Right Censoring in Interdependent Duration Models: The Possibility of Approximating a Joint Survivor Function Using Copulas

Aya Kachi*
University of Illinois

May 1, 2011
Research note prepared for the Visions in Methodology (VIM) Conference at Ohio State University.
Please do not cite without author's permission.

Abstract

In many topics in political science, understanding survival time and failure rates of political processes is crucial. Event history analysis in general offers a set of tools that allows us to assess whether a certain political process is at risk of experiencing some failure at a given time point. In a previous work with Jude Hays, we pointed out that the occurrence of multiple political events or a single political event occurring to multiple units might be interdependent across events or units Hays and Kachi (2009). For example, two political processes–cabinet formation duration and cabinet survival–might be endogenous to each other when the intensity of formation negotiation depends on the parties' anticipation about the durability of the future government. At the same time, the very durability of a formed government also reflects the diversity in parties' preferences in the negotiation phase. To handle this interdependence in the endogenous outcome variables, we developed an interdependent duration model, using a system of simultaneous equations. However, the approach taken in the previous work lacked a procedure to handle right-censored observations. The difficulty of writing the likelihood function with right-censored data comes from the fact that deriving the joint cdf for the likelihood involves multiple integrals with respect to all the duration variables. The joint cdf is necessary in order to define the joint survivor function, through which censored observations can contribute information to the likelihood; however, the aforementioned estimator was derived through an approach that obtained the joint pdf without the joint cdf. In this project, I explore the use of copula functions to recover a joint cdf from a given joint pdf in a relatively simple way. I evaluate the performance of my estimator by Monte Carlo simulations and illustrate the method in a study of the determinants of government formation duration and survival in European parliamentary democracies.

*Ph.D. student, Department of Political Science, University of Illinois. E-mail: akachi2@illinois.edu; URL: https://netfiles.uiuc.edu/akachi2/home.

In many topics in political science, understanding survival time and failure rates of political processes is crucial. Scholars in comparative politics, for example, have explained the survival and dissolution of cabinets in parliamentary democracies (King et al. 1990; Warwick 1992), the duration of political regimes (Chapman and Roeder 2007; Svolik 2008), and the timing of union-friendly labor reforms (Murillo and Schrank 2005). International relations scholars have examined the survival of military alliances (Bennett 1999), post-conflict peace duration (Fortna 2004; Werner and Yuen 2005), and the speed at which policies diffuse around the world (Simmons and Elkins 2004). In American politics, research has explored the time until major pieces of legislation are amended (Maltzman and Shipan 2008), the duration of Supreme Court nominations (Shipan and Shannon 2003), and the timing of issue position taking in Congress (Box-Steffensmeier et al. 1997; Boehmke 2006; Darmofal 2009). And these examples only scratch the surface.

There is now recognition that many of the durations that we want to explain are interdependent in various ways (Boehmke et al. 2006; Boehmke 2006; Quiroz Flores 2008). The first kind of interdependence is when the time to one political event depends on the time to another related event. An example for this type of interdependent durations includes the time it takes to negotiate an international treaty and the survival of that agreement, the time it takes to create a new constitution and the length of its survival, the amount of time it takes an individual to form political attitudes and the stability of those beliefs, the amount of time it takes to confirm a bureaucrat and the duration of his or her tenure in office, and the length of time a peacekeeping mission is in place and the duration of the post-mission peace. The second type of interdependence is when the time to a particular political event for one actor depends on the time to that same event for other actors. For example, the time it takes states to enter wars, alliances, and international organizations depends on the time it takes other states to make these decisions. The entry and exit decisions of political candidates in electoral contests depend on the timing of their opponents. If policies diffuse across countries, the time it takes one country to adopt a particular policy depends on the adoption timing of other states. The time it takes a political candidate to initiate negative campaigning most likely depends on the timing of other candidates' decisions to run negative campaign ads.

In a previous work with Jude Hays, we developed an interdependent duration model, using a system of simultaneous equations (Hays and Kachi 2009). One of the key contributions of this previous work was to place these two kinds of interdependent durations within a single unified econometric framework. More importantly, and more relevantly to the methodological extension that I undertake in this paper, our approach models the structure of dependency among multiple durations as scholars often have in mind. Instead of treating the inter-duration dependency as nuisance in estimating the effects of covariates on the durations (as in the existing multi-duration models), we directly model the dependency by expressing a duration as a function of the other durations and provide a method for estimating the inter-duration dependency, as well as the effects of covariates on each duration.

The previous work on interdependent durations, however, provides only a basic estimation method to deal with multiple duration processes. For example, this estimator is still useful if the data are uncensored and there are no time-variant covariates. However, social scientists are not always fortunate enough to have such "low-maintenance" data. In this paper, I focus on the modification of the likelihood to accommodate data with right-censoring.

