Journal of Economic Literature
Vol. XXVI (June 1988), pp. 646–679

Economic Duration Data and Hazard Functions

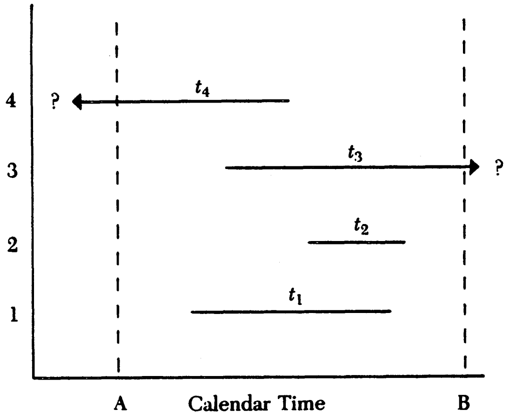
By NICHOLAS M. KIEFER
Cornell University

I would like to thank Terri Devine, George Jakubson, Soren Johansen, Chuck McCulloch, Lars Muus, George Neumann, Geert Ridder, Sunil Sharma, Insan Tunali, and participants in workshops at Cornell, Aarhus, and McMaster Universities for many helpful suggestions and discussions. Bent Jesper Christensen provided invaluable research assistance. This work was supported by the NSF and by a fellowship from the John Simon Guggenheim Memorial Foundation.

I. *Introduction*

HOW MUCH TIME do individuals spend in unemployment? How does this change over the business cycle? How does the duration of unemployment vary across individuals? Answers to questions such as these are needed for several reasons. First, the welfare of the unemployed is surely more closely related to the time they spend without a job than to the fact of their being unemployed. In this sense, the unemployment rate, which involves both the incidence (or occurrence) of unemployment spells and their durations, is a less useful statistic than is the average duration of unemployment. Second, the length of unemployment spells plays a critical role in economic theories of job search, where the distribution of the duration of unemployment depends on the individual's reservation wage. Any careful evaluation of the theory, therefore, requires accurate information on the duration of unemployment.

The regularly available data on duration of unemployment collected by the Current Population Survey is subject to some serious shortcomings as a source of information on the distribution of durations. In this survey, information on the length of unemployment is collected only on those individuals unemployed at the time of the survey. This means that time spent unemployed will not be computed for those individuals employed at the time of each adjacent monthly survey, but unemployed between surveys. For example, suppose in Figure 1 that A is the March survey date and B is the April survey date, and that t_1 and t_2 will not be seen. In general, short spells will be underrepresented in the sample, a problem known as *length-biased sampling*. Another consequence of exclusively sampling the unemployed for information on unemployment duration is that these individuals have not completed their unemployment spells; the survey interrupts spells still in progress. For example, spell t_4 will be seen in the March survey, but



A: Beginning of study period (e.g., March survey)
B: End of study period (e.g., April survey)
t_1, t_2: Completed spells
t_3: Right-censored spell
t_4: Left-censored spell

Figure 1. Duration Data

only the duration from the inception of the spell to the survey date, A, will be recorded. Similarly, spell t_3 will be seen in April, and the elapsed duration will be recorded. These observations on the duration of unemployment are *right-censored.* Thus, both because of right-censoring and of length-biased sampling, the Current Population Survey information of unemployment durations is suspect.[1]

Of course, there are surveys that sample individuals at some date and record their labor market experience until an ending date (A to B in Figure 1), or that sample individuals at some date (for example, B) and ask for information retrospectively to date A. Though the accuracy of such data, particularly retrospective data, is often in doubt, these sampling schemes do result in observations on completed spells, such as t_2 and t_1, as well as observations on right-censored spells such as t_3. Spells such as t_4 may be recorded as a duration beginning at A and lasting until the completion of the spell, in which case the actual duration is unknown because the time from the inception of the spell to the beginning of data collection (A) is unknown. In this case the spell is *left-censored.* Alternatively, the completed duration t_4 might be recorded. In either case an issue of length-biased sampling arises because spells in progress at the inception of the study are more likely to be long spells than short spells. Moreover, even if the sample consisted exclusively of a random sample of completed spells, there are difficulties in using the information on the duration of unemployment as a dependent variable in a regression framework where the determinants of unemployment length are measured by a set of exogenous variables x. The problem that arises in a regression context is how to measure those x's whose values change during the unemployment spell. Therefore, even without the censoring problem, duration data present conceptual problems for economists used to thinking in terms of conventional regression analysis.

For these reasons, a literature has arisen in economics addressing the special problems associated with duration data. This literature has drawn heavily on statistical methods developed largely in industrial engineering where they are used to describe the useful lives of various machines and in the biomedical sciences to describe events such as the survival times of heart transplant recipients. These methods have a natural application to many economic problems. Probably the most widely studied duration data in economics are data on lengths of spells of unemployment. The papers by Tony Lancaster (1979) and Stephen Nickell

[1] For a detailed discussion of the problems of inferring duration distributions using Current Population Survey sampling techniques, see Nicholas M. Kiefer, Shelly Lundberg, and George R. Neumann (1985). David R. Cox and P. A. W. Lewis (1966, ch. 4) discuss length-biased sampling.

(1979) propose and apply hazard function methods for studying unemployment durations. The models proposed can be regarded as reduced forms resulting from behavioral models relying on job search arguments. Of course, other interpretations are also possible. As usual, reduced-form results can serve to rule out some potential structural models, but cannot distinguish between others. Nicholas M. Kiefer and George R. Neumann (1979) and Wiji Narendranathan and Nickell (1985) are among the several papers that attempt to estimate behavioral models. There is an extensive additional literature on unemployment durations; much of it is surveyed in Theresa J. Devine and Kiefer (1987). One focus of many studies is the effect of the "replacement ratio" in the unemployment insurance system—the ratio of benefits to the previous wage—on unemployment durations. Lancaster (1979) remarks that "the final estimate of the elasticity of unemployment duration with respect to the ratio of unemployment income to the last wage received is about 0.6 . . . and an elasticity of this order could now be regarded as established beyond reasonable doubt" (p. 956). As usual in economics, there is dissent; see Tony Atkinson and John Micklewright (1985).

Other actual and potential areas of application are duration of marriages, spacing of births, time to adoption of new technologies, time between trades in financial markets, product durability, geographic or occupational mobility (time between moves), lifetimes of firms, time to invention from research investment, payback periods for overseas loans, durations of wars, time in office for congressmen and other elected officials, time from initiation to resolution of legal cases, spacing of purchases of durable goods (or replacement capital), time in rank, and length of stay in graduate school.

The central concept in these statistical methods is occupied not by the unconditional probability of an event taking place (e.g., the probability of an individual being unemployed exactly 10 weeks), but of its conditional probability (e.g., the probability of an individual leaving unemployment in the tenth week given that he has been unemployed 9 weeks). To understand further the concept of conditional probabilities, consider forecasting the round in which your favorite team will be eliminated in a single-game elimination tournament. A natural way to think about this problem is to consider the probability of losing the first game—the probability of first-round elimination—then to consider the probability of losing the second game, conditional on making it into the second round. Reasoning in terms of conditional probabilities requires only consideration of single games. Pursuing this strategy yields the sequence of conditional probabilities $\lambda(i) = Pr$ (lose in round i, given a win in previous rounds). An alternative approach, more in keeping with usual econometric practice, is to specify the (unconditional) probability function $f(i) = Pr$ (lose in round i) directly. This second approach seems much more complicated, as it involves simultaneous consideration of the possibilities of losing. Why not take advantage of the fact that the event we are considering, loss in round i, can be considered the outcome of a sequence of simpler events? Of course, conditional and unconditional probabilities are related, so the mathematical description of the process is the same in either case. It is the conceptual difference that is important in economic modeling of duration data.

Suppose an unemployed worker spends some time every day looking for employment. We as economists are interested in describing the distribution across individuals of the duration of unemployment—here the number of days unemployed. This information is useful, together with information on the incidence of unemployment, in understanding the unemployment rate. What is the probability that the worker will be unemployed exactly 10 days? This event can be described as the outcome of a sequence of simpler, conditional events. We might specify a model, for example, in which every day the worker looks for a job he has the same probability, λ, of finding one. That is, conditional on being unemployed through yesterday, the probability of finding employment today is λ. Thus the sequence of conditional probabilities is constant. (The analogy in our tournament has games decided at random, so the probability of winning does not depend on the opponent.) The probability that the worker will be unemployed exactly 10 days is $f(10) = \lambda(1 - \lambda)^9$. Now, the assumption that the probability of finding a job is the same every day is strong. It seems possible that this conditional probability varies as the length of the spell of unemployment increases, perhaps due to increased search intensity or to a change in the minimum wage acceptable to the individual. Note that the economics of the problem suggests that we reason in terms of the conditional probabilities $\lambda(i)$, not the unconditional probabilities $f(i)$. Why not maintain a close link between our theoretical notions and the way we interpret data by doing the econometrics the same as the theory?

The functions $\lambda(i)$ are *hazard functions* for the random variables "rounds played" in the tournament example and "days unemployed" in the job-search example. Rounds played and days unemployed are examples of durations.

The special methods of duration analysis are useful and convenient means of organizing, summarizing, and interpreting data for which a representation in terms of a sequence of conditional probabilities is theoretically or intuitively appealing. Hazard function specification emphasizes the conditional probabilities, while specification in terms of a probability distribution emphasizes unconditional probabilities. For any specification in terms of a hazard function there is a mathematically equivalent specification in terms of a probability distribution. The two specifications involve the same parameters and are simply two different ways of describing the same system of probabilities. Thus, *the hazard function approach does not identify new parameters.* As we will see, the likelihood function can be written equivalently in terms of hazard functions or probability distributions, but it is the same likelihood function in either case.

Why then learn about hazard functions when we already know about some familiar probability distributions—say the normal and lognormal? If our model suggests that conditional probabilities are of interest, why not specify the probability distribution, fit it, and calculate the implied conditional probabilities, λ, later? This approach could be taken, but if we are thinking in terms of conditional probabilities, it makes sense to choose a parametrization that allows the estimated sequence of conditional probabilities to behave as we think it should. The normal and lognormal distributions, for example, have complex hazard functions that do not admit the constant hazard as a special case. Because the constant hazard is in many cases a natural special case or null hypothesis, this is an unattractive feature of several probability distributions that are appealing in other applications. Consequently the duration literature relies heavily on other distributions, for example the exponential and Weibull. Some commonly used specifications are given in Section IIB.

Defining a duration precisely requires a time origin (a beginning), a time scale, and a precise definition of the event ending the duration. In a sample consisting of many individuals, different individuals will often have different time origins for the durations they experience. Unemployment spells, for example, could begin at any date; the beginning date is the time origin for the spell. The duration of a spell is its length. Typically, durations, that is, spell lengths, are the dependent variables under study, but it should be kept in mind that these durations are not spells in "real time" unless the time origin is the same for every spell. In practice one would like individuals in the sample to be as homogeneous as possible, after controlling for observable differences, at the time origin specific to the individual. The time scale in economic applications is usually calendar time. The units depend on the application and the precision of measurement. For instance, it is pointless to analyze data measured in months on an hours scale.

