9

Introduction to event history methods

DAVID STRANG

Event history analysis is a set of techniques for the study of the timing of various kinds of events. Events may take the form of either state transitions or event recurrences (Hannan 1989). State transitions are changes in a discrete variable, such as movement from married to unmarried, from colonial dependency to sovereignty, or from a command to a market economy. Event recurrences are distinct happenings such as the outbreak of civil or international wars. Most event history methods assume that the process operates in continuous time (i.e., that an event can occur at any time). Parallel methods exist for the discrete-time case, where events can only occur at particular points in time (see Allison 1982).

The aim of event history analysis is to describe and model the underlying stochastic process that generates events. This goal is usually translated into a regression-like examination of how explanatory variables accelerate or slow the rate at which the event occurs. These variables may characterize the environment or the case "at risk" and may be measured on a categorical or interval level.

The potential applications of event history analysis are very broad. Contemporary methods have roots in actuarial and medical settings (where "survival" analyses examine human mortality) and in industrial engineering (where "failure time" analyses examine product durability). Event history analysis has more recently found wide application in the social sciences, where most work has studied demographic and institutional shifts occurring to people and organizations: the classical trio of birth, death, and marriage – or merger.[1]

There is a growing empirical literature that uses event history methods to study public policy making and change in political structures. Within the United States, researchers have examined change in government structures (Knoke 1982; Tolbert and Zucker 1983), legislative decisions (Pavalko 1989; Berry and Berry 1990), and collective action (Olzak

1989). Cross-national comparisons include the study of shifts in regime type (Hannan and Carroll 1981), the passage of national legislation (Soysal and Strang 1989; Usui, Chapter 10, this volume), executive succession (Bienen and van den Walle 1989), and decolonization (Strang 1990). But the potential applications of event history analysis to macro-sociological and political concerns has just begun to be tapped.

METHODOLOGICAL BASICS

An *event history* is a record of the sequence and timing of events occurring to an individual case during some observation period. For example, an event history of the regime changes for a country would consist of a record of changes in political regime together with the dates of these changes.

Event history data could be analyzed with standard regression techniques – for example, by modeling the length of time between events. But this strategy presents two difficulties. First, some durations are incomplete or "censored"; observation periods may end with a case waiting to experience an event, and the analyst does not know if and when the event will occur. In fact, some cases may experience no events during the period under observation. Second, causal factors are often time varying; their values change over time. Standard regression methods can accommodate neither censoring nor time-varying explanatory factors. Event history methods are designed to do so.

The statistical theory underlying event history analysis characterizes the distribution of the random variable T, the time of the event. While most distributional theory works in terms of the cumulative density function and the probability density function, event history models are usually developed in terms of the *instantaneous transition rate*, or rate for short. The rate is defined as

$$r_{jk}(t) = \lim(dt \rightarrow 0)\ Pr(t,\ t + dt)/dt,$$

where $Pr(\cdot)$ is the limiting probability that a case moves from state j to state k (i.e., an event occurs) at t, given that the event does not occur before t. The rate is analogous to (and can be approximated by) the proportion of cases that experience an event over some interval of time, divided by the length of the interval.

Nonparametric methods permit the description of temporal patterns in the frequency of events. Standard techniques provide graphic displays of the survivor function, the integrated hazard, and the hazard. These techniques are valuable in helping the analyst assess how transition rates vary over time, which is crucial to the development of parametric models. They are also useful when the analyst wishes to compare

differences in event timing across groups – for example, to compare susceptibility of democratic or socialist regimes to coups d'état.

Parametric methods seek to model the underlying process as a function of covariates and time. It is common to specify the rate as an exponential function of covariates multiplied by some function of time:

$$r_{jk}(t) = \exp(BX_{it})q(t).$$

The log-linear form for the covariates is chosen to help ensure that predicted rates are nonnegative, as implied in the rate's definition. X is indexed by i to indicate heterogeneity by case and by t to make clear that the values of explanatory variables may change over time.

Parametric models are generally distinguished by different choices of $q(t)$. Adequate representation of time dependence in the rate helps the analyst obtain reasonable estimates of the effects of measured covariates. Time often serves as a proxy for basic causal factors that are difficult to measure – for example, the effects of age on human mortality. In addition, unobserved heterogeneity (i.e., variation in rates uncaptured by measured covariates) produces systematic forms of time dependence. Popular parametric approaches for modeling time dependence include Gompertz, Weibull, and log-logistic formulations. Cox models employ a partially parametric technique that controls for an unspecified $q(t)$.