The difficulty of writing the likelihood function with right-censored data, particularly for models with interdependent durations, stems from the fact that deriving the joint survivor function included in the likelihood—or equivalently the joint cdf because the survivor function is defined as $S(\mathbf{y}) = 1 - F(\mathbf{y})$—involves multiple integral with respect to all the duration variables. The multiple integral become a challenge only when we have right-censored data, because observations can contribute information to the likelihood through the joint survivor function (and not the joint pdf) only when these observations are right-censored. This multiple integral makes estimation extremely burdensome. Hays (2009) discusses this problem and suggest a possible estimation method using recursive importance sampling (RIS). This strategy uses RIS as simplification of brute-force numerical search for the maximum likelihood. In this paper I take a different approach to approximate the likelihood by using copula functions to recover a joint cdf from a given joint pdf. The major advantage of this approach is the estimability of the likelihood without concerns about the computational burden, especially in the cases where we have a number of interdependent duration processes in the model.

I evaluate the performance of my estimator by Monte Carlo simulations and illustrate the method in a study of the determinants of government formation duration and survival in European parliamentary democracies.[1]

This paper is organized as follows. First, I review the set-up of the interdependent duration model developed in the previous work. Second, I present a general form of the likelihood function with right-censored cases. Third, I demonstrate a way to approximate the joint survivor function using a copula function. To exemplify the use of a copula, I will focus on a specific case where there are only two duration processes and each follows a Weibull distribution. Fourth, I evaluate the performance of this estimator using Monte Carlo experiments.Finally, I estimate a simultaneous durations model of government formation and survival.[2] I conclude with a discussion about the related issues to consider in the next phase of this project.

1 Review of the SEQ Interdependent Duration Models

In this section, I review the basic structure of the simultaneous equations model for interdependent duration processes developed in a previous work (**?**). It should clarify how the full information maximum likelihood estimator was constructed and how the likelihood without right-censoring appears in general.

1.1 Linear Parameterization of Weibull Durations (The AFT Model)

The dependent variables of interest, $\mathbf{y}^*$, are D distinct duration processes that have Weibull distributions with two parameters.

$$y^*_{id} \sim Weibull(\lambda_d, \theta_d), \tag{1}$$

where $i = \{1, \cdots, N\}$ denotes the observational-unit index and $d = \{1, \cdots, D\}$ denotes the duration index, implying that there are $N \times D$ observations in total. The notation λ is the shape parameter and θ is the scale parameter. These distributional parameters take the common values across N observational units; hence they have only one subscript that indicates duration process. A common way to parameterize a Weibull model of D interdependent durations is to log-linearize the model and obtain a log-linear system of D equations (Box-Steffensmeier and Jones 2004). It is also known that the logged Weibull variable turns out a standard Gumbel variable that is scaled by the shape parameter in the original Weibull

[1] At the moment, this is a “forthcoming” section. Also, I would like to note that this particular methodological innovation was originally motivated by the main substantive topic of my own dissertation, interdependence between the duration of democratic transition and the survival of democracy. At the moment, I am not emphasizing this substantive topic due to the on-going definitional and data-collection debates...

[2] Incomplete. Forthcoming.

distribution. [3] For example, in the univariate Weibull case, the log-linear form would look like;

$$\begin{aligned} y = \ln y^* &= \ln\theta + \frac{1}{\lambda}\varepsilon \\ &= \mathbf{X}\boldsymbol{\beta} + \frac{1}{\lambda}\varepsilon, \end{aligned} \tag{2}$$

where $\varepsilon \sim ExtremeValueI\ (StandardGumbel)$ and we define $y = \ln y^*$. The second line of equation (2) shows how we could include covariates, by making the Weibull scale parameter, θ, a function of the covariates, $\theta = e^{\mathbf{X}\boldsymbol{\beta}}$. For further detail regarding the link between a Weibull and an extreme value distribution, see Appendix 2.

1.2 The System

The system of D distinct durations with N observational units in matrix notation is

$$\mathbf{y}_{(ND\times 1)} = \mathbf{A}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{L}\mathbf{u}. \tag{3}$$

The dependent variable, $y_{id} = \ln \mathbf{y}^*_{id}$, is a logged Weibull random variable. The vector $\mathbf{y}$ is a stack of D vectors, each of which contains N observational units.

$$\mathbf{y}_{(ND\times ND)} = \begin{pmatrix} \mathbf{y}_{.1} \\ \vdots \\ \mathbf{y}_{.D} \end{pmatrix}, \text{ where } \mathbf{y}_{.d(N\times 1)} = \begin{pmatrix} y_{1d} \\ \vdots \\ y_{Nd} \end{pmatrix}.$$

