Draft of forthcoming article

Selection Bias and Continuous-Time Duration Models:
Consequences and a Proposed Solution

American Journal of Political Science
Volume 50, Issue 1
(January 2006)

Frederick J. Boehmke
University of Iowa
frederick-boehmke@uiowa.edu

Daniel S. Morey
University of Iowa

Megan Shannon
University of Mississippi

Abstract

This paper analyzes the consequences of non-random sample selection for continuous-time duration analyses and develops a new estimator to correct for it when necessary. We conduct a series of Monte Carlo analyses that estimate common duration models as well as our proposed duration model with selection. These simulations show that ignoring sample selection issues can lead to biased parameter estimates, including the appearance of (non-existent) duration dependence. In addition, our proposed estimator is found to be superior in root mean square error terms when non-trivial amounts of selection are present. Finally, we provide an empirical application of our method by studying whether self-selectivity is a problem for studies of leaders' survival during and following militarized conflicts.

Acknowledgements: The authors would like to thank Scott Bennett, Suzie De Boef, Garrett Glasgow, Joseph Lang, Gary King, Chris Zorn and the anonymous referees for helpful comments and suggestions; Randolph M. Siverson, Giacomo Chiozza, and H.E. Goemans for providing assistance with data. All remaining errors are ours. Statistical code used for Monte Carlo simulations and DURSEL program used for model estimation in STATA are available from the authors' websites.

Disclaimer: This document is the final version of the manuscript, submitted by the corresponding author to *the American Journal of Political Science* Editorial Office. Readers should note that minor corrections or revisions may occur during the copyediting process. Authors citing this paper should consult the final, published version (on line or in paper) when it is available in January 2006.

Introduction

The study of the duration of events by political scientists has increased dramatically in recent years. This has lead to opportunities to test our theories in novel ways and has inspired new, more sophisticated theories that focus on the timing of political phenomena. Substantively, this has improved our understanding of a variety of topics including the lifespan of cabinets in parliamentary democracies, the length of international conflicts and partnerships, and the duration and timing of legislative policy adoptions. Statistically, the widespread application of duration models has garnered attention to a variety of statistical issues, many of which are unique to duration analyses.[1]

In a number of cases, political scientists have been able to develop solutions to problems unique to duration analyses. But in some cases, solutions have not yet been found. One such issue is non-random sample selection, which occurs when unobserved factors that affect the duration of an event also determine whether the event is observed at all. While solutions for sample selection in linear regression and discrete choice models have been developed and employed fruitfully in political science, these solutions are inappropriate for the study of duration data. Yet the same underlying processes that cause sample selection for continuous and binary outcomes can operate just as easily on duration outcomes. In many cases, some of which we discuss in the following section, political scientists express concern about the possibility of selection in their duration analyses, but have had to employ ad hoc fixes or put the problem aside. With no information about the consequences of ignoring non-random selection for duration analyses, it is impossible to know the implications of these decisions.

In this paper we study the consequences of non-random sample selection for duration models and propose an estimator to test and correct for it. In particular, we focus on the effect of sample selection for one increasingly popular class of models: continuous-time duration models such as the exponential, Weibull, and Cox models. Our evidence indicates that the consequences of ignoring selection processes are just as serious for duration models as they are for standard regression models. Most importantly, we find that parameter estimates may be biased in an unknown direction, making conclusions drawn from them tenuous at best.

Our proposed solution to modeling sample selection is a full information maximum likelihood model that simultaneously estimates the duration and selection processes. This approach parallels the standard correction for sample selection in linear regression and discrete choice frameworks, but uses a more appropriate distribution that permits both exponential and Weibull durations. While the bivariate exponential distribution that we use allows us to derive appropriate duration models, it does have the limitation of not permitting maximal error correlation. Yet Monte Carlo analysis indicates that our proposed model is superior to standard duration models that do not account for selection even when the error correlation is outside its assumed range: the estimates are closer to the truth and have smaller root mean square error than standard duration models. We conclude this paper by using our model to explore how war participation affects the tenure of political leaders.

Sample Selection and Duration Models

In this section we provide an overview of the problem of non-random sample selection in general, the problems it causes for continuous and binary choice models, and

discuss duration studies that may suffer from selection bias. Of course, we cannot know whether selection bias is a problem for the analyses discussed below. Rather, we wish to highlight the variety of studies that might benefit from the ability to test and correct for it, including cases where authors have themselves raised the issue of selection bias but have not had a way to address it.

Selection bias has received a fair amount of attention in political science, particularly in recent years.[2] While survey researchers have generally been in the forefront of this issue, international relations and comparative politics scholars have also begun to consider the effects of self-selectivity and strategic decisions on empirical models.[3] Sample selection[4] occurs when unobserved factors that influence the process of interest help determine whether that process is observed at all. When this happens, values of the dependent variable for uncensored observations are systematically unrepresentative of the population being studied.[5] One can not even argue that the dependent variable is representative for the observed sample since selection induces unrepresentative values of the dependent variable, even after controlling for values of the independent variables.

This can be illustrated by an example: if a country's leader has private information about his ability to quickly win a war, then this unobserved factor may make him more likely to initiate a war, and may also make his country more likely to win the war (Fearon 1995).[6] Conducting an analysis of what determines victory in the sample of observed wars leads to inconsistent coefficient estimates under a broad range of circumstances. Selection bias can be avoided by estimating a model that accounts for the decision to enter a war while simultaneously estimating the factors that influence victory. In the example given, since winning a war may be treated as a discrete event, one would

estimate a probit-probit selection model for either censored (Dubin and Rivers 1990; Maddala 1983; Sartori 2003) or truncated data (Boehmke 2003; Maddala 1983).

Now consider the same example in a duration framework: rather than estimating a model of victory, one might wish to model the length of the war. In this situation, one would usually estimate any of a variety of duration models for discrete or continuous-time data (Allison 1984; Box-Steffensmeier and Jones 1997, 2004). Yet these models do not account for the potential relationship between unobserved factors that influence the occurrence of war and unobserved factors that determine the length of war.[7] If this is indeed the case, there are no standard models to control for the selection process when estimating the duration of war.[8]

We believe that there are many cases across the subfields of political science similar to the one just described. Examples from international relations include studies of the duration of security alliances (Bennett 1997) and the longevity of newly democratic regimes (Kadera, Crescenzi, and Shannon 2003). In comparative politics the literatures on cabinet formation (Martin and Vanberg 2003) and stability (King, Alt, Burns, and Laver 1990; Warwick and Easton 1992) raise questions about selection bias. Examples from American politics include studies of the time until appointment of political nominees (McCarty and Razaghian 1999; Shipan and Shannon 2003). All of these works study durations that are only observed for certain individuals, countries, or cabinets. One can easily imagine how unobserved factors influence both the observation and duration of the events under study.

Consider the case of cabinet formation. A variety of studies develop theories and empirical tests about which cabinets form, how long negotiations take, and how long the

resulting cabinets last.[9] Most empirical studies of cabinets focus on one of these three steps, but theoretical arguments suggest unobserved factors that might affect cabinet formation and subsequent duration. As Laver (2002) argues, unobserved shocks should be measured in studies of cabinets, since all anticipated events are accounted for during formation. For example, public opinion shocks are central to Lupia and Strom's (1995) theory of cabinet formation: in countries where only the current government can dissolve the legislature and call new elections, unanticipated and possibly unobserved opinion shocks can lead to renegotiation among coalition members, or to new elections that favor the formation of specific governments. Since the expected duration of these governments influences the timing of the elections and the specific government that forms, there may be a selection process at work.[10]

A related group of studies examine the duration of cabinet formation. Diermeier and van Roozendaal (1998) focus on the roles of uncertainty and private information for the negotiation of cabinets. These unobserved factors may influence the length of the bargaining process, since in some cases parties may extend the negotiation period to send a signal to other participants. This signal may influence both the outcome of negotiations and the duration of the subsequent government. Martin and Vanberg (2003, p.328 fn. 11) explicitly question the assumption that the unobserved factors are independent:

> One can easily imagine how this assumption may be violated – e.g., when personality conflicts between party leaders are important to both government termination and subsequent bargaining delay. To account for this, we would have to move away from standard censoring techniques and examine these processes jointly using a model of stochastic censoring (which would be interesting but beyond the scope of the present study).

Martin and Vanberg are candid in their concern over how selection processes influence their study of cabinet negotiations. Yet the lack of appropriate statistical techniques

ultimately renders them unable to model the link between government termination and delay in cabinet bargaining.

We believe the literature on cabinet duration and negotiations, as well a multitude of other studies in political science, may benefit from a method that allows scholars to link selection processes to duration analyses. The studies described in this section analyze durations that are only observed after political leaders, organizations, or countries decide to take specific actions – actions that may reasonably depend on the subsequent duration. Our method, developed in the next section, allows researchers to directly account for the selection process and to obtain accurate estimates when it is a problem.

