



Iowa Research Online
The University of Iowa's Institutional Repository

Department of Political Science Publications

7-11-2013

Accounting for Right Censoring in Interdependent Duration Analysis

Jude C Hays
University of Pittsburgh

Emily U Schilling
University of Iowa

Frederick J. Boehmke
University of Iowa

DOI: <https://doi.org/10.17077/k09r-q7ra>

Copyright © 2013 the authors

Comments

Subsequently published as "Accounting for Right Censoring in Interdependent Duration Analysis" in *Political Analysis* (Summer 2015) 23 (3): 400-414.
doi: [10.1093/pan/mpv012](https://doi.org/10.1093/pan/mpv012)

Hosted by [Iowa Research Online](https://www.iowa-research-online.org/). For more information please contact: lib-ir@uiowa.edu.

Accounting for Right Censoring in Interdependent Duration Analysis ¹

Jude C. Hays
University of Pittsburgh

Emily U. Schilling
University of Iowa

Frederick J. Boehmke
University of Iowa

July 11, 2013

*Paper prepared for presentation at the 2013 Society for Political Methodology Summer Conference,
University of Virginia, July 18-20, 2013.*

¹Corresponding author: jch61@pitt.edu or emily-schilling@uiowa.edu. We are grateful for funding provided by the University of Iowa Department of Political Science.

Abstract

Duration data are often subject to various forms of censoring that require adaptations of the likelihood function to properly capture the data generating process, but existing spatial duration models do not yet account for these potential issues. Here we develop a method to estimate spatial duration models when the outcome suffers from right censoring, the most common form of censoring in this area. In order to address this issue, we adapt Wei and Tanner's (1991) imputation algorithm for censored (nonspatial) regression data to models of spatially interdependent durations. The algorithm treats the unobserved duration outcomes as censored data and iterates between multiple imputation of the incomplete, i.e., right censored, values and estimation of the spatial duration model using these imputed values. We explore performance of estimators for Weibull and log-normal durations in the face of varying degrees of right censoring via Monte Carlo and provide empirical examples of its estimation by analyzing spatial dependence in states' entry dates into World War I.

Introduction

As with other types of outcome variables, political duration outcomes are often dependent across observations, whether that dependence arises from common exposure, contagion, or selection. Spatial econometrics provides a common method for addressing these various forms of interdependence and models for continuous outcomes are well known and growing in their use in political science. Interdependence is not limited by the nature of the dependent variables, of course, and related estimators for discrete and duration outcomes (e.g., Hays and Kachi's (2009)) have also emerged.

With the explosion of duration models in political science over the last decade or two, scholars have become increasingly aware of the unique challenges that duration data often pose. For example, duration data are often subject to various forms of censoring that require adaptations of the likelihood function to properly capture the data generating process. Existing spatial duration models do not yet account for these potential issues and the goal of this paper is to develop a method to estimate spatial duration models when the outcome suffers from right censoring, the most common form of censoring in this area. Right censoring occurs when we partially observe a duration: we know that an observation began and lasted until at least a given point in time, but we do not know exactly how much longer it survived. This poses a significant challenge for modeling interdependence since the outcome for one observation depends on the outcomes for other observations. But when durations are censored, the analyst is incapable of properly capturing this interdependence since the complete outcome data are not observed.

In order to address this issue, we adapt an imputation algorithm for censored (non-spatial) regression data to models of spatially interdependent durations Wei and Tanner's (1991). The algorithm treats the unobserved duration outcomes as censored data and iterates between multiple imputation of the incomplete, i.e., right censored, values and esti-

mation of the spatial duration models using these imputed values. Imputation of the spatially correlated errors is performed via rejection sampling to obtain draws greater than the censoring point, using results on the distribution of a linear combination of Gumbel or log-normal variables. We explore performance of the estimator in the face of varying degrees of right censoring via Monte Carlo and provide empirical examples of its estimation by analyzing spatial dependence in states' entry dates into World War I. This offers a good example given the great degree of interdependence in entry decisions but also because of the presence of significant amount of right censoring caused by countries that did not join before the conflict ended.

Spatial Duration Models

Spatial analysis fits well into the study of duration processes. It is easy to think of examples of these processes that may actually have units whose durations are interdependent with one another, such as policy diffusion across states or countries or international conflict spreading into other countries. Sometimes this interdependence is believed to be a nuisance and must be removed from the data but often this interdependence is important for the types of political phenomena we seek to study. This dependence can emerge in two different ways. The first is when the time to one political event depends on the time to a related event. The second kind is when the time to a particular political event for one actor is dependent on the timing of that same event for other actors Hays and Kachi's (2009).

Previous attempts at including a spatial dimension into the study of duration processes have relied on the inclusion of a variable indicating the number or proportion of neighbors that have experienced the event of interest. This does not allow for the simultaneous nature of spatial dependence to be properly captured. A large amount of research

seeks to incorporate spatial dependence for continuous dependent variables, indicating the importance of modeling spatial autocorrelation in order to prevent biased estimates Gimpel and Cho; Darmofal; Franzese and Hays's (2004; 2006; 2007).

The traditional models of duration processes do not allow for spatial dependence to occur, thus potentially omitting factors that can affect the hazard. Omitted variables greatly affect the results found in the model and they will reduce the effect of those variables that increase the hazard and increase the effect of variables that reduce the hazard. One way to account for the potential bias that results from omitting variables can be controlled for by including frailty terms, which can help account for the fact that some observations may be more frail than others due to variables not included into the model such as a parameter that controls for the spatial interdependence between observations Box-Steffensmeier and Jones; Darmofal's (2004; 2009).

The inclusion of frailty terms still does not solve the problem of properly modeling the spatial interdependence that can arise in duration processes. The model that forms the basis of part of this analysis is the simultaneous equation full information maximum likelihood estimator that is developed by Hays and Kachi (2009). In this model the dependent variable is based off of a Weibull distribution such that

$$y_i^* \sim Weibull(\lambda, \theta), \tag{1}$$

where λ is the shape parameter and θ is the scale parameter. The errors are assumed to be distributed according to the standard Gumbel distribution.¹ A standard Gumbel variable

¹The distribution and density functions of the standard Gumbel distribution are

$$f(u) = e^u e^{-e^u} \tag{2}$$

$$F(u) = 1 - e^{-e^u}. \tag{3}$$

is equal to a logged Weibull variable. Then it follows that

$$y = \ln(y^*) = \ln(\theta) + \frac{1}{\lambda}, \epsilon \quad (4)$$

$$= X\beta + \frac{1}{\lambda}\epsilon. \quad (5)$$

This leads to a full model in matrix notation that is as follows

$$\mathbf{y} = \mathbf{W}\mathbf{y} + \mathbf{X}\beta + \mathbf{L}\mathbf{u}. \quad (6)$$

\mathbf{W} is the matrix through which the spatial dependency enters this model.² The matrix \mathbf{W} contains not only the spatial weights matrix but also ρ which is the parameter of spatial dependence or how much spatial autocorrelation there is in the model. In political science, spatial relationships are generally expected to have a positive relationship with one another, leading to a positive ρ . The matrix \mathbf{L} has $\frac{1}{\lambda}$ down the center diagonal so that each error term is multiplied by it. We know from equation 4 that the error term in this model is multiplied by the inverse of the shape parameter λ . Following from the structural form, the reduced form of the spatial model is

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\beta + (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{L}\mathbf{u}, \quad (7)$$

$$= \mathbf{\Gamma}\mathbf{X}\beta + \mathbf{\Gamma}\mathbf{L}\mathbf{u}, \quad (8)$$

$$= \mathbf{\Gamma}\mathbf{X}\beta + \mathbf{v}. \quad (9)$$

where $\mathbf{\Gamma} = (\mathbf{I} - \rho\mathbf{W})^{-1}$ and $\mathbf{v} = \mathbf{\Gamma}\mathbf{L}\mathbf{u}$. It is evident from the reduced form equation how important it is to include the potential spatial relationship in the model because now one can see that the spatial matrix actually has influence on all of the independent variables as well as the spatially lagged dependent variable since $\mathbf{\Gamma}$ is now multiplied

²The description of this model follows Hays and Kachi's (2009) notation.