Parametric models are generally estimated by the method of maximum likelihood. A model for the transition rate permits the analyst to write down the probability of sample observations. Maximum likelihood estimation seeks the values of parameters that maximize this (model-specific) joint probability. Maximum likelihood estimation also gives estimates of parameter standard errors and permits tests of improvement in goodness-of-fit (where more complex models are compared to simpler nested models). While maximum likelihood estimation relies on asymptotic theory, Monte Carlo studies have demonstrated a strong resolving power at the sample sizes that comparative researchers typically work with (Tuma and Hannan 1984).

DATA MANIPULATION

The data demands of event history analysis are on the order of those of dynamic analyses of continuous outcomes. The researcher needs data on both dependent and independent variables over time. These demands are of course much greater than those of cross-sectional analysis, where data at only one point in time are needed. Where many events occur to a single case (perhaps terrorist acts in Lebanon), the analysis may be limited to that case, much like time-series analysis. More commonly, researchers compare the timing of events across cases, making event

history analysis most comparable to multiwave panel (i.e., pooled cross-section and time-series) analysis.

From the user's point of view, the most confusing aspect of event history analysis often involves data organization. Of course, formatting schemes vary across estimation programs. But generally the basic set-up is to divide the time each case is at risk into intervals that can be unambiguously characterized in terms of the known timing of events and the values of explanatory variables. These time intervals, often called "spells," can be thought of as the units of analysis. A spell may terminate with an event, or not; but by definition an event cannot occur during a spell. For example, consider a country studied from 1965 to 1975 that experienced coups in 1967 and 1971. This case could be recorded in three spells: from 1965 to 1967 (ending in a coup), from 1967 to 1971 (ending in a coup), and from 1971 to 1975 (ending without a coup).

Explanatory covariates are attached to spells. For example, one might attach the (temporally appropriate) values of national income and regime type to the spells as defined. If the values of covariates are observed to change at times so they cannot be unambigously attached to spells, spells can be subdivided to permit synchronization. For example, with decennial data on national income we might divide the spell from 1967 to 1971 into spells from 1967 to 1970 (now censored, with the national income value for the 1960s) and one from 1970 to 1971 (ending in an event, with the national income value for the 1970s).

There is no standard way to describe the N involved in an event history analysis. The problem is similar to the difficulty of describing the N in a time-series analysis at t time points; whether one has an N of one, t, or something in between depends on how serious autocorrelated error is, and error is always unobserved. The amount of information in event history data depends on the number of events and on variability in explanatory factors across cases and within cases over time. As indices of the size of the problem, it is informative to report the number of observed events, the number of cases studied, and the frequency of observation on exogenous covariates.

RESEARCH STRATEGY

In a general sense, good research strategy for an event history problem is completely parallel to good research strategy in any explanatory enterprise. It involves sensible conceptualization, measurement, and so on. Some of the distinctive choices involved in an event history analysis can be outlined, however.

A first concern is with the conceptualization and measurement of the

"dependent variable," the event whose timing the analyst wants to study. Event history analysis operates under the assumption that individual events form instances of some abstract class of events. The researcher needs to conceptualize and operationalize the definition of the events to make this assumption acceptable and ignore (in the data analysis) residual idiosyncracy in the actual events.

Event history analysis also generally assumes that the event is precisely dated to a point in time. While researchers always work in terms of intervals of time (years, months, days, etc.), one wants to be "precise enough" given the variability in event times, the frequency of change in explanatory factors, and the frequency of measurement of explanatory factors. Accuracy to the year may be adequate for legislative events occurring over a century or more, while accuracy to the month or day may be necessary when modeling job shifts within an organization.

Because event history analyses study change over time, it is necessary to define the observation period, that is, the period when susceptibility to the event is being examined. This decision should be theoretically motivated; the analyst needs to consider over what interval causal relationships may usefully be considered invariant. Of course, practical issues come into play: it may be impossible or too expensive to collect data on relevant variables over a very long period. In comparative research, analysts most often define their observation period by historical era. Where the event is a relatively new social construction, the starting point of the study is sometimes taken to be the first event (e.g., the first instance of a national workmen's compensation law).