A Duration Model With Selection

Following the logic of existing models for non-random sample selection, the best approach to obtaining accurate parameter estimates of duration in the face of selection is to simultaneously model both processes. This allows one to condition on the selection process when estimating the duration model. Thus, our proposed solution parallels the approach taken by Heckman (1976; 1979) for linear regression and Dubin and Rivers (1989) for logit and probit models. In this section we introduce our Full Information Maximum Likelihood (FIML) estimator for exponential and Weibull durations. Our estimator is similar to Prieger's (2002) Flexible Parametric Selection (FPS) model, which uses a generalized bivariate distribution to bind together the selection and duration equations. Our estimator is based on a related distribution and shares many of the features of the FPS model for exponential and Weibull durations, including a similar limitation on the range of the error correlation.

Our objective is to model a continuous duration outcome, Y_i, with associated density $f(Y_i)$. This outcome depends on a set of covariates by letting $y_i = \exp(-\mathbf{x_i}\boldsymbol{\beta})\varepsilon_i$.[11] The selection process leads us to observe Y_i only for observations for which the binary censoring variable c_i takes on the value 1. Modeling the duration process, then, requires us to account for the selection process by writing out the probability of uncensored and censored observations: $\Pr(Y_i = y_i, c_i = 1)$ and $\Pr(c_i = 0)$. Rewriting the former as a conditional probability and combining the two pieces leads to the following likelihood function for all observations:[12]

$$\Pr(\mathbf{Y}, \mathbf{c}) = \prod_{i=1}^{n} \Big(\Pr(c_i = 1 \mid Y_i = y_i) f(y_i) \Big)^{c_i = 1} \Big(\Pr(c_i = 0) \Big)^{c_i = 0}.$$

The next step is to use specific distributions to calculate each of the probabilities. In the Heckman correction for continuous data, one would model $\Pr(c_i = 1 \mid Y_i = y_i)$ and $\Pr(c_i = 0)$ with probit models and $f(y_i)$ with a linear regression model. Since these models are inappropriate for duration data, we must substitute different distributions. In this paper we use a bivariate exponential distribution, as the univariate exponential is a common duration distribution and can be modified to allow for duration dependence in the form of the Weibull distribution.

There are different forms of the bivariate exponential model to choose from, each of which has benefits and drawbacks.[13] We utilize one developed by Gumbel (1960) with the following joint and cumulative density functions:

$$f(x, y) = \exp(-x - y)\big(1 + \alpha(2\exp(-x) - 1)(2\exp(-y) - 1)\big),$$
$$F(x, y) = (1 - \exp(-x))(1 - \exp(-y))\big(1 + \alpha \exp(-x - y)\big).$$

The measure of association between the two variables x and y is given by $-1 \leq \alpha \leq 1$, and the error correlation is given by $\rho = \alpha / 4$.

One limiting characteristic of this particular implementation of the bivariate exponential is the restriction $|\rho| \leq 0.25$. As we mentioned, other bivariate exponentials exist, but this one appears to provide the best combination of empirical appropriateness and generality. One particularly appealing feature is the allowance of both positive and negative correlation, which explicitly tests the hypothesis that sample selection is ignorable. While this is clearly a limitation of our estimator, others impose similar ones: Prieger's FPS model with exponential duration limits $|\rho| \leq 0.282$ (Prieger 2002). Despite this shortcoming, we anticipate that even when ρ is outside this range, our estimator may provide a better approximation to the truth than ignoring the problem altogether. Evidence from our particular Monte Carlo analysis is consistent with this expectation.

We now use this distribution to develop the discrete outcome models in our likelihood. We start with the marginal distribution of the censoring variable. This corresponds to a binary outcome model, much like logit or probit, but with the errors following an exponential distribution. We write this model as a function of systematic and stochastic components as follows:

$$c_i^* = \exp(\mathbf{w}_i \boldsymbol{\gamma}) u_i;$$
$$c_i = \begin{cases} 0 \text{ if } c_i^* \leq 1 \\ 1 \text{ if } c_i^* > 1. \end{cases}$$

We pick one as the threshold mainly for convenience and because the exponential is defined only for non-negative numbers.[14] Using this model implies that the probability

that an observation is censored is given by $\Pr(\exp(\mathbf{w}_i\boldsymbol{\gamma})u_i \le 1)$; since the error follows a standard exponential distribution this is given by $F(1|\exp(-\mathbf{w}_i\boldsymbol{\gamma}))=1-\exp(-\exp(-\mathbf{w}_i\boldsymbol{\gamma}))$.

While an exponential discrete choice model may seem unusual, transforming it by taking logs puts it in a more familiar form: $\ln(c_i^*) = \mathbf{w}_i\boldsymbol{\gamma} + \ln(u_i)$. The transformed error term now follows a Type I extreme value distribution (Johnson and Kotz 1970); its cdf is similar to the standard normal cdf, though it is not symmetric. Monte Carlo results suggest that even if this equation is misspecified and the error term is (for example) normally distributed, the parameter estimates for the duration equation may be unaffected. Note that this approach permits the selection process to be based on either a single decision or the timing of that decision through a discrete-time duration model.

Now we derive the conditional probability that an observation is uncensored given its duration. Using the bivariate exponential distribution given above, the conditional cumulative density is:

$$F(x \mid Y = y) = 1 - \exp(-x)\big(1 + \alpha(2\exp(-y) - 1)(\exp(-x) - 1)\big).$$

Substituting $c_i^* = \exp(\mathbf{w}_i\boldsymbol{\gamma})u_i$ and $y_i = \exp(-\mathbf{x}_i\boldsymbol{\beta})\varepsilon_i$, and defining $\lambda_{1i} = \exp(-\mathbf{w}_i\boldsymbol{\gamma})$ and $\lambda_{2i} = \exp(\mathbf{x}_i\boldsymbol{\beta})$ we write

$$F(u_i \mid \lambda_{1i}, \varepsilon_i = y_i\lambda_{2i}) = 1 - \exp(-\lambda_{1i}u_i)\big(1 + \alpha(2\exp(-\lambda_{2i}y_i) - 1)(\exp(-\lambda_{1i}u_i) - 1)\big).$$

The probability that an observation is uncensored is one minus this quantity evaluated at $u_i = 1$. Lastly, we allow for right-censoring by calculating the joint probability that an observation selects in and has duration greater than the right censoring point, y_i^0:

$$\begin{aligned}\Pr(y_i \ge y_i^0, u_i > \exp(-\mathbf{w}_i\boldsymbol{\gamma})) &= 1 - \mathrm{F}(y_i^0 \mid \lambda_{2i}) - \mathrm{F}(1 \mid \lambda_{1i}) + \mathrm{F}(1, y_i^0 \mid \lambda_{1i}, \lambda_{2i}),\\ &= \exp(-\lambda_{1i} - \lambda_{2i}y_i^0)\big[1 + \alpha(1 - \exp(-\lambda_{2i}y_i^0))(1 - \exp(-\lambda_{1i}))\big].\end{aligned}$$

Letting d_i be an indicator variable for whether an observed duration is right-censored, we combine the components for uncensored observations with an observed failure time, uncensored observations that are right-censored, and censored observations to complete our exponential likelihood function.

$$\ln L(\boldsymbol{\beta},\boldsymbol{\gamma},\alpha \mid \mathbf{X},\mathbf{W},\mathbf{Y},\mathbf{c},\mathbf{d}) =$$
$$\sum_{i=1}^{n} c_i(1-d_i)\left[-\lambda_{1i}-\lambda_{2i}y_i+\ln(\lambda_{2i})+\ln(1+\alpha(1-2\exp(-\lambda_{2i}y_i))(1-\exp(-\lambda_{1i})))\right]$$
$$+c_i d_i\left[-\lambda_{1i}-\lambda_{2i}y_i^0+\ln(1+\alpha(1-\exp(-\lambda_{2i}y_i^0))(1-\exp(-\lambda_{1i})))\right]$$
$$+(1-c_i)\left[\ln(1-\exp(-\lambda_{1i}))\right].$$

While the exponential duration model with selection may be useful for many applications, it does not allow for the possibility of duration dependence, which is a concern in many empirical applications. Yet with a simple modification we can derive the likelihood for the commonly utilized Weibull model.[15] This is straightforward since if a variable u follows a Weibull distribution, then u^p (for $p>0$) follows an exponential distribution (Johnson and Kotz 1970).[16] Using this fact, we can transform a Weibull variable into an exponential and substitute it into the bivariate exponential cumulative distribution function:

$$F(x,y) = (1-\exp(-x^p))(1-\exp(-y))\left(1+\alpha\exp(-x^p-y)\right).$$

The Weibull likelihood with right-censoring is then derived by repeating the same steps as for the exponential.