A distinguishing feature of duration data is the possibility that some of the durations observed will be censored, as emphasized in the discussion of Figure 1. Censoring is an event that occurs at some time, so our data consist of a measured spell length together with the information that the spell was censored (or not). Let T^*, a random variable, be a spell length for an individual in the absence of censoring, and let c be the censoring time measured from the time origin for the spell. Then the random variable which will be observed is the smaller of T^* and c, or $T = \min(T^*, c)$. We also observe an indicator variable $d = 1$ if the observation is censored ($T = c$), $d = 0$ if uncensored ($T = T^*$). Often the censoring times are known constants (given the time origin), for example, the end of a fixed-length panel survey. Another type of censoring is known as *Type I censoring,* in which c is a predetermined constant common across observations as when data measured in weeks from 1 to 26 are followed by a category "more than 26 weeks." Rather general forms of *random censoring* can fit into the framework discussed here, though it is crucial to assume that individuals whose spells are censored at time c are representative of all the individuals who have spell lengths at least equal to c, perhaps after allowance for explanatory variables. Thus, if we regard the censoring time c as a random variable, it must be independent of T^*, after taking account of other factors. This is a maintained assumption in most applications.

II. *Distributions of Durations*

A. *The Hazard Function*

The probability distribution of duration can be specified by the distribution function

$$F(t) = Pr(T < t)$$

which specifies the probability that the random variable T is less than some value t. The corresponding density function is $f(t) = dF(t)/dt$. These are two equivalent ways of specifying a distribution, and the choice of specification depends on convenience.[2] In studying duration data it is useful to define the *survivor function*

$$\begin{aligned} S(t) &= 1 - F(t) \\ &= Pr(T \geq t) \end{aligned}$$

giving the upper tail area of the distribution, that is, the probability that the random variable T will equal or exceed the value t. Of course, specifying S is merely

[2] The convention $F(t) = Pr(T < t)$ as opposed to $Pr(T \leq t)$) is often adopted when discussing duration data. See, for example, David R. Cox and David Oakes (1985) or John D. Kalbfleisch and Robert L. Prentice (1980). Certain formulas involving the hazard function are more cleanly written with this definition.

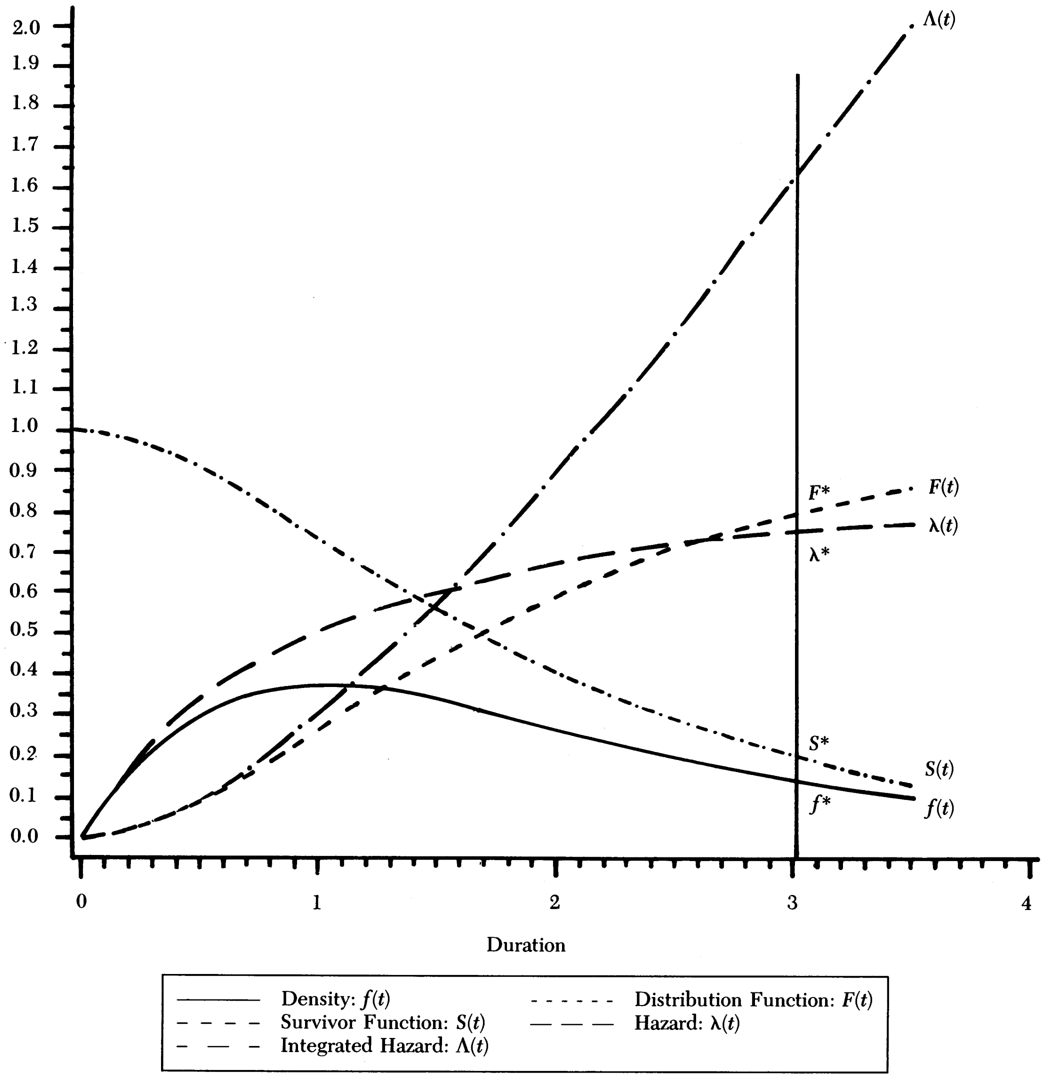


Figure 2. Hazard and Other Functions

an alternative method of specifying the distribution of T, and there are many other functions that could characterize a distribution. A particularly useful function for duration analysis is the *hazard function*

$$\lambda(t) = f(t)/S(t).$$

Roughly, $\lambda(t)$ is the rate at which spells will be completed at duration t, given that they last until t.[3] This is a continuous-time version of the sequence of conditional probabilities illustrated in the tournament example. Figure 2 illustrates

[3] A precise definition in terms of probabilities is

$$\lambda(t) = \lim_{h \to 0} Pr(t \leq T < t + h | T \geq t)/h.$$

Those familiar with the literature on the econometrics of sample selection will recognize the hazard rate as the inverse of Mills' ratio.

$F(t)$, $f(t)$, $S(t)$, and $\lambda(t)$ for one particular distribution. According to the example in Figure 2, the probability of a spell lasting less than 3 "periods" is F^*; equivalently, the probability of a spell lasting 3 periods or more is S^*. The probability that a spell ends between 3 and $3 + \Delta$ periods is $f^*\Delta$, while the probability that a spell ends between 3 and $3 + \Delta$ conditional on having lasted 3 periods is $\lambda^*\Delta$. Given $\lambda = f/S = (dF/dt)/S = (-dS/dt)/S$, we see that

$$\lambda(t) = -d \ln S(t)/dt.$$

The hazard function provides a convenient definition of duration dependence. *Positive duration dependence* exists at the point t^* if $d\lambda(t)/dt > 0$ at $t = t^*$. The hazard graphed in Figure 2 has positive duration dependence for all t. Positive duration dependence means that the probability that a spell will end shortly increases as the spell increases in length. *Negative duration dependence* exists at t^* if $d\lambda(t)/dt < 0$ at $t = t^*$. Negative duration dependence in the tournament example occurs when the team has "momentum": At each round, the probability of losing falls. The rather more straight forward statistical terminology "increasing hazard" and "decreasing hazard" does not seem to have caught on in economics.[4]

The *integrated hazard*

$$\Lambda(t) = \int_0^t \lambda(u)du$$

is also a useful function in practice. It is the basic ingredient in a variety of specification checks. The integrated hazard does not have a convenient interpretation, however. In particular, note that it is not a probability. The relation to the survivor function is

$$S(t) = \exp[-\Lambda(t)].$$

B. *Some Distributions*

The *exponential distribution* is widely used as a model for duration data: The exponential is simple to work with and to interpret, and is often an adequate model for durations that do not exhibit much variation (in much the same way that the linear regression model is simple and adequate if the data do not vary enough to reveal important nonlinearities). For the exponential distribution with parameter $\gamma > 0$,

$$F(t) = 1 - \exp(-\gamma t)$$
$$S(t) = \exp(-\gamma t)$$
$$f(t) = \gamma\exp(-\gamma t)$$
$$\lambda(t) = \gamma$$
$$\Lambda(t) = \gamma t.$$

The survivor and hazard functions corresponding to $\gamma = 1$ are shown in Figures 3 and 4. The exponential distribution is sometimes termed *memoryless*, because the hazard function is constant and so reflects no duration dependence. Because the hazard function uniquely characterizes the distribution, the exponential is the only distribution with this property. It is easy to check that the distribution of the exponential random variable T, conditional on its taking a value at least c, is the same as the unconditional distribution of T. The exponential distribution arises as a prediction from simple stochastic models, as shown below. The

[4] It may be useful to review the discrete versions of the functions we have considered above for continuous random variables. Here the values taken by the random variable T are $t_1, t_2, t_3, \ldots$. The functions are:

$$f(t_k) = Pr(T = t_k)$$
$$S(t_k) = \sum_{j \geq k} f(t_j)$$
$$\lambda(t_k) = f(t_k)/S(t_k).$$

The last expression is easily recognized as the probability a duration ends at t_k, given that it lasts until t_k. The discrete integrated hazard is

