INTERNATIONAL REGIONAL SCIENCE REVIEW 26, 3: 269–288 (July 2003)

SPATIAL PATTERNS AND PROCESSES IN A LONGITUDINAL FRAMEWORK

BRIGITTE WALDORF
Department of Geography and Regional Development,
University of Arizona, Tucson, AZ, bwaldorf@u.arizona.edu

This article explores the conceptual equivalence between hazard models applied to both temporal data and distance data by focusing on the "at-risk" concept, which is central to longitudinal models but has not received sufficient attention in the application of hazard models in spatial settings. A proper conceptualization in a spatial (distance) setting is based on distant-dependent Markovian transition probabilities describing the risk of switching between states. Such a conceptualization is possible for continuous spatial processes, as well as for point-generating processes leading to spatial point patterns. Hazard models for a series of scenarios simulating various point generation trajectories are compared. This process-oriented perspective is further augmented by explicitly accounting for temporal dimensions (speed) of point-generating processes.

Keywords: *spatial hazard models; spatial processes; point patterns*

1. INTRODUCTION

Longitudinal models, also referred to as hazard models and duration models, have become an established method in regional science and related disciplines, applicable in various contexts dealing with the timing of events. Although first applied by engineers concerned about the failure of products, as well as in biomedical research dealing with the timing of deaths following the onset of a disease, longitudinal models have subsequently also been used in the social sciences to understand the temporal dimensions of such diverse phenomena as the length of unemployment spells (e.g., James 1989; Narendranathan and Stewart 1993), duration of residence (e.g., Clark 1992; Odland 1997; Davies Withers 1997; Glavac and Waldorf 1998), and consumer store choice dynamics (Popkowski Leszczyc and Timmermans 1996).

At the core of these models is the nonnegative random variable *T*, measuring the duration of a state, or, equivalently, the length of time prior to a terminating event.

The author would like to thank Noel Cressie, Jim LeSage, two anonymous referees, and the editors for valuable comments and suggestions.

DOI: 10.1177/0160017603253788

Of prime interest is the conditional probability of an event happening at time t, given that the event has not happened up to that time. An elaborate mathematical framework has been developed to model the variations in duration T by specifying conditional failure probabilities or hazard rates of failure. The framework includes nonparametric, semiparametric, and parametric formulations of the hazard model for both continuous and discrete time scales (Lawless 1982; Kiefer 1988; Yamaguchi 1991), as well as the treatment of a variety of econometric issues, such as unobserved heterogeneities (Heckman and Singer 1985) and endogenous interactions in the form of spillover effects and spatial externalities (Irwin and Bockstael 1998).

Just like duration, distance is also a nonnegative random variable. This property has more recently been used to apply the mathematical framework of longitudinal models to spatial patterns, using distance as the endogenous random variable of interest (e.g., Odland and Ellis 1992; Esparza and Krmenec 1996). Applying the mathematical framework of hazard models to distance led to the term *spatial duration models*.

The analogy between duration and distance (spatial duration) is straightforward and justifiable from a mathematical point of view, and issues of estimation, specification, tests, and diagnostics can easily be addressed using the extensive methodological toolbox provided for the analysis of temporal data. It is this analogy that has prompted social scientists to herald spatial duration models as a useful complement (Odland and Ellis 1992) to traditional methods (e.g., Diggle 1983; Boots and Getis 1988) for the analysis of spatial point patterns. However, the analogy between duration and distance raises questions from a conceptual perspective and leads to difficulties in the interpretation of spatial duration models that have not yet been addressed in this emerging field. At the same time, however, regional scientists have not fully taken advantage of the richness of the longitudinal framework, confining their analyses to the description and comparison of spatial point patterns and spatial linkages rather than using the framework for understanding spatial processes.

This article, therefore, critically discusses the conceptual equivalence between spatial duration models and their temporal counterparts by emphasizing the need to focus on spatial processes. It further evaluates the use and interpretation of spatial duration models in the literature and explores avenues of extension. In particular, the article augments the literature on spatial duration models by considering two specific topics. The first issue deals with the transfer of the “at-risk” concept, which is central to longitudinal models but has not been properly considered in the application of hazard models in spatial settings. The second issue deals with the question of how spatial duration models can be used to describe the trajectory of point-generating processes leading up to spatial point patterns. This process-oriented perspective is further augmented by explicitly accounting for temporal dimensions of point generation processes.

The article is organized in four sections. Following this introduction, the second section presents a brief overview of the mathematical foundations of longitudinal models.[1] The third section deals with four issues relevant to the transfer of longitudinal models to a spatial setting. Section 3.1 sets the stage by arguing that a conceptually sound transfer of longitudinal models to a spatial setting exists for spatial phenomena that spread continuously through space and for which, at any distance from an origin, the risk of a transfer from one state to another can be defined. Section 3.2 reviews and evaluates the existing applications of spatial hazard models against the requirement of an underlying continuous spatial process with clearly identifiable states. In section 3.3, simulations of point-generating processes are used to demonstrate how, through the shift in perspective from patterns to processes, spatial hazard models can be fruitfully applied to spatial point pattern. Section 3.4 further elaborates on the process-oriented perspective by explicitly accounting for the timing of point generations and thus the speed of the process. The article concludes with a summary in section 4.

2. Mathematical Foundation of Longitudinal Models

The classic example of longitudinal models refers to the length of time a person is alive, often referred to as the survival time. Implicit in this example are two states, “alive” and “dead,” and of interest is the duration of the life span (i.e., when death occurs or, more generally, when the switch between the two states occurs). Formally, the example includes the following components: two states, s_1 and s_2, recorded for a person or object, and the time T that elapses before the switch from s_1 to s_2 occurs. As shown in Figure 1, the at-risk period for the switch from s_1 to s_2 begins at time $T = 0$. At any point in time during the at-risk period, the person or object may experience the transition to state s_2, and the transition probabilities between states only depend on time and the current state (Markov assumption).[2] The length of time spent in state s_1 or the timing of the terminating event is not constant but varies across objects or persons. Thus, when connected with a random selection process of objects or persons, T becomes a nonnegative random variable for which outcomes are not a priori known.