The matrix $\mathbf{A}$ is the coefficient matrix for the dependence. An element matrix $\boldsymbol{\alpha}^{d'}_{.d}$ contains coefficients representing the effects of the second duration d on the first duration d'. The diagonal elements $\boldsymbol{Sp}$'s in the $\mathbf{A}$ matrix are the matrices that capture the "spatial" dependency. This is the dependency among N observational units within each duration process. We call it "spatial" dependency for convenience, because the linear system captures the among-unit dependency using weights matrices just like in spatial contexts. Note that $\boldsymbol{Sp}^d = \mathbf{0}$ for all d when one assumes no among-unit dependency. Similarly $\boldsymbol{\alpha}^{d'}_{.d} = \mathbf{0}$ when one assumes no dependency between duration d and d'.

$$\mathbf{A}_{(ND\times ND)} = \begin{pmatrix} \boldsymbol{Sp}^1 & \boldsymbol{\alpha}^1_{.2} & \cdots & \boldsymbol{\alpha}^1_{.D} \\ \boldsymbol{\alpha}^2_{.1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{\alpha} \\ \boldsymbol{\alpha}^D_{.1} & \cdots & \boldsymbol{\alpha}^D_{.D-1} & \boldsymbol{Sp}^D \end{pmatrix},$$

[3] The standard Gumbel distribution is a special case of the type-I extreme value (minimum) distribution. The distribution and density functions of the type-I extreme value (minimum) distribution are

$$\begin{cases} f(u) = \frac{1}{b} e^{\frac{u-a}{b}} e^{-e^{\frac{u-a}{b}}} \\ F(u) = 1 - e^{-e^{\frac{u}{b}}}, \end{cases}$$

where a is the location parameter and b is the scale parameter. The distribution and density functions of the standard Gumbel distribution are

$$\begin{cases} f(u) = e^u e^{-e^u} \\ F(u) = 1 - e^{-e^u}. \end{cases}$$

Note that the standard Gumbel distribution is a special case of the type-I extreme value distribution, where $a = 0$ and $b = 1$. A logged Weibull variable has the type-I extreme value distribution in general and only the scaling of the resulting extreme value variable varies depending on how one sets the scale parameter of the extreme value variable. For further details, see Appendix 2.

where

$$\boldsymbol{\alpha}^{d'}_{.d(N\times N)} = \begin{pmatrix} \alpha^{d'}_{.d} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \alpha^{d'}_{.d} \end{pmatrix}, \boldsymbol{Sp}^{d}_{(N\times N)} = \begin{pmatrix} 0 & \alpha^{d}_{(1,2)} & \cdots & \alpha^{d}_{(1,N)} \\ \alpha^{d}_{(2,1)} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \alpha^{d}_{(N-1,N)} \\ \alpha^{d}_{(N,1)} & \cdots & \alpha^{d}_{(N,N-1)} & 0 \end{pmatrix}$$

The vector $\mathbf{x}$ denotes a set of covariates and the supscript indicates to which equation the covariate vector is specific. Each vector $\mathbf{x}$ contains K covariates with coefficients denoted by β. The subscript of $\mathbf{X}$, $.d$, indicates that these x's affect duration d, and the number of covariates, i.e., the number of elements in each $\mathbf{X}_{.d}$ is denoted K_d. The error term u_{it} in this structural form is i.i.d. with the extreme value minimum distribution. The error term is multiplied by $\lambda^{-1}_{.t}$, which is the shape parameter of the original Weibull distribution and the value of λ is allowed to vary across duration processes.

$$\mathbf{X}_{(ND\times(K_0+\cdots+K_D))} = \begin{pmatrix} \mathbf{X}_{.1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{.2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{X}_{.TD} \end{pmatrix}, \text{ where } \mathbf{X}_{.d(N\times K_d)} = \begin{pmatrix} x^{1}_{1d} & \cdots & x^{K_d}_{1d} \\ \vdots & \ddots & \vdots \\ x^{1}_{Nd} & \cdots & x^{K_d}_{Nd} \end{pmatrix}$$

$$\boldsymbol{\beta}_{(K_0+\cdots+K_D\times 1)} = \begin{pmatrix} \beta^{1}_{.1} & \cdots & \beta^{K_1}_{.1} \,|\, \beta^{1}_{.2} & \cdots & \beta^{K_2}_{.2} \,|\, \cdots & \cdots \,|\, \beta^{1}_{.D} & \cdots & \beta^{K_D}_{.D} \end{pmatrix}';$$

$$\mathbf{L}_{(ND\times ND)} = \begin{pmatrix} \mathbf{L}_{.1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{L}_{.D} \end{pmatrix}, \text{ where } \mathbf{L}_{.d(N\times N)} = \begin{pmatrix} \frac{1}{\lambda_{.d}} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \frac{1}{\lambda_{.d}} \end{pmatrix};$$

$$\mathbf{u}_{(ND\times 1)} = \begin{pmatrix} u_{11} \\ \vdots \\ u_{ND} \end{pmatrix}.$$

