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 25, 2014

[1] Corresponding author: jch61@pitt.edu or emily-schilling@uiowa.edu. We are grateful for funding provided by the University of Iowa Department of Political Science. Comments from Chris Zorn, Justin Grimmer, Matthew Blackwell, and participants at a University of Rochester seminar are gratefully acknowledged.

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 an estimator for 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

The use of spatial modeling in the study of politics has grown over the past two decades as the ability to estimate a variety of outcomes, including continuous, discrete, and duration, becomes more readily available. Despite this expansion, there are still issues that have not been addressed. Of particular interest in this paper is the problem of right censoring in spatial duration models. Spatial econometrics provides a common method for addressing these various forms of interdependence and models for continuous outcomes such as spatial regression are well known and growing in their use in political science.

Work on duration models makes clear that they pose their own unique set of challenges. For example, duration data often exhibit various forms of censoring that require adapting 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 when an observation began and that it lasted until at least a given point in time, but we do not know exactly how long it survived beyond this point. This poses a significant challenge for modeling interdependence since the outcome for one observation depends on the outcomes for other observations. Censored durations render the analyst incapable of properly capturing this interdependence since the complete outcome data are not fully observed.

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 iterates between imputing the durations for right censored observations and estimating the spatial duration model given these imputed values. We apply this algorithm to develop an estimator for log-normal durations with right censoring. We then explore the performance of the estimator in the face of varying degrees of right censoring and spatial interdependence via Monte Carlo followed by an empirical illustration the examines 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

As with continuous outcomes, examples abound in which units' durations likely depend on one another. For example, prominent literatures exist that evaluate interstate dependence in the timing of policy adoption or the timing of international conflict. In both cases interdependence constitutes a central question given the strategic nature of conflict and the positive and negative spillovers that policy changes create. In the policy diffusion literature in particular, this interdependence in the timing of adoption has been made exogeneous through the use of temporal lags of spatially connected's units policy choices, despite a longstanding interest in interdependence underscored by recent work that makes it explicit (Volden, Ting and Carpenter 2008). The tendency to address interdependence by assuming it to be exogenous or just treating it as a nuisance may be especially tempting in the context of duration outcomes since they provide an easy shortcut via the use of time lags to make assumptions that imply exogeneity.

Of course, estimating standard duration models without explicitly modeling spatial dependence potentially omits relevant factors that affect the hazard. **?**'s (**?**) show this directly in the case of spatial regression estimators: ignoring spatial correlation by estimating a non-spatial OLS leads to inflated estimates while properly modeling spatial interdependence through spatial lags drastically improves the estimates. Similar results hold when estimating a non-spatial duration model with continuous outcomes. In fact, since a log-normal duration model without any right censoring corresponds exactly to a linear regression estimator with normal errors, their results apply directly.

Previous attempts at capturing spatial dependence for duration outcomes have typically modeled it through the inclusion of frailty terms. The use of frailties has a long tradition in duration modeling; frailties help account for the unobserved differences across units that make some more likely to experience the event of interest sooner than others, thereby playing a role similar to random effects (Box-Steffensmeier and Jones 2004). Recent work captures spatial processes by allowing correlation between the random effects for spatially connected observations. For example, Darmofal (2009) discusses a Bayesian spatial survival model that uses a conditionally autoregressive (CAR) prior to capture spatial autocorrelation in the frailty terms (see also Banerjee, Wall and Carlin 2003). Empirical application to the timing of members of Congress' announcements of their position on the North American Free Trade Agreement demonstrates that these estimators offer useful insight into the dependence structure of these random effects. Further, since the spatial correlation exists in the frailties, they can account for right censoring with only minor adaptation for parametric forms, e.g., the Weibull, or with little adaptation as in the case of the semi-parametric Cox version.

Importantly, though, the inclusion of spatially autocorrelated frailty terms does not address many interesting or theoretically motivated forms interdependence. Spatial dependence emerges not just through the presence of correlated random shocks or unmodeled heterogeneity, but also through explicit interdependence in units' choices. This might arise from strategic choices made by actors about the optimal time to enter a conflict or to announce their position on pending legislation. These forms of interdependence cannot be properly captured with frailties, but rather must occur through the dependent variable as is the case with spatial lag models. Given this, we develop a spatial lag duration estimator that allows for interdependence between duration outcomes rather then only through random errors. For this project we consider a model for a log-normal duration process.

The data generating process that this describes can be written in matrix notation as follows:

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\beta + \mathbf{L}\mathbf{u} \tag{1}$$

where $\mathbf{y}$ represents the natural log of the non-negative duration outcome and $\mathbf{W}$ captures the spatial dependency between observations.[1] ρ is the parameter of spatial dependence or the degree of spatial autocorrelation in the model. In political science, spatial relationships are generally expected to have a positive relationship, leading to a positive ρ. The diagonal matrix $\mathbf{L}$ has $\frac{1}{\lambda}$ down the center diagonal, where λ represents the shape parameter for the error distribution and captures the existence and degree of duration dependence. From this structural form we can derive the reduced form of the spatial model:

$$\begin{aligned}
\mathbf{y} &= (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\beta + (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{L}\mathbf{u}, & (2)\\
&= \mathbf{\Gamma}\boldsymbol{X}\beta + \mathbf{\Gamma}\boldsymbol{L}\boldsymbol{u}, & (3)\\
&= \mathbf{\Gamma}\boldsymbol{X}\beta + \boldsymbol{v}. & (4)
\end{aligned}$$

where $\mathbf{\Gamma} = (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}$ and $\boldsymbol{v} = \mathbf{\Gamma}\boldsymbol{L}\boldsymbol{u}$. The reduced form equation makes evident the importance of including the spatial lag since the outcome for each unit depends on both the observed and unobserved components of other units through the spatial lag matrix. Ignoring this spatial dependence when it is not equal to 0 (i.e., $\rho \neq 0$) leads to biased coefficient estimates as **?**'s (**?**) show in the linear regression context.

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

The reduced form equation also makes evident the additional challenge posed by right censoring. Right censoring occurs when we do not observe the exact moment the event occurs, but only that it occurred after a specific point in time. A standard duration model handles this easily by replacing the density of failure at a specific time with the survival function evaluated at the time of censoring, which captures the probability that the observation lasts longer than the censoring point. In the context of spatial durations however, this becomes more complicated since the probability of surviving past a single point in time depends on the failure time of other observations and therefore the likelihood can not be constructed via the marginal distributions of failure or censoring for each observation. Moreover, since the probability of failure for observed cases will often depend on the true failure time for right censored cases, we lose critical information needed to model the failure time for all observations.

To explicitly introduce right censoring, assume we do not fully observe the duration, y_i, for some subset of observations but rather we observe only that it lasted at least until the censoring point. Thus we divide observations into those cases that are censored, denoted with y_i^C, and those cases that we fully observe, denoted with y_i. In standard duration analysis treating right censored cases as fully observed when they are in fact right censored can lead to bias in the estimates of the parameters (Box-Steffensmeier and Jones 2004).