The random variable T can be captured in a variety of well-known ways. If time is measured on a continuous scale, then the probability density function of T is a nonnegative function $f(x)$ with

$$f : x \rightarrow f(x) \text{ such that } P(T \in (a, b)) = \int_a^b f(x)dx. \tag{1}$$

The distribution function, $F(x)$, specifies the probability that the survival time is less than a value x. It is defined as

$$F(x) = P(T < x) = \int_0^x f(t)dt. \tag{2}$$

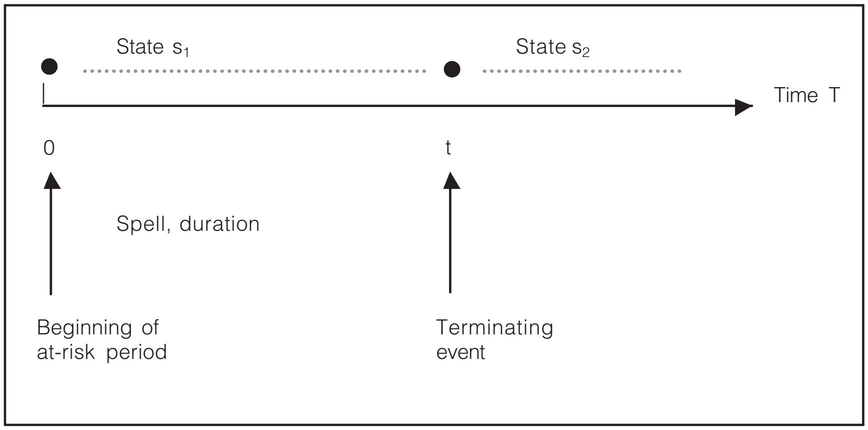

FIGURE 1. Basic Elements of a Longitudinal Design

In the longitudinal framework, two other frequently used specifications describing the distribution of the random variable T are of importance—namely, the survivor function, $S(x)$, and the hazard rate, $h(x)$. The survivor function describes the probability that the terminating event will not occur prior to x. Thus, it takes on the form

$$S : x \to S(x) = P(T \geq x) = \int_x^\infty f(t)dt. \tag{3}$$

The hazard function, $h(x)$, is defined as

$$h{:}x \to h(x) = \lim_{\delta \to 0} \frac{P(T \in [x, x + \delta] | T \geq x)}{\delta} \in (0, \infty), \tag{4}$$

and describes the exit rate or instantaneous rate of the transition occurring during $[x, x + \delta]$, given that it has not happened up to (and including) time x. For $d\,h(x)/dx > (<)$ 0, the hazard increases (or decreases). The increasing (or decreasing) hazard is also referred to as positive (or negative) duration dependence and implies that the conditional probability that the spell will be terminated increases (or decreases) with the increasing duration of the spell.

The focus on the conditional probabilities and hazard functions distinguishes longitudinal models from the conventional regression models that, in contrast, rely on estimating the (unconditional) probability density functions, $f(x)$. Although the probability density function and hazard function are mathematically equivalent, estimating hazard functions has two major advantages. First, information on incomplete spells (the spell lasts for at least time t, but the exact duration is not known) does not need to be discarded in the estimation. Second, changes in exogenous variables that occur during the spell can be taken into account.

The basic framework has been successfully applied in a variety of contexts and with a variety of estimation approaches for the hazard functions. Three types of estimation can be distinguished: nonparametric, parametric, and semiparametric approaches. The nonparametric approach is based on actuarial methods whereby the hazard for each discrete time interval is obtained by relating the number of failures to the duration-dependent at-risk set (those who have not yet failed at duration *t*) and by properly accounting for censored observations.

The parametric formulations use a number of well-known (and well-behaved) distributions (Waldorf and Esparza 1991), such as the exponential, Weibull, log-logistic, uniform, and Gamma distribution for which the parameters are estimated via maximum likelihood techniques. The Weibull distribution with $h(x = \lambda\beta(\lambda x)^{\beta-1}$ takes on a pivotal role as its parameters induce a flexible shape of the hazard function. For $\beta = 1$, the Weibull distribution reduces to the exponential distribution, and its hazard is thus constant (i.e., the exit rates are the same no matter how long the spell has already lasted); for $\beta >$ (or $<$) 1, the exit rates monotonically increase (or decrease) with increasing duration of the spell.

In the social sciences, the effects of exogenous variables X on the hazard are most often accommodated in a semiparametric model such as the frequently used proportional hazard model. The proportional hazard model assumes that the hazard function is separable into two factors: a baseline hazard, $h_o(t)$, and a function $\Phi(X, \beta)$ that is independent of the duration of the spell. Typically, the function Φ is specified as an exponential function of a linear predictor such that

$$h(t \mid X) = h(t \mid X = 0) \exp \sum_{k=1}^{n} \beta_k X_k(t), \tag{5}$$

where $h(t \mid X = 0) = h_o(t)$ is the baseline hazard, and X_k are exogenous variables with associated proportional and duration-independent effects β_k on the conditional probability of terminating the spell. The baseline hazard may remain unspecified, and the model is then estimated via a partial log-likelihood function (Cox 1972, 1975). Alternatively, the baseline hazard may refer to a particular distribution of T (e.g., a Weibull distribution), and estimation of the parameters involves maximum likelihood procedures. Three types of covariates can be incorporated: attributes that do not change over time (e.g., race), time-dependent[3] variables that are measured for an individual but that may change during a spell (e.g., marital status), and time-varying covariates that constitute explicit functions of time, $x(t)$.

Two frequently employed extensions of the longitudinal framework are competing-risks models (e.g., Hachen 1988; Han and Hausman 1990; Narendranathan and Stewart 1993; Thomas 1996) and multiepisode models (e.g., Blossfeld and Hamerle 1989; Popkowski Leszczyc and Timmermans 1996). The competing-risks model is appropriate if a spell can be terminated in two or more ways. Multiepisode models are appropriate if more than one spell is observed for each object or individual in the sample.

3. Applications of the Longitudinal Framework in a Spatial Setting

The above section shows that longitudinal modeling has become a rich methodological toolkit that focuses on the conditional probabilities/hazards for the nonnegative random variable *duration*. Over the past decade, regional scientists have adopted the longitudinal framework in the context of spatial analysis, describing spatial hazards rather than temporal hazards. Common to all these studies is that they use distance[4] between points—a one-dimensional projection of two-dimensional space—to serve as the mathematical equivalent of duration. The emerging models are labeled spatial duration models or spatial hazard models. Because distance, just like time, can be viewed as a nonnegative random variable, the entire mathematical apparatus developed for longitudinal models can also be applied to spatial duration models (see Figure 2). It should be noted that this differs from the transfer typically employed in spatial econometrics. That is, whereas spatial econometrics is concerned about dealing with spatial interdependencies that arise when using spatial rather than temporal units of observation, the focus here is on using a spatial rather than a temporal dependent variable.