$$\ln L(\boldsymbol{\beta},\boldsymbol{\gamma},\alpha,p \mid \mathbf{X},\mathbf{W},\mathbf{Y},\mathbf{c},\mathbf{d}) =$$
$$\sum_{i=1}^{n} c_i(1-d_i)\left[-\lambda_{1i}+\ln(1+\alpha(2\exp(-(\lambda_{2i}y_i)^p)-1)(\exp(-\lambda_{1i})-1))+\ln(p)+\ln(\lambda_{2i})\right.$$
$$\left.+(p-1)\ln(\lambda_{2i}y_i)-(\lambda_{2i}y_i)^p\right]+c_i d_i\left[-\lambda_{1i}-\left(\lambda_{2i}y_i^0\right)^p\right.$$
$$\left.+\ln(1+\alpha(1-\exp(-(\lambda_{2i}y_i)^p))(1-\exp(-\lambda_{1i})))\right]+(1-c_i)\left[\ln(1-\exp(-\lambda_{1i}))\right].$$

Monte Carlo Analysis of Naïve and FIML Duration Estimators

In this section we conduct Monte Carlo analysis to analyze the consequences of selection bias for standard duration models and to demonstrate the effectiveness of our proposed FIML correction. When the data are generated according to the assumptions of our model, we find commonly used duration models such as the exponential, Weibull, and Cox produce parameter estimates that are biased whereas those from our FIML estimator appear to be consistent. Further, even when we generate data from an alternate error distribution that allows correlation outside the assumed range, our FIML estimates are biased, but are still closer to the truth than those obtained from the naïve models. In both cases we find that our estimator always outperforms the alternatives in root mean square error terms when non-random selection is non-trivial.

We rely on Monte Carlo analysis to study these issues because even the simple exponential model does not produce a closed form solution for the coefficient estimates, making it impossible to analytically demonstrate the effect of selection on parameter estimates. When it is impossible to explicitly derive general results, Monte Carlo analysis allows us to study an estimator's properties while controlling or manipulating other components (Mooney 1997). Of course, one should keep in mind that the conclusions depend on the set-up of the particular simulations.

For our Monte Carlo analysis we generate data according to an exponential duration model. We use the exponential duration model because it is the simplest and allows us to focus on the effects of selection without complicating the model by adding duration dependence. Also, popular models such as the Weibull and Cox include the exponential as a special case, which allows us to study how these more general models

fare on the same selected data and whether their estimates indicate that the exponential is, in fact, the correct model. The data are generated in accordance with the model presented in the previous section and use the bivariate exponential distribution to create continuous time duration data that are observed according to a discrete exponential censoring rule:

$$
\begin{aligned}
c_i^* &= \exp(-0.5 + 1 \times w_i) u_i, \\
c_i &= \begin{cases} 0 \text{ if } c_i^* \le 1, \\ 1 \text{ if } c_i^* > 1, \end{cases} \\
y_i &= \exp(-(0 + 0.75 \times x_i)) \varepsilon_i .
\end{aligned}
$$

We assume that y_i is observed if and only if c_i=1.

To assess the performance of our estimator and its robustness to the limitation on the correlation parameter, we conduct two sets of simulations based on different error generating processes. The first set of simulations generates error terms from the assumed bivariate exponential distribution by taking independent and identically distributed draws from the bivariate uniform and then transforming them to exponentials (Devroye 1986). The second set of simulations assesses the performance of the model when the error correlation is outside its assumed range by taking draws from the bivariate normal with correlation ρ and then transforming each to an exponential.[17] As both simulations are run for $|\rho| \le 0.25$, comparing the estimates in this range also assesses the robustness of our approach to the bivariate exponential assumption.

For our simulations we generate 5000 observations of our independent variables $\boldsymbol{X}$ and $\boldsymbol{W}$, which we hold fixed for the entirety of the simulations.[18] A single simulation consists of fixing the value of ρ, generating draws of the errors, calculating the values of c_i and $\underline{y}_i$, then estimating the FIML exponential duration model with selection using all the data and the naïve duration models using only observations where c_i is one, which

happens for about half of our observations.[19] We save the resulting coefficients and standard errors and then repeat the process for different values of ρ.[20]

In practice, political scientists rarely estimate an exponential model due to its assumption of duration independence. Many empirical processes are duration dependent, and ignoring duration dependence leads to inconsistent parameter estimates. Thus, most studies estimate either a parametric model such as the Weibull, Gompertz, gamma or lognormal duration models, or a semi-parametric model such as the Cox, which estimates the coefficients without specifying a functional form for duration dependence (Box-Steffensmeier and Jones 1997, 2004; Zorn 2000). To study whether and how selection affects these more general models, we also estimate the commonly used Weibull and Cox models on the same data.[21] All models are estimated using robust standard errors.

[Figure 1 here.]

Before moving on to a detailed evaluation of the Monte Carlo results, Figure 1 presents a graphical depiction of the average parameter estimates. We omit the estimates from the exponential model as they are essentially equivalent to those from the Weibull model. The results for error correlations between -0.25 and 0.25 are from the simulation using the bivariate exponential distribution and the results outside those points are from the transformed bivariate normal error terms. The plotted values provide a simple demonstration of the bias that ignoring non-random sample selection introduces into parameter estimates. It also shows that our duration model with selection produces estimates that are at the correct values in the permitted range of correlation and closer to the true values outside that range. The next two tables provide additional evidence in favor of our FIML estimator.

Table 1 presents the details of the average estimates for three parameters: the constant term, α; the log of the duration dependence parameter, ln(p), from the Weibull model; and the estimated error correlation from our FIML estimator.[22] Note that the Cox model does not estimate any of these parameters. Results are reported for both sets of simulations for selected values of ρ. In addition to average estimates and their standard deviations, we also report the Root Mean Squared Error (RMSE), which adds the variance to the squared bias and takes the square root.[23]

[Table 1 here.]

It is immediately apparent from Table 1 that ignoring sample selection leads to biased estimates of both the constant term and the duration dependence parameter whenever there is non-zero correlation between the unobserved factors. The constants for the naïve models are biased in the opposite direction of the correlation whereas the duration dependence parameter is biased in the same direction as the correlation.[24] Because the original model contains no duration dependence, this suggests that sample selection can be added to heterogeneity as another cause of apparent duration dependence. Note that the bias is slightly greater for the simulations using the correct error distribution.

By comparison, the average estimates for the exponential duration model with selection show virtually no bias whenever the correlation is between –0.25 and 0.25, though the estimates using the alternate error distribution display increasing bias as the error correlation exceeds its assumed range. Because our estimator restricts the correlation, it is unable to fully correct for sample selection with large correlation.[25] This is illustrated by the fact that while the average estimate of the error correlation is accurate

in the proper range, it is estimated at its boundaries in the second set of simulations when it is outside this range. Yet despite this limitation, the average estimates are still always closer to the truth than those from the naïve duration models, suggesting that our method provides more accurate estimates even when this assumption is not met. Since the true amount of correlation is never known in an empirical application, this result is important for demonstrating the benefits of our correction.

To better illustrate the superiority of our method, compare the RMSEs of the different models for the intercept term. Only when the correlation is zero do the naïve models have smaller RMSEs; once it is non-zero the RMSEs for our model are smaller for both sets of simulations. Further, the difference increases quickly, with the naïve models producing RMSEs that are two to three times as great even when the correlation is relatively small.

Table 2 shows the simulation results for β and includes the estimates from the Cox model. The estimates from our FIML estimator appear to be consistent in the first set of simulations using the true error distribution. They are generally quite close to the truth using the alternate distribution when the correlation is in its assumed range. Some bias occurs near the bounds because the simulations will include errors with greater than assumed correlation. The bias becomes more pronounced once the error is outside its bounds. Again, the presence of non-zero correlation leads to biased estimates from the naïve models.[26] Consider the results for the exponential and Weibull models. For both sets of simulations, the bias in the parameter estimate is in the direction of the correlation, though the bias appears to be slightly greater at comparable values of ρ using the proper

error distribution.[27] Note that the bias for these models is greater than for ours even when the error is outside its assumed range in the second set of simulations.

Comparing the standard deviations of the estimates shows that those from the naïve models are generally the same or slightly smaller than those from the FIML model. Combining these two by comparing the RMSEs, however, shows that our estimator strongly outperforms both the exponential and Weibull models: the RMSEs for the naïve models increase quickly from equality when $|\rho| = 0.05$, to 34% when $\rho = -0.20$, and to 60% when $\rho = 0.20$ using the true error distribution. Similar results are obtained with the alternate error distribution, though the RMSE ratios start to decrease as the correlation approaches its bounds due to the bias in the results for our FIML model.

