Interdependent Duration Models in Political Science*

Jude C. Hays[†]
University of Illinois

Aya Kachi[‡]
University of Illinois

August 30, 2009
Prepared for the American Political Science Association's Annual meeting

Abstract

Interdependent duration processes are common in politics and other strategic settings. The time to one type of political event frequently depends on the time to another related event, and the time to an event for one actor often depends on the time to that same event for others. Put in a slightly different way, politics and strategic behavior generate interdependence across durations and duration interdependence across actors. We present a generalized parametric simultaneous equations model that incorporates these two kinds of interdependent duration processes and derive the corresponding full information maximum likelihood (FIML) estimator based on the Weibull distribution. We show with Monte Carlo experiments that our estimator outperforms the alternatives available to those doing applied empirical research. Naive estimators that either ignore the interdependence among duration processes or treat one as exogenous to the others are badly biased when the true relationships are simultaneous ones. Two stage least squares, while consistent, is highly inefficient relative to the FIML. We illustrate these findings in a study of the determinants of government formation duration and survival in European parliamentary democracies and an analysis of the timing of position taking in the US Congress. The interdependence in these durations is substantively important and suggests strategic bargaining over governments in Europe and free-riding behavior among members of Congress.

Political scientists are frequently interested in understanding when important political events occur. Students of 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, Arnold and Zorn, 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 one of two ways (or possibly both). The first kind of interdependence is when the time to one political event depends on the time to another related event. The second is when the time to a particular political event for one actor depends on the time to that same event for other actors. Examples of the first kind of interdependence include the time it takes to negotiate an international treaty and the survival of that agreement,

*The authors thank Christopher H. Achen, Seden Akcinaroglu, Jake Bowers, Janet Box-Steffensmeier, Damarys Canache, José Cheibub, Tom S. Clark, Songying Fang, Brian J. Gaines, Matt Golder, Sona Golder, Kosuke Imai, Gary King, James Kuklinski, Matthew Lebo, Benjamin E. Lauderdale, Eduardo L. Leoni, Walter R. Mebane Jr., Stephen Meserve, Burt L. Monroe, Kristopher W. Ramsay, Munro Richardson, Anne E. Sartori, Jasjeet S. Sekhon, Jacob Shapiro, Curtis S. Signorino, Aaron B. Strauss, Milan Svolik, Dustin Tingley, Bonnie Weir and Christopher Zorn for helpful comments. Earlier versions of this paper were presented at the Comparative Politics Workshop at the University of Illinois, the 66th MPSA Annual National Conference, the New Faces in Political Methodology Conference at Pennsylvania State University, the Political Methodology Colloquium at Princeton University and the 25th (*Silver Edition!*) Society for Political Methodology Summer Conference.

[†]Assistant professor, Department of Political Science, University of Illinois. E-mail: jchays@illinois.edu; URL: https://netfiles.uiuc.edu/jchays/www/page.html.

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

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. As for the second kind, 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. In our applications we look at the relationship between the time it takes a government to form and its subsequent survival duration and the interdependence among Congresspersons in the timing of their position taking on the North American Free Trade Agreement (NAFTA). One of our key contributions is to place these two kinds of interdependent durations within a single unified econometric framework.

Our paper is organized as follows. First, we review some of the existing empirical strategies for modeling interdependent durations, highlighting the important differences between simultaneous equations (SEQ) and seemingly unrelated regression (SUR) models (or substantive and nuisance models of interdependence more generally) as well as the relative strengths and weaknesses of using copula-based and change-of-variables-based likelihoods. Second, we present a general simultaneous equations model of interdependent durations and derive the corresponding FIML estimator. Our model encompasses both traditional SEQ models and spatial duration models. Third, we compare the performance of several estimators against our FIML using Monte Carlo experiments. Fourth, we estimate a simultaneous durations model of government formation and survival and a spatial duration model for the timing of NAFTA position taking by US Congresspersons. We conclude with a discussion about the ubiquity and importance of interdependent duration processes across political science subfields and mention some possible extensions to the models we present.

1 Existing Strategies for Modeling Interdependent Durations

How should we model interdependent durations? There are two basic approaches to modeling (the two kinds of) interdependence. One approach is to assume that the interdependence arises in the stochastic part of the model only. The second posits full interdependence in both the stochastic and systematic components. Oftentimes, strategies of the first variety are called nuisance approaches, while those of the second are referred to as substantive models of interdependence.[1] Together with the two kinds of interdependent duration processes–interdependence across durations and duration interdependence across actors–this gives a four-fold model typology. Examples include SUR, spatial error, SEQ, and spatial lag models respectively, and each of these can be found in the literature.

Quiroz Flores (2008), for instance, uses copula functions to estimate a SUR model of the tenure of chief executive officers and the median tenure of their ministers. The argument is that there are unobservable common shocks that affect the tenure of both chief executives and their ministers. Flores finds significant correlation in tenures and shows with Monte Carlos that an estimator that accounts for cross-equation correlation in disturbances is more efficient than those that do not. Darmofal (2009) estimates a spatial duration model for issue position taking on NAFTA in the U.S. Congress. This work, which we discuss in more detail below, is representative of the spatial-error approach to interdependence. His model allows for individual and shared frailties. Specifically, it is a model of spatially autocorrelated random effects. Darmofal finds strong evidence for state-level shared frailties in the timing of issue position taking.

An early example of interdependence across durations is found in Lillard (1993). He presents a simultaneous equations model of marriage duration and fertility timing. In his setup, the hazard of marriage dissolution has a direct effect on the fertility hazard, and prior outcomes of the fertility process affect the dissolution hazard. The baseline hazards are a function of piecewise linear splines, which allows for a more flexible form. Honoré and de Paula (2008) provide a recent example of duration interdependence across actors. They derive an interdependent duration model from a strategic two-player game. In their model, agents choose how long to participate in a particular activity before switching to an alternative activity. The

[1]Occasionally, these labels are inaccurate. For example, many scholars treat unobservables as substantively important (see Boehmke 2006). The interdependence is in the disturbances, but it is central to the analysis nonetheless.

utility from switching for one agent depends on whether the other agent has switched. There are many examples of this form of interdependence in microeconomics including the adoption of new technologies and market entry decisions by firms. With Monte Carlo experiments, they show, among other things, that treating endogenous durations as exogenous typically leads to an overestimation of strength of interdependence.

From our perspective, the main problem with nuisance approaches to duration dependence is that they fail to capture the forms of strategic interdependence that we have in mind: that duration *i* (or the duration for unit *i*) is a function of duration *j* (or the duration for unit *j*). The problem with the substance approaches is that they can lead to models that are more difficult to estimate. In other words, we see a potential trade-off between the conceptual match of the empirical models with theory and the ease with which these models can be estimated. Given this trade-off, we develop the simplest possible SEQ FIML estimator.

In answering the question about how to model duration interdependence, one must keep in mind the connection between structure and substance, on the one hand, and structure and estimation, on the other. We discuss these connections over the next several sections, beginning with structural assumptions about outcome dependency. With respect to estimation, we show how the SUR structure leads naturally to copula-based estimators while the SEQ structure makes the change-of-variables approach, because of its relative simplicity, attractive to those doing applied research.

1.1 Assumptions about Dependency Structure: SUR vs. SEQ

The seemingly unrelated regressions (SUR) and simultaneous equations (SEQ) models make very different assumptions about the structure of outcome dependency. Understanding these differences is crucial to choosing the right model and estimator.

First, the association among the outcome variables, $\mathbf{y}$'s, can be driven solely by their stochastic components that are generated from a single joint probability distribution. The SUR model captures this type of dependency among outcomes. In matrix notation, the SUR model takes the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{y}$ is an $ND \times 1$ vector containing N observations on D endogenous variables. The matrix $\mathbf{X}$ contains $ND \times K$ observations on K exogenous variables, where $K = \sum_{d=1}^{D} k_d$, k_d being the number of exogenous variables in the equation for the d^{th} endogenous variable, and $\boldsymbol{\beta}$ is a $K \times 1$ vector of coefficients on them. The final term $\boldsymbol{\varepsilon}$ in equation (1) is an $ND \times 1$ vector of disturbances with covariance matrix $\mathbf{V}_{SUR} = \boldsymbol{\Sigma}\mathbf{I}$. Rewriting (1) in terms of its constituent equations, we have

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_D \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ & & \vdots & \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_D \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_D \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_D \end{bmatrix}, \tag{2}$$

where $\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}_1, ..., \boldsymbol{\varepsilon}_D]'$ is generated from a single joint distribution with the covariance matrix

$$\mathbf{V}_{SUR}(\boldsymbol{\varepsilon}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1D} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2D} \\ & & \vdots & \\ \sigma_{D1} & \sigma_{D2} & \cdots & \sigma_{DD} \end{bmatrix} \mathbf{I}. \tag{3}$$

Due to the jointly-generated disturbances, outcomes y's seem to be related to each other. It should be noted, however, that the dependency structure implied by this model is different from what we have in mind when we say y_i depends on y_j $(i \neq j)$ –there is no component in this model that captures the relationship $y_i = f(y_j)$.

On the contrary, the SEQ approach models the explicit dependency among outcomes. It takes the form

$$\mathbf{y} = \mathbf{AIy} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{4}$$

which written in terms of its constituent equations is

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_D \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \alpha_{12}\mathbf{I} & \cdots & \alpha_{1D}\mathbf{I} \\ \alpha_{21}\mathbf{I} & \mathbf{0} & \cdots & \alpha_{2D}\mathbf{I} \\ & & \vdots & \\ \alpha_{D1}\mathbf{I} & \alpha_{D2}\mathbf{I} & \cdots & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_D \end{bmatrix} + \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ & & \vdots & \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_D \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_D \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_D \end{bmatrix}. \tag{5}$$

The $\mathbf{AI}_{(ND \times ND)}$ matrix that consists of $\alpha_{ij}\mathbf{I}_{(N \times N)}$ $(i \neq j)$ represents the degree of direct dependency among $\mathbf{y}_i$'s. For example, $\alpha_{ij}\mathbf{I}_{(N \times N)}$ denotes the effect of $\mathbf{y}_j$ on $\mathbf{y}_i$.