by $\mathbf{X}\beta$. This equation shows that each β is now a linear combination of what would happen to the individual regardless of the other actors in the model and the sum of that individual's spatial relationship with all of the other actors. This shows that if the spatial dependence is not equal to 0 (i.e. $\rho \neq 0$) all of the β estimates will be biased. The likelihood function is developed from the change of variables theorem which derives the joint pdf of the y 's from the joint pdf of the u 's since the joint distribution of the u 's is easier to obtain as a result of the fact that they are assumed to be i.i.d and the marginal distribution of u is known. Solving the reduced form equation for \mathbf{u} , yields

$$\mathbf{u} = g^{-1}(\mathbf{y}) = \mathbf{\Gamma L}^{-1}\mathbf{y} - \mathbf{L}^{-1}\mathbf{X}\beta. \quad (10)$$

This function along with Jacobian matrix (\mathbf{J}) of $g^{-1}(\mathbf{y})$ are used to develop the likelihood function for the joint density of \mathbf{y} . Note that this likelihood does not account for any censoring in the dependent variable and is why this estimation procedure on its own is not enough to analyze right censored data. The likelihood function is

$$\mathcal{L} = \left(\prod_i^N f(g^{-1}(y_i)) \right) |\det(\mathbf{J})|, \quad (11)$$

$$= \left(\prod_i^N f(u_i) \right) |\det(\mathbf{J})|. \quad (12)$$

Through Monte Carlos, Hays and Kachi compared their estimator with other estimators that political scientists have used to model spatial interdependence. The authors find that even under small sample conditions, the FIML estimates are unbiased and the standard error estimates are correct, neither of which occurs simultaneously in their three other models examined (two stage least squares, ML-AIDM, and ML-AEDM). It is obvious that this model will be useful in conducting the analysis on the interdependent duration process of interest but it has no solution to the problem of right censoring that is

often confronted in event history analysis.

The Complications of Right Censoring

A censored variable occurs when the value of a variable is unknown over the range of some values. When a variable is censored and analyzed using a standard OLS, the estimates that are provided will be inconsistent. The focus of this paper is on right censoring. Right censoring is when an event is observed only if it occurs before some prespecified time. This was frequently observed in clinical trial studies in which as a result of time, cost, or complications with the patients, the study is terminated before all of the patients have failed (or been cured). Those observations who never realized the event of interest are often given a value equal to the end of the study (Moeschberger and Klein 1997). In the case of the application under study in this paper, the right-censoring has occurred in this data since the World War I ended before all of the countries had entered the war.

Thus, the survival function has an upper limit $t_i = C_i$, where C_i denotes the right censoring point for the i th unit. If the censoring point is known, it will be possible to construct a likelihood function that accounts for this censoring. The likelihood of the sample under conditions in which we have n observations where the full duration time, t , is observed is given by

$$\mathcal{L} = \prod_i^n f(y_i). \quad (13)$$

If some observations are right censored the likelihood function will be changed. For censored cases, the duration, y_i , is observed only to the last observation period y_i^C , after which duration continues, but is unobserved. The duration for right censored cases is restricted to y_i^C . Thus, the observations need to be divided based on whether they are cen-

sored or not. The uncensored cases contribute information regarding failure times but the censored observations cannot since they have not failed yet, incorporating information into the survivor function. Thus the likelihood function will consist of two parts:

$$\mathcal{L} = \prod_{y_i \leq y^C} f(y_i) \prod_{y_i > y^C} S(y_i^C), \quad (14)$$

or

$$\mathcal{L} = \prod_i^n f(y_i)^{\delta_i} [S(y_i^C)]^{1-\delta_i}, \quad (15)$$

where

$$\delta_i = \begin{cases} 1 & \text{if } y_i \leq y_C; \\ 0 & \text{if } y_i > y_C. \end{cases} \quad (16)$$

Without the separation of the data into observations that fail before the censoring point and those that survive past it, the model includes the information from the censored values in the hazard function even though these values never actually failed themselves. They will be treated as if they have actually experienced the event of interest when in reality they never did. This can lead to bias in the estimates of the parameters Box-Steffensmeier and Jones's (2004). Thus, the use of the maximum likelihood presented above is essential in properly specifying a model for data that are right censored.

Right censoring occurs in a variety of political processes. In the application under study in this paper, WWI ended before all of the countries had entered the war. There are a variety of other examples of right censoring that one may also assume have some spatial relationship. If one is studying the duration of a regime or the duration of one party's control in the government, this variable could be censored as a result of there not

being a change in government leadership before then end of the data collection process. Another example would be in the study of state policy diffusion, it is likely that a state may not adopt a particular policy by the last time period that is recorded. In both of these last two examples, it is possible that after the data collection process had ended, the government leadership switched or the state did adopt that particular policy. Both of these examples also represent cases in which one might expect there to be some spatial dependence between neighboring or similar states/governments. It is not hard to imagine that as the more states around a particular state adopt an attractive government policy, there will be more pressure for them to do the same. With the changes in government leadership, if a country nearby experiences instability or a shift in their current regime this could cause a ripple effect throughout the area causing other governments to also experience instability.

Without the solution adapted in this paper, censoring in spatial models has been dealt with in two different ways. The first is assigning a value to all of the censored cases; oftentimes this is the threshold cutoff between those cases that are observed and those that are censored. The second option is to just delete the observations that were censored and disregard them in the analysis. This is problematic when dealing with non-spatial data because it causes the dataset to be truncated. Unless one can be sure that the factors that produce the censoring to occur are unrelated to those that affect whether or not the event will occur, truncating the sample is less than ideal. Box-Steffensmeier and Jones's (2004). Deleting observations is even more problematic in the context of spatial modeling.

The issue of dropping censored cases from spatial data is especially problematic because of the fact that the observation of individual i 's duration is dependent on the observation of all of the other individuals $j \neq i$. If it is assumed that a subset of the j 's is censored, denoted j_c , and they were removed from the data set all of the information in the observation of i 's duration that is dependent on those censored cases would be lost.

The fact that their failure was never observed in the time period that is under study does not mean that their lack of failure had no effect on the observation of i 's failure.

An EM Approach for Imputation and Estimation

The solution to right censoring provided in the previous section is not sufficient for dealing with spatial data. To solve the issue of right censoring within spatial data, one may want to determine the the distribution of the error term for the censored values. In order to do this it is necessary to find the conditional CDF for u given that u is greater than the censoring point. This is difficult to do when dealing with spatial data because of the dependence that exists in the model. In this model, each dependent variable is dependent on its own error term but it is also dependent on the observation of all of the other actors' y and each of these ys are dependent on their u , leading each individual y to be dependent on all other ys and us . This issue can be dealt with in one of two ways; it can either be derived mathematically which would be computationally burdensome or through the use of simulation.

In this paper, Wei and Tanner's (1991) imputation method for censored regression data is adapted to account for the spatial nature of the data that this paper is focused on. Wei and Tanner's (1991) original method focuses on two different algorithms for imputing censored regression data: the Poor Man's Data Augmentation algorithm and the Asymptotic Data Augmentation algorithm. For these algorithms the authors make no assumptions regarding the distribution of the error and each of the observations are independent of one another.

The Asymptotic algorithm is similar to the Poor Man's algorithm in that they both rely on the ability to draw from conditional distribution of the error term, subject to the fact that the observed failure time is larger than the time that was actually observed.

There is one large difference between the two algorithms; the imputations derived from the Poor Man's algorithm are dependent on the current estimates rather than specifying random draws of the estimates like the Asymptotic algorithm. This imputation method in its simplest form argues that the algorithm should first estimate the model using all of the cases that are not censored and for which all of the information is known. Using this model, the reduced form residuals are calculated in order to create a distribution. This distribution is then drawn from in order to calculate an imputed value for the duration of the censored cases that would provide them with a duration that is greater than the censoring point. These imputed values are then used in reestimating the model using all of the cases (including those that had previously been censored). This process is then repeated over multiple draws of the error term until the estimates converge.

We adapt the Poor Man's algorithm approach to suit the spatial data that this method will be applied to. It has to be adapted because of the spatial interdependence that exists between observations. It no longer can be assumed that the disturbances are independent of one another as is done in OLS. It is then not possible to separate out the censored and uncensored cases and the first step of the algorithm must be run on the entire model rather than be just a model that is run with the uncensored duration processes. By dropping the censored observations, bias would be introduced into the estimates and could prevent the estimates from approaching their true value. The algorithm is adapted for both log-normal and Weibull duration processes.

Log-Normal Durations

The algorithm adapted for log-normal duration processes starts by estimating a spatial lag model that does not impute the censored cases in order to get values from which to start from as well as the estimates that will be compared against the imputed analysis. Before the start of the imputation process, there are 5 random draws generated from which the

censored errors will come from. The rest of the program is run 5 times using each of these different draws.