[Table 2 here.]

The results for the Cox model are different because it reports a hazard interpretation that combines the coefficient and the duration dependence parameter: $\beta \times p$. Since the Cox is a semi-parametric model that does not estimate *p,* we can not report results for β in isolation. Of course, the true value of p is one, so if the estimate is not biased then the estimates are comparable. Recall from the earlier results, however, that the Weibull model's estimates of the duration dependence parameter are biased. Therefore any bias in the estimates from the Cox model will include biases in both β and p.

The average estimates using the correct error distribution show that the bias in the Cox model is greater and in the same direction as for the two naïve models. The bias is greater because the presence of positive duration dependence when the correlation is positive and negative duration dependence when it is negative pulls the estimate of $\beta \times p$

further away from 0.75. In the second set of simulations, the Cox parameter estimates have similar bias, though it appears to be greater when the correlation is positive and smaller when it is negative. The differences in the estimates of duration dependence across the two sets of simulations account for the difference in the amount of bias in the Cox estimates. The RMSE comparison shows that for both sets of simulations, the Cox model is never preferred to our estimator, with RMSEs ranging from 10% to 300% greater.

In sum, our comparison of the duration model with selection to standard duration models offers strong evidence for our FIML estimator. Not only does it appear to provide consistent estimates when its assumptions are met, it also produces estimates that are closer to the truth even when using the alternate error distribution that allows for correlations outside the assumed range.[28] In both cases, our model also produces smaller RMSEs whenever the correlation is non-negligible.

Empirical Application: War and the Tenure of Leaders

Our Monte Carlo simulations show that ignoring non-random sample selection processes produces biased parameter estimates. In this section we demonstrate the potential consequences of ignoring sample selection by applying our correction to a political science question: does war influence the tenure of leaders? This investigation, in addition to our simulations, allows for a more complete understanding of how selection impacts duration analyses.

How might selection processes affect an analysis of war and leader duration? As discussed before, non-random selection causes bias if unmeasured factors that influence observation of an event are correlated with unmeasured factors in the equation of interest.

In a study of war and leader tenure, factors that affect a leader's expected tenure after war initiation may also affect the leader's decision to go to war. For example, diversionary theory argues that leaders who are vulnerable to being removed from office are more likely to use aggressive foreign policies (DeRouen 1995; Ostrom and Job 1986).[29] Because measuring vulnerability is difficult at best (Miller 1999, 399), its effects are largely consigned to the error term.[30] Diversionary theory therefore implies a negative correlation between expected post-crisis tenure and the decision to enter into a crisis. On the other hand, if only secure leaders enter into crises, we would expect a positive correlation between the error terms.

In the following analysis, we apply our FIML duration model with selection to Bueno de Mesquita and Siverson's (1995) study of war and leader survival. We follow their study in constructing our duration analysis, but we also model the leader's decision to get involved in a conflict. Since Bueno de Mesquita and Siverson use a Weibull duration model for their estimation, we report estimates from our Weibull duration model with selection.

We begin our investigation with the leader's decision to go to war, which determines whether we observe a post-crisis tenure. To estimate the selection model, our analysis shifts from the usual dyadic level to the monadic level of observation. To this end, we employ the state leader-year as our unit of analysis. Our data for the selection model include all state leaders from 1919 to 1975. The dependent variable for the selection equation, *Enter*, is coded one if a leader went to war that year and zero otherwise.[31] War is defined as involvement in any conflict with at least 1,000 battle-related deaths, as reported by the Correlates of War project (Small and Singer 1982).

To explore what influences a leader's decision to go to war, we include state and individual level measures drawn from various sources in the selection equation. We examine the possible monadic effects of regime type (Haas 1965, Rummel 1983) with *Democracy*, an indicator variable constructed by using each state's *Polity IV* score, where states scoring 6 and above are considered democratic (Marshall and Jaggers 2000). *Major Power* status is an indicator variable coded one if the state is a major power in the current year, using standard Correlates of War measures (Small and Singer 1982). *Borders* is the total number of countries that border the state in the given year, derived from the COW Direct Contiguity Data set, version 3.0. To control for neoliberal influences, we include the natural *Log of Trade Openness*, taken from Chiozza and Goemans (2003).

Modeling selection allows us to test additional hypotheses derived by Bueno de Mesquita and Siverson regarding leaders' decisions to enter a crisis, as a significant component of their theory focuses on a leader's decision to enter into war. Specifically, we test Bueno de Mesquita and Siverson's sixth hypothesis: the longer an autocrat is in power, the more likely he is to go to war. To this end, we include an interaction term, *Autocracy*Tenure*, in our selection equation (along with the two main effects of *Autocracy* and *Tenure*).[32]

The second stage of our analysis examines the tenure of all state leaders who went to war between 1919 and 1975. We estimate a duration model of these leaders' time in office using Bueno de Mesquita and Siverson's (1995) heads of state data. The dependent variable in the duration equation is the leader's *Time in Office* following crisis onset, measured in days. Following Bueno de Mesquita and Siverson, we explain leader time in office using measures of *Log Tenure*, *Battle Deaths*, crisis *Winner*, and *Log*

*Tenure*Democracy*. We also include an indicator variable for *Democracy*, since it is part of an interaction term in the analysis.[33]

Table 3 presents the results of our two-staged FIML duration with selection model as well as results from the independent selection and duration models. Column 1 displays the probit results of the leader's decision to enter a crisis; column 2 displays the results of the original Bueno de Mesquita and Siverson (1995) Weibull model for our smaller sample extending from 1919 to 1975; column 3 displays results from our Weibull duration model with selection.[34] The probit and FIML selection equations provide similar conclusions, with statistically significant coefficients for total borders and trade openness. More bordering countries increases the likelihood of a leader initiating war, while increasing trade openness reduces his probability of going to war. None of the other variables achieve statistical significance, including the interaction term testing Bueno de Mesquita and Siverson's (1995) hypothesis regarding autocratic tenure.

[Insert Table 3 here.]

Of more interest for our purposes here are the coefficients of the duration equations. Overall, the results from our duration model with selection are quite similar to those from the independent Weibull model in Column 2, indicating that pre-crisis tenure significantly increases a leader's time in office, whereas democracy significantly decreases leader tenure.[35] The main distinction between the independent Weibull and FIML models is that while battle deaths is significant at the 0.10 level in the Weibull model, it become insignificant once selection is accounted for (p=0.17).

In this case, then, the consequences of selection are not severe for the assessment of statistical significance, with the possible exception of the battle deaths variable. But

there are important ways in which correcting for selection changes or improves our inferences. First, the duration with selection model generates a positive and significant estimate of the correlation between the error terms.[36] The positive value of rho is inconsistent with predictions of diversionary theory, suggesting that if a leader's expected tenure is greater than average, his expected probability of going to war is also greater than average. Second, by accounting for selection, the constant term increases from –1.16 in the independent Weibull model to –0.37 in the Weibull duration model with selection. Consequently, our model indicates a larger hazard rate and therefore shorter times until failure. The shift in the constant term occurs because leaders' tenure is artificially increased by factors that influence the selection of war. Once we account for the decision to go to war, the estimated tenure of a leader decreases. This finding is consistent with our Monte Carlo results, which indicate that positive error correlation biases the estimate of the constant term downward.

Accounting for a leader's selection into war as well as his tenure is important for two reasons. First, our analysis provides information about the factors that determine which leaders go to war, as modeling the decision to go to war decreases the estimated post-war tenure of a leader. Second, the presence of significant correlation and the shift in the intercept term alter the shape of the estimated survival function and hazard rate, as well as the impact of the independent variables on their shapes.

To illustrate these consequences we calculated survival functions for both models. The survival function gives the probability that a leader is still in office at a given moment in time.[37] We are also interested in the marginal effect of independent variables such as democracy on the survival function. Thus we generated survival functions for

autocracies and democracies and report the difference between them, providing an estimate of the effect of democracy on leader tenure after war initiation.[38] These first differences, along with the survival functions for autocracies, are presented in Figure 2.

[Insert Figure 2 here.]

The results indicate a substantively large difference in the estimated survival functions and first differences. The survival function for the duration model with selection is much lower than for the Weibull model, indicating that the probability of a leader lasting five years, for example, is only thirty-nine percent and not fifty-three percent. Overall, the probability of surviving more than two years is about two-thirds smaller once the selection process is accounted for. Put another way, the median tenure for our duration model with selection is 2.8 years whereas for the naïve Weibull model it is 5.75 years. The difference arises from the larger constant term estimated by our model, which implies a greater hazard rate, earlier failure, and consequently, shorter tenure.

