Event History Analysis

CHANGE OF STATES

Contributors: Paul D. Allison
Book Title: Event History Analysis
Chapter Title: "CHANGE OF STATES"
Pub. Date: 1984
Access Date: March 17, 2015
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9780803920552
Online ISBN: 9781412984195
DOI: http://dx.doi.org/10.4135/9781412984195.n7
Print pages: 58-66

http://dx.doi.org/10.4135/9781412984195.n7

CHANGE OF STATES

In this chapter we shall consider a class of models that allows for both multiple kinds of events and repeated events. This class includes all the previously discussed continuous-time models as special cases. Known as Markov renewal or semi-Markov models, these models have been previously described in the social science literature by Coleman (1981), Tuma, Hannan, and Groeneveld (1979), and others.

Markov models describe processes in which individuals are in one of a set of mutually exclusive and exhaustive states at any point in time. For example, in applying their model to the study of marital status, Tuma et al. (1979) distinguished three states: married, unmarried, and attrited. Knoke (1982) used a version of the model to study forms of city government: commission, council-manager, mayor-council. Diprete (1981) studied employment versus unemployment. The set of possible states is often called the state space. For the model to be of any interest, it must be possible to make transitions between at least some of the states, and it is assumed that those transitions can occur at any point in time.

These state transitions are equivalent to the “events” discussed in the preceding chapters. In fact, any kind of event considered thus far can be thought of as a**[p. 58 ↓]** transition between states, although this may sometimes be a bit artificial. For example, an arrest can be thought of as a transition from having had, say, four arrests to having had five arrests. Similarly, the birth of a child is a move from n children to n + 1 children.

Transition Rates

Transitions among these states are controlled by a set of unobserved *transition rates*, which are defined as follows. Let $P_{ij}(t, t + s)$ be the probability that an individual who is in state i at time t will be in state j at time t + s. Then the transition rates, denoted by r *ij*

(t), are given by

$$r_{ij}(t) = \lim_{s \to 0} P_{ij}(t, t + s)/s \qquad [19]$$

Notice the similarity between equation 19 and the definition of the type-specific hazard functions in equation 16. In fact, the type-specific hazard functions are just transition rates for the special case in which all individuals begin in the same origin state. To put it another way, transition rates can be regarded as type-specific hazard functions in which events are distinguished both by origin and destination states.

The transition rates, in turn, are allowed to depend on time and the explanatory variables, with the most common functional form being

$$\log r_{ij}(t) = a_{ij}(t - t') + b_{ij}\, x \qquad [20]$$

where x represents a set (vector) of explanatory variables, b
ij

represents a set (vector) of coefficients, t' is the time of the last transition, and a
ij

(t - t') is some function (as yet unspecified) of the time since the last transition. The function a
ij

(t - t') is often constrained to be a constant a
ij

, which implies that the time between transitions is exponentially distributed. Implicit in the model are the assumptions that time intervals between events are independent and, for each ij combination, identically distributed.

This model, or variations thereof, can be estimated with a variety of forms of data (Coleman, 1981). Nevertheless, the best form of data is clearly an event history in which both the timing and the type of all transitions are recorded. Tuma's RATE program is designed to estimate this model directly using event history data. In fact, however, virtually any computer program for doing event history analysis can estimate at least some version of this model. This chapter will show how to estimate equation 20 using a standard program for estimating proportional hazards models.

[p. 59 ↓] Analysis may be accomplished through the following steps:

- Break each individual's event history into a set of intervals between events.
- Separate these intervals into groups according to origin state.
- For each origin state, estimate models for multiple kinds of events, specifically for class IIa (competing risks). Each destination state corresponds to a different event type. As in Chapter 5, a separate model may be estimated for each destination state, treating all other destinations as censored at the end of the interval.