3.1 Transfer to the Analysis of Continuous Spatial Processes

The longitudinal framework focuses on modeling the hazard of an event happening in $[t, t + \delta]$, given that it has not yet happened at time t. For example, a hazard model shifts the focus from the probability of a person finding employment at time t to the conditional probability of finding employment during $[t, t + \delta]$, given that the job search was unsuccessful up to time t. This shift in emphasis is permissible because every observation inevitably moves through the medium time; that is, a person or object in state s_1 at time $T = 0$ experiences the risk of switching to s_2 at any subsequent time $T > 0$.

Indicative for the methodological focus on conditional probabilities is the data requirement: the data are (ideally) extracted from a panel whereby randomly selected individuals are observed over time, their status (e.g., unemployed, not unemployed) is recorded through time, and the switch in status at any given time follows Markovian transition probabilities. The transfer of hazard models to a spatial setting thus requires *spatial* panel data that refer to a continuous spatial process in which, for every distance from the source, the Markovian transition probability can be defined and for which, consequently, the concept of a distance hazard has a natural interpretation.

Spatial panel data can be collected for spatial processes characterized by a continuous expansion from a fixed source (the spatial analog of the beginning of the "at-risk period"; see Figure 1).[5] At distance $D = 0$, a state s_1 is observed, and at some distance away, a switch occurs to a state s_2 (the spatial analog of the terminating

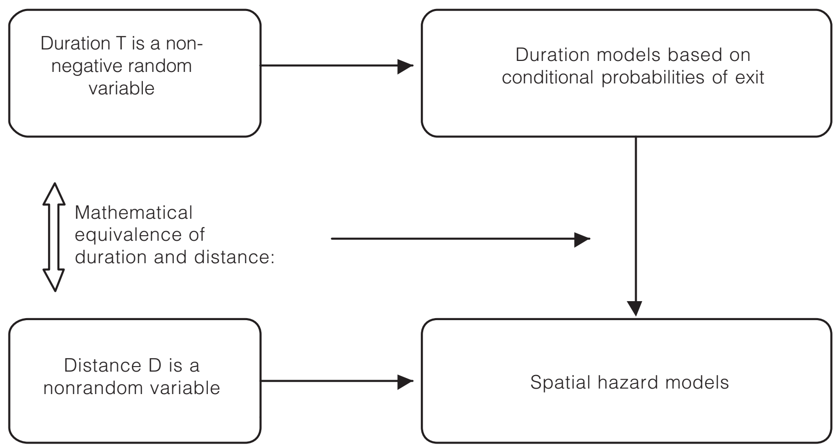

FIGURE 2. Transfer of Hazard Models from a Temporal to a Spatial Setting

event). The distance $D = d$ at which the switch occurs (i.e., the terminating event) becomes the endogenous variable of interest.

Possible applications of casting spatial expansion processes in spatial hazard models are numerous. In the environmental sciences, applications may emerge for analyzing the spread of a forest fire or the diffusion of a pollutant from a single source. For example, for the spread of a forest fire, it is meaningful to determine the probability that, along a linear transect, the fire will cease at distance $[d, d + \delta]$ from its origin, given that it already approached to distance d. Similarly, for the diffusion of a pollutant from a single source, one can specify the distance-dependent hazard of the pollutant concentration falling below a certain threshold.

Applications may also arise in transportation and urban analyses. For example, in congestion studies, one may be interested in specifying the hazard of a traffic jam dissolving at distance (length) $[d, d + \delta]$, given that it extends at least until distance d. Urbanization or urban sprawl is another example of a continuous expansion in space, and the data collection can indeed be thought of as a "spatial panel" data collection. That is, for a randomly selected city and a randomly selected transect originating at the city center, two states are recorded along the transect: s_1 = urban land use, and s_2 = nonurban land use. Analyzing the distance at which the switch from s_1 to s_2 occurs is not only mathematically but also conceptually equivalent to analyzing the time at which a switch from one state to another occurs. Thus, the hazard of urban land use ending at distance $[d, d + \delta]$, given that it persists at least until distance d, has a meaningful interpretation.

Of particular importance for the social sciences is that the applications discussed above can be cast in the form of a proportional hazard model so as to assess the role of exogenous factors that may increase or decrease a baseline hazard. Such factors may be, for example, wind speed, strength, and direction as factors influencing the distance-dependent hazards in the pollution and fire examples, time of the day and traffic volume in the congestion example, and age of the city and zoning regulations in the case of urban sprawl.

Interestingly, the literature on spatial hazard models does not include an application to a continuous spatial expansion process as exemplified above. Instead, the existing applications of spatial hazard models can be split into two types: those dealing with linkages between pairs of points and those dealing with point patterns. Neither type uses spatial panel data, and an underlying process with distance-dependent risks of switching between states is only vaguely defined.

3.2 Spatial Hazard Models for Point Linkages and Point Patterns

The study by Rogerson, Weng, and Lin (1993) uses spatial hazard models to describe linkages between pairs of points. It analyzes the spatial separation between locations of parents and locations of their adult children by fitting Weibull distributions to the observed distances. Similarly, Esparza and Krmenec (1994, 1996) analyzed the spatial extent of producer service markets by fitting Weibull distributions to the observed distances between producer services providers and their clients.[6]

Both applications deal with pairs of points in space that are intricately linked by relocation and trade, respectively. In both cases, the spatial hazard models are solely used for the purpose of descriptive curve fitting rather than explanation. From a data perspective, both situations require knowledge about the spatial source (beginning event, e.g., location of firm) and the terminating event (e.g., a client is found). Nevertheless, when interpreting the estimated hazards, conceptually one has to assume an underlying spatial process that starts at the source and spreads continuously through space until the terminating event occurs.

Unfortunately, it will be impossible to collect the actual spatial panel data that could support the existence of such an underlying spatial process. Assuming the existence of such a process is also debatable from a conceptual point of view. For example, suppose that a spatial hazard model is used to analyze distances between households' old and new residences. Assigning a meaningful interpretation to the hazard rate of the new residence being located at distance $[d, d + \delta]$ from the old residence, given that it is at least d units away, presupposes that the underlying residential search proceeds from the old residence in concentric circles of monotonically increasing radii. Such an interpretation is, however, not consistent with models of residential search behavior (see, e.g., Huff 1986) that suggest that people's search efforts can "jump" between noncontinuous areas and also does not necessarily conform to an expansion process of increasing search distances.