To date censoring in spatial models has been dealt with in two different ways. First, researchers assign a value to all of the censored cases and treat this as the observed failure time; oftentimes this is the threshold cutoff between those cases that are observed and those that are censored. Alternatively, they omit the censored observations from the analysis. Both can lead to biased estimates. The first systematically reduces the dependent variable for censored cases. The second leads to truncation and unless the factors that produce the censoring are unrelated to those that affect when the event occurs, truncating the sample will also tend to produce bias (Box-Steffensmeier and Jones 2004). Spatial modeling has the potential to exacerbate both of these issues since even uncensored observations depend on the realization of the outcome variable in censored cases. We therefore require a method to recover the missing information about censored cases so that we can properly capture the spatial dependence between observations, a task to which we turn in the next section.

An EM Approach for Imputation and Estimation

To develop an estimator that simultaneously addresses spatial interdependence and right censoring we adapt Wei and Tanner's (1991) imputation method for (non-spatial) censored regression data to account for spatial interdependence. The logic underlying their approach lies in alternating between taking draws of the outcome variable for censored cases conditional on the censoring point and estimating the model on the observed values for uncensored cases and the imputed values for censored cases. This approach allows estimation to proceed with computationally nonintensive regression models which at one point would have been easier than doing maximum likelihood with a mix of censored and uncensored observations. We adapt it to spatial durations with censored data for the same reason since calculating a high dimensional multivariate distribution with a mix of observed and censored cases would prove challenging today.

Extending the original approach requires accounting for the interdependence between observations as we iterate between the estimation and imputation stages. In words, we use the most recent estimates to calculate the reduced form residuals. We then translate them back into their structural form counterparts which allows us to take draws in an i.i.d. space. This translation occurs via the Cholesky decomposition of the correlated reduced form errors which, as a lower triangular matrix, allows us to iteratively solve for each of the structural errors as a linear combination of the previous $n-i+1$ reduced form errors. For censored cases we obtain the minimum value that ensures that realized value will exceed the observed censoring point. We use this to take a random draw of the i.i.d. structural form error from a censored normal distribution and use this draw in the solution for the next structural form error. Once this iterative process concludes, we calculate the dependent variable, reestimate the model, and repeat until the parameters stabilize.

To be precise, the algorithm starts by estimating a spatial regression model that treats the censoring point as the observed event time. Using these results, we calculate the reduced form residuals. The data generating process follows from equation 4. The reduced form error $\mathbf{v}$ represents a linear combination of i.i.d. errors. Using these reduced form residuals, we then calculate the Cholesky decomposition, $\boldsymbol{A^{-1}}$, of the expected covariance matrix of these errors.

$$
\begin{aligned}
E[\hat{\mathbf{v}}\hat{\mathbf{v}}'] &= E[(\boldsymbol{\Gamma L}\hat{\mathbf{u}})(\boldsymbol{\Gamma L}\hat{\mathbf{u}})'], && (5)\\
&= (\boldsymbol{\Gamma L})E[\hat{\mathbf{u}}\hat{\mathbf{u}}'](\boldsymbol{\Gamma L})', && (6)\\
&= (\boldsymbol{\Gamma L})(\boldsymbol{I})(\boldsymbol{\Gamma L})', && (7)\\
&= (\boldsymbol{\Gamma L})(\boldsymbol{\Gamma L})'. && (8)
\end{aligned}
$$

For a log-normal distribution $\boldsymbol{L}$ captures the standard deviation as a departure from the standard normal error $\mathbf{u}$ and has σ_u down its diagonal. We can therefore simplify this for the log-normal case to $\sigma_u^2\boldsymbol{\Gamma\Gamma}'$.

We can then write the reduced form errors as a linear combination of i.i.d. normal errors: $\hat{\mathbf{u}} = \boldsymbol{A^{-1}}\eta$. Since $\boldsymbol{A^{-1}}$ is upper triangular, we can iteratively solve for the corresponding value of η one observation at a time starting with the last observation. Assume without loss of generality that we have solved for the k^{th} observation. When we solve for the $(k-1)^{th}$ observation, if the observation is uncensored this results in the implied values. If the $(k-1)^{th}$ observation is censored, we can calculate the necessary value for the duration to exceed the censoring point given the calculated errors for the previous observations and then draw from the distribution conditional censored below at this value. The value of censoring point is calculated in the same way that we solve for the uncensored $\boldsymbol{\eta_k}$. Once the censoring point is determined for $k-1$, we take a random draw from the censored standard normal distribution to obtain a value that is greater than the censored value to impute $\boldsymbol{\eta_{k-1}}$ (see Appendix for example). This method returns the same value for each of the uncensored cases and imputed outcomes for the censored cases.

In order to create the imputed reduced form spatial errors, we multiply the $\boldsymbol{\eta}$ by $\boldsymbol{A^{-1}}$. 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. The imputed Y is then run in a spatial lag model and the parameter vector is saved.

To account for estimation uncertainty in our procedure we multiply impute M draws of the errors for censored cases. 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 estimates using each of the multiply imputed data sets are combined using Rubin's (2009) formula:

$$\bar{\theta} = \frac{1}{M}\sum_{m=1}^{M}\hat{\theta}_m, \tag{9}$$

$$Var(\bar{\theta}) = \frac{1}{M}\sum_{m=1}^{M}Var(\hat{\theta}_m) + \frac{M+1}{M}\left(\frac{1}{M-1}\sum_{m=1}^{M}(\hat{\theta}_m - \bar{\theta})^2\right). \tag{10}$$

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 either ignoring the censoring or ignoring the spatial interdependence as well as to those from the original uncensored data as a benchmark comparison. To evaluate our EM estimator's performance across a wide range of circumstances we vary 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, $\boldsymbol{W}$, based on rook and queen contiguity. We then generate 100 independently and identically distributed observations of a single independent variable, $\boldsymbol{X}$, according to the standard normal distribution. For simplicity's sake we set $\boldsymbol{L} = I$, resulting in the following data generating process:

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}(-1 - 1 \times \mathbf{X}) + (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{u} \tag{11}$$

With 100 observations our approach to generating **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, which produces a matrix in which all of the elements represent 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 varying from 0 to 0.75 by increments of 0.25. 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, 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 -1 to 0.5 by increments of 0.5. The amount of censoring that occurs ranges from fifty to nearly zero percent depending on the degree of censoring and the amount of spatial correlation (see Figure 1 in our supplemental appendix).

For each combination of values of spatial interdependence and censoring points we generate 500 draws of u from a standard normal distribution, calculate the value of $\mathbf{Y}$, then apply our censoring rule so that $Y_i^c = \min\{Y_i, C\}$. We estimate four models: a naïve spatial model that treats all realizations of Y_i^c as uncensored, a naïve log-normal duration model that accounts for the censoring but ignores the spatial dependence, 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.