$$\Lambda(t_k) = \sum_{i=0}^{k} \lambda(t_i).$$

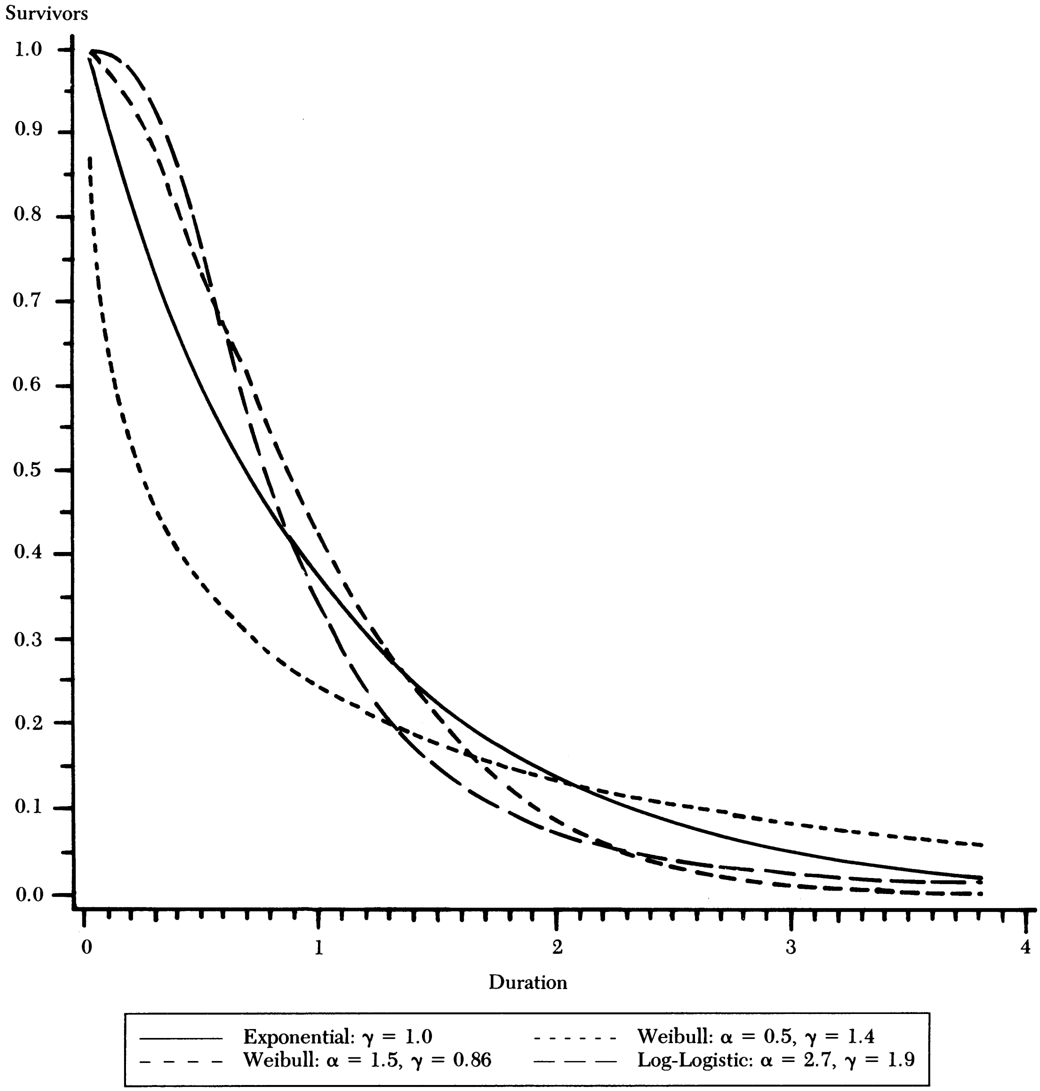


Figure 3. Survivor Functions

fact that the distribution depends only on one parameter, γ, is the drawback in applications because the family of distributions obtained by varying γ is not very flexible. To illustrate, note that $E(T) = 1/\gamma$ and var $(T) = 1/\gamma$ so that the mean and variance cannot be adjusted separately. Thus, the exponential is unlikely to be an adequate description of the data if the sample contains both very long and short durations.

The *Weibull* distribution is a two parameter ($\gamma > 0$ and $\alpha > 0$) family with hazard function

$$\lambda(t) = \gamma\alpha t^{\alpha-1}.$$

This is a simple generalization of the exponential distribution, which is obtained

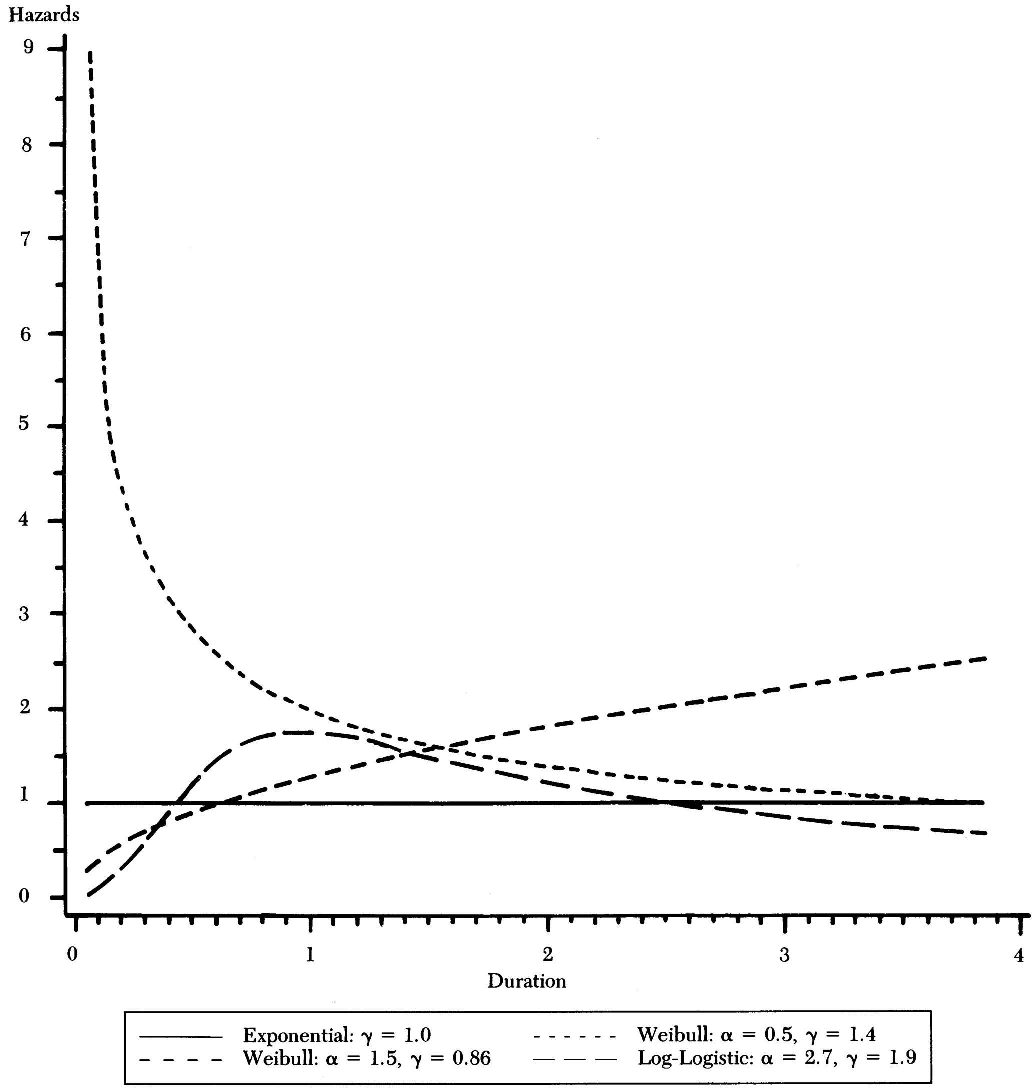


Figure 4. Hazard Functions

by setting $\alpha = 1$. Using our integration formulas leads to expressions for the distribution, survivor, density, and integrated hazard functions

$$F(t) = 1 - \exp(-\gamma t^{\alpha})$$
$$S(t) = \exp(-\gamma t^{\alpha})$$
$$f(t) = \gamma \alpha t^{\alpha - 1} \exp(-\gamma t^{\alpha})$$
$$\Lambda(t) = \gamma t^{\alpha}.$$

See Figures 3 and 4, where in one case $\alpha = 3/2$ and $\gamma = .86$ and in the other $\alpha = 1/2$ and $\gamma = 1.4$. It is easy to check that t^{α} has an exponential distribution with parameter γ. In this sense, the Weibull can be thought of as an exponential distribution on a rescaled time axis. The "trick" of finding a transformation of duration so that the transformed durations are exponentially distributed is quite useful in interpreting models and in developing specification diagnostics. In practice, the parameter γ can depend on explanatory variables. The hazard function is increasing in duration if $\alpha > 1$, decreasing if $\alpha < 1$, and constant if $\alpha = 1$, which is exactly the exponential case. Duration dependence does not depend on the value of γ.

Finally, we consider an example of a duration distribution with a nonmonotonic hazard, the *log-logistic* distribution, with parameters $\gamma > 0$ and $\alpha > 0$ and hazard function

$$\lambda(t) = \gamma\alpha t^{\alpha-1}/(1 + t^{\alpha}\gamma).$$

Applying our formulas we find

$$F(t) = 1 - [1/(1 + t^{\alpha}\gamma)]$$
$$S(t) = 1/(1 + t^{\alpha}\gamma).$$
$$f(t) = \gamma\alpha t^{\alpha-1}/(1 + t^{\alpha}\gamma)^2$$
$$\Lambda(t) = \ln(1 + \gamma t^{\alpha}).$$

For $\alpha > 1$ the hazard first increases with duration, then decreases. If $0 < \alpha \leq 1$ the hazard function decreases with duration. See Figures 3 and 4 where the log-logistic survivor and hazard functions are plotted for $\alpha = 2.7$ and $\gamma = 1.9$.[5]

A research strategy analogous to using "everything and its square" in regression specification search is to specify and fit a quite general nonnegative function for the hazard, for example

$$\lambda(t) = \exp[g(t,\theta)]$$

where θ is a vector of parameters and $g(\cdot)$ is a parametric function, perhaps a polynomial, in duration and any other variables that might be thought important. This approach can be informative in some cases. Alternative approaches to fitting hazard functions of unknown form are discussed below.

The distributions considered here and in most applied work are continuous. Continuous-time models seem appropriate for economic settings because there is typically no natural period in which economic decisions are taken (or economic surprises arrive). With a continuous-time model, parameters are defined and can be interpreted independently of the period of the measure of data. In some settings, however, a discrete time approach is natural: The relevant definitions of the hazard and other related functions are given in Footnote 4.

C. *Economic Theory and Duration Distributions*

This section provides brief examples illustrating that economic theory can be used to obtain implications for duration distributions in particular settings. We begin with a highly stylized job-search model leading to an exponential distribution for durations of spells of unemployment.[6] It is sufficient for our purpose to consider a model where individuals are assumed to occupy only two states, employment and unemployment, because we will focus attention on only one transition. For concreteness, we concentrate on the transition from unemployment to employment. State e is employment and u is unemployment. The formal structure we will use allows the worker to change states at any time t.

The unemployed draw wage offers w from the distribution $p(w)$. Unemployed workers receive offers at constant rate η, that is, the probability that the worker will get an offer in any short interval with length Δ is $\eta\Delta$. Offers are drawn inde-

[5] See Cox and Oakes (1985), Kalbfleisch and Prentice (1980), J. F. Lawless (1982), Rupert G. Miller, Jr. (1981), or Elandt-Johnson (1980). There are many other candidates for duration distributions that might be useful in specific circumstances. These include the *log-normal, Gompertz, inverse Gaussian*, generalized F, and Gamma distributions, as well as mixtures of any of these.

[6] A detailed and insightful recent discussion of the theory of search is given by Dale Mortensen (1986).

pendently from the density $p(w)$. Individuals know the offer distribution $p(w)$ but do not know the location of each firm on that distribution. Assume that v_u, the instantaneous utility associated with being unemployed, does not depend on either the outstanding wage offer w or the offer distribution $p(w)$, and is constant over the duration of the spell of unemployment. We are abstracting, for purposes of illustration, from many important empirical considerations. Assume that the utility associated with being employed is a function of the wage paid, $v_e = v_e(w)$, but not of the offer distribution. It seems reasonable to assume that $dv_e/dw > 0$; higher wages are preferred to lower. The worker receives offers at random intervals and decides, when an offer is received, whether to accept the offer or to continue searching. It is clear that an unemployed worker will accept or decline a job immediately upon receiving a draw; that is, if an offer is acceptable, it is acceptable as soon as it arrives.