The literature provides several examples of the second type of application, that is, the use of spatial hazard models to describe point patterns. Odland and Ellis (1992) used a proportional hazard model with spatially varying covariates to analyze directional variations in the spacing of settlement locations in Nebraska. The variable of interest in their study is the nearest neighbor distance between settlements, and it is found that the distance hazards increase as one moves from East to West. The Odland and Ellis study is of particular importance because it can be credited as the first attempt to use hazard models to study spatial patterns. Moreover, it is the first (and only) study that deals with the identifiability of the "end point,"[7] a problem that results from the projection of two-dimensional *xy*-space to one-dimensional distance space. The nonidentifiability of the end point implies that distance hazards actually refer to an infinite (and uncountable) set of points arranged in a circle around the observation rather than just a specific point in space.

Nearest neighbor distances are also used in the Harvestore diffusion analysis by Pellegrini and Reader (1996). In addition to a proportional hazard model, the authors also estimated fully parameterized models in which the baseline hazard is specified using various distributions, with the Weibull distribution providing the best fit. From a methodological point of view, their study contributes to the literature by demonstrating how censoring can be used to deal with edge effects. Edge effects occur if a point is closer to the edge of the study area than to any other point of the point pattern. Hazard models are particularly apt at dealing with such edge effects by recording the nearest neighbor distance as censored and as being at least as long as the distance to the edge.

A purely descriptive study of a point pattern is provided by Reader (1998). His study tackles the random labeling hypothesis and uses interevent distances (between-point distances) rather than nearest neighbor distances. He argued that the emerging nonparametric survival functions provide a useful complement to *K*-function analyses of comparing spatial point patterns. It should be noted that, by using interevent distances rather than the subset of first-order nearest neighbor distances, his models are equivalent to multiepisode models and thus implicitly assume independence of distances in a spatial system with complex dependencies.

Finally, the study by Pellegrini and Grant (1999) demonstrates that spatial hazard models can be used not only for distances in physical space but also for distances in more abstractly defined spaces. To account for variations in policy coalitions in U.S. Congress, they estimate hazard models for nearest neighbor distances and for extreme-value distances in ideological space.

These studies demonstrate that for descriptive and comparative purposes, distance data extracted from a spatial point pattern can be adequately and fruitfully analyzed in a spatial hazard framework. In particular, through its ability to deal with censored data, the hazard framework can efficiently handle edge effects of point patterns in a bounded area. Moreover, it offers a convenient means of estimating the parameters for a variety of distributions and thus can distinguish between spatial point patterns with positive and negative distance dependence of the spatial hazard.[8]

However, the key problem of using hazard models for spatial point patterns is that it is not clear what the underlying processes are and how to clearly define *states*. As a result, it becomes difficult to conceptualize distance-dependent Markovian transition probabilities. These difficulties are somewhat lessened when thinking about spatial point patterns as the outcomes of point-generating processes. In a point-generating process, a new point can emerge at any distance from the already existing points so that, conceptually, one can think of distance-dependent Markovian transition probabilities for the switch between the two states of “no point generated” and “point generated.” The switch in perspective from patterns to processes also opens up new avenues for the application of spatial hazard models. In particular, the distances used as input into the model depend on the specific trajectory of the point-generating process. This issue is explored in the following section. Moreover, the focus on point-generating processes also allows us to account for the speed of the process and thus add *time* as an exogenous variable affecting the hazard of point generation. This issue is explored in section 3.4.

3.3 *Spatial Hazard Models of Point-Generating Processes*

Spatial point patterns do not emerge instantaneously but evolve over time. This switch in perspective from patterns to processes is of particular importance for developing a sound conceptual underpinning of a spatial hazard model.

Imagine, for example, a point pattern emerging with the location of the initial point $j = 0$. For this initial point, no distance to any other point can be measured (thus, a distance-dependent hazard does not exist). For the next point, $j = 1$, every possible location is at risk of receiving the point, and one can conceptualize the risk of receiving a point as a function of distance to the already existing point. In fact, the location of any subsequent point j, $j > 0$, can be evaluated relative to the distances to the already existing points, with the minimum distance taking on a pivotal role for the assessment of a distance-dependent hazard of receiving a point. For example, in a highly contagious process, the hazard of the jth point being generated in close proximity to any of the earlier generated points k (= 0, . . ., j – 1) will be high and decline with increasing distance.

Let j denote the jth generated point following the initial seed point $j = 0$, and d_j is the minimum distance between point j and point k, $0 \leq k < j$. Furthermore, let d_j^* denote point j’s nearest neighbor distance in the completed point pattern. Then, $d_j^* \leq d_j$. As a consequence, an analysis of nearest neighbor distances derived from the completed pattern will yield different results than those derived from the minimum distance d_j, measured as the point j is generated. Specifically, the hazards estimated from a completed pattern are inflated relative to the distance hazards derived from the underlying point-generating process.

The following example simulates point-generating processes for two types of spatial patterns with $n = 41$ points each. The first pattern gives rise to a positive distance-dependent hazard; the second gives rise to a negative distance-dependent hazard. For both patterns, five point-generating processes are simulated. The first simulated process, labeled *expansion*, generates points in such a way that the point

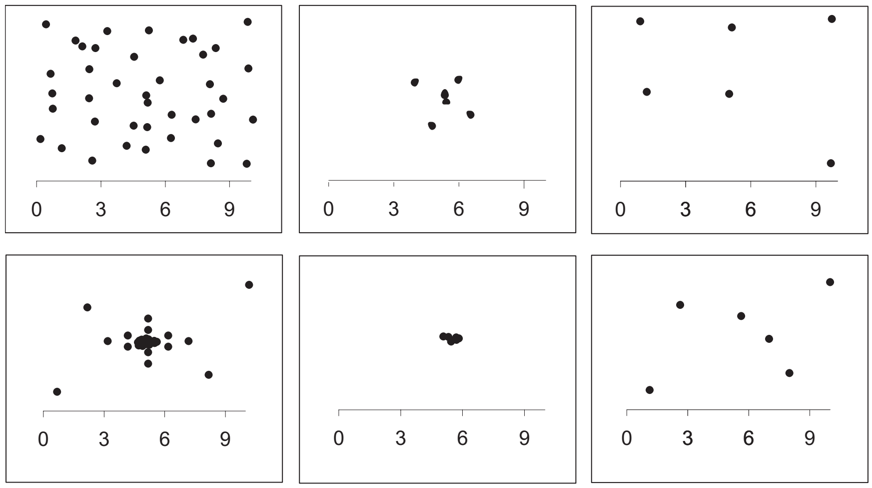

FIGURE 3. Complete Patterns (left) and Expansion (middle) and Infill (right) Point Generation Processes at *j* = 5 for Positive (top) and Negative (bottom) Dependence in the Complete Pattern

selected in the jth step produces a smaller distance d_j than any other not yet selected point. This process will give rise to an expanding pattern that covers a wider area with every step. The second process, labeled *infill*, simulates the opposite extreme; that is, following the initial seed point, it generates points in such a way that the point selected in the jth step produces a greater distance d_j than any other not yet selected point. Whereas the expansion and infill scenarios constitute the extremes at either end of the possible distances, the remaining three simulations take on middle positions, with the points being generated in random order. Figure 3 shows the completed patterns as well as the simulated processes at $j = 5$ for the expansion and the infill processes.