For our EM estimator we set the maximum number of iterations for each draw to 100 and use fifteen imputations for each step of the EM process. With high degrees of censoring, our proposed estimator occasionally fails to produce estimates. This happens about a third of the time when the degree of censoring is high and the spatial correlation was low, circumstances which make it difficult for our estimation procedure to draw strength across observations. In exploring individual draws we found this to be an issue usually resolved by changing the seed and rerunning the estimator, so it should not pose a significant problem for well-behaved applications. Standard errors are calculated according to Equation 10.

[Figure 4 here.]

Figures 4 presents the average bias in the estimates for the intercept, slope coefficient, and spatial dependence parameters, with the first two having true values of -1 and the latter varying from 0 to 0.75.[2] 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 estimator suggests bias for all three parameters, with the deviations from the true values most severe for the coefficient and intercept. These deviations get smaller as the spatial correlation increases or the censoring parameter decreases. Since this estimator ignores censoring it makes sense that it does worse as censoring increases. The magnitude of the apparent bias can be quite severe with average estimates frequently ranging from 80-45% of the true value of -1.

Third, the naïve duration model exhibits the opposite pattern. It does quite well with little interdependence, which we expect since when $\rho = 0$ out data correspond exactly to a standard duration model. When the spatial correlation reaches 0.25, however, the intercept begins to show some apparent bias which quickly becomes much worse: the average bias is over 300% of the magnitude of the true value when $\rho = 0.75$. The slope coefficient exhibits a similar pattern, though on a much smaller scale, with deviations up to 10% of the true effect.

Fourth, the EM estimator produces average estimates for all three parameters near their true values. The results shown here exhibit some potential bias in estimating the intercept and slope parameters for $\rho = 0$ and small values of c, which makes sense since in those cases one has much less information about the outcomes for the censored cases since

[2] We report detailed results for all parameters in our supplemental appendix.

their values do not spatially influence the outcomes in uncensored cases.[3] We also see some slight deviation in the estimate of the spatial parameter relative to the benchmark case with moderate spatial dependence, but not with large or zero dependence. When censoring is very low a simple duration estimator may be preferred and when spatial correlation is high a simple spatial estimator may suffice, but our estimator performs about as well as the alternatives even in these situations and clearly outperforms both in all other cases.

Given the greater complexity of our estimator, including the multiple imputation component, we also want to consider the relative precision of its estimates. We start by evaluating the accuracy of the reported standard errors for our approach and then move to a mean squared error comparison across alternatives. Figure 3 in our supplemental appendix provides a comparison of the average standard errors across the 500 draws to the standard deviation of the sampling distribution of the parameter estimates and indicates that our estimator provides accurate estimates of uncertainty.

[Figure 5 here.]

Moving to a comparison across estimators, Figure 5 plots the square root of the sum of the variance and the squared bias of each parameter for all four estimators. The plots show clear evidence that even with its greater complexity our EM estimator generally equals or outperforms both of the naïve ones and by a wide margin for the spatial estimator with spatial interdependence less than 0.75 and for the duration estimator with spatial dependence equal to or greater than 0.5. The results for the spatial dependence parameter are generally comparable across models.

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

[3]We confirmed this by increasing the amount of censoring but do not report the results given the increasingly low rate at which we obtained estimates.

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.

Illustration: The Diffusion of WWI

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). 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 attempt to model the contagion of conflict that addresses the problem of right-censoring is Melin and Koch (2010). They take a two stage approach to analyze the expansion of militarized interstate disputes (MIDs) over the period 1816-1992. In their first-stage, the expansion stage, they model the time until an initial dyadic militarized dispute expands. This is important because the potential for conflict expansion varies greatly across militarized disputes, and potential joiners take this likelihood into account. Therefore, in the second stage, the joining stage, Melin and Koch model the time until a potential joiner enters the conflict, conditional on the expected duration of the initial dyadic 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. This approach handles right-censoring and the strategic timing of conflict entry decisions in a sophisticated way, but in many instances the triadic nature of the interdependence will be too restricted in scope.

Radil, Flint and Chi (2013) take a different approach, rooted in social network analysis, in their study of WWI. They examine how international political networks and geography influenced the decisions of states to join the conflict in four waves of expansion. The initial wave began with Autria-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. Radil et al. identify the structural position of each state in multiple networks defined by geographical contiguity, alliances, and international rivalries and then relate this position to their decision to join the war. They find a strong and statistically significant correlation (using Quadratic Area Procedure matrix regression) between structural equivalency measures of network positions for two states at one stage of the conflict and their joint involvement at subsequent stages. Additionally, they use convergence of iterated correlations applied to the multiple network matrices to identify eight latent classes of states and show with ANOVA that these groups help to explain the timing decisions of their members.

While these studies represent best practice in the empirical study of conflict diffusion, there are significant limitations in both. Melin and Koch only account for the connections between a potential joiner and the initial target and conflict instigator. Taking WWI as an example, this would mean that 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.

Radil et al. account for the complex structure of interdependence that influenced the war joining decisions of states, and this is an important advance, but they reduce the network relationships to latent classes and measures of structural equivalence that make it impossible to identify which sources of connectivity matter most for the expansion of conflict. Is it geographical relationships, alliances, or rivalries? Moreover,their ANOVA framework is less than ideal for duration analysis given that it cannot address the right-censoring problem. The strengths and weaknesses of these two approaches to interdependent duration analysis mirror those of the simpler naïve strategies that we presented in our Monte Carlo experiments. They are *either* strong with respect to modeling the time to conflict expansion (Melin and Koch and our non-spatial duration model) *or* the interdependence in war-joining behavior (Radil et al. and our naïve spatial model).

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 duration model in Melin and Koch as well as the network analysis in Radil et al. All of the entry timing decisions are potentially linked through multiple connectivity matrices 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 half of the independent states in the international system. Had the war continued, undoubtedly, more states would have been drawn into in the conflict.

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 exposure, and relative military capabilities, but this is approach unsatisfactory. Ultimately, each states 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: geographical distance, rivalry, and defensive alliances. These sources of interdependence suggest that states will be influenced by the participation and entry timing decisions of their neighbors, rivals, and allies.

To capture the role of geography in the spread of war, we use an inverse distance spatial weights matrix, which is standard in the spatial econometrics literature. With this matrix, every state's entry timing decision is influenced by every other state's decision, but the interdependence between geographically proximate states is much stronger than it is for distant ones. Of course, the geographical notion of distance can be extended to social, political and economic contexts (**?**), and this is how we approach our rivalry and alliance weights matrices. Every state's entry timing decision is influenced by every other state's decision, but the interdependence between rivals and allies is much stronger than for states not connected by a rivalry or alliance. To be more specific about how we generate our matrices, consider the case of international rivalry. If state i has no rivals, then its weights for $i \neq j$ are

$$w_{ij} = \frac{1}{n-1}.$$

If state i has both rivals and non-rivals, then its weights for non-rivals $i \neq j$ are

$$w_{ij} = \frac{(1-r)}{n-n_r-1},$$

where r is a parameter on the interval $[n_r/(n-1), 1]$ that determines the relative influence of rivals and non-rivals and n_r is $i's$ total number of rivals. For $i's$ rivals, the weights are

$$w_{ij} = \frac{r}{n_r}.$$

With this function for w_{ij}, everyone influences everyone else equally is a special case where $r = n_r/(n-1)$. This produces a weights matrix for which all of the off-diagonal elements are $1/(n-1)$. At the other extreme, $r = 1$, the function produces rows in the weights matrix where *only* rivals have influence. We report the results for $r = 1$ below, but our findings are robust to the choice of r.