The following reduced form can be derived from the structural form (3);

$$\begin{aligned} \mathbf{y}_{(ND\times 1)} &= (\mathbf{I}-\mathbf{A})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I}-\mathbf{A})^{-1}\mathbf{L}\mathbf{u} \\ &= \boldsymbol{\Gamma}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Gamma}\mathbf{L}\mathbf{u} \\ &= \boldsymbol{\Gamma}\mathbf{X}\boldsymbol{\beta} + \mathbf{v}, \end{aligned} \tag{4}$$

where $\boldsymbol{\Gamma} = (\mathbf{I}-\mathbf{A})^{-1}$ and $\mathbf{v} = \boldsymbol{\Gamma}\mathbf{L}\mathbf{u}$.

1.3 Deriving the Likelihood via Change of Variables

The only task left before writing a likelihood function is to derive the joint density of y's. We do not know the joint distribution of y's, but fortunately it is not hard to obtain the joint distribution of u's, because they are assumed to be i.i.d and we know that the marginal of u has the type I extreme value distribution. We use the change of variables theorem to derive the joint pdf of y's from the joint pdf of u's. By solving equation (4) for $\mathbf{u}$, we have

$$\mathbf{u}_{ND\times 1} = g^{-1}(\mathbf{y}) = (\boldsymbol{\Gamma}\mathbf{L})^{-1}\mathbf{y} - \mathbf{L}^{-1}\mathbf{X}\boldsymbol{\beta}. \tag{5}$$

The Jacobian matrix of $g^{-1}(\mathbf{y})$ is

$$\mathbf{J} = \begin{pmatrix} \frac{\partial g_{11}^{-1}(\mathbf{y})}{\partial y_{11}} & \cdots & \frac{\partial g_{11}^{-1}(\mathbf{y})}{\partial y_{ND}} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_{ND}^{-1}(\mathbf{y})}{\partial y_{11}} & \cdots & \frac{\partial g_{ND}^{-1}(\mathbf{y})}{\partial y_{ND}} \end{pmatrix}.$$

If the inverse vector function, $(u_{11}, \cdots, u_{ND}) = g^{-1}(y_{11}, \cdots, y_{ND})$, exists for all $\mathbf{y} = (y_{11}, \cdots, y_{ND})$ such that $\mathbf{y} \in \{\mathbf{y} = g(\mathbf{u})\}$, the joint density of $\mathbf{Y} = g(\mathbf{U})$ is given by

$$\begin{aligned} h(y_{11}, \cdots, y_{ND}) &= \begin{cases} f(g_{11}^{-1}(y_{11}, \cdots, y_{ND}), \cdots, g_{ND}^{-1}(y_{11}, \cdots, y_{ND}))|det(\mathbf{J})| \\ 0, \text{ otherwise} \end{cases} \\ &= \begin{cases} f(u_{11}, \cdots, u_{ND})|det(\mathbf{J})| \\ 0, \text{ otherwise} \end{cases} \\ &= \begin{cases} f(u_{11})f(u_{12})\cdots f(u_{ND})|det(\mathbf{J})| \\ 0, \text{ otherwise.} \end{cases} \end{aligned} \tag{6}$$

The last line in equation (6) follows from the i.i.d. assumption of u, and each $f(u_{id})$ is the standard Gumbel pdf.

2 Derivation of the Likelihood Function with Right-Censored Observations

If the joint survivor function $S(y_{11}, \cdots, y_{ND})$ is known, the likelihood function with right-censoring is

$$\begin{aligned} L &\propto \Big\{h(y_{11}, \cdots, y_{ND})\Big\}^{\delta_i}\Big\{S(y_{11}, \cdots, y_{ND})\Big\}^{1-\delta_i} \\ &= \prod_{i=1}^{N}\Big\{h(y_{i1}, \cdots, y_{iD})\Big\}^{\delta_i}\Big\{S(y_{i1}, \cdots, y_{iD})\Big\}^{1-\delta_i} \quad \text{b/c we assume } N \text{ interdependent observations} \\ &= \prod_{i=1}^{N}\Big\{\prod_{d=1}^{D} f(g^{-1}(y_{id}))\Big\}^{\delta_i}\Big\{S(y_{i1}, \cdots, y_{iD})\Big\}^{1-\delta_i}|det(\mathbf{J})| \\ &= \prod_{i=1}^{N}\Big\{\prod_{d=1}^{D} f(u_{id})\Big\}^{\delta_i}\Big\{S(y_{i1}, \cdots, y_{iD})\Big\}^{1-\delta_i}|det(\mathbf{J})| \end{aligned} \tag{7}$$

This is the exact expression of the likelihood only if the joint survivor function is also derived from the change-of-variables method as it is the case for the joint pdf. However, unfortunately the survivor function is not derived in a way entirely consistent with the joint pdf, because the joint cdf, which in turn provides the survivor function as $S(\mathbf{y}) = 1 - F(\mathbf{y})$, is not uniquely recoverable from the joint pdf. The consequence is that $|det(\mathbf{J})|$ in equation (7) might carry too much or too little weight for the portion of the likelihood in which the S function, instead of the f function, is contributing. I will further discuss the weight that $|det(\mathbf{J})|$ should carry in a later section.[4]

[4]Note to myself: I will later consider a weighted version, $\frac{\sum \delta_i}{N}|det(\mathbf{J})|$.