To start the imputation, the predicted value of Y , \hat{Y} , is calculated from the previous estimates and then is subtracted from the true Y in order to get predicted errors, \hat{u} . The predicted errors are then used to get the variance-covariance matrix on which a Cholesky decomposition is performed. This means that the variance-covariance matrix, Σ has a Cholesky decomposition such that $\Sigma^{-1} = \mathbf{A}'\mathbf{A}$. \mathbf{A} is an upper triangular matrix such that $\boldsymbol{\eta} = \mathbf{A}\mathbf{u}$ where $\boldsymbol{\eta}$ s are independent random draws from a distribution that is a linear combination of normal random variables. One can then use these i.i.d. $\boldsymbol{\eta}$ s and solve for the spatial errors. If one solves for \mathbf{u} , the result is $\mathbf{u} = \mathbf{A}^{-1}\boldsymbol{\eta}$.

In order to do this, we generate the i.i.d $\boldsymbol{\eta}$ s. This is done by going through each of the observations (starting with the final duration) and calculate the estimated value of $\boldsymbol{\eta}$ or the censoring point. The calculation of the estimated $\boldsymbol{\eta}$ s for the uncensored cases is relatively simple since \mathbf{A}^{-1} is upper triangular. For the final observation, it is just its estimated error, \hat{u}_i , divided by their element in \mathbf{A}^{-1} , or a_{ii} . As it moves up the matrix, it becomes more complicated but only in the respect that each individual's \hat{u} is now subtracted from the linear combination of all the $\mathbf{A}^{-1}\boldsymbol{\eta}$ and is then multiplied by their individual element in \mathbf{A}^{-1} . Since the errors for the censored values are never observed, $\boldsymbol{\eta}$ cannot be solved for directly. It is possible to figure out what value is necessary for the duration to exceed the censoring point given the calculated errors for the previous observations and thus draw from the distribution conditional censored below at this value. The value of censoring is calculated in the same way that we solve for the uncensored $\boldsymbol{\eta}$ s. Once the censoring point is determined for each individual, we take a random draw from the censored standard normal distribution to obtain a value that is greater than the censored value.

In order to create the imputed reduced form spatial errors, the $\boldsymbol{\eta}$ are multiplied by \mathbf{A}^{-1} . Using these imputed values for the errors, an imputed Y is calculated from the

combination of \hat{Y} from the previous spatial lag model and the imputed spatial errors calculated. This method returns the same value for each of the uncensored cases and imputed outcomes for the censored cases. The spatial lag model is then run with these imputed Y s and the estimates from this are saved. The estimates from each of the multiply imputed data sets are combined using Rubin's (2009) formula:

$$\bar{\theta} = \frac{1}{M} \sum_{m=1}^M \hat{\theta}_m, \quad (17)$$

$$Var(\bar{\theta}) = \frac{1}{M} \sum_{m=1}^M Var(\hat{\theta}_m) + \frac{M+1}{M} \left(\frac{1}{M-1} \sum_{m=1}^M (\hat{\theta}_m - \bar{\theta})^2 \right). \quad (18)$$

. The process continues until the difference between the current model's and the previous model's log-likelihood function are different by a very small value (the log-likelihoods are also averaged across the 5 draws).

Weibull Durations

To develop a spatial Weibull duration model that accounts for right censoring requires a slight modification to the one just derived for the lognormal estimator. In the lognormal case we can impute values for right censored cases analytically since a linear combination of normal variables has a normal distribution. This allows us to use the Cholesky decomposition approach to write the reduced form errors as a linear combination of independent and identically distributed standard normal variates and to solve for their values or censoring points one at a time. For censored cases we just take a draw from the standard normal distribution censored from below at the appropriate value. Once we switch to a Weibull distribution, however, we can not move so easily between the reduced form and structural error distributions; we do, however, rely on this logic to develop an alternate numerical approach.

Again, we can write the data generating process as

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\beta + (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{L}\mathbf{u}, \quad (19)$$

$$= \mathbf{\Gamma}\mathbf{X}\beta + \mathbf{\Gamma}\mathbf{L}\mathbf{u}, \quad (20)$$

$$= \mathbf{\Gamma}\mathbf{X}\beta + \mathbf{v}, \quad (21)$$

where \mathbf{u} has a standard Gumbel distribution and \mathbf{L} represents its corresponding shape parameter. The reduced form error \mathbf{v} represents a linear combination of independent Gumbel variables, but this does not produce a Gumbel distribution. Nadarajah (2008) derives the density and cumulative distribution functions for such a linear combination, but the resulting formula proved too complicated to be of practical use. We therefore take a numerical approach.

As before, we proceed by estimating the model to generate an estimate of the parameters. We then calculate the Cholesky decomposition, \mathbf{A}^{-1} , of the expected covariance matrix of these errors:

$$E[(\mathbf{\Gamma}\mathbf{L}\hat{\mathbf{u}})(\mathbf{\Gamma}\mathbf{L}\hat{\mathbf{u}})'] = (\mathbf{\Gamma}\mathbf{L})E[\hat{\mathbf{u}}\hat{\mathbf{u}}'](\mathbf{\Gamma}\mathbf{L})', \quad (22)$$

$$= (\mathbf{\Gamma}\mathbf{L})(\sigma_u^2\mathbf{I})(\mathbf{\Gamma}\mathbf{L})', \quad (23)$$

$$= \sigma_u^2(\mathbf{\Gamma}\mathbf{L})(\mathbf{\Gamma}\mathbf{L})'. \quad (24)$$

We can then write the reduced form errors as a linear combination of standard Gumbel errors: $\hat{\mathbf{u}} = \mathbf{A}^{-1}\eta$. Since \mathbf{A}^{-1} is upper triangular, we can iteratively solve for the corresponding value of η one observation at a time starting with the last observation. For censored cases this produces a value corresponding to the smallest value that will produce an imputed outcome at least as large as the observed censored value. Given the complexity of the distribution for a linear combination of Gumbel variates, we use a numerical approx-

imation technique, which we explain below, to sample from the appropriate distribution. Once we have calculated or imputed values of η for all of the observations we multiply by \mathbf{A}^{-1} to generate draws of the reduced form errors, which we add to the predicted mean for each observation, $\widehat{\Gamma}\mathbf{X}\widehat{\beta}$, to produce a vector of the dependent variable that consists of the observed values for uncensored cases and imputed values for the censored cases.

Our numerical sampling technique works as follows. For each observation we know its reduced form error represents a linear combination of independent and identically distributed Gumbel variables based on the i th row of the matrix $\Gamma\mathbf{L}$. We can simulate this distribution by taking J draws from the estimated distribution $\widehat{\Gamma}\widehat{\mathbf{L}}\nu$, where ν is an $N \times J$ matrix of independent and identically distributed Gumbel variables. This gives us J independent draws from the distribution of the reduced form error. Each of these draws corresponds to an associated value of η via the matrix \mathbf{A}^{-1} which means that if we randomly sample a value from the simulated distribution we can obtain the associated random draw of η . To account for censoring we sample only from values of the reduced form error greater than the censoring point. To generate multiple imputed values of Y we repeat this process M times.

After each step of the EM process, we take the average of the M parameter vectors as our current estimate to begin the next step. The algorithm continues until the results converge, which we assess by whether the log-likelihood changes by less than a small amount, e.g., 0.000001. Once it converges, we calculate the final estimate as the average of the M estimates from the last step and calculate the covariance matrix according to Rubin's (2009) formula as in the log-normal case.

The key to this approach for the spatial Weibull duration estimator lies in the generalization of the Cholesky decomposition approach for non-normal variables. While we can not analytically work through the decomposition to generate random draws as one might for jointly normally distributed variables, we can use the Cholesky decomposition

to generate a system of linear equations that allow us to produce a set of calculated and imputed values for observed and censored cases, respectively, that have the appropriate joint distribution and which we can then use to reproduce the vector of outcomes, Y , with the same values for observed cases and imputed values for censored cases. To our knowledge, this technique has not yet been used so this may represent an additional contribution of this paper.