Since the trajectories of point generation are different in each scenario, the measured minimum distances d_j will differ as well, with the expansion scenario yielding the smallest distances and the infill scenario yielding the largest distances. The three random point-generating processes take on middle positions. Consequently, the hazards will differ across the scenarios. Furthermore, since $d_j^* \le d_j$, the hazard function for each scenario will also differ from that of the nearest neighbor distances in the completed pattern. To allow for a flexible shape of the hazard function, Weibull distributions are fitted to the observed distances for each simulated point generation process, as well as for the nearest neighbor distances in the complete pattern. The estimation results[9] for both patterns are summarized in Table 1.

TABLE 1. Parameter Estimates of Fitted Weibull Distributions

	Complete Pattern	*Point-Generating Processes*				
		Expansion	*Infill*	*Random 1*	*Random 2*	*Random 3*
Pattern 1: Positive distance-dependent hazard						
λ	0.731	0.612	0.415	0.493	0.501	0.489
	(0.045)	(0.033)	(0.061)	(0.056)	(0.055)	(0.060)
β	2.702	3.311	1.401	1.807	1.821	1.674
	(0.379)	(0.423)	(0.256)	(0.264)	(0.265)	(0.288)
Median distance	1.19	1.46	1.85	1.66	1.63	1.64
n	41	40	40	40	40	40
Log-likelihood	–24.289	–16.981	–46.637	–36.331	–36.303	–39.531
Pattern 2: Negative distance-dependent hazard						
λ	2.262	1.840	1.344	1.787	1.691	1.695
	(0.738)	(0.542)	(0.467)	(0.544)	(0.498)	(0.489)
β	0.627	0.695	0.587	0.662	0.675	0.691
	(0.112)	(0.117)	(0.103)	(0.109)	(0.123)	(0.122)
Median distance	0.25	0.32	0.40	0.32	0.34	0.35
n	41	40	40	40	40	40
Log-likelihood	–80.611	–74.598	–81.258	–76.580	–76.038	–74.991

Note: Standard errors are in parentheses.

The upper panel of Table 1 shows the results for the first (complete) pattern with positive distance dependence and for the simulated processes giving rise to the pattern. As expected, the estimated median distance is smallest for the completed pattern, and those of the random processes are sandwiched between the median distances of the infill and the expansion processes. The shape parameter β is estimated to be greater than 1 for all simulated point generation processes as well as the completed pattern, suggesting that the hazards increase with increasing distance. However, the magnitudes of the estimates vary substantially across the scenarios. As expected, the shape parameter is high if the point-generating process proceeds in an expanding fashion since such a process favors short distances over long distances. However, for the infill process, the estimated shape parameter is not significantly ($\alpha = 0.05$) greater than 1, thus not allowing us to reject the null hypothesis of a constant hazard. This suggests that drawing conclusions from an analysis of the completed pattern cannot be transferred to the underlying point generation processes. In fact, as Figure 4 shows, the hazards estimated on the basis of the nearest neighbor distances in the completed pattern are inflated relative to the estimated hazards derived from the minimum distances in any of the simulated point generation processes, with the most drastic inflation occurring relative to the infill process.

The lower half of Table 1 shows the estimation results for the second pattern and for the simulated point-generating processes. As was the case for the first pattern,

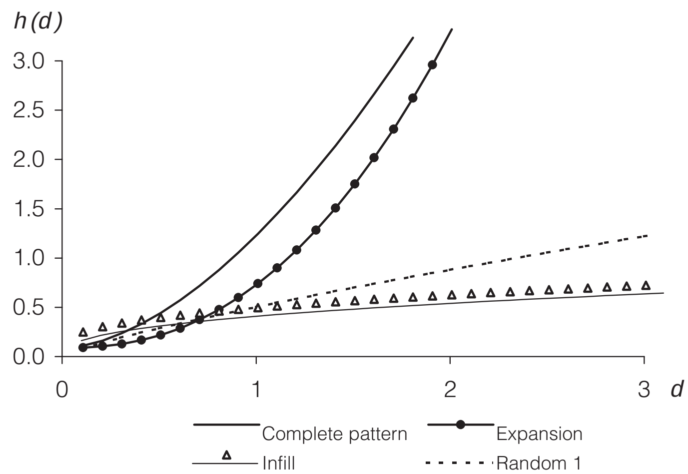

FIGURE 4. Estimated Hazard Functions for the Complete Pattern with Positive Distance Dependence and for Simulated Point Generation Processes

the estimated median distance of the completed pattern is smaller than for any of the point generation processes, thereby signaling that the analysis of the completed pattern is not representative of any process that may have created the completed pattern. However, the hazard model for the completed pattern is similar to the hazard models for the five point generation scenarios in that all shape parameters β are significantly smaller than 1, so that each point-generating process replicates the negative distance dependence of the hazards observed for the completed pattern. As shown in Figure 5, the estimated hazard functions are also substantially closer to each other than was the case for the first pattern. Nevertheless, it is the infill process that once again shows the greatest discrepancy from the hazard function of the completed pattern.

3.4 Temporal Dimension of Point-Generating Processes

The Odland and Ellis (1992) study suggests an east-west trend with a tendency toward regular spacing of settlements. Given that the settlement process in the midwestern and prairie states of the United States by and large advanced from east to west, this result can also be interpreted as a time trend, with younger settlements being further apart than older settlements. As such, the results allude to the necessity of considering the time when the points are generated.[10]

The process-oriented spatial hazard models described in the previous section are based on the minimum distance at the time of a point's emergence and thus account for the sequential nature of the point generation process. However, the time factor and thus the speed of the process are not explicitly taken into account.