3 Approximating a Joint CDF Using a Copula Function

The main methodological innovation of this project is a less burdensome way to obtain a joint cdf for the interdependent duration variables. To recap the process up to the likelihood derivation, we start with duration dependent variables with certain probability distributions (assumed based on conventions and the substantive theory) and log-linearize the dependent variables to construct a linear parametric model of the covariates' effects on the durations. Each error term of the system of log-linearized equations has some (univariate) probability distribution. For example, if the original duration variable is assumed to be Weibull distributed (as in the example that I show in this paper), each error term has the type I extreme value distribution (EVD). To construct the likelihood function of the logged dependent variables, we need to derive the joint density of the dependent variables. This is done through the change-of-variables method, using the fact that each error term has a known distribution *and* these errors are independently distributed. This method allows us to construct a system of duration equations, which maintains the structure of causal theories that scholars often have in mind, but it does not directly provide the joint cdf of the dependent variables. To obtain the joint cdf of durations, $F(\mathbf{y})$, which is necessary to write the survivor function $S(\mathbf{y}) = 1 - F(\mathbf{y})$, one has to take the integrals of the joint pdf with respect to all the duration processes, $y_1, \cdots, y_D$. From equation (**??**), the exact expression for the likelihood would be

$$L = \prod_{i=1}^{N} \Big\{ \prod_{d=1}^{D} f(u_{id}) \Big\}^{\delta_i} \Big\{ 1 - \overbrace{\Big(\int_0^{u_{i1}} \cdots \int_0^{u_{iD}} \prod_{d=1}^{D} f(z_{id}) dz_{i1} \cdots dz_{iD} \Big)}^{F((y))} \Big\}^{1-\delta_i} |det(\mathbf{J})|. \tag{8}$$

The multiple integral makes estimation extremely burdensome. Hays (2009) discusses the difficulty due to multiple integral in detail and suggest a possible estimation method using recursive importance sampling (RIS). This strategy uses RIS as simplification of brute-force numerical search for the maximum likelihood.

To alleviate this estimation burden caused by the multiple integral and make interdependent duration models more usable for applied researchers, I explore the use of copula functions to recover a joint cdf from a given joint pdf.

3.1 Copula Functions

A copula is a function that gives a proper joint distribution function from univariate marginal distribution functions (Nelsen 2006). Several papers in political science use copulas or copula related distributions to derive likelihoods for empirical analysis including Boehmke et al. (2006), Boehmke (2006), Quiroz Flores (2008) and Fukumoto (2009) among others. The primary advantage of using copulas is that one has the joint distribution function, which is necessary to construct many likelihoods, including the likelihoods for interdependent duration models.[5]

Since "copula" is merely a category of functions that gives a proper joint distribution function from univariate marginal distribution functions, a number of different copulas that have different dependency structures among random variables have been studied. To exemplify the use of copula functions to construct the interdependent duration likelihood, I focus on one of the most commonly used copulas, Farlie-Gumbel-Morgenstern (FGM) copula, and the case where we have exactly two interdependent random variables ($D = 2$).[6] First, consider a joint distribution function of random variables y_1^* and y_2^*

[5] Other examples include the likelihoods for qualitative or limited dependent variables models.

[6] The interdependent duration model constructed in the previous section is general enough to include as many duration processes as one wishes in theory (putting aside the estimation burden), and multivariate copulas have been studied and

generated from the following FGM copula

$$F(y_1^*, y_2^*) = F(y_1^*)F(y_2^*)[1 + \alpha\{1 - F(y_1^*)\}\{1 - F(y_2^*)\}], \tag{9}$$

where α, the association parameter, captures the degree of dependence between the two y^*'s and $-1 \leq \alpha \leq 1$. The corresponding joint density function is given as

$$f(y_1^*, y_2^*) = f(y_1^*)f(y_2^*)[1 + \alpha\{2F(y_1^*) - 1\}\{2F(y_2^*) - 1\}]. \tag{10}$$

Note that the degree of dependency in a copula is captured only by a single α. It implies that this association parameter α in the copula is some "combination" (or a function) of the two dependency parameters we defined in our duration model based on our substantive theory, which tells us that the first duration depends on the second duration ($y_1 = f(\alpha_2 y_2)$) and vice versa ($y_2 = f(\alpha_1 y_1)$).