Since the endogenous variable $\mathbf{y}$ is now on the right-hand side of the structural equations (5), one needs to derive the reduced form in order to discuss properties of the stochastic component. Equation (5) written in reduced form is

$$\begin{aligned} \mathbf{y} &= (\mathbf{I} - \mathbf{AI})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \mathbf{AI})^{-1}\boldsymbol{\varepsilon} \\ &= \boldsymbol{\Gamma}\mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \end{aligned} \tag{6}$$

where $\boldsymbol{\Gamma} = (\mathbf{I} - \mathbf{AI})^{-1}$ and the covariance matrix for $\mathbf{u}$, the vector of reduced form disturbances, is

$$\mathbf{V}_{SEQ}(\mathbf{u}) = \boldsymbol{\Gamma}'\mathbf{V}\boldsymbol{\Gamma}. \tag{7}$$

An important difference between the SUR and SEQ models is that the covariances in (7) are a function of the coefficients in $\mathbf{A}$. In other words, the variances and covariances among the reduced form disturbances have to be consistent with the structural relationships among the endogenous variables. This has implications for estimating the two models.

In the next two sections, we construct maximum likelihood (ML) estimators based on the SUR and the SEQ assumption respectively. We employ both the copula approach and the change-of-variables technique to derive the necessary joint densities. These two methods look unrelated at first sight. However, the resulting likelihood functions are comparable and they exemplify the different assumptions about the dependency structure that we make by choosing either the SUR or the SEQ framework. We focus on this comparison below in section 1.4.

1.2 Estimator (1): Copula-Based Likelihood for an Duration SUR Model

A copula is a function that gives a proper joint distribution function from univariate marginal distribution functions. Several papers in political science use copulas or copula related distributions to derive likelihoods for empirical analysis including Boehmke, Morey and Shannon (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–e.g., the likelihoods for qualitative or limited dependent variables models. The main disadvantage of using copula-based distributions is that the covariance structures are constrained. These constraints imply limits on the range of association among the variables, and they also make it difficult to use copulas to derive likelihoods for SEQ models.

First, consider a joint distribution function of random variables y_1^* and y_2^* generated from the following Farlie-Gumbel-Morgenstern (FGM) copula

$$F(y_1^*, y_2^*) = F(y_1^*)F(y_2^*)[1 + \alpha\{1 - F(y_1^*)\}\{1 - F(y_2^*)\}], \tag{8}$$

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

$$f(y_1^*, y_2^*) = f(y_1^*)f(y_2^*)[1 + \alpha\{2F(y_1^*) - 1\}\{2F(y_2^*) - 1\}]. \tag{9}$$

For example, given a univariate Weibull distribution function

$$F(y^*) = 1 - e^{-(\frac{y^*}{\theta})\lambda}, \tag{10}$$

the copula (8) and (9) generates the joint cumulative and the joint density functions of the bivariate Weibull distribution.

$$\begin{aligned}
F(y_1^*, y_2^*) &= (1 - e^{-(\frac{y_1^*}{\theta_1})^{\lambda_1}})(1 - e^{-(\frac{y_2^*}{\theta_2})^{\lambda_2}})(1 + \alpha e^{-(\frac{y_1^*}{\theta_1})^{\lambda_1} - (\frac{y_2^*}{\theta_2})^{\lambda_2}}) \\
f(y_1^*, y_2^*) &= \frac{\lambda_1}{\theta_1}\frac{\lambda_2}{\theta_2}(\frac{y_1^*}{\theta_1})^{\lambda_1 - 1}(\frac{y_2^*}{\theta_2})^{\lambda_2 - 1}e^{-2[(\frac{y_1^*}{\theta_1})^{\lambda_1} + (\frac{y_2^*}{\theta_2})^{\lambda_2}]} \\
&\quad \times [4\alpha - 2\alpha e^{(\frac{y_1^*}{\theta_1})^{\lambda_1}} - 2\alpha e^{(\frac{y_2^*}{\theta_2})^{\lambda_2}} + (1 + \alpha)e^{(\frac{y_1^*}{\theta_1})^{\lambda_1} + (\frac{y_2^*}{\theta_2})^{\lambda_2}}],
\end{aligned} \tag{11}$$

where $y_1^* \geq 0$, $y_2^* \geq 0$, $-1 \leq \alpha \leq 1$, $\theta_1 > 0$, $\theta_2 > 0$, $\lambda_1 > 0$ and $\lambda_2 > 0$. Again, α is a dependence parameter, which induces the correlation between y_1 and y_2, and the λ's are shape parameters that determine the curvature of the distribution. θ's are scale parameters. Note that this becomes the Gumbel (bivariate exponential) distribution when $\theta_1 = \theta_2 = 1$ and $\lambda_1 = \lambda_2 = 1$.

With the FGM bivariate Weibull distribution, the degree of admissible linear association between the variables, Pearson's correlation, is limited to $-0.322409 \leq \rho \leq 0.322409$. For those who are interested, we derive the possible range for ρ and summarize some of the mathematical properties of this bivariate Weibull distribution in Appendix 1. This constraint may or may not be a serious limitation depending on the true strength of interdependence among the durations one is modeling. The more serious concern stems from the difficulty in transforming the copula-based SUR estimator into a SEQ estimator. To see this, we need to compare SUR and SEQ likelihoods.[2] Below is the likelihood function for a SUR model where there are two duration processes ($D = 2$) with N observations for each duration;

$$\begin{aligned}
L(\mathbf{X}, \boldsymbol{\beta}, \lambda_1, \lambda_2 | y_1^*, y_2^*) &= \prod_{i=1}^{N} f(y_{i1}^*, y_{i2}^*) \\
&= \prod_{i=1}^{N} \Big(\frac{\lambda_1}{\theta_1}\frac{\lambda_2}{\theta_2}(\frac{y_{i1}^*}{\theta_1})^{\lambda_1 - 1}(\frac{y_{i2}^*}{\theta_2})^{\lambda_2 - 1}e^{-2[(\frac{y_{i1}^*}{\theta_1})^{\lambda_1} + (\frac{y_{i2}^*}{\theta_2})^{\lambda_2}]} \\
&\quad \times [4\alpha - 2\alpha e^{(\frac{y_{i1}^*}{\theta_1})^{\lambda_1}} - 2\alpha e^{(\frac{y_{i2}^*}{\theta_2})^{\lambda_2}} + (1 + \alpha)e^{(\frac{y_{i1}^*}{\theta_1})^{\lambda_1} + (\frac{y_{i2}^*}{\theta_2})^{\lambda_2}}]\Big),
\end{aligned} \tag{12}$$

where the θ_d's, the scale parameters, are equal to $e^{\mathbf{X}_d\boldsymbol{\beta}_d}$ and the λ_d's are the shape parameters.

1.3 Estimator (2): Change-of-Variable Likelihood for an SEQ Weibull Duration Model

Next, we present a general simultaneous equations model for interdependent duration processes and derive its full information maximum likelihood estimator. We then return to the SUR likelihood for purposes of comparison in section 1.4.

[2]ML estimation using this joint distribution offers a possible solution to the problems caused by unobservables (mainly inefficiencies) from which the existing literature may suffer. Unobservables are much more pernicious in the selection model context because they are a potential source of endogeneity and bias (see Boehmke, Morey and Shannon, 2006). Copulas like the FGM are helpful here because the selection bias correction does not require covariance decomposition.

1.3.1 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.

$$y^*_{id} \sim Weibull(\lambda_d, \theta_d), \tag{13}$$

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

$$\begin{aligned} y = \ln y^* &= \ln \theta + \frac{1}{\lambda}\varepsilon \\ &= \mathbf{X}\boldsymbol{\beta} + \frac{1}{\lambda}\varepsilon, \end{aligned} \tag{14}$$

where $\varepsilon \sim ExtremeValueI\ (StandardGumbel)$ and we define $y = \ln y^*$. The second line of equation (14) 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.

1.3.2 The System

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

$$\mathbf{y}_{(ND\times 1)} = \mathbf{A}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{L}\mathbf{u}. \tag{15}$$

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

[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.

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'.

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

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.

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

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

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

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

The following reduced form can be derived from the structural form (15);

$$\begin{aligned} \mathbf{y}_{(ND\times 1)} &= (\mathbf{I}-\mathbf{A})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I}-\mathbf{A})^{-1}\mathbf{L}\mathbf{u} \\ &= \boldsymbol{\Gamma}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Gamma}\mathbf{L}\mathbf{u} \\ &= \boldsymbol{\Gamma}\mathbf{X}\boldsymbol{\beta} + \mathbf{v}, \end{aligned} \tag{16}$$

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

1.3.3 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 (16) for $\mathbf{u}$, we have

$$\mathbf{u}_{ND\times 1} = g^{-1}(\mathbf{y}) = (\mathbf{\Gamma}\mathbf{L})^{-1}\mathbf{y} - \mathbf{L}^{-1}\mathbf{X}\boldsymbol{\beta}. \tag{17}$$

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

$$\begin{aligned} h(y_{11}, \cdots, y_{ND}) &= \begin{cases} f(g_{11}^{-1}(y_{11}, \cdots, y_{ND}), \cdots, g_{ND}^{-1}(y_{11}, \cdots, y_{ND}))|det(\mathbf{J})| \\ 0, \text{ otherwise} \end{cases} \\ &= \begin{cases} f(u_{11}, \cdots, u_{ND})|det(\mathbf{J})| \\ 0, \text{ otherwise} \end{cases} \\ &= \begin{cases} f(u_{11})f(u_{12})\cdots f(u_{ND})|det(\mathbf{J})| \\ 0, \text{ otherwise.} \end{cases} \end{aligned} \tag{18}$$

The last line in equation (18) follows from the i.i.d. assumption of u, and each $f(u_{id})$ is the standard Gumbel pdf.