An Analysis of Job Changes

To illustrate this procedure, we shall analyze data on job changes of 477 physicists surveyed in 1966. (For a detailed description of the data see Hagstrom, 1974.) Information on job histories of these physicists was obtained from *American Men and Women of Science* (Cattel Press, 1966). Data were available on each job held by each physicist from the receipt of the doctorate to the time of the survey in 1966. Since the physicists were of widely varying ages in 1966, however, the length of the job histories also varied widely across individuals.

For this analysis, jobs were classified by three types of employers: (1) university departments whose “quality” was rated in Cartter's (1966) study; (2) four-year colleges and universities not rated by Cartter (usually lesser known institutions); and (3) nonacademic employers, including government and industry. These three employer types are the three states in a semi-Markov model of employer changes. The objective is to estimate the effects of several explanatory variables on transitions among these three states.

This situation departs somewhat from the typical semi-Markov model in that it is possible to make a transition from any of the three states back into the same state. Thus, one can make a transition from a nonacademic employer to another nonacademic employer. While this creates no analytical difficulties, it does increase the number of kinds of transition. With three origin states and three destination states, there are nine possible transitions, and a separate model will be estimated for each one of them.

The 477 physicists held a total of 1069 jobs during the period of observation. Of these jobs, 477 were censored because they were still in progress when the study was terminated. The remaining 592 jobs ended in transitions from one employer to another.

[p. 60 ↓] The explanatory variables were all treated as constant over time. They include a quality rating of each physicist's undergraduate institution, Cartter's (1966) rating of the doctoral department, number of years between the bachelors and doctoral degrees, a dummy variable indicating whether or not the individual received a postdoctoral fellowship, a dummy variable indicating U.S. citizenship, a dummy variable for “inbreeding” (employer is the physicist's doctoral department), number of previous jobs, Cartter rating of the current department (when available), career age (time since Ph.D.), and the calendar year in which the job began.

The model to be estimated is given by equation 20, where $a_{ij}(t - t')$ is left as an unspecified function of time. Hence, we have nine proportional hazards models, each of which can be estimated using the partial likelihood method. The first step is to divide the 1069 jobs into three groups by the three origin states. This yields 651 rated academic jobs, 212 unrated academic jobs, and 206 nonacademic jobs.

Beginning with the rated academic jobs, three models are estimated, one for each of the three destination states. For transitions to rated academic jobs, any such transition (there were 202) is treated as an event and all other outcomes are treated as censored.

Similarly, in estimating the second model, any change to an unrated academic job is treated as an event and all other outcomes are treated as censored. The same procedure is used for the transitions from rated academic to nonacademic jobs. For all three models, the same subsample is analyzed and the same duration time is specified for each individual. The difference in the models is the specification of which duration times end in events and which are censored. Appendix B shows how this can easily be set up using the SAS program PHGLM.

Table 7 presents results from estimating these models. Examining the first three columns of Table 7, we see that most of the variables have estimated effects which are not significantly different from zero. For transitions from rated academic to rated academic, it appears that higher rates of transition are more characteristic of former postdoctoral fellows, those with more previous jobs, those whose jobs began in later calendar years, and those at younger career ages. Number of previous jobs and career age have particularly strong effects. Exponentiating the coefficient of .253 yields 1.29, which tells us that at each subsequent job the risk of a rated-academic to rated-academic transition increases by 29 percent. This is most likely a consequence of unexplained heterogeneity in the sample rather than a true causal effect of the number of jobs. Thus, we expect that those who have had many jobs in a given period of**[p. 61 ↓] [p. 62 ↓]** time are more likely to change jobs again in the immediate future. The career age coefficient of -.204 exponentiates to .82, indicating that each additional year past the doctorate decreases the hazard of a transition by 18 percent; over a five-year period the hazard would be reduced by 64 percent. The latter figure is obtained by calculating 100 [1 - exp(5 (-.204])]. There are many possible explanations for this effect, but we shall not pursue them here.

