49	0.42	0.57
Twofold:	country effect	τ	0.49	0.42	0.57
	convention effect	α	0.85	0.66	1.12

Considering the confidence interval at the 5% level, we remark that the country effect is not influenced by the inclusion of the convention effect. This last one is significantly more important than the country effect, which means that 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. 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 ratification of a previous convention is significantly positive as expected, while non-ratification influences the hazard negatively as compared to the case where the present convention does not update a previous one. This hints to the presence of dynamics in the ratification process. The models with frailties indicates a non-monotonic impact of the real GDP per capita.[10] Comparing with Boockmann (2001), we remark that we have the same significant cost variables in both studies.[11]

Turning to internal pressure variables, none of them is significant in the specification without frailty terms. Considering two frailties 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

one frailty. We present the corresponding results in appendix C. Still, we are not aware of any alternative for the model with two frailties.

[10]For both, the profile is an inverted $\cup$ with a maximum at a GDP per capita of about \$6000, so that the relationship is essentially increasing and concave (the mean GDP per capita in the sample is \$2280).

[11]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 dramatically slowed down convergence.

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[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
Penalized log-likelihood		-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.

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 when controlling for unobserved heterogeneity.

Development aid has a negative influence on the hazard, in particular if unobserved heterogeneity is controlled for. An explanation may be that countries receiving 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. 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, these results suggest that external pressure is nonexistent in the ratification decision.

We computed penalized log-likelihoods (equal to the log-likelihood minus half the number of parameter of the model times the log of the number of observations) in order to compare the models on the basis of the Bayesian Information Criterium (BIC, Schwarz, 1978). 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 the differentiating penalized log-likelihood, 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.[12] 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 use the convention subject information 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

[12]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.

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 estimated 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.

Switching from the frequentist paradigm to a Bayesian approach does not seem to influence estimation results a lot, because the findings confirm results from previous studies on ILO ratification behaviour, especially Boockmann (2001). Our results confirm that it is important to control for the country under study. Some unobserved explanatory country-specific variables seem to have a large influence on the ratification behaviour. The results also show the presence of an even larger amount of heterogeneity among conventions. Some of them 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.

A Identification: proof

The proof is similar to the proof of the first theorem in Honoré (1993). Consider the case of two conventions submitted to country i. The model is:

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

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

The joint survivor function for country i is:

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

where $\mathcal{L}_\xi$ is the Laplace transform for ξ_i. Notice that $S_i(t_{i1}, t_{i2}|\psi_1, \psi_2)$ is observable by taking a large number of observations for a population homogeneous with respect to ψ_1 and ψ_2 (even if they are unknown, the way the population is clustered is known). By differentiating (A.3) over t_{i1} and t_{i2} and taking the ratio, we obtain:

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

Let us denote by k the quantity $1/\lambda_{0,1}(t_{i0})$. 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 have $k \Lambda_{0,1}(t) + c_1$, where c_1 is obtained by the initial condition $\Lambda_{0,1}(0) = 0$. Then, $\Lambda_{0,1}(t|X_i)$ is identified. 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 ϕ_1 and ϕ_2 are fixed, the survivor $S_i(t_{i1}, t_{i2}|\psi_1, \psi_2)$ 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. We have thus the identification of H_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 a and b and integrate it out:

$$\begin{aligned}
L(\beta) = & \prod_{i=1}^{I}\prod_{j=1}^{J_i} \xi_i\psi_j \exp\left[\beta' x_{ij}(t_{ij})\right] \int_0^\infty \lambda_0(t_{ij}) \exp\left[-\sum_{(k,l)\in R_{ij}} \lambda_{kl}(t_{ij}|x_{kl},\xi_k,\psi_l)\right] \\
& \lambda_0(t_{ij})^{a-1} \exp\left[-b\lambda_0(t_{ij})\right] d\lambda_0(t_{ij}) \\
= & \prod_{i=1}^{I}\prod_{j=1}^{J_i} \xi_i\psi_j \exp\left[\beta' x_{ij}(t_{ij})\right] \int_0^\infty \lambda_0(t_{ij})^{a} \exp\Bigg[-\lambda_0(t_{ij})\Big(b+ \\
& \sum_{(k,l)\in R_{ij}} \xi_i\psi_j \exp\left[\beta' x_{ij}(t_{ij})\right]\Big)\Bigg] d\lambda_0(t_{ij}) \\
\propto & \prod_{i=1}^{I}\prod_{j=1}^{J_i} \frac{\xi_i\psi_j \exp\left[\beta' x_{ij}(t_{ij})\right]}{\left(b+\sum_{(k,l)\in R_{ij}} \xi_i\psi_j \exp\left[\beta' x_{ij}(t_{ij})\right]\right)^{a}}.
\end{aligned} \tag{A.9}$$

And we obtain the partial likelihood (5) by letting $a, b \to 0$, that is assuming the baseline hazard follows a gamma non-informative prior.

C Results obtained by maximum partial likelihood and by penalized partial likelihood

The following table displays estimation results obtained by maximum partial likelihood for the standard Cox model and by penalised partial likelihood, as described in Therneau and Grambsch (2000), for the Cox model with one gamma frailty. They have been obtained using the software R 1.9.1, and the package 'survival' is required for penalised likelihood.[13]

Table 4: Estimates of the β parameters

Variable	**Standard Cox**		**Cox: one frailty**	
	Coef.	S.d	Coef.	S.d
Cost				
Real GDP per capita[a]	**3.91**	1.40	2.80	1.52
Real GDP per capita, squared	**-3.18**	1.50	-2.24	1.64
No explicit update	**0.95**	0.28	**0.96**	0.28
Own past ratification if explicit update	**1.39**	0.36	**1.41**	0.36
Exports[c]	-0.21	1.42	0.22	1.13
Population[b]	0.00	0.04	-0.04	0.05
Internal pressure				
Democracy	**0.38**	0.15	0.25	0.16
Left majority	**-0.68**	0.31	-0.58	0.32
Vote against convention:				
Government	-0.20	0.24	-0.16	0.24
Employers	0.24	0.21	0.23	0.21
External pressure				
Development aid[c]	**-7.21**	2.06	**-8.57**	2.25
Worldbank loans[c]	2.55	1.54	**3.15**	1.60
IMF credits[c]	3.78	1.93	3.93	2.02
Exports into industrialized countries[c]	-0.77	3.73	-2.48	4.02
Exports into industrialized countries (non oil exporting countries)[c]	-0.72	3.46	0.54	3.92
Non oil exporting country	0.13	0.67	-0.04	0.77

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.

[13] R is a free software available at http://www.r-project.org/.

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. (1972): "Discussion of Professor Cox's paper," *Journal of the Royal Statistical Society*, Series B, 34, 216–217.

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.

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.

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 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 Econométriques*, vol. 1-2. Economica, Paris, 2 edn.

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.

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.

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*, Series B, 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., 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," 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, pp. 3381–3463. J. J. Heckman and E. Leamer (eds.), Elsevier, 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. (2001): "Multilevel models for survival analysis with random effects," *Biometrics*, 57, 96–102.