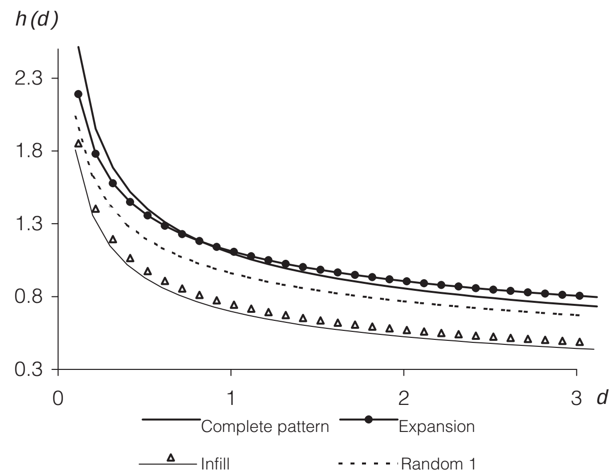

FIGURE 5. Estimated Hazard Functions for the Complete Pattern with Negative Distance Dependence and for Simulated Point Generation Processes

Accommodating time directly into a spatial hazard model of a point generation process can be accomplished with proportional hazard models. That is, the time *T* at which a point is generated becomes an exogenous variable and can be entered into the linear predictor of partially or fully parameterized hazard models:

$$h(d \mid z) = h_o(d)\exp(z\beta), \tag{6}$$

where *d* measures the minimum distance at the time of a point's emergence, and the vector of exogenous variables *z* includes the time of a point's emergence.[11]

The influence of time is simulated using the pattern with positive distance dependence (see Figure 3) and the five point-generating processes described in section 3.4. The hazard rates for each process are estimated under varying assumptions about the speed of the process. The constant-speed scenario assumes that the *j*th point is generated at time $T(j) = j$. In the scenario representing an accelerated process, $T(j) = \sqrt{j}$. Finally, assuming that $T(j) = j^2$ represents a decelerating process. All estimations are performed using a Weibull distribution for the baseline hazard. Table 2 shows the estimation results.

The upper panel of Table 2 shows the results for the constant-speed scenario. Compared to the models that only account for the sequential nature of the point-generating process (see Table 1), including time as an exogenous variable increases the shape parameter β. In fact, when accounting for time as an exogenous variable, all five point-generating processes have a significantly positive distance dependence of the hazard. This is also the case for the infill trajectory that, without accounting for the effect of time, is estimated to have a constant hazard (see Table 1).

TABLE 2. Parameter Estimates of $h(d) = \lambda\beta(\lambda x)^{\beta-1}(\exp\beta_0 + \beta_1 T)$ for Different Point Generation Processes and Different Speeds

	Expansion	*Infill*	*Random 1*	*Random 2*	*Random 3*
Constant speed					
λ	0.614	0.487	0.522	0.517	0.524
	(0.033)	(0.027)	(0.034)	(0.041)	(0.038)
β	3.363	2.966	2.738	2.384	2.346
	(0.420)	(0.405)	(0.378)	(0.365)	(0.343)
β_o	–3.959	–1.688	–1.284	–1.139	–1.292
	(0.102)	(0.991)	(0.098)	(0.112)	(0.128)
β_1	–0.0045	0.047	0.031	0.023	0.032
	(0.0048)	(0.0038)	(0.004)	(0.006)	(0.005)
Log-likelihood	–16.327	–20.089	–23.770	–28.731	–27.588
Acceleration					
λ	0.613	0.500	0.525	0.523	0.529
	(0.033)	(0.027)	(0.033)	(0.037)	(0.038)
β	3.367	3.594	2.894	2.524	2.449
	(0.422)	(0.443)	(0.404)	(0.366)	(0.326)
β_o	–0.322	–2.310	–1.692	–1.487	–1.725
	(0.170)	(0.170)	(0.175)	(0.179)	(0.202)
β_1	–0.039	0.377	0.244	0.196	0.253
	(0.040)	(0.034)	(0.040)	(0.043)	(0.044)
Log-likelihood	–16.299	–13.754	–22.214	–26.753	–25.888
Deceleration					
λ	0.614	0.465	0.516	0.511	0.515
	(0.034)	(0.036)	(0.038)	(0.046)	(0.042)
β	3.369	2.236	2.447	2.176	2.150
	(0.420)	(0.358)	(0.328)	(0.329)	(0.384)
β_o	–0.426	–1.358	–1.056	–0.943	–1.058
	(0.072)	(0.094)	(0.081)	(0.096)	(0.105)
β_1	–0.0001	0.001	0.0007	0.0005	0.0007
	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0001)
Log-likelihood	–16.208	–29.747	–26.783	–31.507	–30.726

Note: Standard errors are in parentheses.

This monotonic behavior of the hazard is, however, disturbed by the time covariate: a positive[12] time parameter β_1 suggests that, for points generated late in the process, the estimated hazard exceeds the baseline hazard, whereas a negative parameter suggests that the hazard decreases over time.

Interestingly, the effect of time differs across the point generation processes. For the expansion process, the effect of time is estimated to be negative: ceteris paribus, the hazard rates at a later stage of the process are smaller than those during the early stages (see Figure 6). However, the effect is not significant. In contrast, for all other processes, time has a highly significant positive influence on the hazard rates and is strongest for the infill process. Thus, during the early stages of

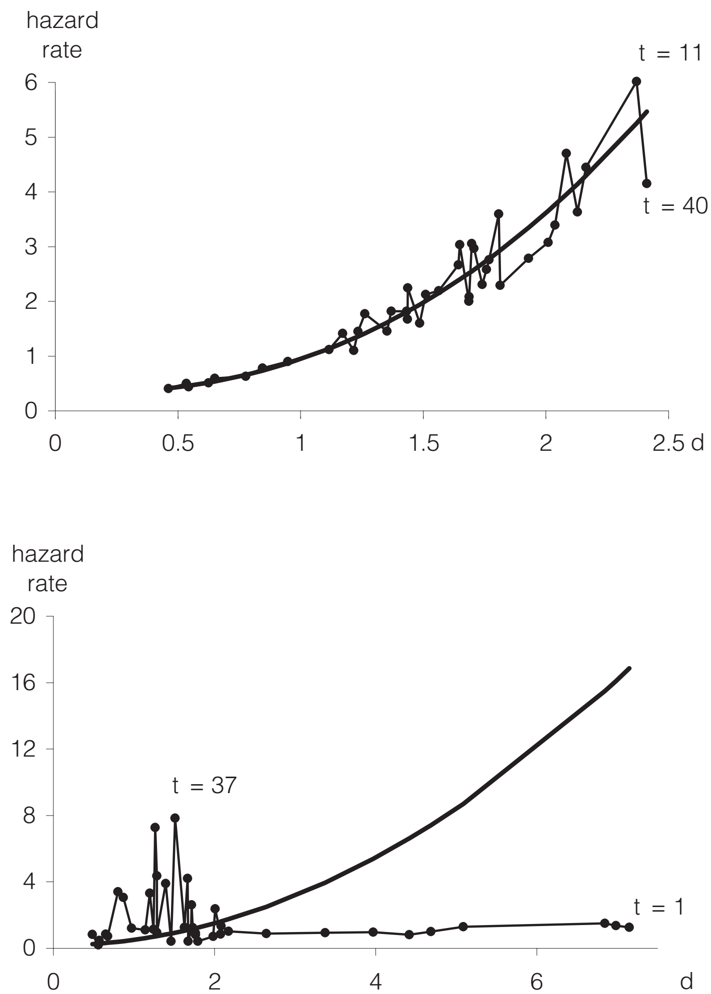

FIGURE 6. Time-Induced Deviations (dotted line) from the Baseline Hazard (straight line) for Expansion (top) and Infill (bottom) Point Generation Processes