The likelihood function with no censoring is[4]

$$\begin{aligned} L &\propto h(y_{11}, \cdots, y_{ND}) \\ &= \Big(\prod_{i=1}^{N}\prod_{d=1}^{D} f(g^{-1}(y_{id}))\Big)|det(\mathbf{J})| \\ &= \Big(\prod_{i=1}^{N}\prod_{d=1}^{D} f(u_{id})\Big)|det(\mathbf{J})|. \end{aligned} \tag{19}$$

The log-likelihood function is

$$\begin{aligned} \ln L &= \sum_{i=1}^{N}\sum_{d=1}^{D}[\ln f(g^{-1}(y_{id}))] + \ln|det(\mathbf{J})| \\ &= \sum_{i=1}^{N}\sum_{d=1}^{D}[\ln f(u_{id})] + \ln|det(\mathbf{J})|. \end{aligned} \tag{20}$$

1.4 Copula-Based Likelihoods vs. Change-of-Variable Likelihoods

It is useful to compare the copula-based likelihood with the change-of-variables likelihood for the simple case of two duration processes in which there are no covariates, implying $\mathbf{X}\boldsymbol{\beta} = 0$ or equivalently $\theta_d = 1$.

[4]Censoring complicates estimation because the interdependence among durations means the likelihood contains a multidimensional integral.

To see the relationship of the two approaches, it is sufficient to consider the case where there is only one observation point (no spatial interdependence), and therefore we will omit the subscript that indicates observation (unit), and include only the duration subscript. First, consider an SEQ model for logged-duration dependent variables with no covariates, which takes the form

$$\begin{cases} \ln y_1^* = \alpha_2 \ln y_2^* + \frac{1}{\lambda_1} u_1 \\ \ln y_2^* = \alpha_1 \ln y_1^* + \frac{1}{\lambda_2} u_2 \end{cases}$$
$$\Leftrightarrow \begin{cases} u_1 = (\ln y_1^* - \alpha_2 \ln y_2^*)\lambda_1 \\ u_2 = (\ln y_2^* - \alpha_1 \ln y_1^*)\lambda_2, \end{cases} \tag{21}$$

where $\ln y_i^*$ denotes a dependent variable. y_i^* measures a spell of time and it is assumed to have the FGM Weibull distribution. The Jacobian for $\mathbf{u}$ can be computed as

$$\mathbf{J} = \begin{pmatrix} \frac{\partial u_1}{\partial y_1^*} & \frac{\partial u_1}{\partial y_2^*} \\ \frac{\partial u_2}{\partial y_1^*} & \frac{\partial u_2}{\partial y_2^*} \end{pmatrix} = \begin{pmatrix} \frac{\lambda_1}{y_1^*} & -\frac{\alpha_2 \lambda_1}{y_2^*} \\ -\frac{\alpha_1 \lambda_2}{y_1^*} & \frac{\lambda_2}{y_2^*} \end{pmatrix}. \tag{22}$$

$$|det(\mathbf{J})| = \frac{\lambda_1 \lambda_2}{y_1^* y_2^*} |1 - \alpha_1 \alpha_2|. \tag{23}$$

From equation (19), the exact expression of the likelihood derived by the change-of-variables approach is

$$\begin{aligned} f_{cvt}(y_1, y_2) &= \Big(\prod_{d=1}^{2} f(u_d) \Big) |det(\mathbf{J})| \\ &= e^{u_1 - e^{u_1}} e^{u_2 - e^{u_2}} \frac{\lambda_1 \lambda_2}{y_1^* y_2^*} |1 - \alpha_1 \alpha_2| \\ &= y_1^{*\lambda_1} e^{-y_1^{*\lambda_1}} y_2^{*\lambda_2} e^{-y_2^{*\lambda_2}} \frac{\lambda_1 \lambda_2}{y_1^* y_2^*} |1 - \alpha_1 \alpha_2| \\ &= \lambda_1 \lambda_2 y_1^{*\lambda_1 - 1} y_2^{*\lambda_2 - 1} e^{-2(y_1^{*\lambda_1} + y_2^{*\lambda_2})} e^{y_1^{*\lambda_1} + y_2^{*\lambda_2}} |1 - \alpha_1 \alpha_2|. \end{aligned} \tag{24}$$

The transformation from the second to the third lines of equation (24) uses the fact that

$$e^{u_i} = y_i^{*\lambda_i}. \tag{25}$$

Recall the likelihood function with the joint pdf constructed from the FGM copula, (12);

$$\begin{aligned} f_{copula}(y_1, y_2) &= \lambda_1 \lambda_2 y_1^{*\lambda_1 - 1} y_2^{*\lambda_2 - 1} e^{-2(y_1^{*\lambda_1} + y_2^{*\lambda_2})} \\ &\quad \times [4\alpha - 2\alpha e^{y_1^{*\lambda_1}} - 2\alpha e^{y_2^{*\lambda_2}} + (1 + \alpha) e^{y_1^{*\lambda_1} + y_2^{*\lambda_2}}]. \end{aligned} \tag{26}$$

By comparing the two likelihoods, (24) and (26),

$$\begin{aligned} & f_{cvt}(y_1, y_2) = f_{copula}(y_1, y_2) \\ \Leftrightarrow\ & \alpha = \frac{|1 - \alpha_1 \alpha_2| - 1}{4e^{-y_1^{*\lambda_1} - y_2^{*\lambda_2}} - 2e^{-y_1^{*\lambda_1}} - 2e^{-y_2^{*\lambda_2}} + 1}. \end{aligned} \tag{27}$$

In order for the covariance structure that is implied by the copula to be consistent with the structural SEQ relationships among the endogenous variables, equation (27) must hold. This equality makes the copula-based estimator for the general SEQ model much more complicated than the change-of-variables-based estimator. Of course, the advantage of working with the copula is that we would have the distribution function. Fortunately, as long as we are working with uncensored duration data, we do not need the distribution function to estimate our models. In the next section we evaluate the change-of-variables FIML estimator against commonly used alternatives.

2 Monte Carlo Evaluations of the FIML and Alternative Estimators

Tables 1-2 present the results of several Monte Carlo experiments in which we evaluate the performance of four estimators: the FIML-SEQ, two-stage least-squares, ML-AEDM, and ML-AIDM estimators.[5] The last two estimators are what political scientists currently use. ML-AEDM stands for maximum likelihood *assumed* exogenous duration model. This is the standard ML applied to a single equation that has an endogenous duration on the right-hand-side. By standard we mean that the estimator treats the endogenous duration as exogenous. In the context of one of our applications (coalition bargaining duration and government survival), this is what most people are using when they put "crisis duration"–the number of days to form a government–on the right-hand-side of a government survival model. ML-AIDM stands for ML *assumed* independent durations model. This is when the analyst fails to recognize that his or her duration of interest is linked in important ways to another or multiple other durations. The ML-AEDM suffers from simultaneity bias while the ML-AIDM suffers from omitted variable bias.

[TABLES 1-2 ABOUT HERE]

The experimental data is generated using the reduced form SEQ model[6]. We assume two durations, each with an exclusive covariate and a unique shape parameter. More specifically, the structural version of the model is

$$y_1 = \alpha_1 y_2 + \beta_1 x_1 + \lambda_1^{-1} u_1$$
$$y_2 = \alpha_2 y_1 + \beta_2 x_2 + \lambda_2^{-1} u_2$$

We are interested in the cases of positive reinforcing interdependence ($\alpha_1 = \alpha_2 > 0$), negative reinforcing interdependence ($\alpha_1 = \alpha_2 < 0$), and mixed interdependence ($\alpha_1 = -\alpha_2$) for small ($N = 100$) and medium-sized samples ($N = 500$). For the naive estimators, reinforcing interdependence should cause inflation bias in the estimated coefficients on the endogenous right-hand-side variables, while mixed interdependence should cause attenuation bias, and the differences in the shape parameters ($\lambda_1^{-1} < \lambda_2^{-1}$) should introduce asymmetries in these biases. The results of our Monte Carlos are mostly as we expected them to be. Starting with Table 1 where we report the results for a small sample (N=100) with positive reinforcing interdependence. The AIDM estimator overestimates β_1 and β_2 as expected. (We get estimates that correspond to the reduced-form parameters instead of the structural parameters.) The estimator also provides inflated estimates of λ_1^{-1} and λ_2^{-1}. The AEDM estimator also inflates the coefficients on the endogenous variables, y_1 and y_2, in this case. These biases, in turn, induce additional biases in the opposite direction for the β_1 and β_2 estimates. Note that the upward bias in the estimate for α_2 is larger than the bias for α_1 because the shape parameter λ_2^{-1} is larger than λ_1^{-1}, and this generates stronger covariance between y_1 and u_2 than exists between y_2 and u_1. The key inequality is $\frac{\partial y_1}{\partial u_2} = \alpha_1 \lambda_2^{-1} > \frac{\partial y_2}{\partial u_1} = \alpha_2 \lambda_1^{-1}$. These patterns are repeated in the experiment with negative reinforcing interdependence. In the case of mixed interdependence, there is attenuation bias in the AIDM estimates for β_1 and β_2. We note that, in the case of the AEDM estimator, the attenuating force is so strong for α_2 that the sign is wrong on average.[7] In all our experiments, the two-stage least-squares estimator performs better than the naive estimators in terms of bias, but occasionally, particularly in the small samples, performs worse in mean-squared-error terms. The two-stage least-squares results are much better for our medium-sized sample.

The FIML estimates are virtually unbiased in all cases even in our smaller sample, and the standard error estimates, calculated with the observed information matrix, are accurate. The standard error accuracy of the other estimators is frequently poor. The AIDM estimator is always overconfident, and the degree of overconfidence increases with the true variance of the sampling distribution. With the AEDM estimator, the degree of overconfidence is higher for the more badly biased coefficients (i.e., α_2 and β_2), a particularly disturbing combination that makes sound inference difficult.

[5]We also evaluated the three-stage least-squares estimator, but its performance was dominated by two-stage least-squares, so we do not report these results. They are available from the authors upon request.

[6]We do not focus on spatial lag models. These are treated extensively in Franzese Jr. and Hays (2007).

[7]Technically, the bias need not be attenuating in the sense that the estimate is, on average, closer to zero than the truth. Small effects in one direction can be overwhelmed by bias inducing forces pushing in the other direction so that the estimate, on average, has the opposite sign and is farther from zero than the truth.