Under these assumptions the worker's optimal behavior is described by a reservation wage policy; that is, there is a wage w^* such that offers greater than or equal to w^* will be accepted and lower offers will be declined. This follows because the expected discounted utility of being employed is an increasing function of the wage received, while the utility of being unemployed does not depend on the outstanding wage offer. If a worker prefers employment at wage w to unemployment, he will prefer employment at higher wages as well. The reservation wage w^* is the wage at which the worker is indifferent between accepting the offer w^* and declining the offer. The probability that an offer is acceptable is given by the probability of a wage offer equaling or exceeding w^*, namely

$$\pi = \int_{w^*}^{\infty} p(w)dw.$$

The transition rate from unemployment to employment is then given by the product of the offer arrival rate and the probability that an offer is acceptable; this transition rate is the probability of leaving unemployment at any moment given that the individual is still unemployed up to that moment. In other words, it is the hazard function for the distribution of unemployment durations

$$\lambda = \eta\pi.$$

Observe that the duration of unemployment does not appear in this expression for the hazard rate. Consequently the implied distribution of durations of spells of unemployment is exponential

$$f(t) = \lambda\exp(-\lambda t).$$

The model has several interesting features. First, it is possible to use the optimality equations to develop restrictions on the effects of interindividual variations in offer distributions $p(w)$ on the reservation wage w^* and therefore on the hazard λ. For example, individuals with higher mean offers can be expected to have higher acceptance probabilities per draw, other things being equal. This kind of prediction can be used to obtain implications for coefficients of a parametrization of λ as a function of explanatory variables. Thus the variable "mean wage offer" or a suitable proxy is expected to have a positive effect on λ, the conditional probability of becoming employed. These predictions can be checked as a sort of informal specification test. Second, the model is extremely simple to estimate, as will be seen. Third, the multistate extension is straightforward. Kenneth Burdett et al. (1984) estimate a three-state dynamic discrete choice model of labor turnover; the present simple model is a special case.

The implication of an exponential distribution of unemployment durations, however, is not robust to economically plausible changes in assumptions. Changes over time in the arrival rate of offers η, the offer distribution $p(w)$, or the utility associated with unemployment (for example, due to exhaustion of unemployment insurance benefits) will lead to a different duration distribution.

A variation leading to nonexponential distributions can be mentioned briefly. The first is a model of turnover in the labor market similar in structure to that just described but focusing on spells of employment. Spells of employment are assumed to end at layoff or on receipt of a better offer. The probability that a worker is laid off is assumed to depend on seniority. In particular, workers face layoffs in inverse order of seniority. This model leads to a prediction of a declining hazard for employment duration. The prediction holds up in a Weibull model fit by Burdett et al. (1985) and in a variety of other studies.

III. *Estimation*

A. *Nonparametric*

Graphical methods are useful for displaying data on durations and for preliminary analyses, perhaps to suggest functional forms, of homogeneous observations, homogeneity being achieved roughly by grouping on explanatory variables. They are also useful for specification analysis of more complicated models. In specification analysis, graphs of generalized residuals, discussed below, are useful, much as residual plots are useful in linear regression model building.

The sample survivor function for a sample of n observations with no censoring is

$$\hat{S}(t) = n^{-1}(\# \text{ of sample points} \geq t)$$

the empirical cumulative distribution function turned around. A modification is required to allow for censoring. Suppose the completed durations in our sample of size n are ordered from smallest to largest, $t_1 < t_2 < \ldots < t_K$. The number of completed durations K is less than n because some observations are censored (i.e., their spells have not ended at the calendar time the study ends) and because of ties. Ties occur when two or more observations have the same duration.

As an example, consider the sample of strike lengths in days reported in Table 1. The data pertain to U.S. manufacturing industries for the period 1968 through 1976 and cover official strikes involving 1,000 workers or more with major issue classified as general wage changes by the Bureau of Labor Statistics. The data are given by John Kennan (1985, table 1).[7] We restrict our attention to strikes beginning in June of each year. This restriction increases the homogeneity of the sample and eliminates the need to consider the monthly dummy variables used by Kennan. We also censor strike lengths longer than 80 days because we are interested in illustrating methods for handling censored data. There were 62 strikes in this data set and there were 12 strikes that lasted 80 days or more, so about one-fifth of the observations are censored. There were a number of ties in the data: There were 4 strikes lasting 2 days, another 4 lasting 3 days, 2 lasting 9 days, and so on. In all the number of distinct completed durations K is 37.

Let h_j be the number of completed spells of duration t_j, for $j = 1, \ldots, K$. In the absence of ties the h_j are all equal to one. The h_j for the strike data are given in Table 1. Let m_j be the number of observations censored between t_j and t_{j+1}; m_K is the number of observations with durations greater than t_j, the

[7] A related application of duration methods to strike lengths is by Tony Lancaster (1972), who studies strike lengths using United Kingdom data.

TABLE 1

STRIKE DATA AND NONPARAMETRIC HAZARD AND SURVIVOR ESTIMATES

Ordered Duration Number j	Duration in days t_j	j_j	n_j	Hazard $\hat{\lambda}(t_j)$	Survivor $\hat{S}(t_j)$
1	1	1	62	.016	.984
2	2	4	61	.066	.919
3	3	4	57	.070	.855
4	4	1	53	.019	.839
5	5	1	52	.019	.823
6	7	1	51	.020	.806
7	8	1	50	.020	.790
8	9	2	49	.041	.758
9	10	1	47	.021	.742
10	11	1	46	.022	.726
11	12	2	45	.044	.694
12	13	1	43	.023	.677
13	14	1	42	.024	.661
14	15	1	41	.024	.645
15	17	1	40	.025	.629
16	19	1	39	.026	.613
17	21	2	38	.053	.581
18	22	1	36	.028	.565
19	23	1	35	.029	.548
20	25	1	34	.029	.532
21	26	1	33	.030	.516
22	27	2	32	.063	.484
23	28	1	30	.033	.468
24	29	1	29	.034	.452
25	32	1	28	.036	.435
26	33	1	27	.037	.419
27	35	1	26	.038	.403
28	37	1	25	.040	.387
29	38	1	24	.042	.371
30	41	1	23	.043	.355
31	42	1	22	.045	.339
32	43	2	21	.095	.306
33	44	1	19	.053	.290
34	49	2	18	.111	.258
35	52	2	16	.125	.226
36	61	1	14	.071	.210
37	72	1	13	.077	.194

Notes: h_j and n_j enter the calculation of $\hat{\lambda}(t_j)$ and $\hat{S}(t_j)$ and are defined in the text. The data are from John Kennan (1985, table 1).
Twelve observations are censored at duration 80 days.

longest complete duration. In a year-long panel, for example, spells of unemployment with lengths in weeks might be observed. Spells beginning 6 weeks before the end of the survey and still in progress at its end are regarded as censored between 6 and 7 weeks. The information content in this censored spell is that the actual duration is longer than 6 weeks. In the strike data the m_j are all equal to zero, except the last, m_{37}, which is equal to twelve. Let n_j be the number of spells neither completed or censored before duration t_j

$$n_j = \sum_{i \geq j}^{K} (m_i + h_i).$$

The n_j are given in Table 1 for the strike data. Now, the estimated survivor function will be a step function, as in the uncensored case. So will the corresponding estimated hazard function. The hazard $\lambda(t_j)$ is the probability of completing a spell at duration t_j, conditional upon the spell's reaching duration t_j. A natural estimator for $\lambda(t_j)$ is

$$\hat{\lambda}(t_j) = h_j/n_j,$$

the number of "failures" at duration t_j divided by the number "at risk" at duration t_j. The corresponding estimator for the survivor function is

$$\hat{S}(t_j) = \prod_{i=1}^{j} (n_i - h_i)/n_i = \prod_{i=1}^{j} (1 - \hat{\lambda}_i)$$

which is the *Kaplan-Meier* or *product-limit* estimator. Essentially, this estimator is obtained by setting the estimated conditional probability of completing a spell at t_j equal to the observed relative frequency of completion at t_j. Both $\hat{\lambda}(t_j)$ and $\hat{S}(t_j)$ are reported in Table 1 for the strike data. The development here is informal; however, it is possible to interpret this estimator as a maximum-likelihood estimator.[8]

The Kaplan-Meier estimator of the survivor function is related to the actuarial estimator of the *life table* familiar to demographers. For a life table, the time line is split up into fixed intervals, typically one-year intervals. A survival rate is then calculated for each interval. Let λ_i be the probability of completing a spell in the ith interval, conditional upon entering the ith interval. A natural estimator for λ_i is the fraction of those entering the interval who complete their spells during the interval. The situation is complicated by censoring, however. The actuarial estimator adjusts for censoring by subtracting one-half of the number of observations censored during the ith interval from the number entering the interval in calculating the fraction of completed spells. Label this estimator λ_i^a. Then the life table is estimated by $S_i^a = \Pi(1 - \lambda_j^a)$. The Kaplan-Meier estimator of the survivor function differs from the life table in that the intervals for which the hazards are calculated depend on the data.[9]

Plots of the hazard, integrated hazard, and log-integrated hazard against duration can be simpler to interpret than plots of the survivor function itself. For the exponential distribution, for example, the hazard is constant and the integrated hazard is linear in duration. The integrated hazard can be estimated by

$$\hat{\Lambda}(t_j) = \sum_{i \leq j} \hat{\lambda}(t_i).$$

Plots of the integrated hazard are typically smoother and therefore easier to interpret than plots of the hazard directly.[10]

Figure 5 illustrates the estimated survivor functions based on samples of size 25 from the exponential, Weibull, and log-logistic distributions shown in Figure 3. Note that, as a practical matter, the

[8] Greenwood's formula $\text{var}[S(t_j)] = [(S(t_j)]^2 \Sigma_{i<j} h_i/[n_i(n_i - h_i)]$ can be used to approximate the variance of the estimated survivor function. This formula is unlikely to be very useful for values of the survivor function near zero or one. See Kalbfleisch and Prentice (1980, section 1.3, pp. 13–15). Bruce Turnbull (1976) gives a method for estimating the distribution function in the presence of more complicated censoring and grouping of observations. Soren Johansen (1978) gives a maximum-likelihood interpretation of the product-limit estimator.

[9] A recent discussion of life tables, with references to the extensive literature, is by Chin-Long Chiang (1984).

[10] Alternatively, the integrated hazard could be estimated as minus the logarithm of the Kaplan-Meier estimator. The two methods give similar results when the λ are small.