These changes in the survival rate also produce changes in the estimated first difference for democracy. While the Weibull model indicates that changing a regime from an autocracy to a democracy decreases the survival function anywhere between twenty-five to fifty percent, the duration model with selection results indicate that the change is never greater than twenty five percent. On average, the corrected first difference is less than half the size of the inaccurate first difference from the Weibull.[39] Thus our proposed correction provides additional information about the decision to go to war, a leader's tenure in office after war initiation, and the conditions under which leaders may be expected to initiate crises.

Discussion and Conclusions

The consequences of sample selection for duration analyses should receive more attention from political scientists. Our simulations provide clear evidence that ignoring non-random selection poses severe problems for analyses using duration models, including the exponential, Weibull and Cox models. These problems include biased coefficient estimates, incorrect evidence of duration dependence, inaccurate predicted hazard and survival functions, and incorrect first differences Studies that potentially suffer from selection bias may produce erroneous conclusions about what factors influence (or do not influence) the duration process of interest.

Despite our findings, further work is needed. Our Monte Carlo study can be modified in a number of ways, including varying the rate of selection, the correlation of the independent variables, the presence and amount of duration dependence, or including the selection equation variable in the duration equation. In addition, our FIML estimator can be further developed to allow for time-varying covariates.

We also think it would be fruitful to pursue alternate estimators for several reasons. First, different estimators such as the log-normal or gamma distributions would permit alternate forms of duration dependence. Second, this would allow for more common specifications of the selection equation error distribution. Third, using bivariate distributions such as the normal would increase the range of allowed correlation between the error terms by producing a probit selection equation with a log-normal duration equation. One could also allow correlation between the two equations by introducing correlated random effects, as in Gordon's (2002) competing risks model, though at the cost of increased estimation complexity.

A more general approach can be developed through the application of copula theory, which allows one to bind together a variety of continuous distributions. While our FIML estimator is a member of the family of Farlie-Gumbel-Morgenstern copulas (Smith 2003), a wider variety of specifications of the two equations is available.[40] Lee (1983) applies the bivariate normal family of copulas to linear regression models with sample selection; Prieger (2002) extends it to duration outcomes. This approach has the advantage of permitting maximal correlation, but is more difficult to work with and does not necessarily preserve characteristics of the marginal distributions in the conditional distributions, which Prieger (2002) argues may be particularly important for duration analyses. Both our FIML estimator and Prieger's FPS model preserve these characteristics. Future work should explore these alternate approaches if sample selection proves to be a relevant consideration for duration studies in political science.

References

Achen, Christopher H. 1986. *The Statistical Analysis of Quasi-Experiments.* Berkeley:The University of California Press.

Allison, Paul D. 1984. *Event History Analysis: Regression for Longitudinal Event Data*. Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-046. Beverly Hills and London: Sage Publications.

Beck, Nathaniel, Jonathan N. Katz and Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42(4): 1260-1288.

Bennett, D. Scott. 1997. "Testing Alternative Models of Alliance Duration, 1816-1984." *American Journal of Political Science* 41(3): 846-878.

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

Berinsky, Adam. 1999. "The Two Faces of Public Opinion." *American Journal of Political Science*: 43(4): 1209-1230.

Berry, Frances Stokes and William D. Berry. 1990. "State Lottery Adoptions as Policy Innovations: An Event History Analysis." *American Political Science Review* 84(2): 395-415.

Boehmke, Frederick J. 2003. "Using Auxiliary Data to Estimate Selection Bias Models, With an Application to Interest Groups' Use of the Direct Initiative Process." *Political Analysis* 11(3): 234-254.

Box-Steffensmeier, Janet M. and Bradford D. Jones. 1997. "Time Is of the Essence: Event History Models in Political Science." *American Journal of Political Science* 41(4): 1414-1461.

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

Brambor, Thomas, William Roberts Clark, and Matt Golder. Forthcoming, 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14(1).

Brehm, John. 1993. *The Phantom Respondents: Opinion Surveys and Political Representation*. Ann Arbor: University of Michigan Press.

Brehm, John. 1999. "Alternative Corrections for Sample Truncation: Applications to the 1988, 1990, and 1992 Senate Election Studies." *Political Analysis* 8(2): 147-165.

Bueno de Mesquita, Bruce, and Randolph M. Siverson. 1995. "War and the Survival of Political Leaders: A Comparative Study of Regime Types and Political Accountability." *American Political Science Review* 89(4): 841-855.

Chiozza, Giacomo and Hein E. Goemans. 2003. "Peace through Insecurity: Tenure and International Conflict." *Journal of Conflict Resolution* 47 (4): 423-443.

Clark, David H. and William Reed. 2003. "A Unified Model of War Onset and Outcome." *The Journal of Politics* 65(1): 69-91.

Correlates of War 2 Project. 2003. "State System Membership List, v2002.1." http://cow2.la.psu.edu.

DeRouen, Karl R. Jr. 1995. "The Indirect Link: Politics, the Economy, and the Use of Force." *The Journal of Conflict Resolution* 39(4): 671-695.

Devroye, Luc.1986. *Non-Uniform Random Variate Generation*. New York: Springer-Verlag.

Diermeier, Daniel and Peter van Roozendaal. 1998. "The Duration of Cabinet Formation Processes in Western Multi-Party Democracies." *British Journal of Political Science* 28(4): 609-626.

Dubin, Jeffrey A. and Douglas Rivers. 1989. "Selection Bias in Linear Regression, Logit and Probit Models." *Sociological Methods and Research* 18(2): 360-390.

Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49(3): 379-414.

Gordon, Sanford C. 2002. "Stochastic Dependence in Competing Risks." *American Journal of Political Science* 46(1): 200-217.

Gumbel, E.J. 1960. "Bivariate Exponential Distributions." *Journal of the American Statistical Association* 55(292): 698-707.

Haas, Michael. 1965. "Societal Approaches to the Study of War." *Journal of Peace Research* 2(4): 307-323.

Heckman, James J. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models." *Annals of Economic and Social Measurement* 5/4: 475-492.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47(1): 153-161.

Johnson, Norman L. and Samuel Kotz. 1970. *Distributions in Statistics: Continuous Univariate Distributions - 1*. Boston: Houghton Mifflin Company.

Johnson, Norman L. and Samuel Kotz. 1972. *Distributions in Statistics: Continuous Multivariate Distributions*. New York: Wiley.

Jones, Bradford S. and Regina Branton. 2003. "Event History Modeling in State Politics Research." Paper presented at the 2003 Annual Meetings of the Midwest Political Science Association.

Kadera, Kelly M., Mark J.C. Crescenzi, and Megan L. Shannon. 2003. "Democratic Survival, Peace, and War in the International System." *American Journal of Political Science* 47(2): 234-247.

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

King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Ann Arbor: The University of Michigan Press.

Klein, John P., Shih-Chang Lee and M.L. Moeschberger. 1990. "A Partially Parametric Estimator of Survival in the Presence of Randomly Censored Data." *Biometrics* 46(3): 795-811.

Laver, Michael. 2002. "Government Termination." *Annual Review of Political Science* 6(1): 23–40.

Lee, Lung-Fei. 1983. "Generalized Econometric Models with Selectivity." *Econometrica* 51(2): 507-512.

Lupia, Arthur and Kaarne Strom. 1995. "Coalition Termination and the Strategic Timing of Parliamentary Elections." *American Political Science Review* 89(3): 648-666.

Maddala, G.G. 1983. *Limited Dependent and Qualitative Variables in Econometrics.* Cambridge: Cambridge University Press.

Martin , Lanny W. and Georg Vanberg. 2003. "Wasting Time? The Impact of Ideology and Size on Delay in Coalition Formation." *British Journal of Political Science* 33(2): 323-333.

Marshall, Monty G. and Keith Jaggers. 2000. Polity IV dataset. http://www.bsos.umd.edu/cidcm/inscr/.

McCarty, Nolan, and Rose Razaghian. 1999. "Advice and Consent: Senate Responses to Executive Branch Nominations, 1885-1996." *American Journal of Political Science* 43(4): 1122-1143.

Miller, Ross A. 1999. "Regime Type, Strategic Interaction, and Diversionary Use of Force." *The Journal of Conflict Resolution* 43(3): 388-402.

Mooney, Christopher. 1997. *Monte Carlo Simulation.* Sage Publications.

Ostrom, Charles W. Jr. and Brian L. Job. 1986. "The President and the Political Use of Force." *The American Political Science Review* 80(2): 541-566.

Prieger, James E. 2002. "A Flexible Parametric Selection Model for Non-Normal Data with Application to Health Care Usage." *Journal of Applied Econometrics*: 17(4): 367-392.