3 Applications

3.1 Interdependence Across Durations: Government Formation and Tenure Durations

Scholarly interest in the empirical determinants of government formation and dissolution in parliamentary democracies is longstanding, and these topics remain among the most central in the comparative study of developed democracies. Two of the most popular topics in this literature are explaining the durations of both coalition bargaining over ministerial portfolios and government survival. There is good reason for this focus. The failure of parliamentary parties to form governments quickly (e.g., the recent crisis in Belgium) and chronic government instability (e.g., Italy for much of the postwar period) have significant social costs and are viewed as symptoms of dysfunctional democracy.

The quantitative empirical literature in this area is large.[8] Typically, the empirical studies explore a set of contextual and cabinet specific factors that determine both kinds of durations. The effects are estimated separately (e.g., for the coalition bargaining duration, Diermeier and van Roozendaal 1998; Martin and Stevenson 2003; and for the cabinet survival, Warwick and Easton 1992; Alt and King 1994, and Diermeier and Stevenson 1999). This is not to say that the interdependence has been completely ignored.King et al. (1990) and Warwick (1992), among others, put government formation duration, what they call crisis duration, and the number of formation attempts on the right-hand side of their government survival models.

In the more theoretically oriented literature, Strøm, Budge and Laver (1994) highlight the importance of cabinet termination and dissolution rules for government formation. Fearon (1998) also formalizes the effects of expected enforcement levels of bargained outcomes on the bargaining stage itself, in the context of international agreements. His formulation suggests that a longer shadow of the future can give states an incentive to bargain harder, delaying agreement in hope of getting a better deal. Diermeier, Eraslan and Merlo (2003) also formalize explicitly the interdependence of government formation bargaining and the bargained outcome -cabinet survival. The main purpose of Diermeier, Eraslan and Merlo (2003) is to analyze the conditions under which certain types of coalitions are formed. As an empirical matter, their interest lies in estimating the probability that a particular type of coalition is chosen. Durations of bargaining and government survival still play important roles in their model, but those durations are not the primary focus of their analysis. In their model, the inefficient delay of bargaining is generated mainly by a stochastic factor, the state of the world that is either favorable or unfavorable for a cabinet's survival, while the inefficient delay in Fearon (1998) is mainly due to the dichotomous bargaining choices and (or) uncertainty.

There are fewer theoretical studies of government termination. Laver and Shepsle (1996) stress that the ending of one cabinet begins the formation process for the next and that dissolution and formation are conceptually nonseparable, though their own emphasis is more on the making than breaking of governments. Lupia and Strøm (1995) show that majority governments may dissolve and call early elections when the expected payoff is high enough. Their model explains why a cabinet, which is an "equilibrium" of the earlier bargaining process, might find it worthwhile to terminate its tenure and call an election. All of these studies make important contributions, but fall short of the kind of systematic integration that we see as necessary.

We argue that the lengths of coalition bargaining and government survival are interdependent duration processes. Unfortunately, to this point, the two have been studied largely in isolation. The single equation studies suffer from multiple sources of bias. One potential problem is omitted variable bias in regressions that leave out the important "right-hand-side" duration. Simultaneity is a concern for studies that do connect government formation and dissolution in single equation models by putting variables like crisis duration or the number of cabinet formation attempts on the right-hand-side of government survival regressions. The simultaneity problem is obvious from the structures of these models. Bargaining and survival durations are clearly related, but the causal arrow points both ways. If we put one duration on the right hand side of a model explaining the other-as is frequently done in studies of government survival-our estimates will be biased by the reverse causal relationship. The clear empirical implication of these formal models is that we should not estimate coalition bargaining and government survival durations sep-

[8] See reviews in Laver (1998, 2003) for more extensive treatments of both literatures.

arately or naively put one duration on the right-hand-side of a single-equation regression that has the other duration on the left-hand-side.

Our dataset consists of 475 cabinets from sixteen Western European countries -Austria, Belgium, Denmark, Finland, France (Fourth Republic), Germany, Iceland, Ireland, Italy, Luxembourg, The Netherlands, Norway and Sweden. The data run between 1945 and 1998.

Interestingly, there is a positive relationship between the time it takes for government formation and the length of government survival in our sample (see Figure 1). Governments that formed in less than fifty days survived, on average, 580 days whereas coalitions that took more than 100 days to reach agreement lasted 818 days. This is a bit perplexing since we might expect long delays in government formation to be indicative of the inability of parliamentary parties to work together effectively (King et al., 1990). There is another way to look at this relationship, however. Parties that anticipate long-lasting governments may bargain harder over coalition agreements since these ”contracts” will determine the balance of executive power, distribution of benefits from holding office, and overall course of policy for a significant period of time into the future. Is this relationship spurious or causal? And what implications, if any, does this interdependence have for empirical analyses of government formation and survival durations?

[Figure 1 ABOUT HERE]

The FIML and 2SLS estimators rely on instruments to identify the causal effects of cabinet formation duration on government survival and vice versa. We use the continuation and maximum duration variables as instruments. In both cases, we think it highly plausible on theoretical grounds that the instruments satisfy the necessary exclusion restrictions. We present two sets of results for each estimator. The first set is from a covariate sparse specification, and the second is from a covariate rich specification. The sparse specification includes, in addition to the endogenous variables, the instruments needed for the FIML and 2SLS estimators. We focus primarily on the differences between the AEDM and FIML estimators and the simultaneous relationship between cabinet formation duration and government survival.

[Table 3 ABOUT HERE]

With the FIML estimator we find robust evidence that the positive correlation between bargaining duration and government survival seems to be driven by the latter causing the former. The covariate-sparse and rich comparison highlights the unbiasedness and efficiency advantages of the FIML in small samples. The AEDM estimator finds a positive and statistically significant relationship between cabinet duration formation and government survival in both equations. This is not surprising given the positive covariance, but keep in mind that the *causal* argument that prolonged formation processes (what many scholars refer to as the ”crisis” duration) lead to longer-lived governments is viewed by most as dubious. We expect the opposite relationship, and this is what we find with the FIML estimator. Turning to the covariate rich specification, the relationships between the government formation and survival durations is wiped away in the AEDM estimates. By contrast, with the FIML estimates, we continue to find a statistically significant and positive effect of government survival on formation duration. This is due in large part to the efficiency of the estimator. The estimated standard error for the FIML coefficient on the government survival variable is almost one-third the size of the AEDM standard error.

Overall, we interpret the FIML results as strong evidence that parties anticipate the length of the future government's tenure and this affects how they bargain. This is the idea of strategic interdependence that comes out of the game theoretic literature on the topic developed by Diermeier, Merlo and others. We do not find evidence of the reverse causal relationship. In other words, although there are studies that suggest that the duration of formation processes affects government survival, we do not find robust evidence that this is the case. These theories maintain that longer bargaining indicates the difficulty in reaching agreements among the coalition members in general and hence portends a shorter lifespan for the formed government.

3.2 Duration Interdependence Across Actors: Strategic Timing of Issue Position Taking

It is often said that timing is everything in politics. This is certainly true when it comes to the behavior of elected politicians, and position taking on legislation is one of the clearest examples. Drawing on the logic of formal signaling models, Box-Steffensmeier, Arnold and Zorn (1997) argue in their seminal paper that issue position taking in Congress will be strategically timed. Members of Congress (MCs) who receive clear signals about the policy preferences of their constituents will announce early, while those who receive mixed signals will delay. They also contend that constituency preferences will interact with individual-level factors in either cross-cutting or reinforcing ways, and that institutional factors such as leadership status and committee membership will influence the timing of issue position taking. In their empirical analysis, Box-Steffensmeier et al. examine issue position taking on NAFTA. They find that MCs from border districts took early positions, as did Republican leaders, ceteris paribus. Conservative MCs from highly unionized districts delayed their position taking.

Boehmke (2006) extends this analysis by linking the timing of issue position taking by MCs and the content of their positions through unobservables. He argues that factors that cause delay in position taking on NAFTA also make it more likely that members will support the legislation. Legislators that hold out, whatever the reason may be, ultimately decide to vote in favor, either because of presidential pressure or out of concern for party interests. Darmofal (2009) develops the model further by allowing for spatial correlation in the timing of issue position taking among representatives. He models spatially connected individual and shared (state-level) frailties. In his preferred specification, the state-level shared frailties specification, MCs from the same state have a common random effect, and these effects are geographically correlated. Frailties cluster among states that share borders.

The strategic nature of issue position taking is clear from this literature. One form of interdependence that is left out of these models is strategic interdependence across members of Congress. Darmofal's models come the closest to capturing this interdependence, but his model is better interpreted as capturing omitted variables that cluster geographically rather than true interdependence among MCs since the correlation is only in the disturbance term. There is good reason to expect strategic interdependence across members, particularly members from districts in close proximity. These members represent overlapping constituencies, and therefore may have incentives to take early positions to signal their commitment and resolve on a particular issue. If true, this would spark competitive dynamics among representatives. It is also possible that members of Congress free-ride off of the early position taking of their colleagues. Early position taking and the political responses it provokes provide valuable information to other members who, at some point, will be expected to take a stance. These relationships can be modeled using our interdependent durations model.

The dependent variable is the number of days after August 11, 1992, the date when the first MC (Peter Visclosky) announced his NAFTA position, before the other MCs took a pro or con position on the legislation. The constituency variables included in the analysis are the district-level Perot vote, union membership, and average household income. The interest group factors include the contributions from corporate and labor PACs. The institutional variables are NAFTA committee membership and party leadership indicators. Ideology is the individual-level variable, which is interacted with constituency-level variables to model cross-cutting pressures on MCs. Interestingly, there is a substantial amount of within state variation in the data. In the case of logged durations, the within variance is larger than the between, or, more specifically, the average within state variance is larger than the variance in state-level means. This fact is consistent with negative interdependence or free-riding behavior.

Since we are analyzing a single duration, our interdependent durations model simplifies to a spatial-lag model. We use Darmofal's adjacency matrix based on queen contiguity to connect MCs and include state-level fixed effects for all states with more than one representative. The same four estimators used previously–the assumed independent, assumed exogenous, 2SLS, and FIML estimators–can be applied to this model. We report the results in Table 4. The first column gives the Box-Steffensmeier et al. results from their Cox proportional hazards model. The remaining four columns give the estimates for our Weibull accelerated failure time models.[9] In all of the models that allow for duration interdependence, we find evidence of free-riding (i.e., the coefficient estimate for rho is always negative). When one MC for whatever reason-including constituency, institutional, and individual-level factors-announces an early position on NAFTA, his or her colleagues from bordering districts are more likely to delay their position taking. As expected, the 2SLS estimator has the largest standard errors, and the assumed exogenous estimator overstates the strength of interdependence relative to the FIML.