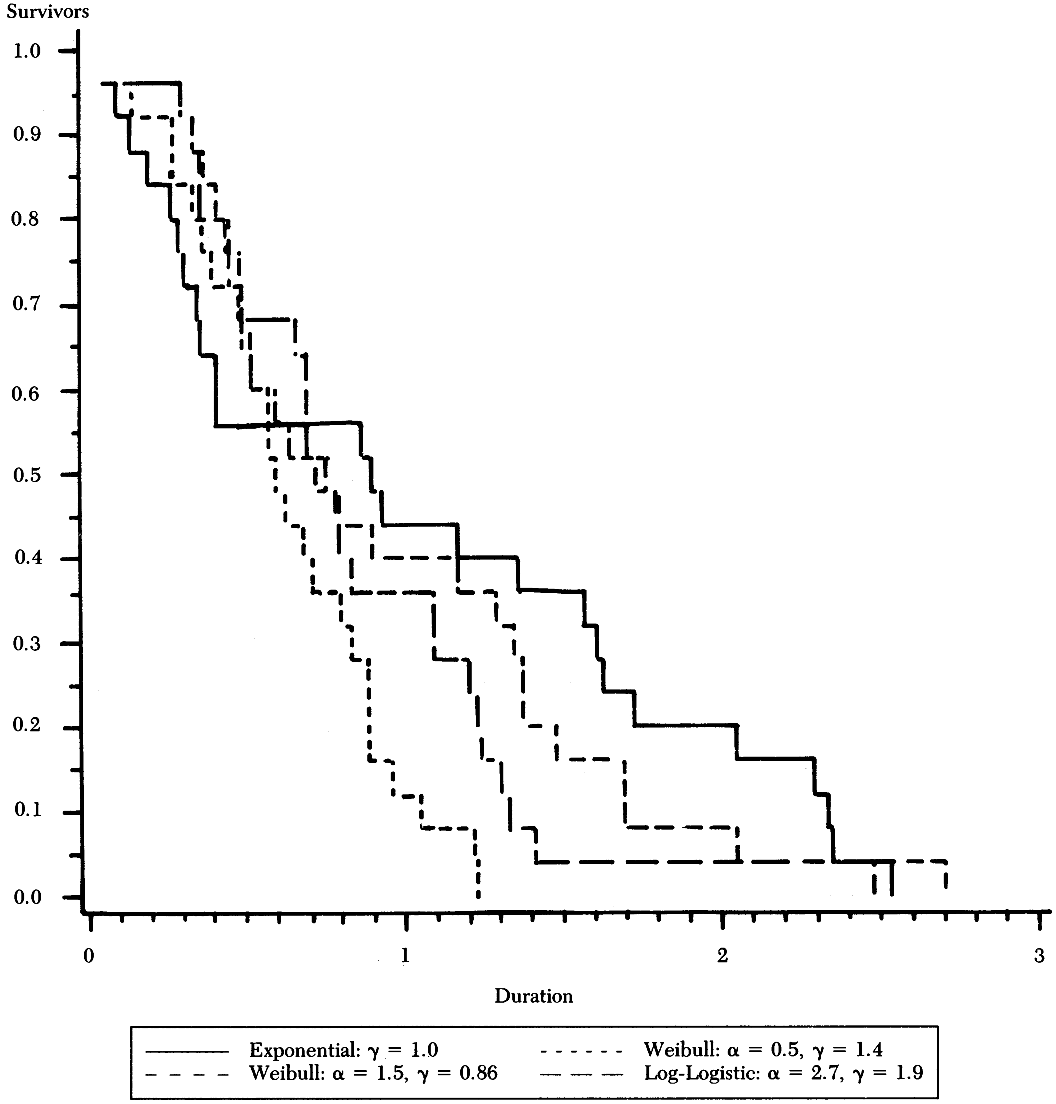


Figure 5. Sample Survivor Functions

accuracy of the estimator is better for shorter durations; inferences about very long durations are based on fewer observations. A comparison of the estimated survivor functions with the actual survivor functions of the distributions generating the data given in Figure 3 demonstrates this point.

The estimated survivor function for our sample of strike lengths in days is graphed in Figure 6. Figure 7 shows the estimated hazard function for our data. The hazard appears to have a slight upward slope, though it should be kept in mind that values for long durations are less precisely estimated than values for short durations. This estimate of the hazard function also exhibits an implausible

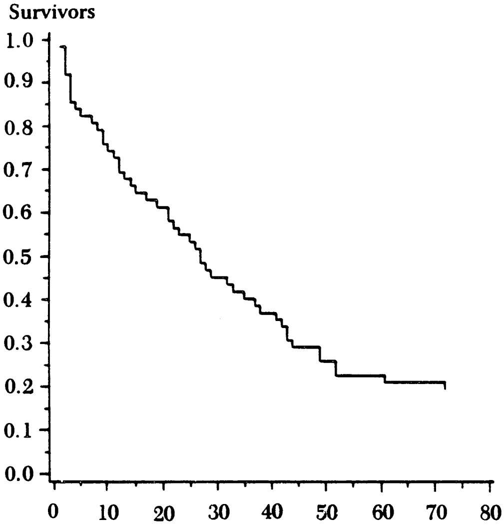


Figure 6. Strike Data Survivor

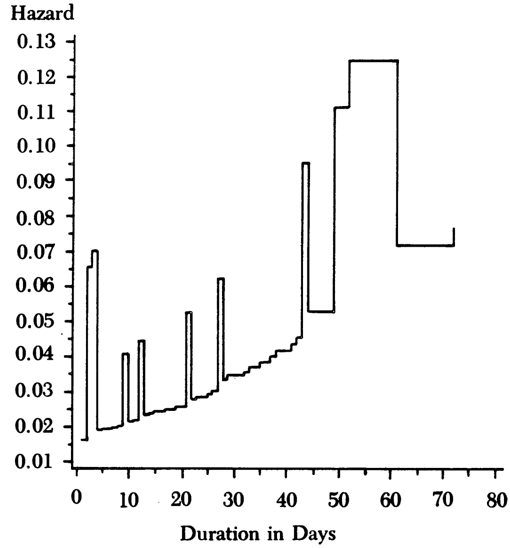


Figure 7. Strike Data Hazard

amount of variation over time (do we really think that the conditional probability a strike will end varies that much from day to day?). This is a problem with interpreting graphs of the empirical hazard function. A natural response by an economist is to smooth the plot somehow. A frequent device is to turn to a plot of the integrated hazard, which does involve smoothing. Figure 8 shows the integrated hazard for the strike data. To interpret Figure 8 keep in mind that the integrated hazard for the exponential distribution is a straight line. A convex integrated hazard implies that the hazard itself is increasing—positive duration dependence—while a concave integrated hazard (note that it can never decrease) implies a decreasing hazard or negative duration dependence. Figure 8 suggests that the exponential model may be adequate, and is certainly a sensible starting point for a parametric analysis.

B. *Parametric Methods*

Suppose that the family of duration distributions under consideration has been specified, so that the data distribution is known up to a vector of parameters θ. The family may have been chosen on the basis of a particular economic theory, convenience, and perhaps some preliminary plotting of data. Of course, specification analysis after the model has been fit may reveal that the family of distributions specified cannot adequately describe the data. For the present, we concentrate on estimating the parameter θ, and not on the issue of specification.

Write the density of a duration of length t as $f(t,\theta)$. If a sample of n completed spells were available and each individual's spell independent of the others, the likelihood function is

$$L^*(\theta) = \prod_{i=1}^{n} f(t_i,\theta)$$

as usual. In other words, the likelihood function is the joint probability distribution of the sample as a function of parameters θ. When a spell is censored, at duration t_j for example, the only information available is that the duration was at least t_j. Consequently the contribution to like-

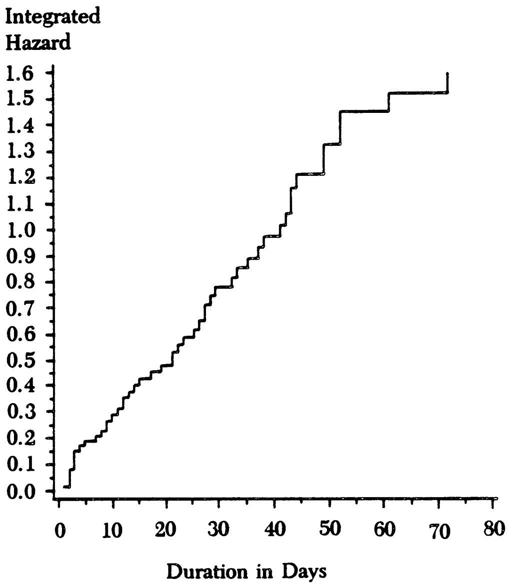


Figure 8. Integrated Hazard for Strike Data

lihood from that observation is the value of the survivor function, $S(t_j,\theta)$, the probability that the duration is longer than t_k. Let $d_k = 1$ if the kth spell is uncensored, $d_k = 0$ if censored. Then the likelihood function $L(\theta) = \ln L^*(\theta)$ is[11]

$$L(\theta) = \sum_{i=1}^{n} d_i \ln f(t_i,\theta) + \sum_{i=1}^{n} (1-d_i)\ln S(t_i,\theta)$$

which has completed spells contributing a density term $f(t_i,\theta)$ and censored spells contributing a probability $S(t_i,\theta)$. Using the fact that the density is the product of the hazard and the survivor function, $f(t,\theta) = \lambda(t,\theta)S(t,\theta)$, and the fact that the log of the survivor function is minus the integrated hazard $\ln S(t,\theta) = -\Lambda(t,\theta)$, the log-likelihood function can be written in terms of the hazard function

$$L(\theta) = \sum_{i=1}^{n} d_i \ln\lambda(t_i,\theta) - \sum_{i=1}^{n} \Lambda(t_i,\theta).$$

In practice it is usual to estimate the parameters by maximum likelihood.[12] Under a variety of well-known sets of sufficient conditions the maximum-likelihood estimator $\hat{\theta}$ is consistent for θ and $\sqrt{n}(\hat{\theta}-\theta)$ is asymptotically normally distributed with mean zero and a variance which can be consistently estimated by

$$V[\sqrt{n}(\hat{\theta}-\theta)] = -[n^{-1}\partial^2 L(\theta)/\partial\theta\partial\theta']^{-1}.$$

The variance can also be estimated using expected second derivatives of the log-likelihood, but this is problematic when censoring is present.[13] Hypothesis testing and construction of confidence intervals can be done in the usual way, based on the asymptotic distribution of the maximum-likelihood estimator.[14] Consider as an example the exponential model with $\lambda(t,\gamma) = \gamma$ and $\Lambda(t,\gamma) = \gamma t$. The log-likelihood function is then

$$L(\gamma) = \sum_{i=1}^{n} d_i \ln\gamma - \gamma \sum_{i=1}^{n} t_i$$

with first derivative

$$\partial L(\gamma)/\partial\gamma = \gamma^{-1}\sum_{i=1}^{n} d_i - \sum_{i=1}^{n} t_i.$$

When this is set to zero we derive the maximum-likelihood estimator

$$\hat{\gamma} = \sum_{i=1}^{n} d_i / \sum_{i=1}^{n} t_i$$

with approximate variance

[11] In writing this likelihood function we have used the assumption of independent censoring. In general, we would write down the joint distribution of the failure times and the censoring times and base the likelihood on that distribution. With independent censoring or fixed censoring times, the result is the likelihood function in the text.

[12] With the likelihood function in hand it is in principle possible to do a Bayesian analysis of the data, forming a posterior distribution for θ by combining the likelihood and a prior distribution.

[13] Taking expectation requires exact specification of the censoring mechanism.