We repeat this process iteratively as we work through the EM algorithm process of producing a vector of uncensored outcomes, Y , and estimating the model on the resulting vector. This approach can be quite time intensive. Since we have right censoring we need many observations on the upper tail of the simulated distribution of the reduced form errors. This involves hundreds of thousands and ideally millions of draws, so J is quite large. Further, the process must be repeated for each censored observation and each iteration since the matrix, $\hat{\Gamma}\hat{L}$, from which the linear combination emerges changes each iteration. And since we multiply impute values of Y for each iteration we must repeat everything M times.

Monte Carlo

In order to evaluate our approach we conduct a series of Monte Carlo simulations. These allow us to study its properties, both statistically and computationally, and to compare the estimates that it produces to those that one would obtain from ignoring the censoring. We also use the estimates that one would obtain from the original uncensored data, were it available, as a benchmark comparison. We evaluate our EM estimator's performance across a wide range of circumstances by varying both the amount of spatial interdependence and the amount of censoring.

Our data generating process proceeds as follows. We start with one hundred units

spread out evenly across a ten by ten grid. We construct a spatial dependence matrix, \mathbf{W} , based on rook and queen contiguity. We then generate 100 independently and identically distributed observations of a single independent variable, \mathbf{X} , according to standard normal distributions. For simplicity's sake we set $\mathbf{L} = \mathbf{I}$, resulting in the following data generating process:

$$\mathbf{Y} = (\mathbf{I} - \rho\mathbf{W})^{-1}(-1 - 1 \times \mathbf{X}_1) + (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{u}. \quad (25)$$

With 100 observations our approach to generating \mathbf{W} results in about around 7% connectivity between units. As is common, we row-standardized the spatial weights matrix by dividing each element by the sum of the elements in its row. This creates a matrix where all of the elements are proportions and each row sums to 1. This ensures that the spatial parameters are comparable across different models (Anselin and Bera 1998). Since most spatially dependent relationships in political science have positive spatial autocorrelation we run simulations with ρ equal to 0, 0.25, and 0.5. We hold the independent variables constant across all of the simulations.

We introduce censoring through a common censoring point for all observations. This mimics what occurs when some observations have not failed by the end of the study, e.g., such as when all units have not adopted a policy in an event history analysis, or when the event of interest becomes infeasible at certain point in time, as in our application to the timing of countries' entry in World War I. Based on the distribution of the dependent variable that results from our data generating process, we selected censoring points that vary from -2 to 2 by an increment of 0.5. The amount of censoring that occurs ranges from none of the data being censored to over almost 80% of the data being censored (See Figure 1).

For each combination of values of spatial interdependence and censoring points we

generate 500 draws of Y and estimate all three models on each draw. We set the maximum number of iterations of our EM estimator for each draw to 200 and used five imputations for each step of the EM process. Standard errors are calculated according to Equation 18. The estimator that we employ depends on the distribution of u : when it follows a normal distribution we use the log-normal estimator and when it follows a Gumbel distribution we use the Weibull estimator. We (eventually) evaluate both since the Weibull imputation process differs significantly from that used for the log-normal version.

Log-Normal Results

For this set of Monte Carlos we assume that u follows a normal distribution, which corresponds to a log normal duration process. We generate draws of u from a standard normal distribution, calculate the value of Y , then apply our censoring rule so that $Y_i^c = \min\{Y, C\}$, where C ranges from -2 to 2 by increments of 0.5. We estimate three models: a naive model that treats all realizations of Y_i^c as uncensored, our EM spatial duration model with imputation of censored values, and a log-normal spatial duration model using the uncensored value Y_i . The latter serves as a best-case scenario against which to compare our EM approach since spatial models often show some degree of bias in parameter estimates with relatively small samples.

[Figure 2 here.]

Figures 2 and 3 present the average parameter estimates for the slope coefficient and the intercept, which have true values of -1 and the spatial dependence parameter, which varies from 0 to 0.25 to 0.5. Three patterns emerge quite clearly. First, the benchmark estimates evidence a potential, though slight, negative bias, especially for the intercept and spatial dependence parameters. This is consistent with other simulations of spatial estimators and should be kept in mind in evaluation the performance of the other estimators since it represents the best case scenario in which no censoring occurs.

Second, the naïve spatial model produce estimates far from the correct values for both the slope coefficient and the intercept, though not for the spatial dependence parameter. The apparent bias appears quite bad even with relatively small amounts of censoring and appears to get worse as spatial dependence decreases. For example, censoring rates generally average below 20% once $C \geq 0$ and the average estimate of the slope coefficient for the naïve model deviates by up to 30% from the true value of -1 . Once censoring drops below 5% the apparent bias appears to disappear.

[Figure 3 here.]

Third, the EM estimator produces average estimates for all three parameters near their true values, except for a slight deviation in the slope parameter when $\rho = 0$ and $C \leq -0.5$ which gets worse when $C = -2$ and in the intercept when $C = -2$. The difficulty estimating these parameters when $\rho = 0$ make sense since in those cases one has much less information about the outcomes for the censored cases since their values do not spatially influence the outcomes in uncensored cases.

The average parameter estimates provide solid evidence that our EM spatial estimator produces results much closer to the true values than ignoring the problem. Given the greater complexity of the model, including the multiple imputation part, we also need to evaluate the bias-uncertainty tradeoff. We did this in two steps. First, we evaluate our approach's ability to recover correct measures of uncertainty, then we combine these with bias to perform a mean squared error comparison.

[Figure 4 here.]

Figure 4 provides a comparison of the average standard errors across the 500 draws and the standard deviation of the sampling distribution of the parameter estimates. While both estimators appear to underestimate the standard errors with high censoring rates, the EM estimator does so to a much greater extent, with average standard errors up to 50% too small in the most extreme case. This issue requires further attention, perhaps by

increasing the number of multiply imputed data sets. Note also that the EM approach results in larger average standard deviations with even modest censoring rates (that they match when censoring reaches almost zero is not surprising since the two models are become identical as censoring goes to zero), up to three times as big. While we do not show the results here, a similar but less pronounced underestimation occurs for the intercept while no such problems occur for the spatial dependence parameter. This offers a clear motivation for a root mean squared error comparison.

[Figure 5 here.]

We make such a comparison in Figure 5. This plots the square root of the sum of the variance and the squared bias of each parameter for all three estimators. The plots show clear evidence that even with its greater level of uncertainty our EM estimator clearly outperforms the naïve one whenever $C \leq 0$, which corresponds to censoring rates of greater than about 20%. The slope coefficient and intercept evidence the greatest improvement while the results appear somewhat unclear for the spatial dependence parameter, with the EM approach doing better when $\rho = 0$ and the naïve model better when $\rho = 0.5$.

Overall, then, we take these results as providing solid evidence in favor of our EM spatial duration model for censored data. With even modest levels of censoring it appears to outperform a naïve duration model that ignores the censoring in root mean square error terms. The results also indicate that it provides estimates that do deviate much from the benchmark model with fully observed data. With extremely high rates of censoring, some apparent bias does emerge, but given the size of the standard deviations it may not be meaningful. These results also point to a need to further investigate the generally modest underestimation of the standard errors by our EM estimator.

Weibull Results

Authors' note: This model takes forever to estimate so we do not yet have Monte Carlo results to

present.

Illustration: The Diffusion of WWI

There is a large literature in international relations on the diffusion or contagion of war (Davis, Duncan and Siverson 1978; Most and Starr 1980; Levy 1982; Siverson and Starr 1991; Gartner and Siverson 1996; Werner and Lemke 1997; Kedera 1998; Melin and Koch 2010). According to one line of thought, the spread of war is analogous the spread of infectious disease. War is theorized to diffuse through geographical proximity, rivalries, and military alliances among other mechanisms. Duration models provide a natural framework for empirically evaluating theories of conflict diffusion. Given a particular level of conflict “exposure,” the question is how long will it take before a country succumbs to the scourge of war. Right-censoring presents a significant methodological challenge to duration analyses of war diffusion, however. Wars end before all the potential joiners have entered the conflict. An armistice is like a vaccine or the end of a clinical trial.

One recent and sophisticated attempt to model the contagion of conflict that addresses the problem of right-censoring is Melin and Koch. They take a two stage approach in their analysis. In the first-stage, the expansion stage, they model the time until an initial dyadic militarized dispute expands. This is important because many conflicts remain dyadic. In the second stage, the joining stage, they model the time until a potential joiner enters the conflict, taking into account the expected duration of the initial dispute from their first-stage estimates. They find, *inter alia*, that capabilities, contiguity, and alliances reduce the time it takes for potential joiners to enter a conflict.