We identify all of the defensive alliances that were in force at the onset of the war using **?**, being careful to exclude alliances that were formed during the conflict. In total, there were sixteen defensive alliances connecting eleven of the countries in our sample. We code all of the rivalries, both proto-rivalries (short-term) and enduring ones, that existed at the onset of the war using **?**. Again, we do not include rivalries that emerged during the conflict. At the onset of the war, there were thirty-one active rivalries involving twenty-six of the states in our sample.

Following Radil et al., our dependent variable is the number of days before entering WWI. Of the 44 sample countries, half eventually joined the conflict. All the spatial weights matrices are row-standardized. We include three important covariates in the analysis. National capabilities are the COW CINC index scores (Singer, Bremer and Stuckey 1972); democracy is the 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 the three types of models evaluated in our Monte Carlo: a non-spatial duration model, a naïve spatial model that treats the time the censoring as an observed failure time, and our multiple imputation model. Based on the Monte Carlo results, we expect the estimates from non-spatial duration model to overstate the effects of the covariates on war joining. This is particularly true for variables such as national capabilities that cluster among states that are linked by the mechanisms or "vectors" through which conflict diffuses. For example, national capabilities in our sample are more than twice as high among the states that are connected by alliance networks. Additionally, because the naïve spatial model fails to account for right-censoring, we expect its estimates to understate both the strength of interdependence and as well as the direct effects (i.e., those prior to any spatial or network feedback) of the covariates on war joining.

We report the results in Table ?. Some clear patterns emerge. First, among the covariates, national capabilities is the only one that has a robust statistically significant effect on the war-participation timing of states. Military power is associated with early entry into the war. Second, among the spatial lags, only the rivalry and alliance lags are statistically significant. We do not find that geography matters for the participation timing decisions of states. This may seem surprising at first; however, the simple fact that this was a *world* war means that geography was less significant than would otherwise be the case. For WWI, military power, rivalry, and alliances were more important determinants of participation than a state's geographical location.[4]

When we compare the estimators, we find very similar patterns to those in the Monte

[4]This is true whether we use an inverse distance or contiguity spatial weights matrix, once we control for national capabilities and trade. The contiguity spatial lag is significant in a model without these two covariates.

Carlo experiments. Focusing on national capabilities, we see that the estimated coefficient from the non-spatial duration model is much larger than the estimates from the other models. The difference is particularly stark when we use the alliance weights matrix (national capabilities and alliance membership are strongly correlated) and cannot be explained by the fact that the spatial model decomposes the total effect of national capabilities into a direct effect and an indirect effect through spatial feedback. If we calculate the spatial multipliers from our imputation model, $(\mathbf{I} - \rho\mathbf{W})^{-1}$, we find that for the defensive alliance weights matrix the average multiplier is 1.05 and the maximum is 1.29. Inflating the direct effect estimates by these multipliers gives $1.05 \times -18.91 = -19.86$ and $1.29 \times -18.91 = -24.39$. The estimated effect from the non-spatial model, -31.45, is 29% larger than the maximum effect from the alliance spatial model and 58% larger than the average effect. Additionally, the naïve spatial model underestimates the direct effect of national capabilities as well the strength of interdependence relative to the imputation model. Thus, as in our Monte Carlo, the estimates from the non-spatial duration model seems to overstate the effect of national capabilities while the estimates from the naïve spatial model understate this effect.

To sum, our imputation approach to interdependent duration analysis leads to a more accurate understanding of the diffusion of WWI. When compared with the non-spatial and naïve spatial approaches, ours produces more accurate estimates of the total effects of covariates as well as, in the case of the naïve spatial model, a better decomposition into direct effects and interdependence-driven multipliers. We also believe our approach improves on the implicit triads strategy in Melin and Koch and the network analysis in Radil et al. for essentially the same reason: our method is the only one that allows for complex interdependence and addresses the right-censoring problem.

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.

A Example of Imputation Procedure

$$\begin{bmatrix} \widehat{v}_1 \\ \widehat{v}_2 \\ \widehat{v}_3 \\ \widehat{v}_4 \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ 0 & b_{22} & b_{23} & b_{24} \\ 0 & 0 & b_{33} & b_{34} \\ 0 & 0 & 0 & b_{44} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{bmatrix}$$

Say case 4 is not censored. We can write $\eta_4 = \widehat{v}_4/b_{44}$.

Say case 3 is censored. We can write

$$\begin{aligned} \widehat{v}_3 &= b_{33}\eta_3 + b_{34}\eta_4, \\ \eta_3 &= (\widehat{v}_3 - b_{34}\eta_4)/b_{33}, \end{aligned}$$

then multiply impute M standard normals $\eta_3^{(m)}$ given that $\eta_3^{(m)} \geq \eta_3$.

Say case 2 is not censored. We can write $\eta_2 = (\widehat{v}_2 - b_{23}\eta_3 - b_{24}\eta_4)/b_{22}$.

Say case 1 is censored. We can write

$$\begin{aligned} \widehat{v}_1 &= b_{11}\eta_1 + b_{12} - \eta_2 b_{13}\eta_3 + b_{14}\eta_4, \\ \eta_1 &= (\widehat{v}_1 - b_{12}\eta_2 - b_{13}\eta_3 - b_{14}\eta_4)/b_{11}, \end{aligned}$$

then multiply impute M standard normals $\eta_1^{(m)}$ given that $\eta_1^{(m)} \geq \eta_1$.