[Table 4 ABOUT HERE]

4 Conclusion

Politics generates interdependence across durations and duration interdependence across actors. There are many examples of this interdependence in prominent areas of political science research. In order to analyze interdependent durations empirically, one has to make assumptions about the structure of dependence. We believe that, in many instances, the simultaneous equations framework provides a better match to theory. Therefore, we developed a generalized parametric simultaneous equations model for interdependent duration processes and derived the corresponding full information maximum likelihood (FIML) estimator based on the Weibull distribution. Along the way, we attempted to demonstrate why a copula-based estimator, while possibly desirable for other applications where the distribution function is needed, is unnecessary for our purposes, particularly given the added complications. We demonstrated with Monte Carlo experiments that our estimator outperforms the alternatives. Naive estimators that either ignore the interdependence among duration processes or treat one as exogenous to the others are badly biased when the true relationships are simultaneous ones. Two-stage-least-squares is highly inefficient relative to the FIML. We illustrated these findings in a study of the determinants of government formation duration and survival in European parliamentary democracies and an analysis of the timing of position taking in the US Congress. The interdependence we uncovered in these durations is substantively important and suggests strategic bargaining over governments in Europe and free-riding behavior among members of Congress. In future work, we hope to address the issue of censored durations, possibly using copula functions, expand the model to allow for time-varying covariates, relax our parametric assumptions, and estimate models where both forms of interdependence (interdependence across durations and duration interdependence across actors) are present simultaneously.

[9]Note that, to be consistent, the coefficient estimates from the proportional hazards and accelerated failure time (AFT) models should have the opposite signs. The proportional hazards model gives the effects of covariates on the hazard rate, while the AFT model gives the effects of covariates on the expected time until failure.

Appendix 1:
Mathematical properties of the bivariate Weibull distribution

In the following, we briefly describe some mathematical properties of the bivariate Weibull distribution presented above. Although it was already proven that any bivariate distribution that belongs to the Farlie-Gumbel-Morgenstern family satisfies the axioms of probability (Gumbel, 1959, 1960), the following mathematical properties that lead to the derivation of ρ are useful.

As shown in Gumbel (1959) and Gumbel (1960), a bivariate distribution function can be constructed from two marginal probability functions, $F(y_1)$ and $F(y_2)$, by the following copula with a constraint on parameter α, $-1 \leq \alpha \leq 1$;

$$F(y_1, y_2) = F(y_1)F(y_2)[1 + \alpha\{1 - F(y_1)\}\{1 - F(y_2)\}], \tag{28}$$

where $-1 \leq \alpha \leq 1$.

The associated joint density function is given as follows;

$$f(y_1, y_2) = f(y_1)f(y_2)[1 + \alpha\{2F(y_1) - 1\}\{2F(y_2) - 1\}]. \tag{29}$$

Using the above findings, a bivariate Weibull distribution can be constructed from the following marginal distributions and densities;

$$F(y_i) = 1 - e^{-(\frac{y_i}{\theta_i})^{\lambda_i}},$$

$$f(y_i) = \frac{\lambda_i}{\theta_i}(\frac{y_i}{\theta_i})^{\lambda_i - 1}e^{-(\frac{y_i}{\theta_i})^{\lambda_i}}; i = 1, 2,$$

where $\lambda_i > 0$ and $\theta_i > 0$.

The joint probability and density are given by

$$F(y_1, y_2) = (1 - e^{-(\frac{y_1}{\theta_1})^{\lambda_1}})(1 - e^{-(\frac{y_2}{\theta_2})^{\lambda_2}})(1 + \alpha e^{-(\frac{y_1}{\theta_1})^{\lambda_1} - (\frac{y_2}{\theta_2})^{\lambda_2}}) \tag{30}$$

$$f(y_1, y_2) = \frac{\lambda_1}{\theta_1}\frac{\lambda_2}{\theta_2}(\frac{y_1}{\theta_1})^{\lambda_1 - 1}(\frac{y_2}{\theta_2})^{\lambda_2 - 1}e^{-2[(\frac{y_1}{\theta_1})^{\lambda_1} + (\frac{y_2}{\theta_2})^{\lambda_2}]}[4\alpha - 2\alpha e^{(\frac{y_1}{\theta_1})^{\lambda_1}} - 2\alpha e^{(\frac{y_2}{\theta_2})^{\lambda_2}} + (1 + \alpha)e^{(\frac{y_1}{\theta_1})^{\lambda_1} + (\frac{y_2}{\theta_2})^{\lambda_2}}], \tag{31}$$

where $y_i \geq 0$, $-1 \leq \alpha \leq 1$, $\theta_i > 0$ and $\lambda_i \geq 0$.

The joint cdf must satisfy the following boundary conditions

$$\begin{cases} F(0, y_2) = F(y_1, 0) = 0 \\ F(\infty, \infty) = 1, \end{cases}$$

and the joint density has to be nonnegative, $f(y_1, y_2) \geq 0$. Any bivariate distribution that belongs to the Farlie-Gumbel-Morgenstern family satisfies these conditions.

Another condition a bivariate distribution always has to satisfy is Fréchet's inequality,

$$F(y_1, y_2) \leq F_1(y_i); i = 1, 2 \tag{32}$$

for all y_1 and y_2. Since the dependence of the joint distribution and density on y_1 and y_2 is symmetric, it is sufficient to explore the performance of one variable y_1. This applies to all the calculations in the rest of this appendix. From (30) it follows after a simplification that

$$\alpha e^{-(\frac{y_1}{\theta_1})^{\lambda_1}}(1 - e^{-(\frac{y_2}{\theta_2})^{\lambda_2}}) \leq 1. \tag{33}$$

In sum, the function (30) satisfies all the required axioms of probability function, under the conditions of $-1 \leq \alpha \leq 1$, $\theta_i > 0$ and $\lambda_i \geq 0$.

The followings are the relevant computations to derive the correlation coefficient ρ. By definition, the correlation coefficient of two random variables, Y_1 and Y_2 can be obtained as

$$\rho = \frac{E(y_1 y_2) - E(y_1)E(y_2)}{\sigma_{y_1}\sigma_{y_2}}. \tag{34}$$

For our marginal probabilities, $F(y_1)$ and $F(y_2)$, the means and variances are

$$\begin{aligned} E(y_i) &= \theta_i \Gamma(1 + \frac{1}{\lambda_i}) = \theta_i \frac{1}{\lambda_i}\Gamma(\frac{1}{\lambda_i}) \\ Var(y_i) &= \theta_i^{\frac{2}{\lambda_i}}[\Gamma(1+\frac{2}{\lambda_i}) - \Gamma^2(1+\frac{1}{\lambda_i})]; \quad i = 1,2. \end{aligned} \tag{35}$$

Now the only term we need to compute to obtain ρ is $E(y_1 y_2)$. From (15), the marginal densities are

$$f(y_i) = \int_0^\infty f(y_1, y_2) dy_i = \frac{\lambda_i}{\theta_i}(\frac{y_i}{\theta_i})^{\lambda_i - 1} e^{-(\frac{y_i}{\theta_i})^{\lambda_i}}; i = 1, 2, \tag{36}$$

which is, of course, the Weibull densities.

The conditional expectation of y_1 can be obtained as

$$\begin{aligned} E(y_1|y_2) &= \int_0^\infty y_1 f(y_1|y_2) dy_1 \\ &= \frac{1}{\lambda_1}\theta_1 \Gamma(\frac{1}{\lambda_1}) 2^{-\frac{1}{\lambda_1}} e^{-(\frac{y_2}{\theta_2})^{\lambda_2}} [-2\alpha(2^{\frac{1}{\lambda_1}} - 1) + e^{(\frac{y_2}{\theta_2})^{\lambda_2}}(2^{\frac{1}{\lambda_1}}(1+\alpha) - \alpha)], \end{aligned} \tag{37}$$

where

$$f(y_1|y_2) = \frac{f(y_1, y_2)}{f(y_2)}. \tag{38}$$

The expectation of the cross-product can be computed as

$$\begin{aligned} E(y_1 y_2) &= \int_0^\infty y_2 E(y_1|y_2) f(y_2) dy_2 \\ &= \frac{\theta_1}{\lambda_1}\frac{\theta_2}{\lambda_2}\Gamma(\frac{1}{\lambda_1})\Gamma(\frac{2}{\lambda_2})\left[1 + \alpha(1 - 2^{-\frac{1}{\lambda_1}} - 2^{-\frac{1}{\lambda_2}} + 2^{-\frac{\lambda_1 + \lambda_2}{\lambda_1 \lambda_2}})\right]. \end{aligned} \tag{39}$$

Substituting (39) into (34), we get

$$\rho = \frac{2^{-\frac{\lambda 1 + \lambda 2}{\lambda 1 \lambda 2}}\left(2^{\frac{1}{\lambda 1}} - 1\right)\left(2^{\frac{1}{\lambda 2}} - 1\right)\alpha \Gamma\left[\frac{1}{\lambda 1}\right]\Gamma\left[\frac{1}{\lambda 2}\right]}{\lambda 1 \lambda 2 \sqrt{-\Gamma^2\left[1 + \frac{1}{\lambda 1}\right] + \Gamma\left[\frac{2+\lambda 1}{\lambda 1}\right]}\sqrt{-\Gamma^2\left[1 + \frac{1}{\lambda 2}\right] + \Gamma\left[\frac{2+\lambda 2}{\lambda 2}\right]}}. \tag{40}$$

Note that the scale parameter θ does not affect the dependence of y_1 and y_2, ρ. As mentioned in the previous sections, ρ is increasing in α and the maximum range is $-.322409 \leq \rho \leq .322409$ when $\lambda_1 = \lambda_2 = 3.29035$.