While this represents best practice in the empirical study of conflict diffusion, there are a couple of significant limitations. First, this approach only accounts for the connections between a potential joiner and the initial target and conflict instigator. Taking WWI as an

example, the time it takes Britain to join the conflict between Austria-Hungary and Serbia would depend only on its relationship to Austria-Hungary and Serbia. However, clearly, Britain's decision to join the conflict when it did depended on Germany's decision to enter the conflict three days earlier. In short, this approach (implicit triads) does not account for the entire structure of interdependence that exists between all potential joiners. Second, it does not address the simultaneity between the expansion and joining outcomes. The Melin and Koch approach would make Germany's decision to join the conflict between Austria-Hungary and Serbia a function of the expected time to expansion, but the time to expansion also depends on Germany's decision to intervene.

In this section, we model the WWI entry timing decisions of states using a spatial lag model of interdependent durations. Our approach addresses the shortcomings of the two-stage model in Melin and Koch. All of the entry timing decisions are potentially linked through the spatial weights matrix and treated as simultaneously determined. WWI is an excellent case for studying the diffusion of war. It began as a localized conflict that over the course of four years expanded to include a majority of the independent states in the international system (Flint et al. 2009; Radil, Flint and Chi 2013). Had the war continued, undoubtedly, more states would have been drawn into in the conflict. Radil et al. identify four waves of war expansion. The initial stage began with Austria-Hungary's declaration of war on Serbia, with entries following soon after by Germany, Russia, Belgium, France, and the United Kingdom. Japan joined the war approximately a month later, and Turkey followed suit within one-hundred days of the outbreak. The middle stage (May 23, 1915 - August 27, 1916) included entries by Italy, Bulgaria, Portugal, and Romania. The late joiners, in order, were the United States, Cuba, Panama, Bolivia, Greece, Thailand, China, Peru, Uruguay, Brazil, Ecuador, Guatemala, Nicaragua, and Honduras.

One could treat the entry timing decisions of these states as independent and driven purely by domestic and international structural factors such as regime type, trade ex-

posure, and relative military capabilities, but this is approach unsatisfactory. Ultimately, each state's decision about when to enter the war was heavily influence by the entry timing decisions of others, and any empirical analysis should take this interdependence into account. We incorporate three forms of interdependence into our models: contiguity, rivalry, and targeted alliances. These sources of interdependence suggest that a state will be influenced by the participation and entry timing decisions of its neighbors, rivals, and the targets of its military alliances.

Our dependent variable is the number of months before entering WWI. Of the 44 sample countries, 27 eventually enter the War, which gives a censoring rate of 39%. All the spatial weights matrices are row-standardized. National capabilities are the COW CINC index scores (Singer, Bremer and Stuckey 1972); democracy is Polity measure of regime type (Marshall and Jaggers 2002); and trade is the value of total trade in current US dollars (Barbieri 2002). We estimate three types of models: a non-spatial model, a naive spatial model that treats the time to censoring as an observed failure time, and our multiple imputation model. Our main concern is that variables such as national capabilities that have been found to determine war joining behavior cluster among states that are linked by the mechanisms or "vectors" through which conflict diffuses. If this is the case, and we do not fully account for the structure of interdependence, we are likely to draw faulty inferences causes of conflict contagion.

We report the estimates for our log-normal and Weibull duration models in Tables 1 and 2 respectively. Starting with the non-spatial models, we find that national capabilities reduce the time until entry in both the log-normal and Weibull specifications, but neither trade nor democracy determine the entry timing decisions of states. With respect to interdependence, our results are mixed at best. There is no evidence that war diffuses through contiguity or rivalry. We find limited support for the idea that targeted alliances are conduits for the spread of war (Table 2). If a state has a military alliance that targets another

state or is itself the target of a military alliance, its entry timing decisions are positively correlated with the entry timing decisions of its adversaries.³ However, this result is not robust, as it depends on distributional assumptions about the hazard function.

[Tables 1 and 2 here.]

The initial diffusion of WWI was contained mostly to great powers, which could mean that capabilities played an important role in the war joining decisions of states. But the great powers were also clustered geographically and connected by formal alliances and informal rivalries, and these linkages represent channels through which war has been theorized to spread. Ideally, our spatial lag models would help discriminate between these two explanations for the pattern of diffusion that we observe. Unfortunately, our results depend on distributional assumptions. In our log-normal models, we find strong evidence that national capabilities determine the entry timing decision of states and no evidence of interdependence. In our Weibull models, we find some evidence of interdependence and very little evidence that capabilities matter. If anything, our results suggest that alliances are important for transmitting the spread of war, but this is preliminary. The fact that we find such limited evidence of interdependence, given what we know historically about WWI, suggest that we need to think harder about our spatial weights matrices.

Overall, it is clear from our analysis that one has to be careful modeling the structure of interdependence across potential war joiners. Approaches that address interdependence through dyadic or triadic units of analysis scale well (Franzese, Hays and Kachi 2012), but they do not fully capture the structures of connectivity through which wars diffuse. Spatial lag models, on the other hand, are only as good as the weights matrices upon

³Interestingly, the naive spatial model with the targeted alliance spatial lag produces estimates that imply a non-stationary process. This is likely due to the artificial clustering created by treating the time to censoring as an observed failure time. If so, this problem is solved by the multiple imputation approach, and, indeed, the MI estimate of ρ is inside the upper bound for stationarity.

which they are built, and importantly, in models that allow for censoring, the structures of interdependence implied by these weights matrices seem to interact strongly with distributional assumptions.

Conclusion

Interdependent duration processes are common in politics and other strategic settings. The time to an event for one actor often depends on the time to that same event for others. For example, the time it takes states to enter wars, alliances, and international organizations depends on the time it takes other states to make these decisions. The entry and exit decisions of political candidates in electoral contests depend on the timing of their opponents. If policies diffuse across countries, the time it takes one country to adopt a particular policy depends on the adoption timing of other states. Simply put, politics and strategic behavior generate duration interdependence across actors.

One challenge for studying interdependent durations in politics is that right censoring prevents us from fully observe the consequences of this interdependence. Either the opportunity to take a particular action ends before such decisions are made, as with our war joining illustration, or our studies end before the events of interest are observed. Unfortunately, methods for analyzing interdependent duration processes are underdeveloped, particularly when there is right-censoring. We have adapted Wei and Tanner's imputation algorithm for censored (nonspatial) data to models of spatially interdependent durations and shown via Monte Carlo that this approach performs reasonably well and is preferable to simply ignoring the censoring problem.

Much work remains to be done. First, our illustration suggests that imputation-based estimation of spatial lag duration models with right-censored data may be sensitive to distributional assumptions. Ideally, we would like to offer a semi-parametric approach

along the lines of Wei and Tanner. Their approach, sampling from a Kaplan-Meier estimate of the distribution of residuals, hinges critically on the assumption that these residuals are independent and identically distributed. We have not yet determined a feasible way to sample from a distribution of spatially interdependent disturbances without making parametric assumptions. Second, we know that it is a bad idea to ignore the censoring problem with spatial lag models, but perhaps ignoring the interdependence problem is less problematic. We need to compare the performance of our model and estimator against non-spatial models that address right-censoring in more traditional ways, but fail to account for interdependence. Finally, we need to determine why our standard error estimates are highly overconfident when the degree of censoring is high.

References

- Anselin, Luc and Anil K Bera. 1998. "Spatial dependence in linear regression models with an introduction to spatial econometrics." *STATISTICS TEXTBOOKS AND MONOGRAPHS* 155:237–290.
- Barbieri, Katherine. 2002. *The liberal illusion: Does trade promote peace?* University of Michigan Press.
- Box-Steffensmeier, Janet M. and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*. Cambridge University Press.
- Darmofal, David. 2006. "The Political Geography of Macro-Level Turnout in American Political Development." *Political Geography* 25(2):123–150.
- Darmofal, David. 2009. "Bayesian Spatial Survival Models for Political Event Processes." *American Journal of Political Science* 53(1):241–257.
- Davis, William W, George T Duncan and Randolph M Siverson. 1978. "The dynamics of warfare: 1816-1965." *American Journal of Political Science* pp. 772–792.
- Flint, Colin, Paul Diehl, Jürgen Scheffran, John Vasquez and Sang-hyun Chi. 2009. "Conceptualizing conflict space: Toward a geography of relational power and embeddedness in the analysis of interstate conflict." *Annals of the Association of American Geographers* 99(5):827–835.
- Franzese, Robert J., Jr. 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):140–164.
- Franzese, Robert J, Jude C Hays and Aya Kachi. 2012. "Modeling History Dependence in Network-Behavior Coevolution." *Political Analysis* 20(2):175–190.