In other words, if one is not interested in a substantive-theoretical dependency structure between the two durations, and decides to treat such dependency as nuisance, then it is possible to use a copula (which captures the "overall" dependency by a single α) to construct a joint pdf and write the likelihood for the two durations simply as;

$$L(\mathbf{X}, \boldsymbol{\beta}, \lambda_1, \lambda_2 | y_1^*, y_2^*) = \prod_{i=1}^{N} f(y_{i1}^*, y_{i2}^*). \tag{11}$$

In fact, this is the duration (seemingly unrelated) SUR model used in the existing multi-duration studies (Boehmke et al. 2006; Boehmke 2006; Quiroz Flores 2008).

3.2 Copula-Based Likelihood vs. Change-of-Variable Likelihood

Naturally, one might wonder about the potential connection between the likelihood written with the joint pdf from a copula and the likelihood written with the pdf derived through the change of variables theorem. It is useful to compare the copula-based likelihood with the change-of-variables likelihood for the simple case of two duration processes that follow specific probability distributions (Weibull in this case), in which there are no covariates, implying $\mathbf{X}\boldsymbol{\beta} = 0$ or equivalently $\theta_d = 1$. To see the relationship of the two approaches, it is sufficient to consider the case where there is only one observation point ($N = 1$). For this reason, I will omit the subscript that indicates observational unit (i) and include only the duration subscript (d).

First, consider an SEQ model for logged-duration dependent variables with no covariates, which takes the form

$$\begin{aligned} &\begin{cases} \ln y_1^* = \alpha_2 \ln y_2^* + \frac{1}{\lambda_1} u_1 \\ \ln y_2^* = \alpha_1 \ln y_1^* + \frac{1}{\lambda_2} u_2 \end{cases} \\ \Leftrightarrow &\begin{cases} u_1 = (\ln y_1^* - \alpha_2 \ln y_2^*)\lambda_1 \\ u_2 = (\ln y_2^* - \alpha_1 \ln y_1^*)\lambda_2, \end{cases} \end{aligned} \tag{12}$$

where $\ln y_d^*$ denotes a dependent variable. The variable y_d^* measures a spell of time and it is assumed to have the FGM Weibull distribution. It is known that a logged Weibull is a type I EVD. The distribution

characteristics are well known. Therefore it would be possible to generalize the likelihood for the cases in which we have more than two interdependent duration variables, and the estimation results should be compared using different copulas in the future research.

and density functions of the type-I extreme value (minimum) distribution are

$$\text{Type I EVD} \begin{cases} f(u) = \frac{1}{b} e^{\frac{u-a}{b}} e^{-e^{\frac{u-a}{b}}} \\ F(u) = 1 - e^{-e^{\frac{u}{b}}}, \end{cases} \tag{13}$$

where a is the location parameter and b is the scale parameter.

The Jacobian for $\mathbf{u}$ is

$$\mathbf{J} = \begin{pmatrix} \frac{\partial u_1}{\partial y_1^*} & \frac{\partial u_1}{\partial y_2^*} \\ \frac{\partial u_2}{\partial y_1^*} & \frac{\partial u_2}{\partial y_2^*} \end{pmatrix} = \begin{pmatrix} \frac{\lambda_1}{y_1^*} & -\frac{\alpha_2 \lambda_1}{y_2^*} \\ -\frac{\alpha_1 \lambda_2}{y_1^*} & \frac{\lambda_2}{y_2^*} \end{pmatrix}. \tag{14}$$

$$|det(\mathbf{J})| = \frac{\lambda_1 \lambda_2}{y_1^* y_2^*} |1 - \alpha_1 \alpha_2|. \tag{15}$$

From equation (**??**), the following is the exact expression of the likelihood (without right-censored cases) derived by the change-of-variables approach.[7]

$$\begin{aligned} f_{cvt}(y_1, y_2) &= \Big(\prod_{d=1}^{2} f(u_d) \Big) |det(\mathbf{J})| \\ &= \lambda_1 \lambda_2 y_1^{*\lambda_1 - 1} y_2^{*\lambda_2 - 1} e^{-2(y_1^{*\lambda_1} + y_2^{*\lambda_2})} e^{y_1^{*\lambda_1} + y_2^{*\lambda_2}} |1 - \alpha_1 \alpha_2|. \end{aligned} \tag{16}$$