the process, the hazards are lower than what can be predicted from the baseline hazard, whereas in the later stages, the hazards are higher than the baseline hazard suggests (see Figure 6).

Keeping the trajectory of point generation constant but altering the speed of the process leads to changes in the magnitudes of the deviations from the baseline hazard. But altering the speed does not substantially affect the baseline hazard or the overall shape of the estimated hazard function. Accelerating the speed of point generation during the process implies that the (absolute) magnitude of the time parameter increases, thereby exaggerating the deviations from the baseline hazard. However, in comparison to the constant-speed scenario, the direction of the effect does not change, with time having a negative but insignificant influence on the hazard rate in the expansion scenario and a significantly positive effect for all other scenarios.

If the point generation proceeds at a decelerating speed, the (absolute) magnitude of the time parameter becomes smaller, and thus time loses its influence on the hazard rate. In fact, as shown in the bottom panel, the time parameters for all scenarios are very close to zero. Nevertheless, the direction of the influence remains the same as for the constant and accelerating speed scenarios. That is, for the expansion process, the time parameter is estimated to be negative but insignificant, whereas the effect of time on the distance hazard is significantly positive for the infill and random point generation processes.

4. CONCLUSIONS

Over the past decade, the transfer of longitudinal models to spatial settings has entered the literature in the form of spatial hazard models. At the core of this transfer is a switch from “duration” to “distance” as the endogenous variable. The mathematical equivalence of duration and distance—both are nonnegative continuous variables—has been used as justification to apply the rich methodological toolkit of longitudinal models to the analysis of distances in space.

This article focuses on the conceptual equivalence between hazard models applied to both duration data and distance data. It elaborates on three main arguments. First, the article argues that a conceptually sound and straightforward transfer of hazard models to a spatial (distance) setting is applicable to spatial processes that spread continuously through space. In those cases, it is possible to define the state of the process for any given distance from an origin, conceptualize the risk of switching from one state to another, and define distance-dependent Markovian transition probabilities. A discussion of possible applications in ecology, transport, and urban analyses further exemplifies the suitability of spatial hazards models for continuous spatial processes.

Second, it is argued that the existing applications of spatial hazard models (i.e., models of distances between pairs of points and distances in point patterns) respond to the mathematical equivalence of duration and distance but ignore the conceptual issue of properly defining “states” and probabilities of switching between states. As a result, the usefulness of spatial hazard models in the context of point linkages and point patterns is limited to describing and comparing patterns and dealing with technical problems such as edge effects.

Third, spatial hazard models do, however, have a conceptually sound basis when applied to point-generating processes rather than completed point patterns. The process-oriented perspective allows a conceptual link to be established between the point generation process and a distance-dependent risk of receiving a new point. A series of simulated point generation processes demonstrates that the specific trajectory strongly influences the parameters of the fitted hazard models. Moreover, the parameters also differ from those derived from hazard models of nearest neighbor distance in the completed pattern. The simulations of the point generation processes are further augmented by explicitly accounting for the temporal dimension or process speed. Toward that end, the time at which a point is generated is specified as an exogenous variable in a proportional hazard model. Three scenarios with differing speed assumptions and applied to various spatial trajectories of point generation demonstrate that the time effect can substantially distort the regularity of distance-dependent hazards. This result further underscores the need to shift the perspective from static patterns to dynamic processes when analyzing spatial phenomena in a hazard framework.

Notes

1. For a detailed (and very good) review, the reader is referred to Kiefer (1988).

2. The Markov formulation is equivalent to the survival model discussed below and can be extended to more than two states (Sherris 2002; Hosgood 2002).

3. The terminology employed in the literature is inconsistent. For example, Greene (1998, 2000) used the terms *time-dependent covariate* and *time-varying covariate*, which are opposite to the definitions provided in this article.

4. The longitudinal framework used to model distance can also be employed for other nonnegative random variables describing space—most notably, area. In many applied research questions, such as those dealing with economic trade, residential search, or ecology, the focus on area promises to provide an insightful and as yet unexplored avenue.

5. In his comments on an earlier version of this article, James LeSage (personal communication, 2002) pointed out that the hazard function can be combined with a spatial weight matrix such that the hazard function depends on the average distance from all points rather than a single point in space (the origin).

6. It should be noted that since a firm may have more than one client, the data structure gives rise to a multiepisode model. Moreover, the sample is based on randomly selected firms rather than randomly selected trade linkages, and thus the implicit assumption of independence of multiple trade linkages measured for one firm is quite strong. However, it should be noted that this dependency problem is not confined to spatial applications of hazard models.

7. Very elegantly, Odland and Ellis (1992) used one of the advantages of spatial duration models (i.e., the models' ability to incorporate spatially varying covariates). Specifically, they used all spatial coordinates along the shortest line separating two nearest neighbor settlements and incorporated the coordinates as spatially varying covariates (i.e., as a function of distance) in a proportional hazard model. The disadvantage of this more sophisticated approach is, of course, that the parameters of spatially varying covariates are not easily interpreted beyond sign and significance.

8. In the past, the shape of the hazard function has also been used to categorize spatial point patterns as clustered, random, or uniform. For example, Pellegrini and Reader (1996, 237) and Pellegrini and Grant (1999, 61) interpreted positive duration dependence of the hazard as an indication of clustering, and a constant hazard of the nearest neighbor distances has been associated with a random point pattern (Odland and Ellis 1992, 101; Pellegrini and Reader 1996, 224). However, it is important to emphasize that positive duration dependence can occur for the nearest neighbor distances of all three types of patterns, and thus the shape of the hazard function of nearest neighbor distances is not suitable for a

unique categorization of spatial point patterns. It is, however, possible to distinguish random from nonrandom point patterns using the hazard function for the squared nearest neighbor distances. That is, in a random pattern, the squared nearest neighbor distances, $u = d^2$, are exponentially distributed (Cliff and Ord 1981) and thus have a constant hazard.