Reed, William. 2000. "A Unified Statistical Model of Conflict Onset and Escalation." *American Journal of Political Science* 44(1): 84-93.

Rosecrance, Richard. 1963. *Action and Reaction in World Politics*. Boston: Little, Brown.

Rummel, R.J. 1983. "Libertarianism and International Violence." *Journal of Conflict Resolution* 27(1): 27-71.

Sartori, Anne E. 2003. "An Estimator for Some Binary-Outcome Selection Models without Exclusion Restrictions." *Political Analysis* 11(2): 111-138.

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.

Signorino, Curtis S. 1999. "Strategic Interaction and the Statistical Analysis of International Conflict." *American Political Science Review* 93(2): 279-98.

Small, Melvin and J. David Singer. 1982. *Resort to Arms*. Beverly Hills, CA: Sage.

Smith, Murray D. 2003. "Modeling Sample Selection Using Archimedean Copulas." *Econometrics Journal* 6(1): 99–123.

Timpone, Richard J. 1998. "Structure, Behavior, and Voter Turnout in the United States." *American Political Science Review* 92(1): 145-158.

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

Zorn, Christopher J.W. 2000. "Modeling Duration Dependence." *Political Analysis* 8(3): 367-380.

Table 1.
Monte Carlo Simulation Results for Intercept (α=0), Duration Dependence (ln(p)=0) and Correlation Parameters

	FIML					Exponential				Weibull					
	Intercept			Correlation		Intercept				Intercept				Log Duration Dependence	
Rho	Mean	SD	RMSE	Mean	SD	Mean	SD	RMSE	RMSE ratio	Mean	SD	RMSE	RMSE ratio	Mean	SD
Bivariate Exponential Errors															
-0.20	0.008	0.071	0.071	-0.194	0.041	0.252	0.036	0.255	3.584	0.279	0.037	0.282	3.961	-0.057	0.022
-0.15	-0.003	0.074	0.074	-0.151	0.045	0.183	0.034	0.186	2.526	0.206	0.034	0.209	2.831	-0.049	0.021
-0.10	0.004	0.063	0.063	-0.097	0.044	0.120	0.032	0.124	1.965	0.136	0.034	0.140	2.222	-0.035	0.020
-0.05	0.001	0.056	0.056	-0.048	0.041	0.057	0.028	0.063	1.141	0.065	0.030	0.072	1.288	-0.019	0.023
0.00	0.003	0.050	0.050	0.000	0.034	0.003	0.031	0.031	0.616	0.003	0.031	0.032	0.633	0.001	0.020
0.05	-0.003	0.046	0.047	0.048	0.033	-0.054	0.031	0.062	1.335	-0.063	0.032	0.071	1.525	0.024	0.022
0.10	0.000	0.041	0.041	0.103	0.028	-0.108	0.028	0.111	2.729	-0.127	0.029	0.130	3.202	0.051	0.022
0.15	0.003	0.039	0.039	0.149	0.028	-0.151	0.028	0.154	3.903	-0.181	0.030	0.184	4.658	0.080	0.023
0.20	0.010	0.040	0.041	0.207	0.030	-0.198	0.026	0.200	4.902	-0.239	0.027	0.241	5.899	0.115	0.023
Bivariate Normal Errors Translated to Exponentials															
-0.80	0.629	0.032	0.630	-0.250	0.000	0.924	0.031	0.925	1.468	0.955	0.033	0.955	1.517	-0.064	0.023
-0.60	0.329	0.029	0.331	-0.250	0.001	0.634	0.029	0.634	1.918	0.678	0.031	0.679	2.053	-0.091	0.023
-0.40	0.104	0.050	0.115	-0.231	0.023	0.396	0.035	0.398	3.454	0.435	0.036	0.436	3.786	-0.080	0.021
-0.20	0.025	0.066	0.070	-0.128	0.043	0.180	0.031	0.182	2.599	0.202	0.033	0.205	2.922	-0.049	0.024
-0.10	0.003	0.069	0.069	-0.071	0.049	0.087	0.028	0.091	1.328	0.098	0.028	0.102	1.492	-0.026	0.022
0.00	-0.003	0.053	0.053	0.001	0.038	-0.002	0.029	0.029	0.547	-0.003	0.029	0.029	0.554	0.002	0.021
0.10	-0.008	0.047	0.048	0.071	0.034	-0.083	0.029	0.088	1.836	-0.097	0.029	0.101	2.116	0.035	0.021
0.20	-0.032	0.042	0.053	0.125	0.030	-0.162	0.027	0.165	3.089	-0.187	0.029	0.189	3.549	0.064	0.020
0.40	-0.062	0.027	0.068	0.246	0.010	-0.308	0.024	0.309	4.573	-0.362	0.025	0.363	5.366	0.156	0.020
0.60	-0.194	0.025	0.196	0.250	0.000	-0.443	0.024	0.444	2.263	-0.521	0.024	0.522	2.662	0.258	0.020
0.80	-0.318	0.023	0.319	0.250	0.000	-0.568	0.023	0.569	1.783	-0.667	0.023	0.668	2.093	0.380	0.022

Note: Bivariate exponential entries based on 200 draws for each value of the correlation. Transformed bivariate normal entries based on 100 draws. See text for details of the data generating and estimation processes.

Table 2.
Monte Carlo Simulation Results for Slope Coefficient (β=0.75)

	FIML			Exponential				Weibull				Cox			
Rho	Mean	SD	RMSE	Mean	SD	RMSE	RMSE ratio	Mean	SD	RMSE	RMSE ratio	Mean	SD	RMSE	RMSE ratio
Bivariate Exponential Errors															
-0.20	0.751	0.033	0.033	0.708	0.031	0.052	1.604	0.707	0.031	0.052	1.610	0.656	0.032	0.099	3.039
-0.15	0.750	0.036	0.036	0.717	0.033	0.047	1.288	0.717	0.033	0.047	1.293	0.673	0.034	0.084	2.323
-0.10	0.749	0.031	0.031	0.730	0.029	0.036	1.159	0.729	0.029	0.036	1.161	0.698	0.033	0.061	1.997
-0.05	0.753	0.030	0.030	0.743	0.029	0.030	0.998	0.743	0.029	0.030	0.997	0.727	0.034	0.041	1.377
0.00	0.750	0.030	0.030	0.750	0.029	0.029	0.960	0.750	0.029	0.029	0.960	0.751	0.033	0.033	1.093
0.05	0.751	0.028	0.028	0.759	0.027	0.029	1.038	0.759	0.027	0.029	1.035	0.781	0.035	0.046	1.674
0.10	0.750	0.030	0.030	0.767	0.030	0.034	1.138	0.767	0.030	0.034	1.132	0.812	0.037	0.073	2.400
0.15	0.749	0.025	0.025	0.772	0.025	0.034	1.335	0.772	0.025	0.033	1.314	0.844	0.035	0.100	3.964
0.20	0.746	0.029	0.029	0.778	0.029	0.040	1.371	0.776	0.029	0.039	1.340	0.880	0.039	0.136	4.691
Bivariate Normal Errors Translated to Exponentials															
-0.80	0.632	0.028	0.121	0.586	0.028	0.166	1.376	0.584	0.028	0.168	1.393	0.541	0.030	0.211	1.748
-0.60	0.688	0.029	0.068	0.637	0.030	0.117	1.713	0.635	0.029	0.119	1.744	0.573	0.028	0.179	2.624
-0.40	0.735	0.035	0.038	0.684	0.034	0.074	1.935	0.683	0.034	0.075	1.963	0.626	0.033	0.129	3.362
-0.20	0.747	0.032	0.032	0.720	0.030	0.043	1.324	0.720	0.030	0.043	1.333	0.683	0.033	0.075	2.318
-0.10	0.754	0.030	0.030	0.740	0.028	0.030	0.972	0.740	0.028	0.030	0.975	0.721	0.035	0.045	1.492
0.00	0.756	0.032	0.032	0.756	0.031	0.031	0.960	0.756	0.031	0.031	0.958	0.755	0.035	0.035	1.094
0.10	0.748	0.029	0.029	0.759	0.028	0.029	1.017	0.759	0.028	0.029	1.016	0.787	0.034	0.050	1.729
0.20	0.756	0.028	0.028	0.776	0.026	0.037	1.297	0.775	0.026	0.036	1.285	0.827	0.032	0.083	2.948
0.40	0.769	0.023	0.030	0.805	0.024	0.060	2.006	0.803	0.024	0.058	1.944	0.936	0.037	0.189	6.357
0.60	0.794	0.024	0.050	0.829	0.023	0.083	1.661	0.825	0.023	0.078	1.573	1.062	0.041	0.315	6.321
0.80	0.822	0.020	0.075	0.855	0.020	0.107	1.426	0.847	0.020	0.099	1.321	1.227	0.041	0.478	6.377

Note: Bivariate exponential entries based on 200 draws for each value of the correlation. Transformed bivariate normal entries based on 100 draws. See text for details of the data generating and estimation processes.