Appendix 2:

The Logged Weibull and the Standard Gumbel Variables

In the derivation of the Weibull FIML estimator (Section 1.3), we claimed that a logged Weibull random variable is a Standard-Gumbel –a special case of the type-I extreme value (minimum)– variable that is scaled by the inverse of the original Weibull shape parameter. This section demonstrates the log transformation of a Weibull variable. Recall the density and distribution functions of the Standard Gumbel distribution and the Weibull distribution;

$$\text{Standard Gumbel distribution} \begin{cases} f(u) = e^u e^{-e^u} \\ F(u) = 1 - e^{-e^u}, \end{cases}$$
$$\text{Weibull} \begin{cases} f(y) = \frac{\lambda}{\theta}(\frac{y}{\theta})^{\lambda-1} e^{-(\frac{y}{\theta})^\lambda} \\ F(y) = 1 - e^{-(\frac{y}{\theta})^\lambda}, \end{cases}$$

where θ is a scale parameter and λ is a shape parameter.

Consider a Weibull random variable Y that is scaled by θ. The log of the Weibull variable is a Standard Gumbel variable, U, scaled by the inverse of the Weibull shape parameter, $\frac{1}{\lambda}$. If this statement is true, then the following holds;

$$\frac{1}{\lambda} u = \ln{(\frac{y}{\theta})}$$
$$\Leftrightarrow y = \theta e^{\frac{u}{\lambda}}.$$

Since $Y \sim Weibull(\lambda, \theta)$,

$$\begin{aligned} F(y) &= 1 - e^{-(\frac{y}{\theta})\lambda} \\ &= 1 - e^{-(\frac{\theta e^{\frac{u}{\lambda}}}{\theta})\lambda} \\ &= 1 - e^{-e^u} = G(u). \end{aligned} \tag{41}$$

$G(u)$ is the cdf of the Standard Gumbel distribution.

The moments for this extreme value distribution are given as follows;

$$E[u] = \gamma,$$

where γ is the Euler-Mascheroni constant, and

$$Var(u) = \frac{\pi^2}{6}.$$

Appendix 3:
Stata Code

```
*Program to Estimate Weibull SEQ Duration Model

clear
pr drop _all
set more off

***********************
*Likelihood Evaluator
***********************

program define seq_dur_ll

args lnf mu1 mu2 alpha1 alpha2 lambda1 lambda2
tempvar J ay1 ay2
scalar a1 = `alpha1'
scalar a2 = `alpha2'
gen `ay2' = `alpha1'*$ML_y2
gen `ay1' = `alpha2'*$ML_y1
matrix IA = [1, -(a1) \ -(a2), 1]
```

```
scalar l1 = `lambda1'
scalar l2 = `lambda2'
matrix L = [l1, 0 \ 0, l2]
matrix IAL = IA*L
qui gen double `J' = ln(det(IAL))
scalar J = `J'
qui replace `lnf'= J + `lambda1'*($ML_y1-`ay2'-`mu1') - exp(`lambda1'*($ML_y1-`ay2'-`mu1')) + /*
*/ `lambda2'*($ML_y2-`ay1'-`mu2') - exp(`lambda2'*($ML_y2-`ay1'-`mu2'))
end

**************************
*Open Data for Regression
**************************
drop _all
use "ADD PATH HERE", clear

gen lnform = ln(formation)
gen lndur = ln(duration)
global Y1 lnform
global Y2 lndur
global X1 invest continuation neffp polar ideo_div2 returnability postelect caretaker
global X2 invest max_dur neffp polar ideo_div2 returnability postelect caretaker


************************
*Produce starting values
************************

stset formation
streg $X1, dist(weibull) time

matrix stregbp1=e(b)
local col1 = colsof(stregbp1)
matrix stregb1=stregbp1[1,1..`col1'-1]
matrix coleq stregb1 = mu1
local stregp1=exp(stregbp1[1,`col1'])

stset duration
streg $X2, dist(weibull) time

matrix stregbp2=e(b)
local col2 = colsof(stregbp2)
matrix stregb2=stregbp2[1,1..`col2'-1]
matrix coleq stregb2 = mu2
local stregp2=exp(stregbp2[1,`col2'])


***************************
*Estimate SEQ model
***************************

ml model lf seq_dur_ll (mu1: $Y1=$X1) (mu2: $Y2=$X2) (alpha1:) (alpha2:) (lambda1:) (lambda2:)
ml init stregb1
ml init stregb2
ml init alpha1:_cons=0
ml init alpha2:_cons=0
ml init lambda1:_cons=`stregp1'
ml init lambda2:_cons=`stregp2'
ml max

*Program to Estimate Weibull Spatial Duration Model
```

```
clear
pr drop _all
set more off
set matsize 800

**********************
*Likelihood Evaluator
**********************

program define splag_ll_dur

args lnf mu rho lambda
qui replace `lnf'= ln(ones - `rho'*EIGS1) + ln(`lambda') + `lambda'*($ML_y1-`rho'*SL1-`mu') - /*
*/ exp(`lambda'*($ML_y1-`rho'*SL1-`mu'))
end


*********************
*Open Data For Weights
*********************

use "ADD PATH HERE", clear
qui sum var1
global nobs = r(N)
mkmat var1-var$nobs, matrix(W)
matrix I_n = I($nobs)
matrix eigenvalues eig1 imaginaryv = W
matrix eig2 = eig1'
matrix ones=J($nobs,1,1)


drop _all

*************************
*Open Data for Regression
*************************
drop _all
use "ADD PATH HERE", clear

gen lntiming = ln(timing)
global Y lntiming
global X pscenter pecenter perotsq mexbordr hhcenter corptpct labtpct ncomact rleader dleader inter1

mkmat $Y, matrix(Y)
matrix SL = W*Y
svmat SL, n(SL)
svmat eig2, n(EIGS)
svmat ones, n(ones)


***********************
*Produce starting values
***********************
stset timing
streg $X, dist(weibull) time
matrix stregbp=e(b)
local col = colsof(stregbp)
matrix stregb=stregbp[1,1..`col'-1]
matrix coleq stregb = mu
local stregp=exp(stregbp[1,`col'])
```

```
***************************
*Estimate spatial lag model
***************************

ml model lf splag_ll_dur (mu: $Y=$X) (rho:) (lambda:)
ml init stregb
ml init rho:_cons=0
ml init lambda:_cons=`stregp'
ml max
```

References

Alt, James E. and Gary King. 1994. "Transfers of Governmental Power: The Meaning of Time Dependence." *Comparative Political Studies* 27(2):190–210.

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

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

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

Box-Steffensmeier, Janet M. and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*. Cambridge University Press.

Box-Steffensmeier, Janet M., Laura W. Arnold and Christopher J.W. Zorn. 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.

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

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

Diermeier, Daniel, Hülya. Eraslan and Antonio Merlo. 2003. "A Structural Model of Government Formation." *Econometrica* pp. 27–70.

Diermeier, Daniel and Peter van Roozendaal. 1998. "The Duration of Cabinet Formation Process in Western Multi-Party Democracies." *British Journal of Political Science* 28:609–626.

Diermeier, Daniel and Randy T. Stevenson. 1999. "Cabinet Survival and Competing Risks." *American Journal of Political Science* pp. 1051–1068.

Fearon, James. 1998. "Bargaining, Enforcement, and International Cooperation." *International Organization* 52(2):269–306.

Fortna, Virginia. 2004. "Interstate Peacekeeping: Causal Mechanism and Empirical Effects." *World Politics* 56:481–519.

Franzese Jr., Robert J. and Jude C. Hays. 2007. "Spatial Econometric Models of Cross-Sectional Interdependence in Political Science Panel and Time-Series-Cross-Section Data." *Political Analysis* 15(2):140–64.

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

Gumbel, Emil J. 1959. "Multivariate Distributions with Given Margins." *Revista da Faculdade de Ciencias* 7(2):179–218.

Gumbel, Emil J. 1960. "Bivariate Exponential Distributions." *Journal of the American Statistical Association* pp. 698–707.

Honoré, Bo. and Áureo de Paula. 2008. "Interdependent Durations." Working Paper .

King, Gary, James E. Alt, Nancy E. Burns and Michael Laver. 1990. "A Unified Model of Cabinet Dissolution in Parliamentary Democracies." *American Journal of Political Science* 34(3):846–871.

Laver, Michael. 1998. "Models of Government Formation." *Annual Review of Political Science* 1(1):1–25.

Laver, Michael. 2003. "Government Termination." *Annual Review of Political Science* 6(1):23–40.

Laver, Michael and Kenneth A. Shepsle. 1996. *Making and Breaking Governments: Cabinets and Legislatures in Parliamentary Democracies*. Cambridge University Press.

Lillard, Lee A. 1993. "Simultaneous Equations for Hazards." *Journal of Econometrics* 56:189–217.

Lupia, Arthur and Kaare Strøm. 1995. "Coalition Termination and the Strategic Timing of Parliamentary Elections." *American Political Science Review* 89(3):648–665.

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

Martin, Lanny and Randy T. Stevenson. 2003. "Government Formation in Parliamentary Democracies." *American Journal of Political Science* 45(1):33–50.

Murillo, M. Victoria and Andrew Schrank. 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.

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

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

Simmons, Beth A. and Zachary Elkins. 2004. "The Globalization of Liberalization: Policy Diffusion in the International Political Economy." *American Political Science Review* 98(1):171–189.

Strøm, Kaare, Ian Budge and Michael J. Laver. 1994. "Constraints on Cabinet Formation in Parliamentary Democracies." *American Journal of Political Science* pp. 303–335.

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

Warwick, Paul. and Stephen T. Easton. 1992. "The Cabinet Stability Controversy: New Perspectives on a Classic Poblem." *American Journal of Political Science* 36(1):122–146.