B Detailed Monte Carlo Results

Table 1: Monte Carlo Results for Coefficient on X

		Estimate				Bias				Standard Error				Standard Deviation				RMSE			
Cens	ρ	Both	Spat	Dur	BM	Both	Spat	Dur	BM	Both	Spat	Dur	BM	Both	Spat	Dur	BM	Both	Spat	Dur	BM
-1	0	−0.926	−0.443	−1.009	−1.013	0.074	0.557	−0.009	−0.013	0.155	0.071	0.168	0.115	0.167	0.075	0.178	0.121	0.183	0.563	0.178	0.122
-.5	0	−0.960	−0.589	−0.992	−0.991	0.040	0.411	0.008	0.009	0.144	0.085	0.145	0.115	0.139	0.078	0.138	0.113	0.144	0.418	0.138	0.113
0	0	−0.981	−0.725	−0.997	−0.991	0.019	0.275	0.003	0.009	0.133	0.096	0.133	0.115	0.136	0.088	0.137	0.121	0.138	0.289	0.137	0.121
.5	0	−0.990	−0.835	−0.998	−0.993	0.010	0.165	0.002	0.007	0.125	0.104	0.125	0.115	0.126	0.096	0.128	0.117	0.127	0.191	0.128	0.117
-1	.25	−0.999	−0.564	−1.011	−0.995	0.001	0.436	−0.011	0.005	0.125	0.066	0.125	0.090	0.128	0.067	0.122	0.096	0.128	0.441	0.122	0.096
-.5	.25	−1.007	−0.694	−1.016	−1.002	−0.007	0.306	−0.016	−0.002	0.114	0.074	0.113	0.090	0.125	0.071	0.121	0.096	0.125	0.314	0.122	0.096
0	.25	−1.009	−0.802	−1.021	−1.005	−0.009	0.198	−0.021	−0.005	0.105	0.080	0.106	0.091	0.109	0.076	0.107	0.094	0.109	0.212	0.109	0.094
.5	.25	−1.006	−0.882	−1.019	−1.002	−0.006	0.118	−0.019	−0.002	0.099	0.084	0.100	0.091	0.099	0.077	0.097	0.091	0.099	0.141	0.099	0.091
-1	.5	−1.012	−0.718	−1.047	−1.002	−0.012	0.282	−0.047	−0.002	0.122	0.082	0.127	0.097	0.128	0.085	0.123	0.098	0.128	0.294	0.132	0.098
-.5	.5	−1.004	−0.815	−1.046	−0.998	−0.004	0.185	−0.046	0.002	0.111	0.087	0.120	0.098	0.114	0.087	0.115	0.101	0.114	0.205	0.124	0.101
0	.5	−1.003	−0.886	−1.051	−0.998	−0.003	0.114	−0.051	0.002	0.105	0.091	0.115	0.098	0.112	0.090	0.116	0.102	0.112	0.145	0.127	0.102
.5	.5	−0.999	−0.938	−1.051	−0.998	0.001	0.062	−0.051	0.002	0.101	0.093	0.111	0.097	0.099	0.086	0.106	0.096	0.099	0.106	0.118	0.096
-1	.75	−1.008	−0.992	−1.128	−1.007	−0.008	0.008	−0.128	−0.007	0.101	0.099	0.138	0.101	0.100	0.097	0.125	0.099	0.100	0.097	0.179	0.100
-.5	.75	−1.000	−0.993	−1.121	−1.000	−0.000	0.007	−0.121	0.000	0.101	0.100	0.138	0.100	0.102	0.100	0.127	0.101	0.102	0.100	0.176	0.101
0	.75	−1.002	−0.999	−1.121	−1.001	−0.002	0.001	−0.121	−0.001	0.101	0.100	0.138	0.101	0.106	0.105	0.128	0.105	0.106	0.105	0.176	0.105
.5	.75	−1.004	−1.003	−1.119	−1.004	−0.004	−0.003	−0.119	−0.004	0.101	0.101	0.136	0.101	0.103	0.102	0.132	0.103	0.103	0.103	0.178	0.103

Notes. Results represent the averages from 500 simulations, excluding cases for which the EM algorithm did not converge within 100 iterations. EM algorithm performed with 25 imputations. Duration coefficients estimated in time to failure format.

Table 2: Monte Carlo Results for Intercept

		Estimate				Bias				Standard Error				Standard Deviation				RMSE			
Cens	ρ	Both	Spat	Dur	BM	Both	Spat	Dur	BM	Both	Spat	Dur	BM	Both	Spat	Dur	BM	Both	Spat	Dur	BM
-1	0	−1.108	−1.615	−1.013	−1.073	−0.108	−0.615	−0.013	−0.073	0.236	0.264	0.132	0.183	0.220	0.274	0.131	0.195	0.245	0.673	0.132	0.209
-.5	0	−1.058	−1.373	−1.016	−1.048	−0.058	−0.373	−0.016	−0.048	0.210	0.225	0.112	0.182	0.206	0.250	0.116	0.181	0.214	0.449	0.117	0.188
0	0	−1.034	−1.231	−0.997	−1.035	−0.034	−0.231	0.003	−0.035	0.194	0.203	0.104	0.181	0.196	0.213	0.110	0.189	0.199	0.315	0.110	0.192
.5	0	−1.033	−1.131	−0.999	−1.034	−0.033	−0.131	0.001	−0.034	0.186	0.190	0.101	0.180	0.188	0.199	0.104	0.186	0.191	0.238	0.104	0.189
-1	.25	−1.095	−1.475	−1.335	−1.054	−0.095	−0.475	−0.335	−0.054	0.240	0.271	0.125	0.204	0.240	0.266	0.153	0.220	0.258	0.544	0.368	0.227
-.5	.25	−1.087	−1.314	−1.339	−1.068	−0.087	−0.314	−0.339	−0.068	0.224	0.243	0.112	0.206	0.228	0.250	0.134	0.213	0.244	0.402	0.364	0.224
0	.25	−1.060	−1.191	−1.327	−1.044	−0.060	−0.191	−0.327	−0.044	0.214	0.223	0.107	0.203	0.227	0.241	0.138	0.220	0.234	0.307	0.355	0.224
.5	.25	−1.038	−1.106	−1.324	−1.033	−0.038	−0.106	−0.324	−0.033	0.207	0.211	0.104	0.202	0.203	0.207	0.130	0.202	0.206	0.233	0.349	0.205
-1	.5	−1.108	−1.330	−1.995	−1.078	−0.108	−0.330	−0.995	−0.078	0.249	0.271	0.120	0.230	0.272	0.297	0.211	0.266	0.293	0.444	1.017	0.277
-.5	.5	−1.100	−1.229	−1.993	−1.083	−0.100	−0.229	−0.993	−0.083	0.240	0.253	0.115	0.231	0.258	0.272	0.206	0.253	0.276	0.356	1.014	0.266
0	.5	−1.063	−1.138	−1.956	−1.055	−0.063	−0.138	−0.956	−0.055	0.231	0.238	0.113	0.227	0.243	0.244	0.203	0.241	0.251	0.280	0.977	0.247
.5	.5	−1.082	−1.118	−1.989	−1.078	−0.082	−0.118	−0.989	−0.078	0.232	0.235	0.111	0.230	0.241	0.239	0.198	0.240	0.254	0.266	1.009	0.252
-1	.75	−1.190	−1.198	−4.378	−1.185	−0.190	−0.198	−3.378	−0.185	0.374	0.373	0.139	0.371	0.412	0.410	0.413	0.413	0.454	0.455	3.403	0.452
-.5	.75	−1.163	−1.175	−4.362	−1.169	−0.163	−0.175	−3.362	−0.169	0.368	0.368	0.139	0.367	0.437	0.394	0.390	0.396	0.467	0.431	3.384	0.431
0	.75	−1.177	−1.178	−4.391	−1.176	−0.177	−0.178	−3.391	−0.176	0.370	0.369	0.139	0.369	0.412	0.411	0.422	0.412	0.448	0.448	3.417	0.448
.5	.75	−1.228	−1.229	−4.375	−1.228	−0.228	−0.229	−3.375	−0.228	0.378	0.378	0.138	0.378	0.443	0.442	0.385	0.443	0.498	0.498	3.397	0.498