Now let's turn to the copula-based likelihood with Weibull durations. To construct the joint pdf of two Weibull distributions, one can simply plug the Weibull marginal pdf and cdf into the FGM copula (equation (10)). The univariate Weibull cdf and pdf are

$$\text{Weibull} \begin{cases} F(y_d) = 1 - e^{-(\frac{y_d}{\theta_d})^{\lambda_d}} \\ f(y_d) = \frac{\lambda_d}{\theta_d} (\frac{y_d}{\theta_d})^{\lambda_d - 1} e^{-(\frac{y_d}{\theta_d})^{\lambda_d}}; d = 1, 2, \end{cases} \tag{17}$$

where $\lambda_d > 0$ and $\theta_d > 0$. The resulting likelihood function with the joint pdf derived using equation (10) is

$$\begin{aligned} f_{copula}(y_1, y_2) &= \lambda_1 \lambda_2 y_1^{*\lambda_1 - 1} y_2^{*\lambda_2 - 1} e^{-2(y_1^{*\lambda_1} + y_2^{*\lambda_2})} \\ &\quad \times [4\alpha - 2\alpha e^{y_1^{*\lambda_1}} - 2\alpha e^{y_2^{*\lambda_2}} + (1 + \alpha) e^{y_1^{*\lambda_1} + y_2^{*\lambda_2}}]. \end{aligned} \tag{18}$$

Finally, by comparing equation (16) and (18),

$$\begin{aligned} & f_{cvt}(y_1, y_2) = f_{copula}(y_1, y_2) \\ & \Leftrightarrow \alpha = \frac{|1 - \alpha_1 \alpha_2| - 1}{4e^{-y_1^{*\lambda_1} - y_2^{*\lambda_2}} - 2e^{-y_1^{*\lambda_1}} - 2e^{-y_2^{*\lambda_2}} + 1}. \end{aligned} \tag{19}$$

In order for the covariance structure that is implied by the copula to be consistent with the structural SEQ relationship among the endogenous variables, equation (19) must hold.

[7] The transformation in equation (16) uses the fact that $e^{u_i} = y_i^{*\lambda_i}$.

3.3 Approximating the Change-of-Variable Likelihood with Right-Censored Observations Using a Copula

Recall that the difficulty in estimating the distribution parameters in the likelihood comes from the multiple integral in the survivor function for censored cases (see equation (8)). I suggest to use copula based joint cdf derived from (9) with the α replaced by the expression in (19).

First, the bivariate Weibull cdf derived from (9) and (17) is

$$F(y_1^*, y_2^*) = (1 - e^{-(\frac{y_1^*}{\theta_1})^{\lambda_1}})(1 - e^{-(\frac{y_2^*}{\theta_2})^{\lambda_2}})(1 + \alpha e^{-(\frac{y_1^*}{\theta_1})^{\lambda_1} - (\frac{y_2^*}{\theta_2})^{\lambda_2}}) \tag{20}$$

where $y_1^* \geq 0$, $y_2^* \geq 0$, $-1 \leq \alpha \leq 1$, $\theta_1 > 0$, $\theta_2 > 0$, $\lambda_1 > 0$ and $\lambda_2 > 0$.

Replacing the association parameter α with the expression in (19), the joint cdf becomes

$$F_{hybrid}(y_1^*, y_2^*) = (1 - e^{-(\frac{y_1^*}{\theta_1})^{\lambda_1}})(1 - e^{-(\frac{y_2^*}{\theta_2})^{\lambda_2}})(1 + \frac{|1 - \alpha_1\alpha_2| - 1}{4e^{-y_1^{*\lambda_1} - y_2^{*\lambda_2}} - 2e^{-y_1^{*\lambda_1}} - 2e^{-y_2^{*\lambda_2}} + 1} e^{-(\frac{y_1^*}{\theta_1})^{\lambda_1} - (\frac{y_2^*}{\theta_2})^{\lambda_2}}) \tag{21}$$

Since the survivor function is $1 - F_{hybrid}(\mathbf{y})$, the approximated (by a copula) likelihood becomes

$$L = \prod_{i=1}^{N} \Big\{ \prod_{d=1}^{D} f(u_{id}) \Big\}^{\delta_i} \Big\{ 1 - F_{hybrid}(y_{i1}, \cdots, y_{iD}) \Big\}^{1-\delta_i} |det(\mathbf{J})| \tag{22}$$

4 Monte Carlo Evaluations of the Hybrid Estimators

[Forthcoming.] To evaluate the performance of this estimator, I need to conduct a set of Monte Carlo simulations. I am currently facing some difficulty finding the correct way to general data with interdependent durations with right-censoring.

5 Discussion: Short- and Mid-Term Tasks

This is a preliminary draft of the initial step of my project that attempts to provide a relatively simple statistical inference tool for interdependent duration data with right censoring. There are some important issues to be considered further.