[14] Alternatively, the likelihood ratio can be calculated; this has the advantage over the Wald procedure of being invariant to reparametrization. In some cases, particularly in calculating specification diagnostics, the Lagrange multiplier or score statistic will be more convenient.

$$V(\hat{\gamma}) = -[\partial^2 L(\gamma)/\partial\gamma^2]^{-1}$$
$$= \gamma^2/\sum_{i=1}^{n} d_i.$$

Note the effect of censoring. If censored spells were treated as complete, the maximum-likelihood estimator would be $\gamma_c = n/\Sigma t_i \geq \Sigma d_i/\Sigma t_i = \hat{\gamma}$. Ignoring censoring leads to upward asymptotic bias in the estimated hazard, or overstatement of the conditional probability of ending a spell.

The maximum-likelihood estimate of γ for the strike data is .0236 with asymptotic standard error .0033. So, the daily chance of settlement, given that the strike is still in progress, is almost 2.5 percent. The implied expected duration of a strike is 1/.0236 or about 42 days (with an approximate standard error of about 6 days). For comparison, the maximum-likelihood estimate obtained by treating censored spells as complete spells is .0432.

For the Weibull distribution with $\lambda(t,\gamma,\alpha) = \gamma\alpha t^{\alpha-1}$ and $\Lambda(t,\gamma,\alpha) = \gamma t^{a}$ the log-likelihood function is

$$L(\gamma,\alpha) = \sum_{i=1}^{n} d_i \ln\gamma + \sum_{i=1}^{n} d_i \ln\alpha + (\alpha - 1)\sum_{i=1}^{n} d_i \ln t_i - \gamma \sum_{i=1}^{n} t_i^{\alpha}.$$

Consideration of the derivatives of this function does not lead to closed-form expressions for the estimators of γ and α; however, numerical maximization is not difficult.[15]

Fitting a Weibull model to the strike data gives estimates γ = .033 with standard error .013 and α = .921 with standard error .091. Noting that the exponential distribution is Weibull with $\alpha = 1$ we conclude that the distribution of the strike data differs little from the exponential. This confirms our initial impression based on Figure 8.

IV. *Explanatory Variables in the Proportional Hazards Model*

Explanatory variables can affect the distribution of durations in many ways. In ordinary regression models it is natural to assume, at least as a starting point, that explanatory variables affect the distribution of the dependent variable by moving its mean around. There is no analogous clear-cut starting point for including explanatory variables in duration models. The proportional hazard specification (to be discussed below) is popular and simple to interpret: The effect of regressors is to multiply the hazard function itself by a scale factor. The accelerated failure time model (also discussed below) has seen less use in economics but is also easy to interpret: The effect of regressors is to rescale the time axis. More general models allow interaction between regressors and duration.

The interpretation of the coefficients of the explanatory variables depends on the specification. In the general case, the coefficient does not have a clean interpretation as a partial derivative analogous to the interpretation of coefficients in the linear regression model. The sign of the coefficient indicates the direction of the effect of the explanatory variable on the conditional probability of completing a spell. The numerical value of this effect, which is the partial derivative, depends on duration and in general on other included variables. Of course, this is true in nonlinear models whether or not the hazard function approach is taken.

In the important special cases of the proportional hazard model and the accelerated lifetime model the coefficients can be given partial-derivative interpreta-

[15] There is no fixed dimension sufficient statistic for γ and α. It is possible to concentrate the likelihood function using the fact that $t\alpha$ is exponential, but there is little to be gained in the way of convenience.

tions, analogous to that given to regression coefficients in the linear model.

These interpretations of coefficients are "purely statistical." Assigning economic meanings to coefficients is a matter of modeling and judicious use of prior information. In many cases it may be easier and more natural to model duration data in terms of hazard functions than in terms of densities; hence it may be easier to give parameters economic interpretations in the hazard function approach than in the more usual setting. Interpreting the coefficients of a model as behavioral parameters is a matter of judgment, however.

A. *The Proportional Hazard Specification*

A class of models that has been widely used in economics and other disciplines is the *proportional hazard* model. In this model the hazard function depending on a vector of explanatory variables x with unknown coefficients β and λ_0 is factored as

$$\lambda(t,x,\beta,\lambda_0) = \phi(x,\beta)\lambda_0(t)$$

where λ_0 is a "baseline" hazard corresponding to $\phi(\cdot) = 1$. It is common, and sensible, practice to measure the regressors so that $\phi(\cdot) = 1$ at the mean value of the regressors. Then λ has an interpretation as the hazard function for the mean individual in the sample. This baseline hazard is an unknown parameter which will (normally) require estimation. Note that the coefficients designated θ previously have been separated into β and λ_0. In this specification the effect of explanatory variables is to multiply the hazard λ_0 by a factor ϕ which does not depend on duration t. A specification of ϕ in general use is

$$\phi(x,\beta) = \exp(x'\beta).$$

This specification is convenient because nonnegativity of ϕ does not impose restrictions on β and estimation and inference are straightforward. As we will see, estimation of β in this model does not require specification of the baseline hazard λ_0.

With the proportional hazard specification we have

$$\partial \ln\lambda(t,x,\beta,\lambda_0)/\partial x = \partial \ln\phi(x,\beta)/\partial x$$

so the proportional effect of x on the conditional probability of ending a spell does not depend on duration. In the important special case $\phi(x,\beta) = \exp(x'\beta)$

$$\partial \ln\lambda(t,x,\beta,\lambda_0)/\partial x = \beta$$

so the coefficient can be interpreted as the constant proportional effect of x on the conditional probability of completing a spell. This is the analog, in a hazard function setting, of the usual partial-derivative interpretation of a linear regression coefficient.

The proportional hazard model with $\phi(x,\beta) = \exp(x'\beta)$ admits a convenient interpretation as a linear model. With

$$\lambda(t,x,\beta,\lambda_0) = \exp(x'\beta)\lambda_0(t)$$

the survivor function for t is given by (using formulas from Section IIa)

$$S(t) = \exp[-\Lambda_0(t)\exp(x'\beta)]$$

where $\Lambda_0(t) = \int \lambda_0(u)du$ is the integrated baseline hazard. To obtain a linear model interpretation we consider the random variable ϵ defined by $\epsilon = -\ln\Lambda_0(t) - x'\beta$. The motivation for this bizarre transformation is that the distribution of the resulting ϵ does not depend on Λ_0 or on $x'\beta$, yielding a linear model for the transformed dependent variable $t^* = -\ln\Lambda_0(t)$. To calculate the distribution of ϵ, write

$$\begin{aligned} Pr(\epsilon < E) &= Pr[-\ln\Lambda_0(t) < E + x'\beta] \\ &= Pr[\ln\Lambda_0(t) > -E - x'\beta] \\ &= Pr[\Lambda_0(t) > \exp(-E - x'\beta)] \\ &= Pr\{t > \Lambda_0^{-1}[\exp(-E - x'\beta)]\}. \end{aligned}$$

This probability can be evaluated using the survivor function for t, giving

$$Pr(\epsilon < E)$$
$$= \exp\left(-\Lambda_0\{\Lambda_0^{-1}[\exp(-E - x'\beta)]\}\exp(x'\beta)\right)$$
$$= \exp[-\exp(-E)]$$

which is the cumulative distribution function for the type 1 extreme value distribution (see Norman Johnson and Samuel Kotz 1970, p. 272). Thus we can write the proportional hazard model in the form

$$-\ln\Lambda_0(t) = t^* = x'\beta + \epsilon,$$

a linear model for t^* in which the error term has a fully specified (but not normal) distribution.

This representation of the model suggests the possibility that least-squares regression methods can be used to estimate the coefficients of explanatory variables. If the data are not heavily censored, this is indeed possible although it requires knowledge of the baseline integrated hazard. (Sometimes Λ_0 involves parameters that must be estimated.) Of course, test statistics must be viewed skeptically because of the nonnormality of the error term. Further, a correction to the intercept estimate must be made to account for the nonzero mean of the error term. For this reason an intercept must be included if least squares is applied even when the explanatory variables are measured in deviations. The regression estimator is illustrated below. Censoring can be accommodated using Tobit-like methods and adjustments, based on the extreme value distribution instead of the normal; however, this makes the estimation problem nonlinear and maximum likelihood is no more difficult and preferred on efficiency grounds.

To illustrate the linear model interpretation we consider the exponential model with parameter $\gamma = \exp(x'\beta)$ and $\lambda_0(t) = 1$. The latter specification is a normalization. As long as an intercept is included in x, the baseline hazard λ_0 is identified only up to a constant factor. Measuring explanatory variables in deviations from means, adding an intercept, and defining $\lambda_0 = 1$ means that exp(constant term) is an estimate of the baseline hazard, that is, the hazard function for the individual with mean values of the explanatory variables. Then we have the linear model

$$-\ln t = t^* = x'\beta + \epsilon.$$

In the Weibull model with $\gamma = \exp(x'\beta)$ and $\lambda_0(t) = t^\alpha$,

$$-\alpha\ln t = t^* = x'\beta + \epsilon.$$

This representation of the two models gives a hint about the bias to be expected in the absence of censoring when a Weibull model is misspecified as exponential. When α is greater than one, so that Weibull time is faster than calendar time, coefficients are likely to be underestimated. When α is less than one, coefficients are likely to be overestimated.

The fact that the distribution of ϵ is known can be used as the basis of specification tests once the parameters are estimated by whatever means. For example, if regressors are omitted, the model is

$$t^* = x'\beta + \epsilon + v$$

where v captures the omitted regressors. The error term $\epsilon + v$ will have more variance than the specification predicts. Similarly, we see that if the wrong transformation of the dependent variable is used the explanatory variables may appear to enter nonlinearly.

If ϕ does not take the simple form $\exp(x'\beta)$ an interpretation as a nonlinear regression model with additive errors with known distribution can be developed. For the specification $\lambda(t,x,\beta) = \phi(x,\beta)\lambda_0(t)$ we can repeat previous arguments to obtain

$$-\ln\Lambda_0(t) = t^* = \ln\phi(x,\beta) + \epsilon$$

where ϵ has an extreme value distribution (again, not a normal distribution).

B. *Parametric Estimation*

Parametric estimation of proportional hazard models is a straightforward extension of the techniques discussed in Section IIIB. The baseline hazard can be chosen from a parametric family and written $\lambda_0 = \lambda_0(t,\alpha)$. The hazard function is then

$$\lambda(t,x,\alpha,\beta) = \phi(x,\beta)\lambda_0(t,\alpha)$$

and the log-likelihood function is

$$L(\alpha,\beta) = \Sigma d_i \ln\lambda(t_i,x_i,\alpha,\beta) - \Sigma\Lambda(t_i,x_i,\alpha,\beta)$$

where $d_i = 1$ if the ith spell is uncensored and Λ is the integrated hazard corresponding to λ. As a concrete example of the estimation technique, consider the popular *exponential regression* specification

$$\lambda(t,x,\beta) = \exp(x'\beta)$$

in which we have incorporated the assumption that $\lambda_0(t)$ is constant and the normalization that it is equal to one. This specification is a natural starting point for analysis in the absence of specific information on the shape of the baseline hazard. Residual analysis and diagnostics may then suggest respecification. The coefficients in the exponential regression specification can be given an interpretation in terms of effects on expected duration. The mean of the exponential $f(t) = \gamma\exp(-\gamma t)$ is $1/\gamma$ so expected duration for this model is $\exp(-x'\beta)$. Hence,

$$\partial\ln(\text{expected duration})/\partial x = -\beta.$$

The log-likelihood function is

$$L(\beta) = \Sigma d_i x_i'\beta - \Sigma t_i \exp(x_i'\beta)$$

with derivatives

$$\partial L/\partial\beta = \Sigma d_i x_i' - \Sigma t_i \exp(x_i'\beta)x_i'$$
$$\partial^2 L/\partial\beta\partial\beta' = -\Sigma t_i \exp(x_i'\beta)x_i x_i'.$$

The log-likelihood function is concave and numerical maximization is straightforward.[16] The negative inverse of the second derivative matrix can be used as an approximate covariance matrix for the estimator of β.