9. LIMDEP 7.0 (Greene 1998) is used for all estimations.

10. The importance of time in the longitudinal analysis of a spatial point pattern was also mentioned, albeit not accounted for, in the diffusion study by Pellegrini and Reader (1996).

11. Note that this model specifies distance D as the endogenous variables and time T as the exogenous variable. It is, of course, possible to reverse the assignment and specify a model in which time (e.g., in the form of duration since the beginning of the process) is the endogenous variable of interest and the spatial dimension becomes the exogenous variable. When conceptualizing both the spatial and temporal dimensions as endogenous, a proper model needs to specify the joint probabilities $\Pr(D \in (a, b) \cap T \in (c, d))$ and the associated joint hazard function $h(d, t)$. However, joint hazard functions can only be handled under rather restrictive assumptions (Heckman and Honoré 1989).

12. The parameter estimates reported in this article refer to $h(d) = h_o(d)\exp(\beta_0 + \beta_1 d)$ rather than the more awkward specification $h(d) = h_o(d)\exp(-\beta_0 - \beta_1 d)$ used by Greene (1998).

REFERENCES

Blossfeld, H.-P., and A. Hamerle. 1989. Using Cox models to study multiepisode processes. *Sociological Methods & Research* 17: 432-48.

Boots, B. N., and A. Getis. 1988. *Point pattern analysis*. Newbury Park, CA: Sage.

Clark, W. A. V. 1992. Comparing cross-sectional and longitudinal analyses of residential mobility and migration. *Environment and Planning A* 24: 1291-1302.

Cliff, A. D., and J. K. Ord. 1981. *Spatial processes: Models and applications*. London: Pion.

Cox, D. R. 1972. Regression models and life tables. *Journal of the Royal Statistical Society, Series B* 34: 187-220.

———. 1975. Partial likelihood. *Biometrika* 62: 269-76.

Davies Withers, S. 1997. Methodological considerations in the analysis of residential mobility: A test of duration, state dependence, and associated events. *Geographical Analysis* 29: 354-72.

Diggle, P. J. 1983. *Statistical analysis of spatial point patterns*. New York: Academic Press.

Esparza, A., and A. Krmenec. 1994. Business services in the space economy: A model of spatial interaction. *Papers in Regional Science* 73: 55-72.

———. 1996. The spatial extent of producer service markets: Hierarchical models of interaction revisited. *Papers in Regional Science* 75: 375-95.

Glavac, S., and B. Waldorf. 1998. Segregation and residential mobility of Vietnamese immigrants in Brisbane, Australia. *The Professional Geographer* 50: 344-57.

Greene, W. H. 1998. *User's manual LIMDEP version 7.0*. Rev. ed. Plainview, NY: Econometric Software, Inc.

———. 2000. *Econometric analysis*. 4th ed. Upper Saddle River, NJ: Prentice Hall.

Hachen, D. S. 1988. The competing risks model: A method for analyzing processes with multiple types of events. *Sociological Methods & Research* 17: 21-54.

Han, A., and J. Hausman. 1990. Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics* 5: 1-28.

Heckman, J., and B. Honoré. 1989. The identifiability of the competing risks model. *Biometrika* 76: 325-30.

Heckman, J., and B. Singer. 1985. Social science duration analysis. In *Longitudinal analysis of labor market data*, edited by J. Heckman and B. Singer, 39-110. Cambridge, UK: Cambridge University Press.

Hosgood, G. 2002. Markov models to estimate and describe survival time and experience in cohorts with high euthanasia frequency. Ph.D. diss., Louisiana State University.

Huff, J. 1986. Geographic regularities in residential search behavior. *Annals of the Association of American Geographers* 76: 208-27.

Irwin, E. G., and N. E. Bockstael. 1998. Endogenous spatial externalities: Empirical evidence and implications for the evolution of exurban residential land use patterns. Paper presented at the North American Meetings, Regional Science Association International, November, Santa Fe, NM.

James, I. R. 1989. Unemployment duration: Modeling and estimation with particular reference to the Australian Longitudinal Survey. *Australian Journal of Statistics* 31A: 197-212.

Kiefer, N. M. 1988. Economic duration data and hazard functions. *Journal of Economic Literature* 26: 646-79.

Lawless, J. F. 1982. *Statistical models and methods for lifetime data*. New York: John Wiley.

Narendranathan, W., and M. B. Stewart. 1993. Modelling the probability of leaving unemployment: Competing risks models with flexible base-line hazards. *Applied Statistics* 42: 63-83.

Odland, J. 1997. Longitudinal approaches to analysing migration behaviour in the context of personal histories. In *Recent developments of spatial analysis*, edited by M. M. Fischer and A. Getis, 149-70. New York: Springer-Verlag.

Odland, J., and M. Ellis. 1992. Variations in the spatial patterns of settlement locations: An analysis based on proportional hazard models. *Geographical Analysis* 24: 97-109.

Pellegrini, P., and T. Grant. 1999. Policy coalition in the U.S. Congress: A spatial duration modeling approach. *Geographical Analysis* 31: 45-66.

Pellegrini, P., and S. Reader. 1996. Duration modeling of spatial point patterns. *Geographical Analysis* 28: 219-43.

Popkowski Leszczyc, P. T. L., and H. J. P. Timmermans. 1996. An unconditional competing risk hazard model of consumer store-choice dynamics. *Environment and Planning A* 28: 357-68.

Reader, S. 1998. Using survival analysis to study spatial point patterns in geographical epidemiology. Paper presented at the North American Meetings, Regional Science Association International, November, Santa Fe, NM.

Rogerson, P., R. Weng, and G. Lin. 1993. The spatial separation of parents and their adult children. *Annals of the Association of American Geographers* 83: 656-71.

Sherris, M. 2002. The two-state Markov model. Retrieved from http://actuarial.web.unsw.edu.au/courses/actl5104/lectures.html

Thomas, J. 1996. On the interpretation of covariate estimates in independent competing-risks models. *Bulletin of Economic Research* 48: 27-39.

Waldorf, B., and A. Esparza. 1991. A parametric failure time model of international return migration. *Papers in Regional Science* 70: 419-38.

Yamaguchi, K. 1991. *Event history analysis*. Newbury Park, CA: Sage.