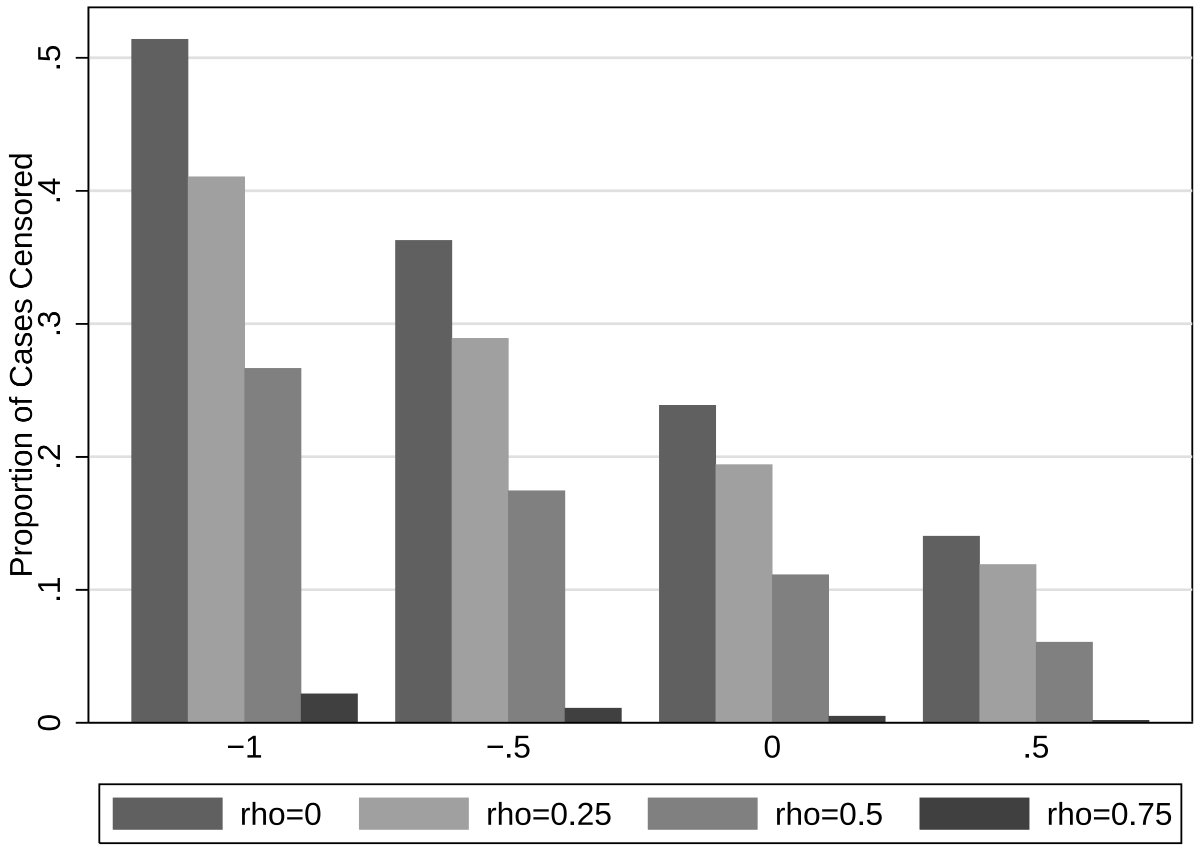
Notes. Results represent the averages from 500 simulations, excluding cases for which the EM algorithm did not converge within 100 iterations. EM algorithm performed with 25 imputations. Duration coefficients estimated in time to failure format.

Table 3: Monte Carlo Results for Spatial Lag Parameter

		Estimate			Bias			Standard Error			Standard Deviation			RMSE		
Cens	ρ	Both	Spat	BM	Both	Spat	BM	Both	Spat	BM	Both	Spat	BM	Both	Spat	BM
-1	0	−0.045	−0.062	−0.069	−0.045	−0.062	−0.069	0.197	0.173	0.163	0.176	0.174	0.170	0.182	0.185	0.183
-.5	0	−0.030	−0.044	−0.036	−0.030	−0.044	−0.036	0.181	0.167	0.161	0.173	0.183	0.162	0.175	0.188	0.166
0	0	−0.033	−0.058	−0.037	−0.033	−0.058	−0.037	0.172	0.166	0.161	0.167	0.172	0.164	0.170	0.181	0.168
.5	0	−0.034	−0.045	−0.037	−0.034	−0.045	−0.037	0.166	0.163	0.161	0.162	0.166	0.162	0.166	0.172	0.167
-1	.25	0.180	0.172	0.213	−0.070	−0.078	−0.037	0.157	0.147	0.134	0.144	0.144	0.137	0.160	0.164	0.142
-.5	.25	0.187	0.186	0.205	−0.063	−0.064	−0.045	0.148	0.143	0.135	0.144	0.149	0.136	0.158	0.163	0.144
0	.25	0.201	0.201	0.216	−0.049	−0.049	−0.034	0.142	0.139	0.134	0.154	0.157	0.149	0.162	0.165	0.153
.5	.25	0.216	0.217	0.222	−0.034	−0.033	−0.028	0.137	0.136	0.134	0.135	0.136	0.134	0.139	0.140	0.137
-1	.5	0.446	0.410	0.464	−0.054	−0.090	−0.036	0.114	0.116	0.105	0.122	0.131	0.118	0.133	0.159	0.124
-.5	.5	0.452	0.428	0.462	−0.048	−0.072	−0.038	0.110	0.112	0.106	0.115	0.121	0.112	0.125	0.141	0.119
0	.5	0.461	0.445	0.466	−0.039	−0.055	−0.034	0.107	0.109	0.105	0.111	0.114	0.109	0.118	0.127	0.114
.5	.5	0.460	0.453	0.463	−0.040	−0.047	−0.037	0.106	0.107	0.105	0.110	0.110	0.110	0.117	0.120	0.116
-1	.75	0.707	0.707	0.708	−0.043	−0.043	−0.042	0.080	0.079	0.079	0.090	0.090	0.090	0.100	0.100	0.099
-.5	.75	0.713	0.711	0.711	−0.037	−0.039	−0.039	0.079	0.079	0.079	0.096	0.085	0.085	0.103	0.093	0.093
0	.75	0.711	0.711	0.711	−0.039	−0.039	−0.039	0.079	0.079	0.079	0.087	0.087	0.087	0.096	0.096	0.096
.5	.75	0.699	0.699	0.699	−0.051	−0.051	−0.051	0.081	0.081	0.081	0.093	0.093	0.093	0.106	0.106	0.106