Parametric estimation of other specifications is similar. Of course, when the log-likelihood function is not concave, special care should be taken to insure that the global maximum is attained. In this case, the behavior of the log-likelihood may not be adequately described by local characteristics at the maximum. Unusual features of the log-likelihood function should be examined and interpreted.

In his study of strike duration Kennan (1985) controlled for the effects of variations in general economic conditions by using an index of industrial production. The index chosen is the residual from a regression of the logarithm of industrial production in manufacturing on time, time squared, and monthly dummy variables. The values of this index are given in Kennan (1985, appendix). A fixed value of the industrial production index is associated with each strike. Fitting the exponential model with $\lambda(x,\beta) = \exp(\beta_0 + \beta_1 x)$ where x is the industrial production index yields estimates $\beta_0 = -3.72$ with standard error .143 and $\beta_1 = 10.21$ with standard error 3.34. The estimate of β_0 implies that expected strike length when industrial production is on its trend line (that is, when $x = 0$) is about 41 days. High values of the production index are associated with high values of the conditional probability of

[16] Newton's method iterating according to $\beta^{n+1} = \beta^n - (\partial^2 L/\partial\beta\partial\beta')^{-1}\,\partial L/\partial\beta$ with derivatives evaluated at β^n, appears to work well. Sometimes it is useful to multiply the correction to β^n by a factor less than one. A practical suggestion for starting values is to measure the x variables as deviations from means, start their coefficients at zero, and include a constant term with starting value minus the log of mean duration.

ending a strike. Our estimate of β_1 implies that a 1 percent unexpected increase in industrial production decreases expected strike length by about 10 percent, or about 4 days. Fitting the Weibull model $\lambda(t,x,\beta,\alpha) = \exp(\beta_0 + \beta_1 x)t^{\alpha-1}$ yields $\beta_0 = -3.79$ (standard error, .443), $\beta_1 = 9.38$ (standard error, 3.10), and $\alpha = 1.003$ (standard error, .101). The Weibull alternative provides support for the exponential specification, because α is insignificantly different from one, and confirms the significant countercyclical pattern of strike durations. This is an interesting finding (and it holds up in the full data set; see Kennan 1985); especially in view of the well-known finding that strike *incidence* is procyclical. Canadian data also provide evidence that strike durations are countercyclical; see Alan Harrison and Mark Stewart (1987).

By way of comparison we consider the ordinary least-squares regression of the logarithm of duration on the industrial production index. Regressions were run on two samples, one with the censored observations included at their censored values (namely 80 days) and the other omitting the censored observations. We are primarily concerned with the slope estimates giving the effect of the state of the economy on duration. Intercepts are included in both specifications. Including the censored observations results in the estimated model

$$\ln(t) = 3.11 - \underset{(3.39)}{8.12}x \qquad R^2 = .041 \quad n = 62.$$

Omitting the censored observations leads to

$$\ln(t) = 2.78 - \underset{(3.19)}{4.83}x \qquad R^2 = .098 \quad n = 50.$$

Recall that, in the hazard rate formulations, β_1 represents the effect of increases in x on the conditional probability of ending a strike whereas in the ordinary least-square regressions the coefficient on x measures the effect of increases in industrial production on the length of the strike. In other words, the slope coefficient in these least-squares regressions should be opposite in sign to the coefficient in the hazard rate formulation. The numbers in parentheses are standard errors—inappropriate, of course, because their computation has not taken account of censoring.

The maximum-likelihood estimate of β_1 in the exponential formulation of the hazard rate was 10.21, whereas the least-squares regression excluding the censored observations results in an estimate (4.83) less than one-half in absolute value of the maximum-likelihood estimate. The regression including the censored observations at their censored values gives an estimate of 8.12 with estimated standard error 3.39. Clearly the treatment of the censored values is critical to the result. In this example the regression including the censored observations results in estimates closer to the maximum-likelihood estimates. It is possible to construct examples in which the regression excluding the censored observations work better in this sense.

C. *Partial Likelihood*

The partial-likelihood approach suggested by Cox (1972, 1975) can be used to estimate β in the proportional hazard model without specifying the form of the baseline hazard function λ_0. Suppose the completed durations are ordered, $t_1 < t_2 < \ldots < t_n$. For the present, suppose there is no censoring and there are no ties in the durations. The conditional probability that observation 1 concludes a spell at duration t_1, given that any of the n observations could have been concluded at duration t_1, is

$$\frac{\lambda(t_1,x_1,\beta)}{\sum_{i=1}^{n} \lambda(t_1,x_i,\beta)}.$$

With the proportional hazard assumption $\lambda(t,x,\beta) = \phi(x,\beta)\lambda_0(t)$, this expression reduces to

$$\frac{\phi(x_1,\beta)}{\sum_{i=1}^{n} \phi(x_i,\beta)}$$

and this quantity is the contribution of the shortest duration observed to the partial likelihood. Similarly, the contribution of the jth shortest duration is $\phi(x_j,\beta)/\Sigma_{i=j}^{n} \phi(x_i,\beta)$. In each case, the contribution to likelihood is the ratio of the hazard for the individual whose spell was completed at duration t divided by the sum of the hazards for individuals whose spells were still in progress just prior to time t (i.e., those whose spells could have ended at duration t). The likelihood is formed as the product of the individual contributions, and the resulting log-likelihood function is

$$L(\beta) = \sum_{i=1}^{n} \{\ln\phi(x_i,\beta) - \ln[\sum_{j=i}^{n} \phi(x_j,\beta)]\}.$$

The intuition here is that, in the absence of all information about the baseline hazard, only the order of the durations provides information about the unknown coefficients.

Censoring is easily handled in the partial-likelihood framework. An individual whose spell is censored between duration t_j and t_{j+1} appears in the summation in the denominator of the contribution to log-likelihood of (ordered, uncensored) observations 1 through j, but not in any others. Censored spells do not enter the numerator of a contribution to likelihood at all. Ties can be handled by the common device of including a contribution to likelihood for each of the tied observations, using the same denominator for each. The negative inverse second derivative matrix of the log-likelihood function can be used to approximate the variance of the coefficient estimator.[17]

Given estimates of β, we may ask whether it is possible to use these in constructing a sensible nonparametric estimate of the baseline hazard $\lambda_0(t)$. Define $D(t_k)$ to be the denominator of the likelihood contribution of the k^{th} ordered, uncensored observation. When no observations are censored $D(t) = \Sigma_{i=k}^{n} \phi(x_i,\beta)$; in general the sum will include additional terms for observations not censored by duration t_k but censored later. A natural nonparametric estimator for the integrated hazard function, analogous to that described in Section III, is given by

$$\hat{\Lambda}_0(t_k) = \sum_{j=1}^{k} d(t_j)/D(t_j)$$

where $d(t_j)$ is the number of spells ended at duration t_j; $d(t_j) = 1$ in the absence of ties. $D(t)$ is evaluated at the estimated value of β. This is essentially the estimator suggested by Cox and Oakes (1985). If the baseline $\hat{S}_0$ survivor function itself is desired, it can be estimated by $\hat{S}_0(t_k) = \exp[-\hat{\Lambda}_0(t_k)]$.

The partial-likelihood method applied to the strike data gives an estimate of β_1, the coefficient on the industrial production index, equal to 9.08 (with standard error 3.40). The constant term is not estimated in this procedure. The partial-likelihood result, which does not depend on assumptions about the shape of the baseline hazard function, is in agreement with the results from the exponential and Weibull specifications. Strike duration is significantly countercyclical.

[17] It can be shown that the partial likelihood can be treated as an ordinary likelihood or as a likelihood function concentrated with respect to λ_0. See Per Kragh Andersen and Richard Gill (1982) and Johansen (1983). The efficiency loss from using partial rather than full likelihood is discussed by Brad Efron (1977).

V. Explanatory Variables in Other Models

A. Accelerated Lifetimes

In the *accelerated lifetime* model the effect of explanatory variables is to rescale time directly. If the baseline survivor function is $S_0(t)$, then the survivor function for an individual with characteristics x is

$$S(t,x,\beta) = S_0[t\phi(x,\beta)]$$

in which essentially time is rescaled by multiplication by ϕ.

The hazard function associated with S is easily seen to be

$$\lambda(t,x,\beta) = \lambda_0[t\phi(x,\beta)]\phi(x,\beta)$$

where $\lambda_0 = -d\ln S_0/dt$ is the hazard function for the distribution S_0. The density is $f(t,x,\beta) = f[t\phi(x,\beta)]\phi(x,\beta)$.

In the important special case $\phi(x,\beta) = \exp(x'\beta)$ the accelerated lifetime model can be given a linear model interpretation. Changing variables to $v = -\ln t - x'\beta$, using an argument parallel to that in Section IVA, allows us to rewrite the model as

$$-\ln t = x'\beta + v$$

where v has the density $f_0[\exp(-v)]\exp(-v)$, which does not depend on explanatory variables x. This representation of the model can lead to specification checks and suggestions for informative plots, as in the discussion following the linear model representation of the proportional hazard model. It also gives a convenient interpretation of the coefficients of explanatory variables:

$$\partial\ln t/\partial x = -\beta.$$

It is instructive to compare the linear model representations of the proportional hazard and accelerated lifetime models. In the former the model is linear in x when the dependent variable is $-\ln\Lambda_0(t)$; because the class of possible specifications $\Lambda_0(t)$ is large, the model is quite general in this regard in comparison with the accelerated lifetime model, which is linear in x for dependent variable $-\ln t$. On the other hand, the distribution of ϵ, the error term in the proportional hazard model, is in a specific form, the type 1 extreme value distribution. The form of the distribution of the error term v in the accelerated lifetime model is not restricted to a single distribution. Indeed, the error distribution can be normal, although this specification has not seen much use in applications where the hazard functions—or conditional probabilities generally—are of particular interest because the hazard takes a complicated form involving the incomplete normal integral. Thus, the two specifications are general in different ways. The proportional hazard model restricts the distribution of the additive error but allows fairly general transformations of the duration variable to achieve linearity in regressors. The accelerated lifetime model restricts the transformation of duration but allows fairly general error distributions.

The exponential specification with λ_0 exponential and $\phi(x,\beta) = \exp(x'\beta)$ is both a proportional hazard and an accelerated lifetime model, because $\Lambda_0(t) = t$. As a general proposition, with ϕ exponential in $x'\beta$, the Weibull family is the only family in the proportional hazard class in which $\ln t$ is linear in x.