- Gartner, Scott Sigmund and Randolph M Siverson. 1996. "War expansion and war outcome." *Journal of Conflict Resolution* 40(1):4–15.
- Gimpel, James G. and Wendy K. Tam Cho. 2004. "The Persistence of White Ethnicity in New England Politics." *Political Geography* 23:987–1008.
- Hays, Jude C. and Aya Kachi. 2009. "Interdependent Duration Models in Political Science." *working* .
- Kedera, Kelly M. 1998. "Transmission, Barriers, and Constraints A Dynamic Model of the Spread of War." *Journal of Conflict Resolution* 42(3):367–387.
- Levy, Jack S. 1982. "The contagion of great power war behavior, 1495-1975." *American Journal of Political Science* pp. 562–584.
- Marshall, Monty G and Keith Jaggers. 2002. "Polity IV project: Political regime characteristics and transitions, 1800-2002.".
- Melin, Molly M and Michael T Koch. 2010. "Jumping into the Fray: Alliances, Power, Institutions, and the Timing of Conflict Expansion." *International Interactions* 36(1):1–27.
- Most, Benjamin A and Harvey Starr. 1980. "Diffusion, reinforcement, geopolitics, and the spread of war." *The American Political Science Review* pp. 932–946.
- Nadarajah, Saralees. 2008. "Exact distribution of the linear combination of p Gumbel random variables." *International Journal of Computer Mathematics* 85(9):1355–1362.
- Radil, Steven M, Colin Flint and Sang-Hyun Chi. 2013. "A relational geography of war: Actor-context interaction and the spread of World War I.".
- Rubin, Donald B. 2009. *Multiple imputation for nonresponse in surveys*. Vol. 307 Wiley. com.

- Singer, J David, Stuart Bremer and John Stuckey. 1972. "Capability distribution, uncertainty, and major power war, 1820-1965." *Peace, war, and numbers* pp. 19–48.
- Siverson, Randolph M and Harvey Starr. 1991. *The diffusion of war: A study of opportunity and willingness*. University of Michigan Press.
- Wei, Greg C. G. and Martin A. Tanner. 1991. "Applications of Multiple Imputation to the Analysis of Censored Regression Data." *Biometrics* 47(4):1297–1309.
- Werner, Suzanne and Douglas Lemke. 1997. "Opposites do not attract: The impact of domestic institutions, power, and prior commitments on alignment choices." *International Studies Quarterly* 41(3):529–546.

Table 1: Comparison of Log-Normal Duration Models of the Timing of Entry into World War I

<i>Spatial Lag</i>	None	Contiguity		Alliance		Rivalry	
		Naive	Imp.	Naive	Imp.	Naive	Imp.
constant	4.844*** (1.131)	3.448*** (0.825)	4.004** (1.242)	3.985*** (0.733)	4.630** (1.072)	3.967*** (0.756)	4.612** (1.090)
capabilities	-27.597** (11.290)	-17.798** (7.375)	-26.769* (10.636)	-15.390** (7.742)	-24.911 (11.333)	-18.678** (7.560)	-27.735* (11.105)
democracy	0.012 (0.074)	-0.005 (0.047)	0.013 (0.070)	-0.009 (0.048)	0.015 (0.072)	0.002 (0.048)	0.026 (0.072)
trade	-0.157 (0.240)	-0.148 (0.159)	-0.082 (0.245)	-0.198 (0.155)	-0.141 (0.237)	-0.183 (0.162)	-0.123 (0.243)
σ^2	2.670 (0.144)	3.405*** (0.825)	6.093** (1.378)	3.459*** (0.741)	6.246** (1.435)	3.567*** (0.762)	6.406** (1.480)
ρ		0.184 (0.120)	0.162 (0.142)	0.647 (0.484)	0.457 (0.614)	0.214 (0.327)	0.161 (0.370)
Likelihood	-76.40	-89.61	-102.3	-90.02	-102.8	-90.52	-103.3

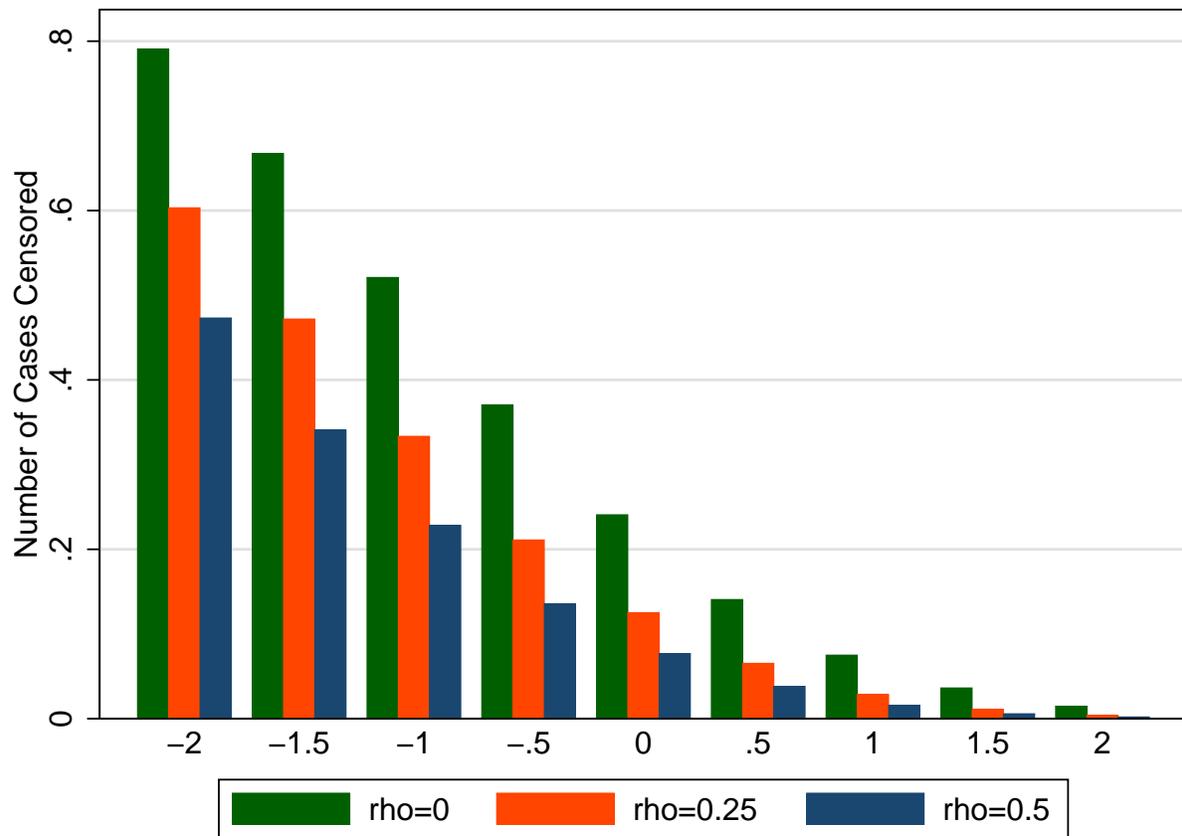
Notes. N=44. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 2: Comparison of Weibull Duration Models of the Timing of Entry into World War I

<i>Spatial Lag</i>	None	Contiguity		Alliance		Rivalry	
		Naive	Imp.	Naive	Imp.	Naive	Imp.
constant	4.963*** (1.039)	3.678*** (0.516)	4.066*** (1.133)	3.991*** (0.462)	4.39*** (0.727)	3.965*** (0.505)	4.417*** (0.716)
capabilities	-16.175*** (7.488)	-6.587* (3.783)	-9.935 (6.588)	-5.523 (3.499)	-5.442 (4.916)	-6.407* (3.777)	-4.378 (5.223)
democracy	0.041 (0.06)	0.005 (0.028)	0.009 (0.054)	0.011 (0.026)	0.021 (0.042)	0.014 (0.028)	0.012 (0.051)
trade	-0.086 (0.216)	-0.058 (0.099)	-0.111 (0.21)	-0.088 (0.097)	-0.184 (0.139)	-0.077 (0.107)	-0.206 (0.145)
shape	0.551*** (0.096)	0.897*** (0.125)	0.63*** (0.09)	0.951*** (0.134)	0.737*** (0.111)	0.900*** (0.127)	0.745*** (0.114)
ρ		0.094 (0.074)	0.134 (0.111)	1.024*** (0.282)	0.931*** (0.367)	0.272 (0.272)	0.123 (0.368)
Likelihood	-75.882	-80.116	-90.689	-78.729	-85.818	-80.48	-85.366