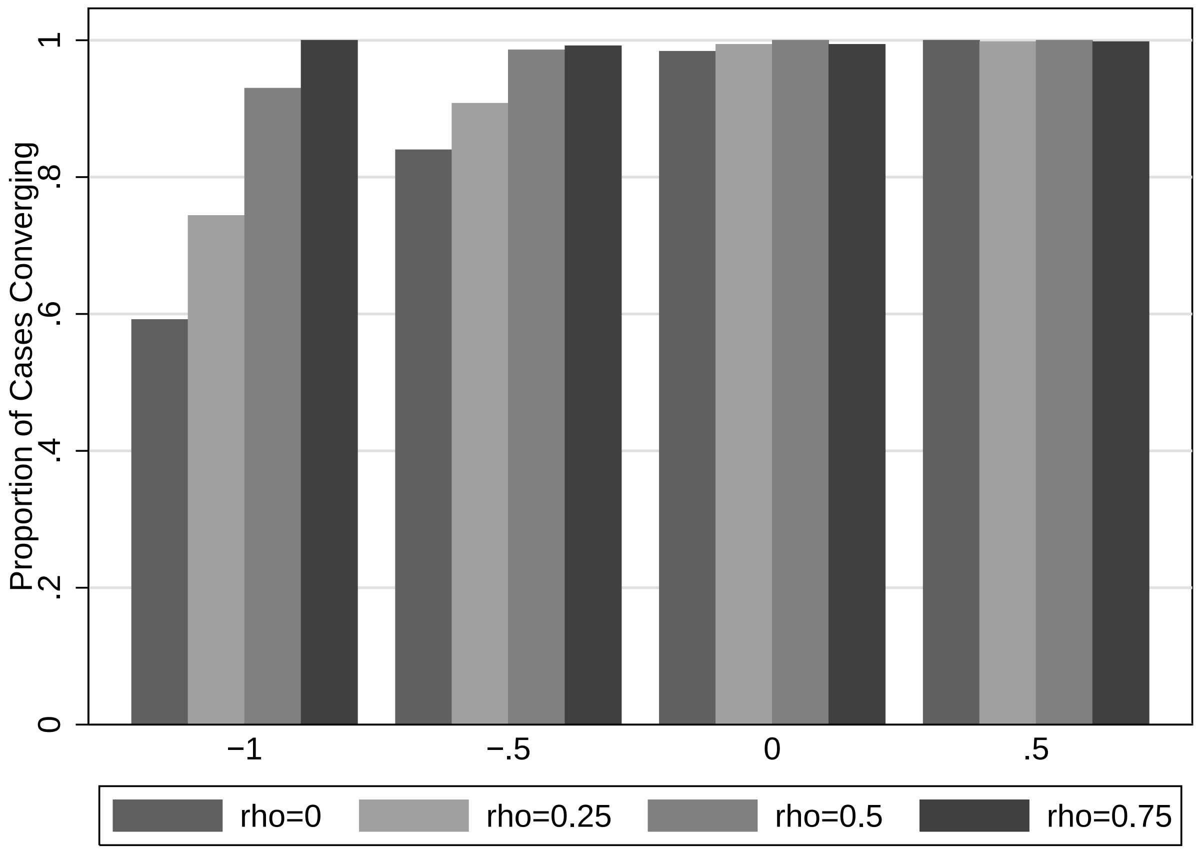
Notes. Results represent the averages from 500 simulations, excluding cases for which the EM algorithm did not converge within 100 iterations. EM algorithm performed with 15 imputations. Duration coefficients estimated in time to failure format.

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



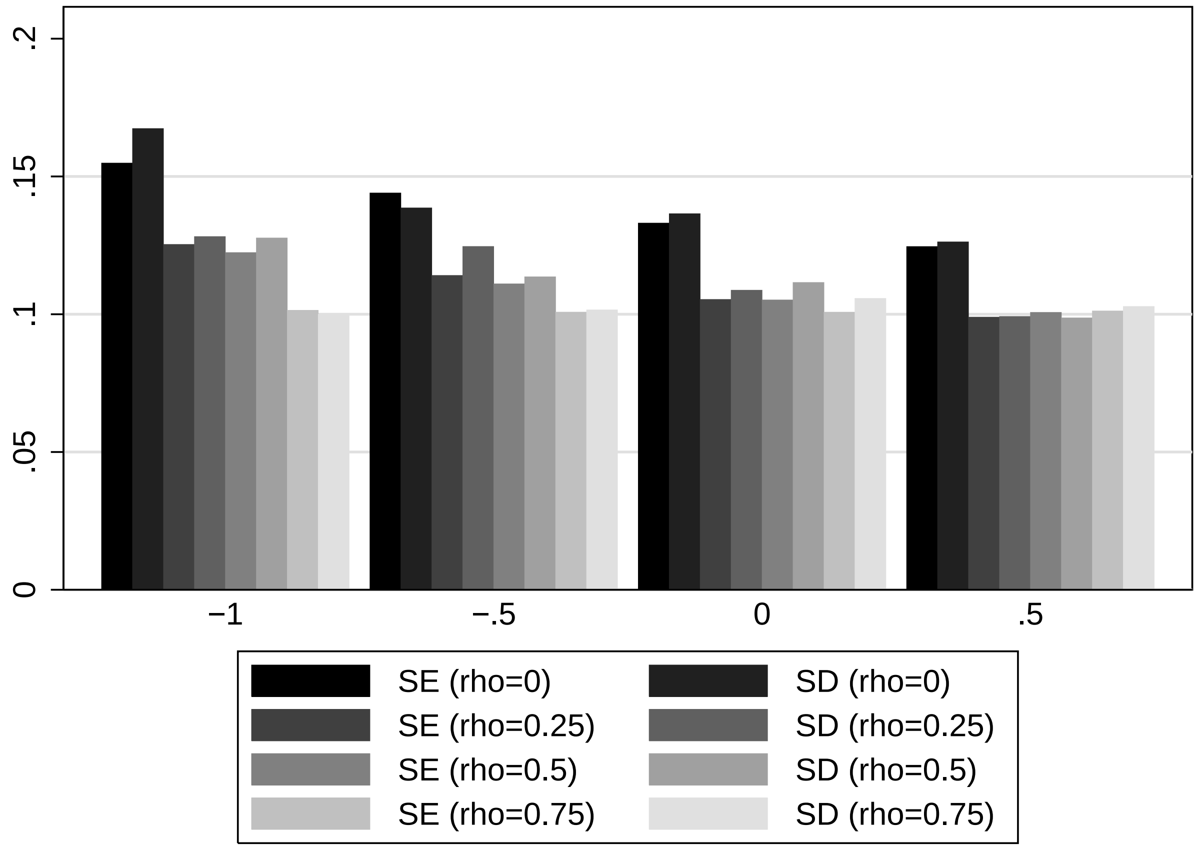
Notes: Results represent averages across 500 simulations, excluding cases for which the EM algorithm did not converge within 100 iterations. EM algorithm performed with 15 imputations.

Figure 2: Proportion of Draws that Converged by Spatial Correlation and Censoring Point



Notes: Results from 500 simulations. EM algorithm performed with 15 imputations and 100 maximum iterations.