Table 3.
Analysis of War Participation and Leader Tenure: Comparison of Naïve Weibull Model with Weibull Duration Model with Selection

	Column 1 Probit Model		Column 2 Naïve Weibull Model		Column 3 Weibull with Selection	
Selection	β	Robust S.E.	β	Robust S.E.	β	Robust S.E.
Democracy	-0.125	0.168			-0.087	0.109
Autocracy	-0.105	0.168			-0.075	0.096
Log(Tenure)	0.019	0.062			0.011	0.036
Log(Tenure) * Autocracy	0.057	0.086			0.038	0.048
Borders	0.040**	0.020			0.028*	0.015
Major Power	0.275	0.190			0.184	0.147
Log(Trade Openness)	-0.097*	0.052			-0.055*	0.033
Constant	-2.663	0.158			-1.739**	0.122
Duration						
Log(Tenure)			-0.800**	0.236	-0.808**	0.240
Democracy			1.885*	1.071	1.959*	1.133
Log(Tenure) * Democracy			-0.450	0.783	-0.514	0.797
Log (Battle Deaths/10K)			0.163	0.099	0.147	0.106
Winner			-0.381	0.548	-0.334	0.565
Constant			-1.160**	0.486	-0.374	0.670
ln p			-0.387**	0.108	-0.652**	0.148
p (Duration Dependence)			0.679**	0.073	0.521**	0.077
rho (Error Correlation)			--	--	0.198**	0.068
N (Uncensored)	5148		69		5148 (69)	
Log Likelihood	-344.33		-114.23		-470.12	
Wald	52.14**		23.00**		18.49**	

Notes: * indicates $p \leq 0.10$ ** indicates $p \leq 0.05$. Robust standard errors cluster on country.

Figure 1.
Average Parameter Estimates from Monte Carlo Simulations

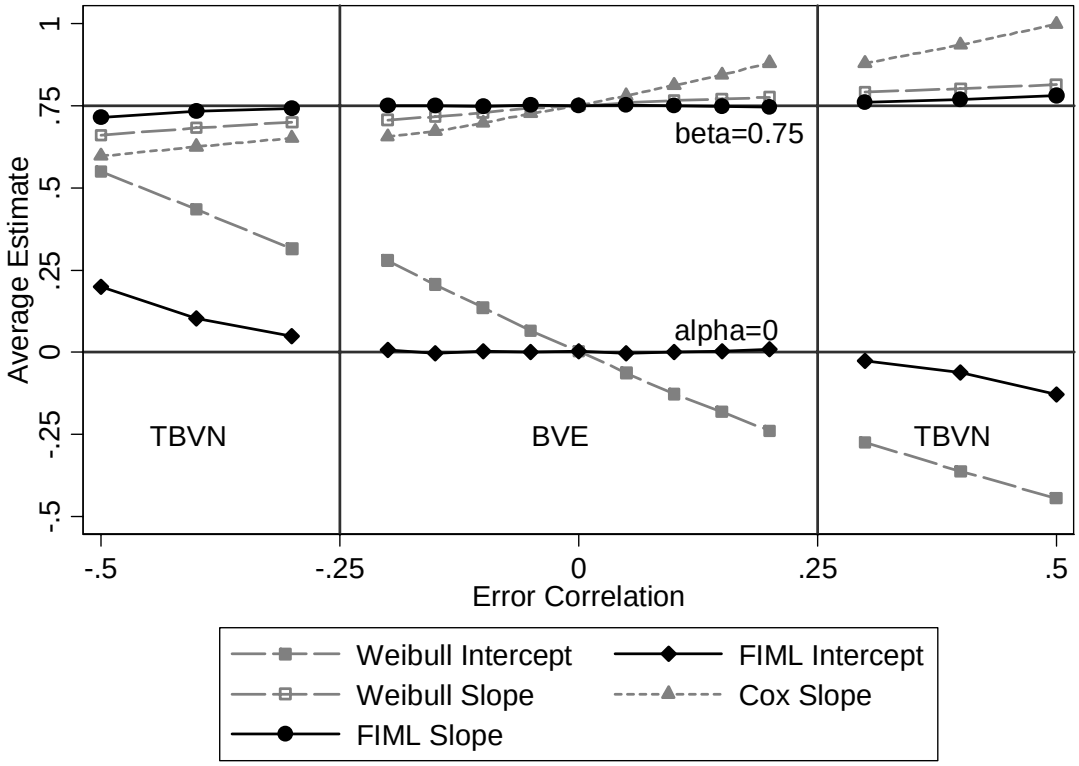
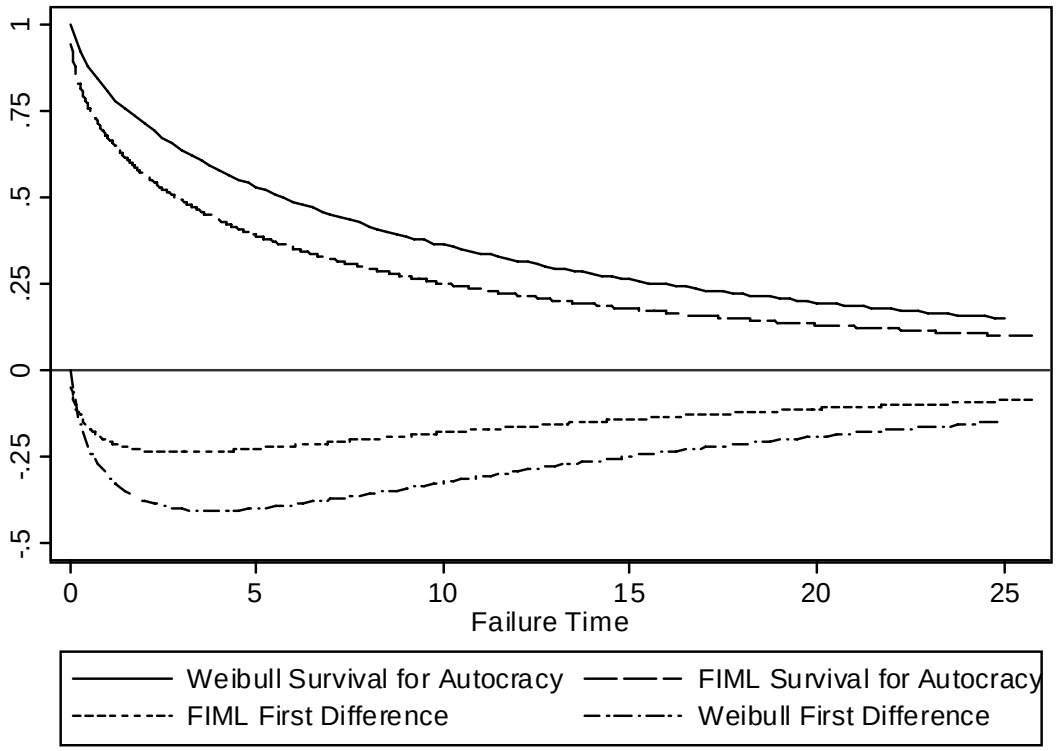


Figure 2.
Impact of Model Choice on Estimated Leadership Survival:
Comparison of Survival Function and Marginal Effect of Autocracy



Notes: Survival functions indicate the probability that a leader is still in office at a given time, given that a crisis was initiated. All variables are set to their average values for autocracies using the 69 observations with uncensored durations. Estimates for democracies are obtained by changing the appropriate indicator variables, including those in interactions. For the duration with selection model, this includes changing their values in the selection equation. First differences represent the difference between the estimated survival functions for democracies and autocracies

[1] We discuss many of these examples and related issues in detail later. Readers looking for more information are referred to Box-Steffensmeier and Jones' (2004) thorough and comprehensive discussion of duration models in political science. Issues that have received extensive attention include controlling for duration dependence (Beck, Katz and Tucker 1998; Bennett 1999; Zorn 2000) and, more recently, accounting for repeated failures (Box-Steffensmeier and Zorn 2002; Branford and Jones 2003).

[2] See, for example, Achen (1986) Berinsky (1999), Brehm (1990, 1999), Boehmke (2003), King (1989), Reed (2000), Signorino (1999), Timpone (1998) and Sartori (2003).

[3] See, for example, Bennett (1999), Hug (2003); Reed (2000), Signorino (1999) and Clark and Reed (2003).

[4] A note on terminology: the specific form of selection that we are studying is known as stochastic censoring, a form of non-ignorable missingness. This corresponds to situations where characteristics of all observations are observed, but the variable of interest is only observed for some observations. In various situations it may be referred to as selectivity, self-selection, non-random sample selection, or selection bias.