1. **"Weight" on the $|det(\mathbf{J})|$ term**

   In equation (**??**), if the joint pdf and the joint survivor functions are both derived through the change-of-variables method, then it is correct to leave in the $|det(\mathbf{J})|$ term, which directly comes from the change-of-variables theorem. However, my suggestion in this paper is to derive the survivor function—the portion in the likelihood function that contributes only when the observation is a right-censored case ($\delta_i = 0$)—using a copula because the change of variables theorem does not provide us with the joint cdf. The consequence is that the $|det(\mathbf{J})|$ term carries too much weight in the likelihood function. There needs to be a weight correction for the *det* term. How?

One possibility is to weight $|det(\mathbf{J})|$ straightforwardly by the proportion of uncensored cases as $\frac{\sum \delta_i}{N}|det(\mathbf{J})|$. Obviously this is a rough approximation.

Another possibility that I would like to explore is the manipulation of the Jacobian matrix itself. An element of the Jacobian matrix in the change-of-variable method captures the amount of the change in the new variable induced by a unit change in the old variable. If i'th observation is right-censored and not contributing to the likelihood, then the i'th row in the Jacobian matrix should not be contributing either. If this is true, then we force 0 in the row i of the Jacobian matrix whenever ith observation is right-censored.[8]

2. **Choice of copula**

   I used the FGM copula as the first example here to recover a joint cdf, but there are several more well-known and well-studied copula functions available. I will need to conduct simulations to show the difference in the estimator performance based on different copulas.

References

Bennett, D. S. (1999). Parametric Models, Duration Dependence, and Time-Varying Data Revisited. *American Journal of Political Science*, 43(1):256–270.

Boehmke, F. J. (2006). The Influence of Unobservable Factors on Position Timing and Content in the NAFTA Vote. *Political Analysis*, 14(4):421–438.

Boehmke, F. J., Morey, D. S., and Shannon, M. L. (2006). Selection Bias and Continuous-Time Duration Models: Consequences and a Proposed Solution. *American Journal of Political Science*, 50(1):192–207.

Box-Steffensmeier, J. M., Arnold, L. W., and Zorn, C. J. (1997). The Strategic Timing of Position Taking in Congress: A Study of the North American Free Trade Agreement. *American Political Science Review*, 91(2):324–338.

Box-Steffensmeier, J. M. and Jones, B. S. (2004). *Event History Modeling: A Guide for Social Scientists*. Cambridge University Press.

Chapman, T. and Roeder, P. G. (2007). Partition as a Solution to Wars of Nationalism: The Importance of Institutions. *American Political Science Review*, 101(4):677–691.

Darmofal, D. (2009). Bayesian Spatial Survival Models for Political Event Processes. *American Journal of Political Science*, 53(1):241–257.

Fortna, V. (2004). Interstate Peacekeeping: Causal Mechanism and Empirical Effects. *World Politics*, 56:481–519.

Fukumoto, K. (2009). What Happens Depends on When It Happens: Continuous or Ordered Event history Analysis Using Copula. Working Paper.

Hays, J. (2009). Bucking the System: Using Simulation Methods to Estimate and Analyze Systems of Equations with Qualitative and Limited Dependent Variables. In *2nd Annual St. Louis Area Methods Meeting (SLAMM), Washington University in St. Louis (https://netfiles. uiuc. edu/jchays/www/Hays_SLAMM09_Revised. pdf)*.

[8] I need to give more thought on this approximation.

Hays, J. and Kachi, A. (2009). Interdependent Duration Models in Political Science. *Working Paper*.

King, G., Alt, J. E., Burns, N. E., and Laver, M. (1990). A Unified Model of Cabinet Dissolution in Parliamentary Democracies. *American Journal of Political Science*, 34(3):846–871.

Maltzman, F. and Shipan, C. R. (2008). Change, Continuity, and the Evolution of the Law. *American Journal of Political Science*, 52(2):252–267.

Murillo, M. V. and Schrank, A. (2005). With a Little Help from My Friends: Partisan Politics, Transnational Alliances, and Labor Rights in Latin America. *Comparative Political Studies*, 38(8):971–999.

Nelsen, R. (2006). *An Introduction to Copulas*. Springer Verlag.

Quiroz Flores, A. (2008). Copula Functions and Bivariate Distributions for Survival Analysis: An Application to Government Survival. NYU Department of Politics Working Paper.

Shipan, C. R. and Shannon, M. L. (2003). Delaying Justice (s): A Duration Analysis of Supreme Court Confirmations. *American Journal of Political Science*, 47(4):654–668.

Simmons, B. and Elkins, Z. (2004). The globalization of liberalization: Policy diffusion in the international political economy. *American Political Science Review*, 98(01):171–189.

Svolik, M. (2008). Authoritarian Reversals and Democratic Consolidation. *American Political Science Review*, 102(2):153–168.

Warwick, P. V. (1992). Economic Trends and Government Survival in West European Parliamentary Democracies. *American Political Science Review*, 86(4):875–887.

Werner, S. and Yuen, A. (2005). Making and Keeping Peace. *International Organization*, 59(2):261–292.