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

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

5 Tables and Figures

Figure 1: Government Formation and Duration

Time for Formation	Average Survival	Standard Deviation	Minimum	Maximum
Less than 50 days	580	481	1	1818
Between 50 and 100 days	649	515	10	1941
More than 100 days	818	529	36	1616

Data source: Warwick (1994), Golder (2005), Keesing's World News Archive

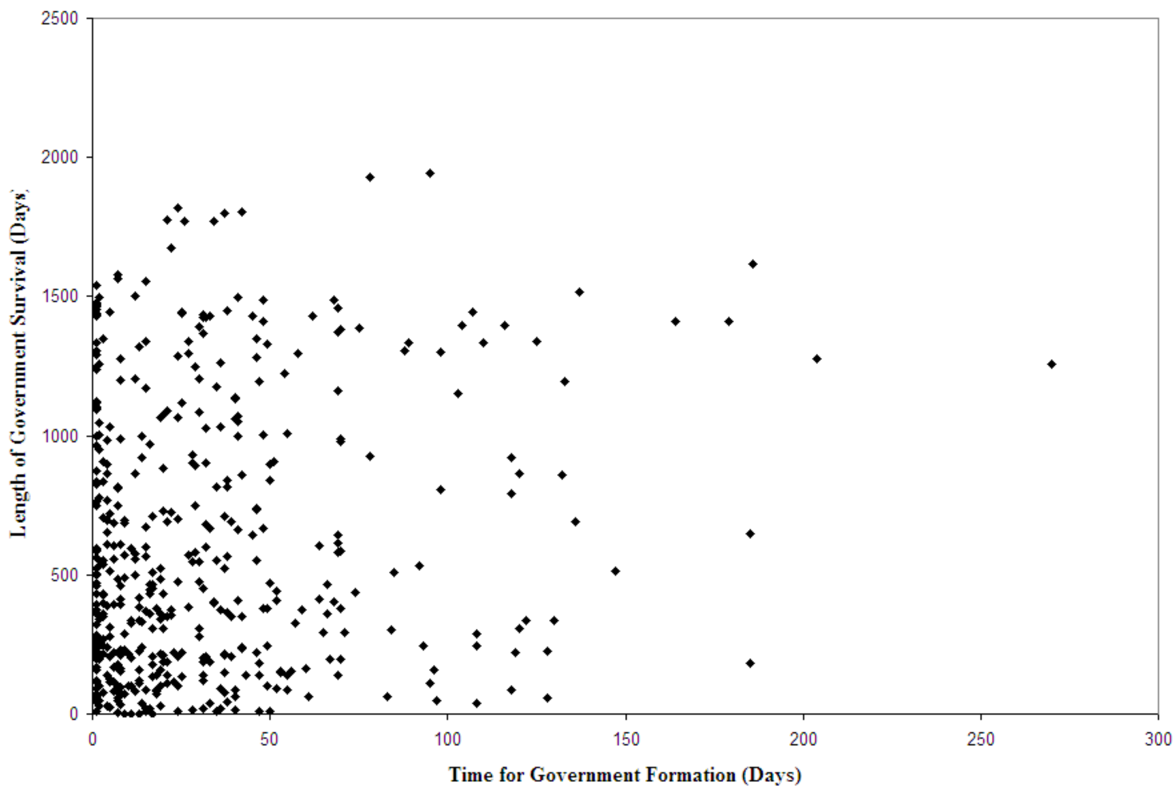

Table 1: Monte Carlo Results for Small Sample ($N = 100$), 1000 Trials

Parameter	Result	Independent	Exogenous	2SLS	FIML
$\hat{\alpha}_1$, $\hat{\alpha}_2$ (0.5,0.5)	Mean	-	0.55, 0.93	0.50, 0.49	0.50, 0.49
	S.D.	-	0.01, 0.05	0.06, 0.12	0.02, 0.05
	RMSE	-	0.05, 0.43	0.06, 0.12	0.02, 0.05
	Mean S.E.	-	0.01, 0.03	0.06, 0.11	0.01, 0.05
	Overconfidence	-	1.19, 1.60	1.04, 1.03	1.04,1.03
$\hat{\beta}_1$, $\hat{\beta}_2$ (1,1)	Mean	1.33,1.34	0.97, 0.72	1.00,1.01	1.00,1.01
	S.D.	0.13, 0.17	0.06, 0.12	0.09,0.17	0.06, 0.12
	RMSE	0.36, 0.38	0.07,0.31	0.09,0.17	0.06, 0.12
	Mean S.E.	0.12, 0.16	0.06, 0.10	0.08, 0.17	0.06, 0.12
	Overconfidence	1.07, 1.01	1.05, 1.17	1.01, 1.00	1.03, 1.00
$\hat{\lambda}_1^{-1}$, $\hat{\lambda}_2^{-1}$ (1,2)	Mean	2.08, 2.85	0.98, 1.74	- *	1.00, 2.00
	S.D.	0.07, 0.10	0.03, 0.07	-	0.04, 0.10
	RMSE	1.08, 0.85	0.04, 0.27	-	0.04, 0.10
	Mean S.E.	0.07, 0.10	0.03, 0.06	-	0.04, 0.10
	Overconfidence	1.04, 1.06	1.01, 1.23	-	1.00, 1.01
$\hat{\alpha}_1$, $\hat{\alpha}_2$ (-0.5,-0.5)	Mean	-	-0.55, -1.06	-0.48, -0.43	-0.50, -0.49
	S.D.	-	0.03, 0.09	0.23, 0.89	0.04, 0.13
	RMSE	-	0.06, 0.57	0.23, 0.90	0.04, 0.13
	Mean S.E.	-	0.03, 0.08	0.23, 0.50	0.03, 0.12
	Overconfidence	-	1.00, 1.14	0.99, 1.80	1.05, 1.09
$\hat{\beta}_1$, $\hat{\beta}_2$ (1,1)	Mean	1.34, 1.37	0.96, 0.64	1.02, 1.07	1.01, 1.03
	S.D.	0.46, 0.40	0.13, 0.23	0.29, 0.90	0.14, 0.28
	RMSE	0.57, 0.54	0.14, 0.43	0.29, 0.90	0.14, 0.29
	Mean S.E.	0.32, 0.37	0.13, 0.22	0.31, 0.61	0.13, 0.27
	Overconfidence	1.41, 1.08	1.05, 1.06	0.95, 1.48	1.04, 1.03
$\hat{\lambda}_1^{-1}$, $\hat{\lambda}_2^{-1}$ (1,2)	Mean	2.48, 2.85	0.97, 1.63	- *	0.99, 1.98
	S.D.	0.34, 0.24	0.08, 0.13	-	0.08, 0.24
	RMSE	1.51, 0.88	0.08, 0.40	-	0.08, 0.24
	Mean S.E.	0.17, 0.22	0.08, 0.12	-	0.08, 0.23
	Overconfidence	1.95, 1.07	1.01, 1.05	-	1.01, 1.05
$\hat{\alpha}_1$, $\hat{\alpha}_2$ (0.5,-0.5)	Mean	-	0.41, 0.25	0.53, -0.60	0.51, -0.51
	S.D.	-	0.05, 0.17	0.38, 0.53	0.06, 0.22
	RMSE	-	0.10, 0.76	0.38, 0.54	0.06, 0.22
	Mean S.E.	-	0.05, 0.11	0.39, 0.49	0.06, 0.20
	Overconfidence	-	1.00, 1.48	0.99, 1.09	1.05, 1.07
$\hat{\beta}_1$, $\hat{\beta}_2$ (1,1)	Mean	0.80, 0.82	0.96, 0.72	1.02, 1.06	1.01, 1.03
	S.D.	0.18, 0.24	0.13, 0.27	0.29, 0.41	0.14, 0.28
	RMSE	0.27, 0.30	0.14, 0.39	0.29, 0.42	0.14, 0.28
	Mean S.E.	0.16, 0.22	0.13, 0.23	0.31, 0.41	0.13, 0.27
	Overconfidence	1.10, 1.08	1.05, 1.18	0.95, 1.01	1.04, 1.02
$\hat{\lambda}_1^{-1}$, $\hat{\lambda}_2^{-1}$ (1,2)	Mean	1.24, 1.71	0.97, 1.69	- *	0.99, 1.97
	S.D.	0.10, 0.14	0.08, 0.15	-	0.08, 0.24
	RMSE	0.26, 0.32	0.08, 0.35	-	0.08, 0.24
	Mean S.E.	0.09, 0.13	0.08, 0.13	-	0.08, 0.23
	Overconfidence	1.03, 1.08	1.01, 1.14	-	1.01, 1.03

* The estimates for λ's can be computed using the estimated α and the estimated variance of the error terms. We have not done the computations yet.