Research strategy also involves consideration of whether and how transition rates depend on time. A first question is, what way of measuring time is relevant? Commonly used "clocks" are historical time (which tends to reflect the impact of external conditions), age (which tends to reflect the impact of internal processes), and time since the last event (which tends to represent the way previous events either suppress or facilitate additional ones). Of course, all of these may be operating at once: the rate of social revolution is probably a function of historical era (some eras are more revolutionary), age (older polities may be more or less flexible in handling challenges), and duration (revolutions often pave the way for counter-revolutions). Event history analysis provides many tools for describing and controlling for time dependence, but it is up to the researcher to conceptualize the possibilities.

Finally, a research strategy involves an observation scheme for the explanatory variables. Ideally, one would like to know the exact values of the explanatory variables at all times when the case is at risk of an event. For some variables this is quite possible: only one measurement is needed on time-invariant characteristics like an individual's gender

and race. Repeated measurement is needed to approximate the values of variables that change erratically over time, as most national characteristics do. More frequent observation is advisable when the variable changes more quickly and unpredictably; when change is slow and/or steady, a smaller number of observations can capture its time-path in a satisfactory way.

THE PROMISE OF EVENT HISTORY ANALYSIS

Event history analysis provides a powerful set of tools that fit well with the interest in qualitative change that marks both macrosociology and political science. Its greatest advantages, relative to alternative methodological approaches to the same problems, lie in its superior resolving power. Event history analysis permits a simultaneous examination of the individual-, network-, and system-level factors that affect the propensity to experience events. Event history analysis also helps the analyst synchronize causes and effects, so outcomes are connected to contemporaneous conditions. Alternative approaches generally average over both outcomes and explanatory variables and are less well equipped to deal with the temporal basis of social processes.

A second promise is shared between event history analysis and other forms of dynamic methods, such as time-serial methods for continuous outcomes as well as methods for the study of event counts and event sequences (which may be viewed as special cases of event history analysis). Dynamic methods explicitly model change in outcomes over time, rather than the level of a variable at a point in time. This shift often provides a different and substantively important perspective on old problems. For example, we may learn something different when we ask why and how command economies turn into market economies than when we ask why some economies are organized by command and others by the market.

A GUIDE TO THE LITERATURE

Books and articles

Applications of event history analysis to macrosociological and macropolitical research have been cited. There are a number of excellent methodological overviews that explain the basic statistical concepts involved in event history analysis, and provide examples from sociology, including:

Allison, Paul D. 1984. *Event History Analysis: Regression for Longitudinal Data Analysis.* Newbury Park, Calif.: Sage.

Blossfeld, Hans-Peter, Alfred Hamerle, and Karl Ulrich Mayer. 1989. *Event History Analysis.* Hillsdale, N.J.: Erlhausen.
Especially strong on software examples and applications.

Carroll, Glenn R. 1983. "Dynamic Analysis of Discrete Dependent Variables: A Didactic Essay." *Quality and Quantity* 17:425–60.

Lawless, J. F. 1982. *Statistical Models and Methods for Lifetime Data.* New York: Wiley.
Extended treatment of parametric techniques and types of censoring.

Miller, Rupert G. 1981. *Survival Analysis.* New York: Wiley.
Focuses on nonparametric and semiparametric methods.

Tuma, Nancy B. In press. "Event History Analysis: An Introduction." In A. Dale and R. Davies (eds.), *Analyzing Social and Political Change.* Newbury Park, Calif.: Sage.

More advanced presentations include the following:

Tuma, Nancy B., and Michael T. Hannan. 1984. *Social Dynamics: Models and Methods.* New York: Academic Press.
Presentation oriented to regression-like analysis, coupled with substantive discussion of the application of event history modeling to the social sciences.

Computer software

There is a large and growing set of statistical packages that perform event history analysis. Event history modules within general purpose statistical packages inlude SURVIVAL in SPSSX, LIFEREG and LIFETEST procedures in SAS, and 1L and 2L in BMDP. The routines in SPSSX are quite limited, while BMDP and SAS provide more features. Dedicated programs provide additional flexibility, but are relatively difficult to use. RATE permits parameters and covariates to vary over time in more general ways. CTM provides expanded routines for the analysis of repeated events. See Goldstein et al. (1989) for an extensive review of software options for personal computers.

BMDP. 1985. *BMDP Statistical Software.* Berkeley: University of California Press.

SAS. 1985. *SAS User's Guide: Statistics, Version 5 Edition.* Cary, N.C.: SAS.

SPSSX. 1990. *SPSSX Reference Guide.* Chicago: SPSS.

Tuma, Nancy B. 1980. *Invoking RATE.* Menlo Park, Calif.: SRI International.