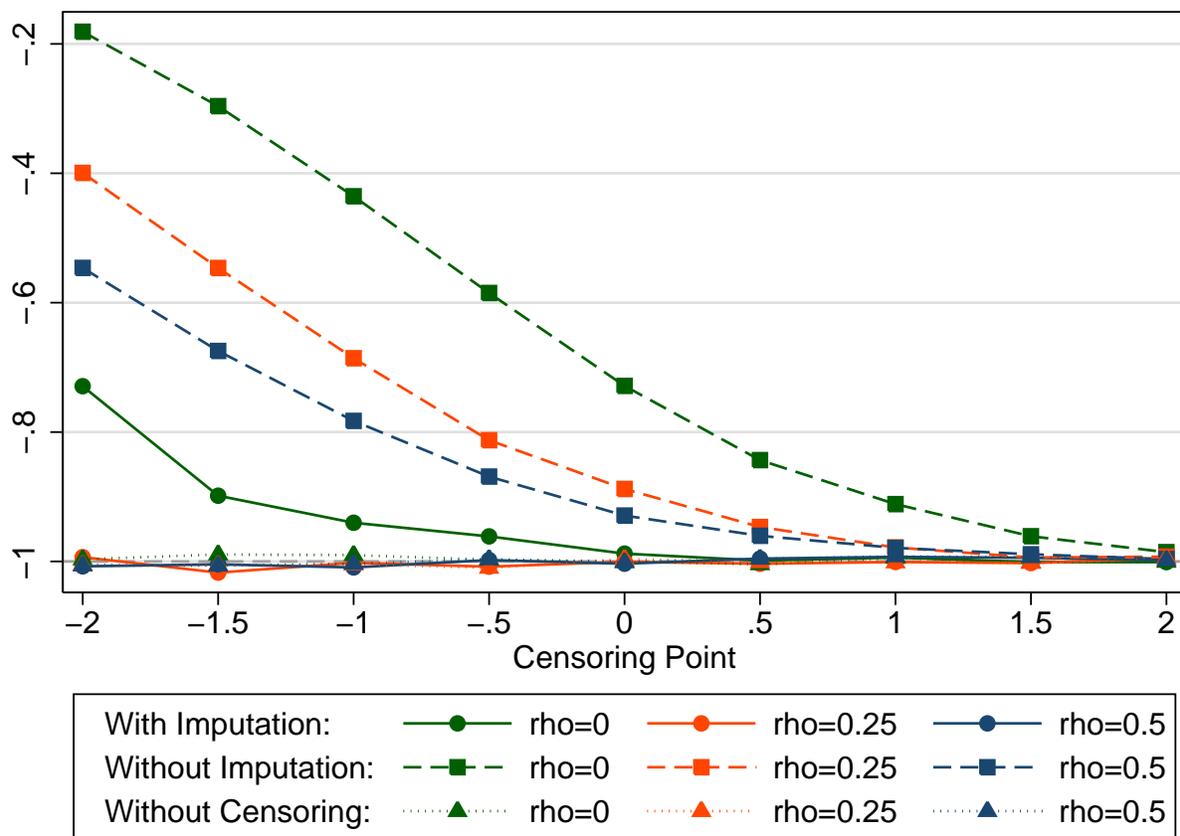
Notes. N=44. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Figure 1: Average Proportion of Censored Cases by Spatial Correlation and Censoring Point



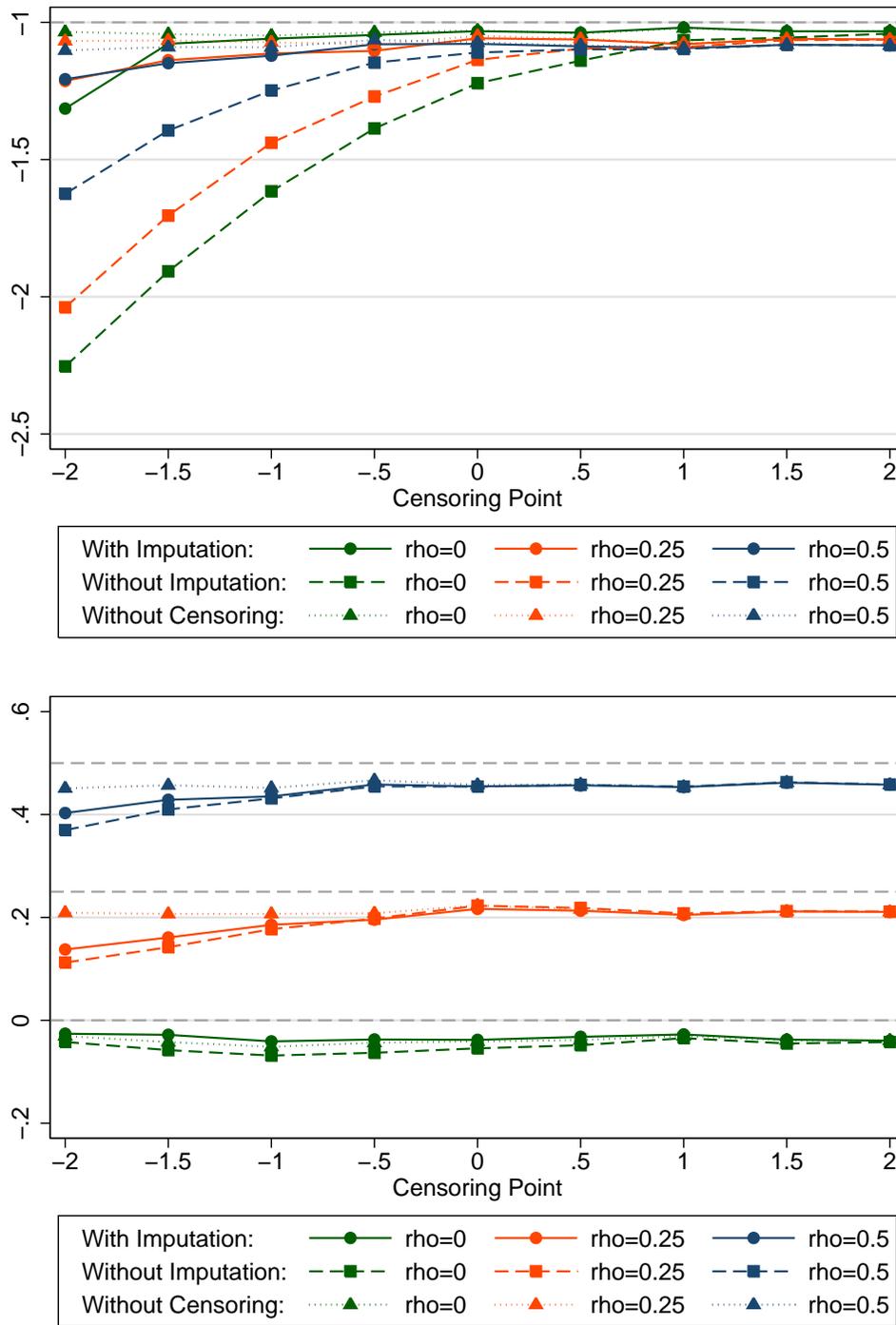
Notes: Results represent the average across 500 simulations, excluding cases for which the EM algorithm did not converge within 100 iterations.