Table 2: Monte Carlo Results for Large Sample ($N = 500$), 1000 Trials

Parameter	Result	Independent	Exogenous	2SLS	FIML
	Mean	-	0.55, 0.93	0.50, 0.49	0.50, 0.50
	S.D.	-	0.01, 0.05	0.06, 0.12	0.01, 0.05
$\hat{\alpha}_1$, $\hat{\alpha}_2$	RMSE	-	0.05, 0.43	0.06, 0.12	0.01, 0.05
(0.5,0.5)	Mean S.E.	-	0.01, 0.03	0.06, 0.12	0.01, 0.05
	Overconfidence	-	1.17, 1.59	1.03, 1.01	1.01, 1.02
	Mean	1.33, 1.34	0.97, 0.71	1.00, 1.01	1.00, 1.00
	S.D.	0.13, 0.17	0.06, 0.12	0.08, 0.17	0.06, 0.12
$\hat{\beta}_1$, $\hat{\beta}_2$	RMSE	0.36, 0.38	0.07, 0.31	0.08, 0.17	0.06, 0.12
(1,1)	Mean S.E.	0.12, 0.16	0.06, 0.10	0.08, 0.17	0.06, 0.12
	Overconfidence	1.08, 1.04	1.02, 1.15	0.98, 1.00	1.00, 1.00
	Mean	2.08, 2.85	0.98, 1.74	- *	1.00, 1.99
	S.D.	0.07, 0.10	0.04, 0.07	-	0.04, 1.10
$\hat{\lambda}_1^{-1}$, $\hat{\lambda}_2^{-1}$	RMSE	1.08, 0.85	0.04, 0.27	-	0.04, 0.10
(1,2)	Mean S.E.	0.07, 0.10	0.03, 0.06	-	0.04, 0.10
	Overconfidence	1.02, 0.98	1.04, 1.25	-	1.02, 1.01
	Mean	-	-0.56, -1.06	-0.50, -0.49	-0.50, -0.50
	S.D.	-	0.01, 0.04	0.06, 0.12	0.02, 0.05
$\hat{\alpha}_1$, $\hat{\alpha}_2$	RMSE	-	0.06, 0.56	0.06, 0.12	0.02, 0.05
(-0.5,-0.5)	Mean S.E.	-	0.01, 0.04	0.06, 0.11	0.01, 0.05
	Overconfidence	-	0.98, 1.09	1.04, 1.02	1.04, 1.02
	Mean	1.34, 1.34	0.96, 0.63	1.00, 1.01	1.00, 1.00
	S.D.	0.25, 0.18	0.06, 0.10	0.09, 0.17	0.06, 0.12
$\hat{\beta}_1$, $\hat{\beta}_2$	RMSE	0.42, 0.38	0.07, 0.38	0.09, 0.17	0.06, 0.12
(1,1)	Mean S.E.	0.15, 0.17	0.06, 0.10	0.08, 0.17	0.06, 0.12
	Overconfidence	1.63, 1.07	1.04, 1.05	1.03, 1.01	1.03, 1.00
	Mean	2.60, 2.90	0.98, 1.65	- *	1.00, 1.99
	S.D.	0.20, 0.11	0.03, 0.06	-	0.04, 0.10
$\hat{\lambda}_1^{-1}$, $\hat{\lambda}_2^{-1}$	RMSE	1.61, 0.91	0.04, 0.36	-	0.04, 0.10
(1,2)	Mean S.E.	0.08, 0.10	0.03, 0.06	-	0.04, 0.10
	Overconfidence	2.56, 1.13	1.00, 1.07	-	0.99, 1.04
	Mean	-	0.41, 0.21	0.51, -0.52	0.50, -0.51
	S.D.	-	0.02, 0.08	0.10, 0.20	0.03, 0.09
$\hat{\alpha}_1$, $\hat{\alpha}_2$	RMSE	-	0.10, 0.72	0.10, 0.20	0.03, 0.09
(0.5,-0.5)	Mean S.E.	-	0.02, 0.05	0.10, 0.19	0.02, 0.09
	Overconfidence	-	0.98, 1.60	1.04, 1.03	1.04, 1.02
	Mean	0.80, 0.80	0.96, 0.72	1.00, 1.01	1.00, 1.01
	S.D.	0.08, 0.11	0.06, 0.12	0.09, 0.17	0.06, 0.12
$\hat{\beta}_1$, $\hat{\beta}_2$	RMSE	0.21, 0.22	0.07, 0.31	0.09, 0.17	0.06, 0.12
(1,1)	Mean S.E.	0.07, 0.10	0.06, 0.10	0.08, 0.17	0.06, 0.12
	Overconfidence	1.07, 1.07	1.04, 1.17	1.03, 1.00	1.03, 1.00
	Mean	1.25, 1.74	0.98, 1.74	- *	1.00, 2.00
	S.D.	0.04, 0.07	0.03, 0.07	-	0.04, 0.10
$\hat{\lambda}_1^{-1}$, $\hat{\lambda}_2^{-1}$	RMSE	0.25, 0.27	0.04, 0.27	-	0.04, 0.10
(1,2)	Mean S.E.	0.04, 0.06	0.03, 0.06	-	0.04, 0.10
	Overconfidence	1.04, 1.13	1.00, 1.23	-	0.99, 1.01

* The estimates for λ's can be computed using the estimated α and the estimated variance of the error terms. We have not done the computations yet.

Table 3: Estimation Results for the Cabinet Formation and Survival Duration

	Independent Durations		Exogenous Durations		2SLS		FIML	
Formation duration ($\mathbf{y}_1$)								
$\boldsymbol{\theta}_1$ (Scale parameter 1)								
Constant	3.473***	1.877***	2.602***	1.944***	-1.31	0.171	2.356***	1.319***
	(0.063)	(0.249)	(0.157)	(0.281)	(0.998)	(2.604)	(0.265)	(0.416)
Continuation	-0.958***	-0.865***	-0.944***	-0.871***	-1.23***	-0.732***	-1.01***	-0.862***
	(0.14)	(0.135)	(0.134)	(0.136)	(0.214)	(0.175)	(0.138)	(0.135)
Investiture		0.019		0.025		0.266		0.036
		(0.104)		(0.105)		(0.188)		(0.105)
Effective Parties		0.102**		0.1**		0.201***		0.105***
		(0.041)		(0.041)		(0.062)		(0.041)
Polarization		0.882*		0.9*		1.15		1.168**
		(0.462)		(0.463)		(1.068)		(0.494)
Returnability		0.507*		0.505*		0.711*		0.515*
		(0.277)		(0.277)		(0.386)		(0.276)
Post-Election		1.173***		1.215***		1.147***		1.126***
		(0.103)		(0.130)		(0.279)		(0.106)
Caretaker		0.317		0.301		0.241		0.414*
		(0.214)		(0.216)		(0.503)		(0.22)
$\boldsymbol{\alpha}_1$ Dependency 1								
Survival			1.019***	-0.07	0.704***	0.068	1.036***	0.086*
			(0.083)	(0.136)	(0.172)	(0.383)	(0.088)	(0.051)
$\boldsymbol{\lambda}_1^{-1}$ (Shape parameter 1)								
Constant	1.187***	1.019***	1.143***	1.019***			1.168***	1.015***
	(0.043)	(0.038)	(0.042)	(0.038)			(0.043)	(0.038)
Cabinet survival ($\mathbf{y}_2$)								
$\boldsymbol{\theta}_2$ (Scale parameter 2)								
Constant	5.207***	6.251***	5.13***	6.271***	5.875***	6.302***	5.302***	6.251***
	(0.106)	(0.182)	(0.108)	(0.197)	(0.420)	(0.264)	(0.118)	(0.182)
Max Duration	0.998***	0.488***	0.719***	0.485***	1.368***	0.473***	0.191***	0.488***
	(0.084)	(0.103)	(0.123)	(0.104)	(0.204)	(0.14)	(0.044)	(0.103)
Investiture		-0.192***		-0.194***		-0.385***		-0.193***
		(0.069)		(0.069)		(0.121)		(0.070)
Effective Parties		-0.026		-0.026		-0.071		-0.026
		(0.024)		(0.024)		(0.053)		(0.025)
Polarization		-1.686***		-1.694***		-2.495***		-1.685***
		(0.25)		(0.252)		(0.501)		(0.25)
Returnability		-0.258		-0.273		-0.312		-0.261
		(0.167)		(0.177)		(0.337)		(0.172)
Post-Election		0.314***		0.315***		0.379		0.312***
		(0.09)		(0.090)		(0.249)		(0.094)
Caretaker		-1.092***		-1.096***		-1.022***		-1.092***
		(0.148)		(0.149)		(0.219)		(0.148)
$\boldsymbol{\alpha}_2$ Dependency 2								
Formation			0.251***	-0.023	-0.618***	-0.031	-0.053*	0.002
			(0.092)	(0.091)	(0.206)	(0.186)	(0.027)	(0.026)
$\boldsymbol{\lambda}_2^{-1}$ (Shape parameter 2)								
Constant	0.785***	0.691***	0.781***	0.691***			0.791***	0.691***
	(0.03)	(0.026)	(0.03)	(0.026)			(0.031)	(0.026)
Log-Likelihood	-1499.24	-1369.39	-1479.32	-1369.22			-1491.02	-1367.53

Note: We could compute the Weibull shape parameter λ for the 2SLS model by using the other parameter estimates and the estimated variances. We have not done the computation yet. Significance levels : $*$: 10% $**$: 5% $***$: 1%

Table 4: The Timing of NAFTA Position Taking

	BSAZ (1997)	Independent	Exogenous	2SLS	FIML
Constituency Factors					
Union Membership	3.21**	-0.33	-0.34	1.72**	-0.33
	(1.19)	(0.28)	(0.28)	(0.87)	(0.28)
Perot Vote, %	-4.91	0.34	0.39	1.69	0.39
	(4.27)	(0.57)	(0.56)	(1.88)	(0.56)
Perot Vote, % Squared	15.64	-1.18	-1.48	-5.36	-1.44
	(11.72)	(1.55)	(1.55)	(5.32)	(1.55)
Mexican Border	1.84**	-0.24***	-0.27***	-0.64***	-0.27***
	(0.32)	(0.05)	(0.05)	(0.15)	(0.05)
Household Income	0.01	0.01	0.01	0.02	0.01
	(0.09)	(0.01)	(0.01)	(0.04)	(0.01)
Interest Group Factors					
Corporate Contributions	-1.44**	0.10	0.09	0.25	0.09
	(0.52)	(0.07)	(0.07)	(0.22)	(0.07)
Labor Contributions	1.09*	-0.09	-0.10	-0.02	-0.09
	(0.50)	(0.06)	(0.06)	(0.20)	(0.06)
Institutional Factors					
NAFTA Committee	0.04	-0.0004	0.002	0.02	0.002
	(0.11)	(0.0130)	(0.013)	(0.04)	(0.013)
Republican Leadership	0.56**	-0.06*	-0.05	-0.05	-0.05*
	(0.26)	(0.03)	(0.03)	(0.10)	(0.03)
Democratic Leadership	0.08	-0.02	-0.02	-0.03	-0.02
	(0.23)	(0.03)	(0.03)	(0.09)	(0.03)
Individual Factors					
Interaction Effect of Ideology and Union Membership	-4.39**	0.44**	0.42*	0.61	0.42*
	(1.78)	(0.22)	(0.22)	(0.73)	(0.22)
Interaction Effect of Ideology and Household Income	0.16	-0.02	-0.01	-0.006	-0.01
	(0.13)	(0.01)	(0.01)	(0.049)	(0.01)
Shape Parameter λ		8.92***	8.95***		8.95***
		(0.38)	(0.38)		(0.38)
Spatial Parameter ρ			-0.09**	-0.69**	-0.08*
			(0.05)	(0.32)	(0.04)
Log Likelihood		194.277	196.300		196.041

This table compares our results with Table 2 in Box-Steffensmeier et al. (1997). Table 2 in Box-Steffensmeier et al. (1997) is for the model that explains the timing of position taking by the House members. "Stars" represent the following; $*$: 10% $**$: 5% $***$: 1%. All of our models were estimated with state fix effects.