TABLE 7 Estimates of Proportional Hazards Models for Transitions Among Three Types of Jobs

TABLE 7
Estimates of Proportional Hazards Models for Transitions Among Three Types of Jobs

	1 From Rated Academic to:			2 From Unrated Academic to:			3 From Nonacademic to:		
Explanatory Variables	*Rated Academic*	*Unrated Academic*	*Nonacademic*	*Rated Academic*	*Unrated Academic*	*Nonacademic*	*Rated Academic*	*Unrated Academic*	*Nonacademic*
Undergraduate rating	.009	−.010	.013	.035	.014	.049	.003	.032	−.018
Graduate department rating	−.123	−.011	−.228	−.039	−.341	−.433	.121	−.269	.224
Time for Ph.D.	−.045	.037	−.017	−.106	.041	−.025	−.026	.157	−.106
Postdoctoral Fellow (D)[a]	.338*	−.009	.234	.467	.592	.415	−0.15	.157	−.058
U. S. Citizen (D)	−.138	−.340	−.191	−.222	−.370	.269	−.293	.153	−.468
Inbred (D)	.210	.538	.746*	−.749	−.484	.150	–	–	–
Number of previous jobs	.253***	.212	.284*	.338*	.146	.115	.237*	−.038	.248
Year job began	.021**	.025	.006	.041**	.002	.036	.045**	.001	−.014
Career age	−.204***	−.161***	−.163***	−.147***	−.123***	−.160***	−.089***	−.105*	−.075**
Cartter rating	.015	−.141	−.193	–	–	–	–	–	–
N of cases	651	651	651	212	212	212	206	206	206
N of job changes	202	42	48	55	43	20	109	25	48
−2 X log-likelihood	1934.97	426.54	486.57	436.52	336.24	148.97	960.77	218.82	413.41

a. (D) indicates dummy variable.
*Significant at .05 level.
**Significant at .01 level.
***Significant at .001 level.

Even fewer variables have significant effects in the next two columns representing transitions to unrated academic and nonacademic jobs. Although this is partly due to a decline in magnitude of some of the estimated effects, it is largely a consequence of the reduced number of observed transitions of these two types—there are only 42 changes from rated academic to unrated academic, and only 48 changes from rated academic to nonacademic. It is generally true in event history analysis that, while censored observations do contribute some information to the analysis, they do not add nearly as much information as uncensored observations. As a result, significance levels and standard errors will depend as much on the actual number of events as on the total number of observations.

The effect of career age is strong and roughly the same magnitude for all three types of transitions. The coefficient for number of previous jobs is also roughly the same magnitude for all three types, but its level of statistical significance varies greatly. The calendar year in which the job began has a significant effect only for rated to rated transitions, but its numerical magnitude is actually larger for rated to unrated moves. One surprising result is the substantial coefficient of "inbreeding" for rated academic to nonacademic transitions. When exponentiated, the coefficient of .746 becomes 2.11, indicating that physicists whose current job is in the university department in which they got their doctorate are more than twice as likely as others to move to a nonacademic job.

The next phase of the analysis is to examine the 212 jobs in unrated academic settings. Again the procedure is to estimate three models for this subsample. In each model, the event is a transition to a particular destination state and all other outcomes are treated as censored data. Cartter rating could not be included as an explanatory variable because, by definition of the origin state, it was not measured for these jobs. Results are shown in the three middle columns of Table 7. The overall picture is quite similar to that for transitions out of rated academic positions, except that even fewer variables have effects that are statistically significant. Again, this is partly due to the fact that both the overall**[p. 63 ↓]** number of cases and the number of observed transitions of each type are substantially reduced.

The final phase of the analysis is to repeat the process one more time for transitions out of nonacademic jobs. Both Cartter rating and "inbreeding" had to be omitted as explanatory variables because they were not meaningful for nonacademic jobs. Results in Table 7 are again quite similar to those for the other origin states.