Yi, Kei-Mu, Bo Honore, and James Walker. 1987. *CTM: A Program for the Estimation and Testing of Continuous Time Multi-State Multi-Spell Models.* Chicago: ERC/NORC, University of Chicago Press.

NOTE

1. Recent collections of these sorts of applications include Hannan and Freeman (1989) and Mayer and Tuma (1990).

REFERENCES

Allison, Paul D. 1982. "Discrete Time Methods for the Analysis of Event Histories." In S. Leinhardt (ed.), *Sociological Methodology*, pp. 61–98. San Francisco: Jossey-Bass.

1984. *Event History Analysis: Regression for Longitudinal Data Analysis.* Newbury Park, Calif.: Sage.

Berry, Frances Stokes, and William D. Berry. 1990. "State Lottery Adoptions as Policy Innovations: An Event History Analysis." *American Political Science Review* 84:395–416.

Bienen, Henry S., and Nicholas van den Walle. 1989. *Of Time and Power.* Stanford, Calif.: Stanford University Press.

Blossfeld, Hans-Peter, Alfred Hamerle, and Karl Ulrich Mayer. 1989. *Event History Analysis.* Hillsdale, N.J.: Erlbaum.

BMDP (1985). *BMDP Statistical Software.* Berkeley: University of California Press.

Carroll, Glenn R. 1983. "Dynamic Analysis of Discrete Dependent Variables: A Didactic Essay." *Quality and Quantity* 17:425–60.

Goldstein, Richard, Jennifer Anderson, Arlene Ash, Ben Craig, David Harrington, and Marcello Pagano. 1989. "Survival Analysis Software on MS/PC-DOS Computers." *Journal of Applied Econometrics* 4:393–414.

Hannan, Michael T. 1989. "Macrosociological Applications of Event History Analysis: State Transitions and Event Recurrences." *Quality and Quantity* 23:351–83.

Hannan, Michael T., and Glenn Carroll. 1981. "The Dynamics of Formal Political Structure: An Event-History Analysis." *American Sociological Review* 46:19–35.

Hannan, Michael T., and John Freeman. 1989. *Organizational Ecology.* Cambridge, Mass.: Harvard University Press.

Knoke, David. 1982. "The Spread of Municipal Reform: Temporal, Spatial, and Social Dynamics." *American Journal of Sociology* 87:1314–39.

Lawless, J. F. 1982. *Statistical Models and Methods for Lifetime Data.* New York: Wiley.

Mayer, Karl Ulrich, and Nancy B. Tuma (eds.). 1990. *Event History Analysis in Life Course Research.* Madison: University of Wisconsin Press.

Miller, Rupert G. 1981. *Survival Analysis.* New York: Wiley.

Olzak, Susan. 1989. "Labor Unrest, Immigration, and Ethnic Conflict: Urban America, 1880–1915." *American Journal of Sociology* 94:1303–33.

Pavalko, Eliza K. 1989. "State Timing of Policy Adoption: Workmen's Compensation in the United States, 1909–1929." *American Journal of Sociology* 95:592–615.

SAS. 1985. *SAS User's Guide: Statistics, Version 5 Edition.* Cary, N.C.: SAS.

Soysal, Yasemin Nuhoglu, and David Strang. 1989. "Construction of the First Mass Educational Systems in Nineteenth Century Europe." *Sociology of Education* 62:277–88.

SPSSX. 1990. *SPSSX Reference Guide.* Chicago: SPSS.

Strang, David. 1990. "From Dependency to Sovereignty: An Event History Analysis of Decolonization." *American Sociological Review* 55:846–60.

Tolbert, Pamela, and Lynne Zucker. 1983. "Institutional Sources of Change in the Formal Structure of Organizations: The Diffusion of Civil Service Reform." *Administrative Science Quarterly* 28:22–39.

Tuma, Nancy B. 1980. *Invoking RATE.* Menlo Park, Calif.: SRI International.

In press. "Event History Analysis: An Introduction." In A. Dale and R. Davies (eds.), *Analyzing Social and Political Change*, Newbury Park, Calif.: Sage.

Tuma, Nancy B., and Michael T. Hannan. 1984. *Social Dynamics: Models and Methods.* New York: Academic Press.

Yi, Kei-Mu, Bo Honore, and James Walker. 1987. *CTM: A Program for the Estimation and Testing of Continuous Time Multi-State Multi-Spell Models.* Chicago: ERC/NORC, University of Chicago Press.