Figure 2: Average Coefficient Estimates from EM Approach, Naïve Duration Model, and Uncensored Benchmark Spatial Duration Model



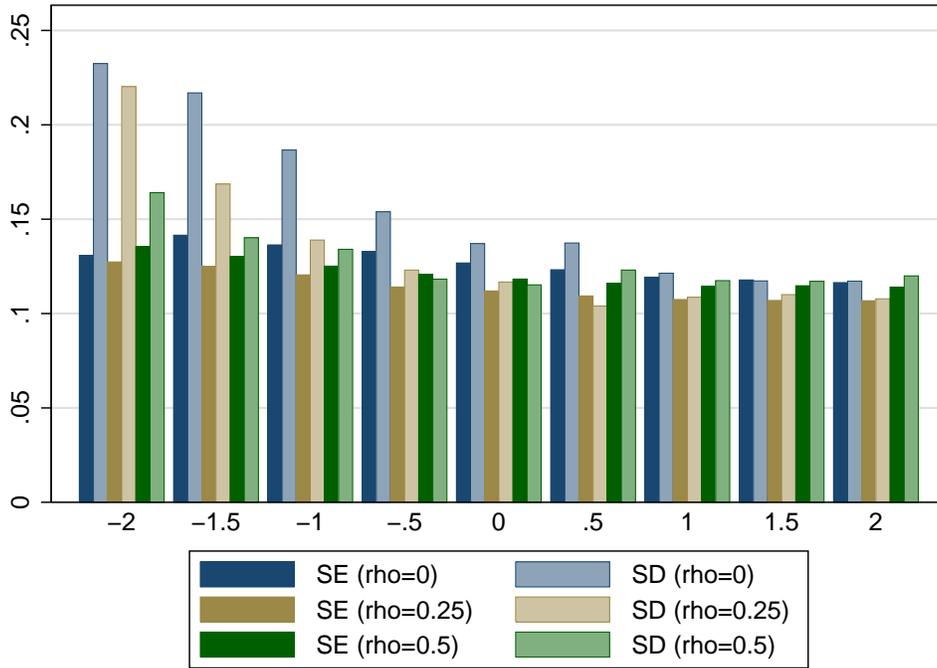
Notes: Results represent the average across 500 simulations, excluding cases for which the EM algorithm did not converge within 100 iterations. Coefficients estimated in time to failure format.

Figure 3: Average Intercept and Spatial Parameter Estimates from EM Approach, Naïve Duration Model, and Uncensored Benchmark Spatial Duration Model

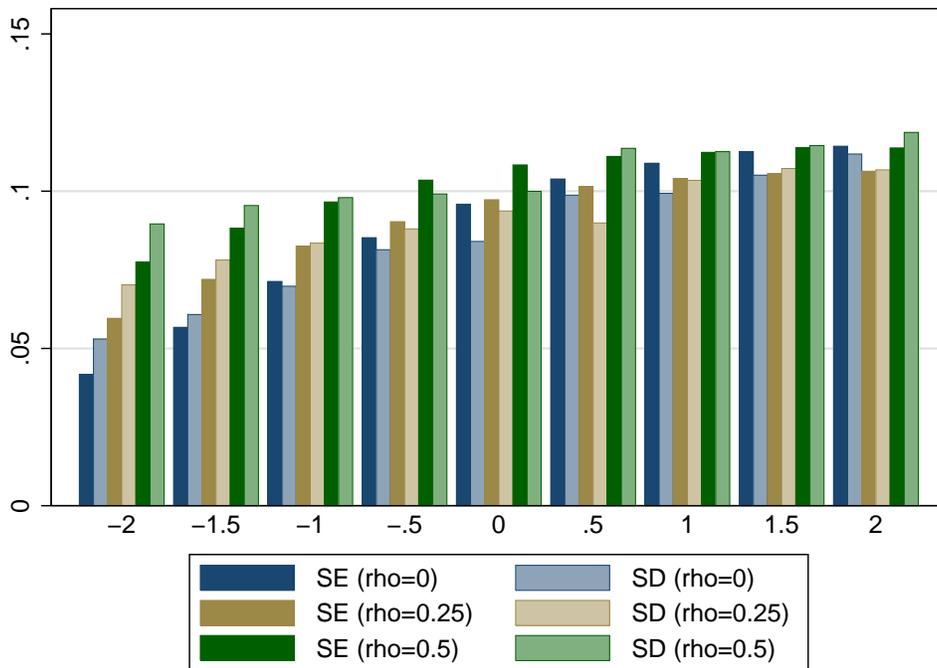


Notes: Results represent the average across 500 simulations, excluding cases for which the EM algorithm did not converge within 100 iterations. Coefficients estimated in time to failure format.

Figure 4: Comparison of Standard Deviations and Standard Errors
EM Estimator

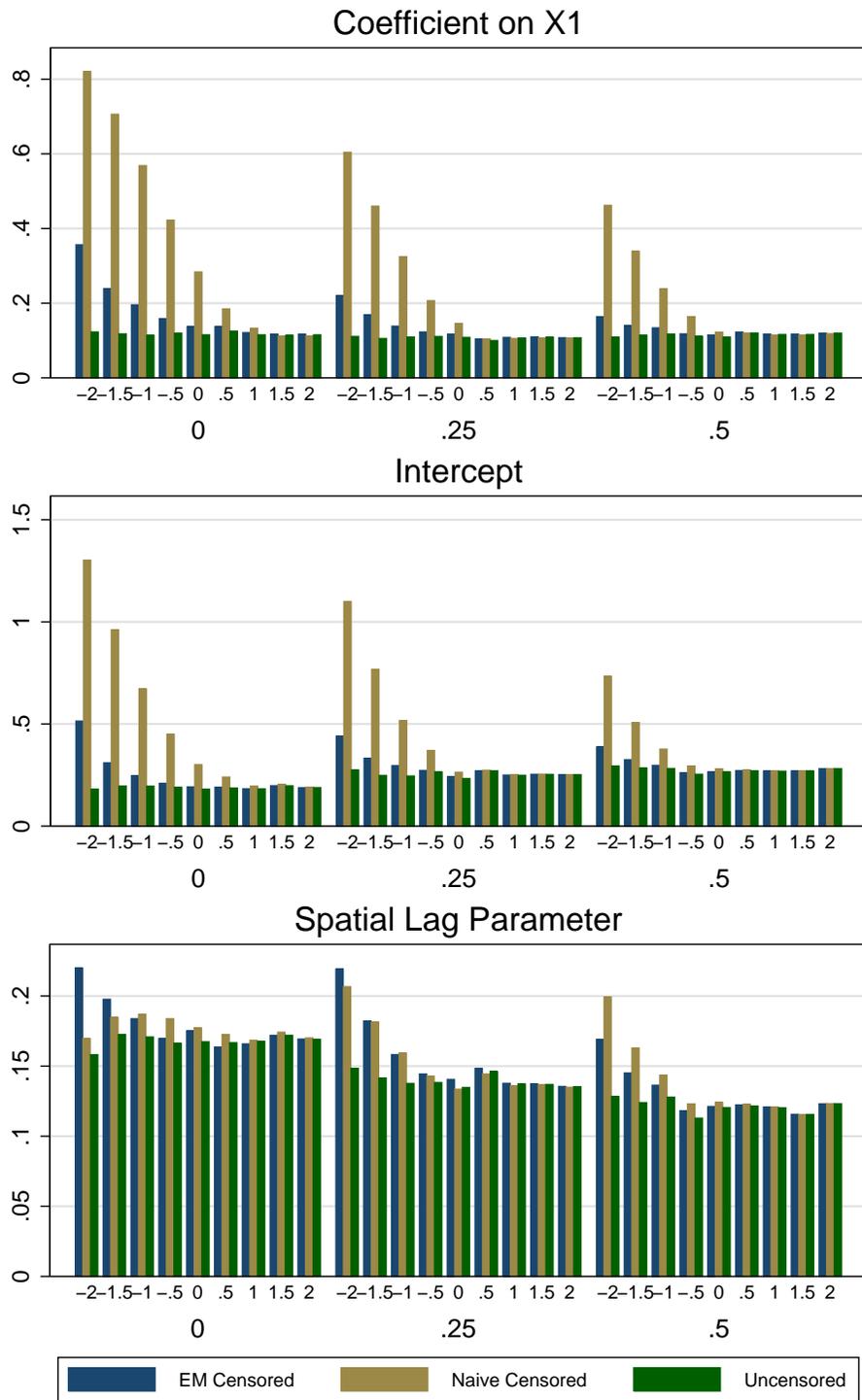


Naïve Duration Model



Notes: Results represent the average across 500 simulations, excluding cases for which the EM algorithm did not converge within 100 iterations. Standard errors represents the average value across the 500 iterations while the standard deviation comes from the sampling distribution of the parameter estimates. 34

Figure 5: Comparison of Root Mean Standard Errors from EM Approach, Naïve Duration Model, and Uncensored Benchmark Spatial Duration Model



Notes: Results represent the average across 500 simulations, excluding cases for which the EM algorithm did not converge within 100 iterations. $RMSE^2(\hat{\theta}) = (\hat{\theta} - \theta_0)^2 + Var(\hat{\theta})$.