What does this analysis tell us? The negative effect of career age is quite consistent across all the transition types. To a lesser degree, so are the effects of the number of previous jobs. The year the job began seems to be important in predicting transitions to rated academic jobs, but not to other destinations. The effects of postdoctoral fellowship and inbreeding show up for two types of transition but not for any others.

Simplifying the Model

Given the similarities across transition types, it is reasonable to inquire whether the results could be simplified in some way. In fact, there is no necessary reason why one should estimate separate models for all nine transition types. Both theory and empirical evidence can suggest combining either origin states or destination states or both. Table 8 gives results from estimating proportional hazards models in which the three destination states are combined while the distinction among origin states is maintained. The estimation procedure for these models was quite straightforward. As with the models in Table 7, intervals between events were subdivided by the three origin job types, and a separate model was estimated for each origin type. No distinction was made among destination states, however.

The coefficient estimates in Table 8 are approximately what one would expect just from averaging coefficients in Table 7. There is one notable change, however. For unrated academic jobs, the effect of a postdoctoral fellowship is statistically significant in Table

8 but not in Table 7. This is undoubtedly a consequence of the greater statistical power obtained by combining destination job types.

It is also possible to test whether the simplification obtained by combining destination job types is statistically justified. The null hypothesis is that the explanatory variables have identical coefficients across destination types but may differ by origin type. For each model estimated, both Tables 7**[p. 64 ↓]** and 8 report the log-likelihood (multiplied by -2). A test statistic is obtained by adding all the log-likelihoods in Table 7, doing the same in Table 8, and then taking the difference between the two sums. Thus, we have

TABLE 8 Estimates of Proportional Hazards Models Combining Destination Job Types

TABLE 8
Estimates of Proportional Hazards Models Combining Destination Job Types

	Origin Job Type		
Explanatory Variables	*Rated Academic*	*Unrated Academic*	*Nonacademic*
Undergraduate rating	.006	.030	.001
Graduate department rating	−.128	−.189	.081
Time for Ph.D.	−.028	−.017	−.017
Postdoctoral fellow (D)[a]	.271*	.492*	−.036
U. S. citizen (D)	−.175	−.181	−.274
Inbred (D)	.349*	−.494	–
Number of previous jobs	.247***	.234*	.213**
Year job began	.019**	.027**	.020*
Career age	−.188***	−.141***	−.086***
Career rating	−.044	–	–
N of cases	651	212	206
N of job changes	292	118	182
−2 X log-likelihood	2859.5	937.4	1614.8

a. (D) indicates dummy variable.
*Significant at .05 level.
**Significant at .01 level.
***Significant at .001 level.

$$48.8 = 5411.7 - 5362.9$$

Under the null hypothesis this statistic has a large-sample chi-square distribution with 54 degrees of freedom, which is the difference in the number of coefficients estimated in Table 7 and Table 8. Since this is far from significant, we may conclude that Table 8 is an acceptable simplification of Table 7.

An examination of the coefficients in Table 8 suggests that differences across origin job types are also small. To test whether further simplification is possible, a single model combining both origin job types and destination job types was estimated.[9] (Coefficient estimates are not shown.) For this model, the log-likelihood (multiplied by -2) was 6624.6. Taking the difference between this log-likelihood and the sum of the log-likelihoods in Table 8 yields a chi-square value of 1212.9 with only 17 degrees of freedom. Clearly, it is statistically unacceptable to collapse origin states.

[p. 65 ↓] One must also be aware that the approaches taken in this chapter, because they are based on observation of repeated events on each individual, suffer from an important limitation discussed in the last chapter. Specifically, it must be assumed that whatever dependence exists among the multiple intervals for a single individual is a consequence of the effects of the explanatory variables that are included in the model. Although the consequences of violating this assumption are not yet well understood, one can always take the conservative approach discussed in Chapter 6 of modifying the standard errors so that they reflect the number of individuals rather than the number of intervals or the number of transitions.

http://dx.doi.org/10.4135/9781412984195.n7