Figure 3: Comparison of Standard Deviations and Standard Errors, varying the Amount of Spatial Correlation and the Amount of Censoring



Notes: Results represent the average across 500 simulations, excluding cases for which the EM algorithm did not converge within 100 iterations. EM algorithm performed with 15 imputations. Standard errors represents the average value across the 500 iterations while the standard deviation comes from the sampling distribution of the parameter estimates.

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.

Banerjee, Sudipto, Melanie M. Wall and Bradley P. Carlin. 2003. "Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota." *Biostatistics* 4(1):123–142.

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. 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.

Gartner, Scott Sigmund and Randolph M Siverson. 1996. "War expansion and war outcome." *Journal of Conflict Resolution* 40(1):4–15.

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.

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, Randolf M and Harvey Starr. 1991. *The diffusion of war: A study of opportunity and willingness*. University of Michigan Press.

Volden, Craig, Michael M. Ting and Daniel P. Carpenter. 2008. "A Formal Model of Learning and Policy Diffusion." *American Political Science Review* 102(03):319–332.

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

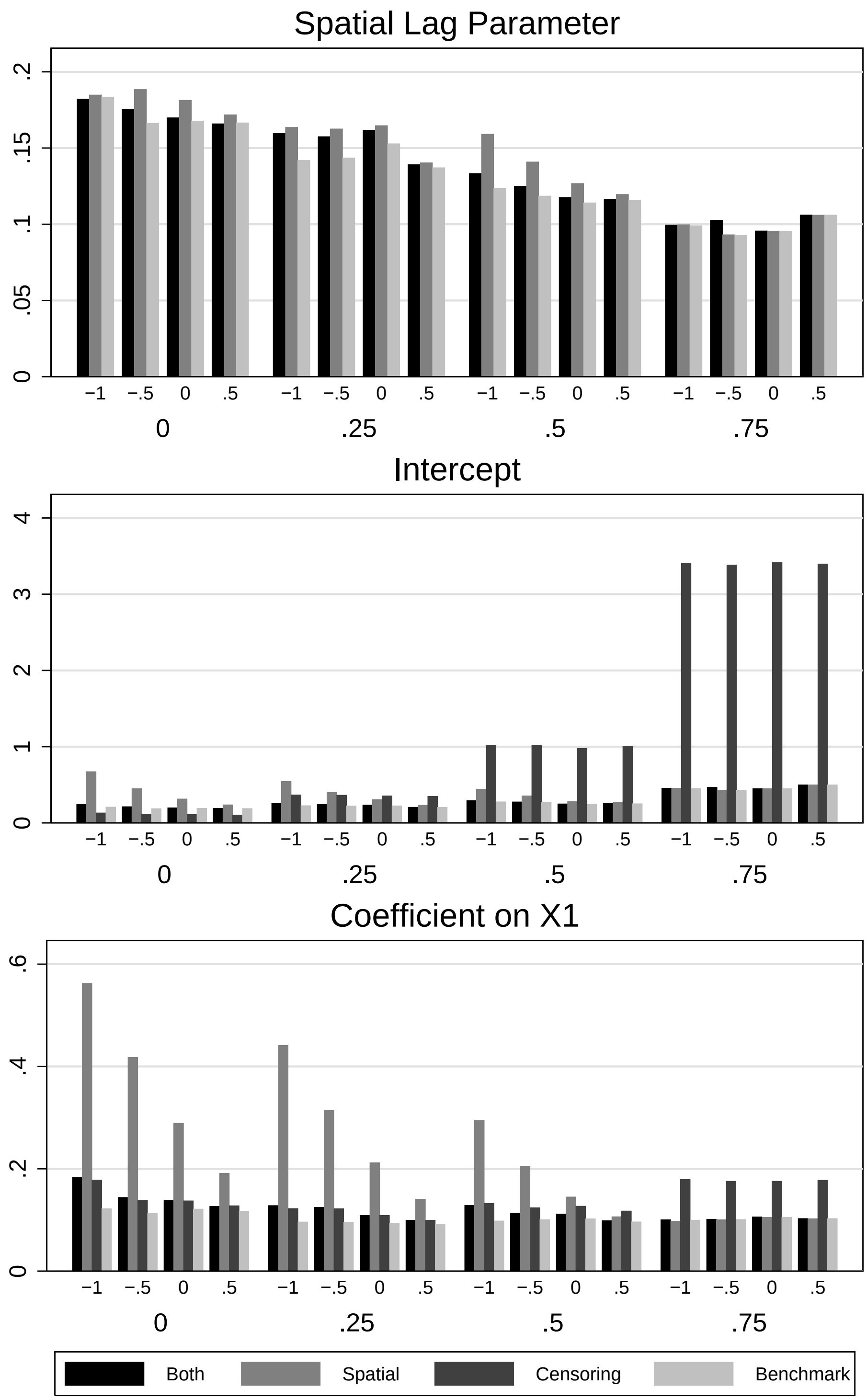
Spatial Lag	**None**	**Contiguity**		**Alliance**		**Rivalry**	
		Naive	Imp.	Naive	Imp.	Naive	Imp.
constant	9.211***	3.523***	7.898**	6.018***	6.707***	4.581***	4.586**
	(1.366)	(0.752)	(2.309)	(1.031)	(1.492)	(1.147)	(1.876)
capabilities	−31.451**	−17.975**	−27.851**	−14.056**	−18.907*	−14.157**	−30.306*
	(13.283)	(7.495)	(11.617)	(7.428)	(10.724)	(6.819)	(18.397)
democracy	0.017	0.001	−0.007	0.013	0.024	−0.000	0.061
	(0.087)	(0.047)	(0.082)	(0.046)	(0.078)	(0.043)	(0.116)
trade	−0.208	−0.186	−0.250	−0.262*	−0.389	−0.138	−0.090
	(0.281)	(0.159)	(0.270)	(0.152)	(0.252)	(0.143)	(0.358)
σ^2	3.070	3.523***	7.897**	1.801***	6.442***	1.683***	12.944**
	(0.500	(0.752)	(2.309)	(0.193)	(1.786)	(0.182)	(4.972)
ρ		0.120	0.214	0.313**	0.471***	0.447***	0.675***
		(0.139)	(0.163)	(0.155)	(0.158)	(0.142)	(0.118)

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

Figure 4: Average Bias from EM Approach, Naïve Duration Model, Naïve Spatial Model, and Uncensored Benchmark Spatial Duration Model, varying the Amount of Spatial Correlation and the Amount of Censoring

Notes: Results represent the average deviation from the true parameter value across 500 simulations, excluding cases for which the EM algorithm did not converge within 100 iterations. EM algorithm performed with 15 imputations. Duration coefficients estimated in time to failure format.

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



Notes: Results represent the average across 500 simulations, excluding cases for which the EM algorithm did not converge within 100 iterations. EM algorithm performed with 15 imputations. $RMSE^2(\hat{\theta}) = (\bar{\hat{\theta}} - \theta_0)^2 + Var(\hat{\theta})$.