[5] An additional note on terminology: in the duration literature censoring refers most commonly to left, right and middle censoring – situations when an observation fails during a time when it is not under observation (i.e., after a study was completed for right censoring). In the selection literature, censoring refers to situation where we have measures of independent variables but no measure of the dependent variable (i.e., a voter may answer demographic questions on a survey but fail to indicate who she voted for).

We use the term censoring to refer to the latter situation and specify right or left censoring if we mean the other.

[6] International relations scholars have argued that private information may play in important role in the decision to initiate wars. Leaders may have better knowledge than potential opponents about factors such as their country's resolve or its military capabilities. It is not a great step to assume that these unobserved factors may also influence the outcome or duration of any wars that occur.

[7] Note that the problem does not arise because the occurrence of war and the outcome or duration of the war are influenced by the same factors. Non-random selection is only a problem when the unobserved factors are related.

[8] The one exception we have found is due to Prieger (2002), which develops an estimator similar to ours as well as one using Lee's (1983) generalized selection model based on the bivariate normal family of copulas (Smith 2002).

[9] The literature is too extensive to discuss every study – a nice overview is provided by Laver (2002).

[10] If these opinion shocks were perfectly measured, the selection problem could be dealt with. It seems reasonable to expect that government officials may have better (and perhaps more idiosyncratic) measures and more information than scholars attempting to study cabinet duration.

[11] Note that this is the hazard interpretation – large coefficients correspond to greater hazards and shorter durations

[12] In most cases, of course, we are substantively interested in the quantity $f(y_i \mid c_i = 1)$. We write out our likelihood using the conditional distribution of the censoring variable because it is easier to derive. The two approaches are mathematically equivalent.

[13] See Johnson and Kotz (1972) for a discussion of various implementations of the bivariate exponential. The drawbacks are generally in terms of restrictions on the values that the correlation can take on, but we also had more difficulty obtaining estimates with some versions.

[14] The choice of threshold is without loss of generality: if the true threshold is not one, the only consequence is that the intercept term shifts the appropriate amount, much the same as with probit and logit.

[15] The choice to allow for Weibull duration dependence in the duration part of our model might suggest that it would also be desirable to extend the discrete part in the same way. While this turns out to be impractical (a Weibull binary choice model is not identified), duration dependence can be controlled for in the selection equation in a discrete-time duration model interpretation through the usual methods (time dummies; linear or quadratic time trends; or cubic splines). See Beck, Katz and Tucker (1998) for discussion.

[16] Note that one can allow for log-normal durations with a probit selection model by estimating a Heckman model with the log duration as the dependent variable (assuming no right censoring). As Prieger (2002) notes, the correlation of u_i with $\ln(\varepsilon_i)$ may range from negative to positive 1, but the correlation between it and ε_i depends on the latter's standard deviation σ – when $\sigma = 1$, for example, $|\rho| \leq 0.76$.

[17] We ran a series of simulated draws of the resulting error terms and found that the correlation between them is about the same as the bivariate normal correlation, though generally a little closer to zero, particularly when the correlation is negative.

[18] These independent variables are generated from a bivariate normal distribution with standard deviations one and correlation 0.5. The values of the coefficients and independent variables were set so that the variation of the systematic portion of the model is a little greater than the variation of the stochastic component.

[19] We estimate the exponential duration model with selection because it corresponds to the underlying data generating process. We also estimated the Weibull duration model with selection on the same data and found the results were indistinguishable from the exponential. We therefore only report one set of results.

[20] All simulations were conducted in Stata 8.2 on a Windows machine with dual 731 MHz processors. Generating the data and estimating each model once took 30 seconds on average. During estimation, we use the inverse of Fisher's Z transformation, $\alpha = Z(\alpha^*) = \left(\exp(2\alpha^*) - 1\right) / \left(\exp(2\alpha^*) + 1\right)$, so that the unbounded α^* generates an estimate of α in the appropriate interval.

[21] We estimated other parametric models that allow for duration dependence and they produced estimates that were virtually identical to those from the Weibull, so we omit them for presentation purposes.

[22] To save space and to focus attention on the parameters of interest (i.e., those of the duration process), we do not report the average values of the selection equation

parameters. We note that our results indicate they are unbiased, or at least consistent. These and other results are available from the authors upon request.

[23] $\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + (\text{bias}(\hat{\theta}))^2$, where the variance is taken from the sampling distribution of the parameter and the bias is calculated by taking the average estimate in the draws and subtracting the true value.

[24] The intuitions behind these findings are as follows: when the correlation is positive, the conditional expected value of the error term for uncensored observations is greater than its unconditional mean (since it is harder for observations with small errors to select in), leading to longer times until failure. This corresponds to a smaller baseline hazard rate. For the duration dependence parameter, positive correlation between the error terms results in a greater proportion of "large" errors in the selected sample, resulting in an increasing failure rate among surviving observations as time increases.

[25] This limitation on the correlation forced the model to push the estimate to the appropriate boundary and occasionally lead to instances where the model failed to converge. Overall, convergence failure occurred in less than three percent of the draws, though it was closer to ten or fifteen percent when the alternate error distribution correlation was at the two extremes.

[26] Note that since the hazard ratio interpretation just reports $\exp(\beta)$, our results have the same implication for reported hazard ratios.

[27] Recall that the explanatory variables in the selection and duration equations are positively correlated. Thus observations with small values of $\boldsymbol{W}$ require large errors to select in, inducing a negative correlation between $\boldsymbol{W}$ and $\boldsymbol{u}$. When the errors and

independent variables are positively correlated, this then leads to a negative correlation between $\boldsymbol{X}$ and $\boldsymbol{\varepsilon}$. This results in relatively early failures when $\boldsymbol{X}$ is large and late failures when it is small, which biases the hazard ratio upwards. Further Monte Carlo analyses indicate that the direction of the bias in the coefficient depends on the sign of the product of the correlation between the independent variables and the coefficient in the selection equation.

[28] Additional simulations also indicate that our findings are robust to the assumption that the censoring rule corresponds to a binary exponential model – if we generate data using a probit censoring rule instead, the duration model with selection still produces estimates with less bias and smaller RMSEs.

[29] While most of the diversionary theory literature focuses on democracies, Rosecrance (1963) finds evidence for non-democratic states engaging in diversionary behavior.

[30] One could model these two decisions as strategic selection in the sense of Signorino (1999). While we think that this could prove fruitful in the future, our first objective was to develop a general duration model with selection rather than one tailored for a specific strategic circumstance.

[31] To conduct our analysis, we merged the duration data with information on all leaders, restricting our analysis to the years 1919 to 1975. Prior to 1919, differences between leader data sets make it difficult to merge the duration data with information on all leaders. The original Bueno de Mesquita and Siverson (1995) analysis examined all leaders between 1816 and 1975.

[32] As we required data on all leaders (not just those who went to war), we used Chiozza and Goemans's (2003) leader tenure data for the selection model.

[33] In general one should not include interactions without main effects since it complicates the interpretation and may lead to false conclusions (i.e., the interaction may be significant because it captures the effect of the omitted main effect) (Brambor, Clark, and Golder, forthcoming). While we would like to replicate Bueno de Mesquita and Siverson's model as closely as possible, in this case the inclusion of the democracy variable influenced the results.

[34] While the naïve Weibull results cohere with predictions made in the original work, some variables decrease in significance. This is most likely due to the reduced sample size, though changes in the underlying data generating process could also cause these differences.

[35] While these results suggest that the effect of tenure is not different for democratic leaders, we should note that even if we omit the democracy indicator the democracy-tenure interaction remains insignificant, possibly due to our reduced sample size.

[36] While the estimate of the correlation is significant, the corresponding unbounded, transformed value narrowly missed significance at the 0.10 level.

[37] For our duration model with selection, the conditional survival function is the probability of an observation being right-censored divided by the probability of going to war. We use the conditional survival function rather than the unconditional survival function (the joint probability of going to war and surviving until a certain time) since we

are interested in leaders' expected tenure given war initiation, as it is this quantity that a leader uses to determine whether to go to war.

[38] All the independent variables were set to their mean values for autocracies. To estimate the impact of democracy, we then changed the value of the relevant indicator variables (in both equations for the selection model) and for the appropriate interaction terms (tenure*democracy in the duration equation and tenure*autocracy in the selection equation).

[39] A similar pattern emerged when we calculated the predicted hazard rates as well, with the marginal effect of democracy on the hazard from the Weibull model about twice as great as the marginal effect from the Weibull duration model with selection.

[40] This family also includes Prieger's (2002) FPS model. See Smith (2002) for a discussion of the application of copula theory to models for sample selection.