At this point it is useful to reconsider the regression results reported in Section II. If the exponential regression specification is correct, then in the absence of censoring, we would expect the ordinary least-squares slope estimate to be close to the maximum-likelihood estimate. They are both consistent in this circumstance, though the least-squares estimator is inefficient. On the other hand, if the exponential regression specification is incorrect, the least-squares estimator may still be consistent, if the true specification is an accelerated lifetime model. Again, we are abstracting from the difficult problem of censoring. Thus, in the absence of censoring, a substantial difference between the least-squares slope estimate and the maximum-likelihood slope estimate is cause for concern about the assumed specification. The matter can be settled, perhaps, by an examination of the regression residuals: Recall that under the exponential regression hypothesis the error term has an extreme value distribution. When censoring is present, the regression results are unreliable whether the exponential specification is satisfactory or not; however, methods for checking the specification using residuals based on maximum-likelihood estimates can be devised. One method is illustrated in Section VIE.

Estimation of the accelerated lifetime model proceeds by choosing functional forms for λ_0 and ϕ and maximizing the resulting log-likelihood function. For example λ_0 could be Weibull or log-logistic and a natural choice for ϕ is $\exp(x'\beta)$. Specification analysis can proceed informally by plots meant to detect departures from the linear model assumption. Nonlinear functions $\phi(x,\beta)$ can also be specified and estimated by maximum likelihood. The linear model interpretation of the accelerated lifetime model generalizes naturally to a nonlinear model with additive errors interpretation.

B. *Other Specifications*

Both the proportional hazard and the accelerated lifetime specifications substantially restrict interdependence between the explanatory variables and duration in determining the hazard. In many cases the data available and the problem at hand will not require more generality. In others, more flexible interaction between x and t must be allowed. If the explanatory variable thought to interact with t is discrete or can be conveniently allocated to separate groups, it may be possible to group the data and use a proportional hazard (or accelerated lifetime) model within groups. Comparison of the estimated integrated hazard or survivor functions may then yield insight into the appropriate specification of the dependence of the hazard on x and t. Because the hazard function must be nonnegative, a convenient specification is

$$\lambda(t,x,\beta) = \exp[g(t,x,\beta)]$$

where the function g is somewhat arbitrary but can be specified to include polynomials and step functions in t and x as necessary. The likelihood function based on this hazard can be used to obtain parameter estimates. It should be stressed that headstrong specification and estimation of complicated models is typically to be avoided; without considerable care the resulting estimates may be just as difficult to interpret as the raw data. A sensible research strategy is to start with simple models and use residual plots and other diagnostics to detect problems. A residual plot is considered below (Section VIE).

C. *Time Varying Explanatory Variables*

Regressors whose values change over the course of spells are conceptually straightforward to handle in the hazard function framework, though experience with these models is limited. Suppose the regressor x is a function of time, $x(t)$, where t is measured from the beginning of the spell (a simple change of notation accommodates the case in which x depends on calendar time rather than duration). Write the hazard function as $\lambda[t,x(t),\theta]$. Using our integration formulas we can write the integrated hazard, survivor, and density functions. These will in general depend on the entire time path (up to t) of the regressor. For example, the integrated hazard is

$$\Lambda(t) = \int_0^t \lambda[u,x(u),\theta]du.$$

The log-likelihood function for n independent observations is

$$L(\theta) = \sum_{i=1}^{n} d_i \ln\lambda[t_i,x(t_i),\theta] - \sum_{i=1}^{n} \int_0^{t_i} [u,x_i(u),\theta]du.$$

Estimation of the parameter θ typically requires numerical maximization of the log-likelihood function. In practice, the regressor may change only once or a few times over the course of a spell and the integral may therefore be simplified into a summation of a few terms.

No apparent simple interpretation of this model in terms of linear models is available. Identification is tricky, in that the effect of trending regressors is difficult to separate from possible duration dependence. Thus, for estimates to be precise, the time paths of regressors must vary substantially across individuals. The practical problems of specifying, estimating, and interpreting models with time-varying explanatory variables are still important areas of active current research. In a recent study Tony Atkinson and John Micklewright (1985) note that transfer payments received during spells of unemployment in the United Kingdom vary over the course of a spell. Allowing for this in estimating the disincentive effects of unemployment insurance turns out to be important.

VI. *Specification Checking*

A. *Functional Form Misspecification and Heterogeneity*

Heterogeneity arises in a population when different individuals in the population have potentially different distributions of the dependent variable. Explanatory variables are included in econometric models to control for heterogeneity. If the control for the effects of related explanatory variables is incomplete and so some heterogeneity remains, problems can arise in interpreting the data. Control can be incomplete as a result of functional form misspecification: For example, if the effect of an explanatory variable x enters as $\exp(\beta_0 + \beta_1 x + \beta_2 x^2)$, but instead the model $\exp(\beta_0 + \beta_1 x)$ is specified, heterogeneity is present as a result of the omitted term $v = \beta_2 x^2$. Heterogeneity can also arise when unobservable variables are important: For example, "spunk" in job-search efforts could affect an individual's transition rate from unemployment to employment.

We will use the term *heterogeneity* to refer to differences remaining in distributions after controlling for the effect of observable variables. Thus, heterogeneity occurs as a problem if individuals have differing duration distributions after control for explanatory variables. Functional form misspecification leads to heterogeneity, so our discussion of the effects of heterogeneity encompasses the effects of misspecification. As in the simple linear regression model, omission of important variables or misspecification of included variables often leads to inaccurate inferences. Heterogeneity in duration models leads to misleading inferences about duration dependence and, potentially, to misleading inferences about the effects of included explanatory variables. We will discuss these in turn.

B. *Heterogeneity and Inference About Duration Dependence*

The effects of heterogeneity on apparent duration dependence can be illustrated simply in a model with no explanatory variables and a simple form of heterogeneity. Suppose that the fraction p of the individuals in a population being studied have hazard function $\lambda_1(t) = \gamma_1$, and fraction $(1 - p)$ have hazard function $\lambda_2(t) = \gamma_2$, where both γ_1 and γ_2 are constants. For example, men and women might have different distributions for duration. The density functions for each of the two groups are $f_1(t) = \gamma_1 e^{-\gamma_1 t}$ and $f_1(t) = \gamma_2 e^{-\gamma_2 t}$. Suppose we assume in specifying a model that the population is made up of two subgroups. Then, we are sampling from the *mixture* distribution $f(t) = pf_1(t) + (1 - p)f_2(t)$. A randomly selected observation has probability p of being a draw from $f_1(t)$ and probability $(1 - p)$ of being a draw from $f_2(t)$. Because we do not know which subgroup each observation is drawn from, or we do not choose to use this information if it is available, we are effectively sampling from the mixture $f(t)$.

Before studying duration dependence in the mixture distribution, it is useful to note that the misspecification we are considering can be regarded as a left-out regressor problem. Define $x = 0$ if the observation is from subgroup 1, $x = 1$ if the observation is from subgroup 2. Then the hazard function for the ith observation can be written $\lambda(t,x_i) = \gamma_1 + x_i (\gamma_2 - \gamma_1)$. Of course, we would be unlikely to specify a hazard function linear in a regressor, so we might instead write $\lambda(t,x) = \exp(\beta_0 + \beta_1 x)$ where the new coefficients are $\beta_0 = \ln\gamma_1$ and $\beta_1 = \ln\gamma_2 - \ln\gamma_1$. Treating the population as homogenous is the same as leaving out the regressor x.

The hazard function for each individual in the population is a constant. The hazard function for the mixture distribution is what we estimate. Using the formula $\lambda(t) = f(t)/S(t)$ we find

$$\lambda(t) = \frac{p\gamma_1 e^{-\gamma_1 t} + (1 - p)\gamma_2 e^{-\gamma_2 t}}{pe^{-\gamma_1 t} + (1 - p)e^{-\gamma_2 t}}$$

which is not constant in duration (t). Some tedious calculating shows that $d\lambda(t)/dt$ is negative—it has the sign of $-(\gamma_1 - \gamma_2)^2$. If we fit a model allowing duration dependence to these data we will find evidence of negative duration dependence.

The result that heterogeneity leads to a downward biased estimate of duration dependence holds quite generally and is intuitively sensible. Consider a random sample of individuals from our mixture distribution. When the sample is taken, about fraction p of the individuals will be from subgroup 1 and $1 - p$ from subgroup 2. Suppose that the hazard is higher for members for group 1. As time elapses, individuals in group 1 will complete their durations at a higher rate than individuals in group 2. Thus, as time passes, the fraction of individuals from group 1 remaining in the sample falls. Because group 2 individuals have a lower hazard function, the decline in the fraction of individuals from group 1 shows up as a decline in the hazard function over time.

C. *Heterogeneity and Inference About Regressors*

Suppose that we fit an exponential regression model $\lambda(x,\beta) = \exp(x_1'\beta_1)$ when the true hazard function is $\exp(x_1'\beta_1 + x_2'\beta_2)$. In this setup neither the true nor the specified model allows duration dependence, allowing a focus on the effect of heterogeneity on coefficients of explanatory variables. Of course, if we did look for duration dependence within the framework of our model based on x_1 alone, then for reasons similar to those outlined above we would find evidence of negative duration dependence. An approximation to the asymptotic bias in the maximum likelihood estimate of β_1 is

$$\hat{\beta}_1 - \beta_1 \approx \left(\frac{\partial^2 L}{\partial \beta_1' \partial \beta_1}\right)^{-1} \left(\frac{\partial^2 L}{\partial \beta_1' \partial \beta_2}\right) \beta_2.$$

In the absence of censoring, the expectations of these second derivatives can be evaluated under the hypothesis that $\beta_2 = 0$. The resulting formula for the approximation bias in $\hat{\beta}_1$ is

$$\text{bias} \approx \left(\sum_{i=1}^{n} x_{1i}x_{1i}'\right)^{-1} \left(\sum_{i=1}^{n} x_{1i}x_{2i}'\right)\beta_2$$

exactly the same as the effect of an omitted regressor in the ordinary linear model. Here, however, the formula holds locally: It is a linear approximation to the bias.[18]

D. *Heterogeneity and Simultaneous Inference on Duration Dependence and Regression Coefficients*

Misspecification of the hazard function leads quite generally to downward bias in the effect of duration on the probability of completing a spell even when regression coefficients are estimated simultaneously. The effect of misspecification on the coefficients of explanatory variables when duration dependence is allowed is much more complicated. Tony Lancaster (1985) studies the effects of heterogeneity in the Weibull model $\lambda(t,x,\beta,\alpha) = t^{\alpha-1} \exp(x'\beta)$. He finds that the maximum likelihood estimator of α is biased downward, as expected. Under the additional assumption that the heterogeneity term is independent of x, the maximum-likelihood estimator of β is biased toward zero. Sunil Sharma (1987) finds, in the same setting, that heterogeneity induces dependence between the effects of included regressors and duration, so that the proportional hazard hypothesis appears to be violated as a result of misspecification. These are asymptotic calculations. Note that heterogeneity arising from functional misspecification is unlike to satisfy this independence condition.

C. A. Struthers and John D. Kalbfleisch (1986) study the effects of misspecification in the proportional hazard model estimated by the partial likelihood method. They find that fitting a proportional hazard model when the true process generating the data is an accelerated lifetime model leads to bias, but that the relative effects of the regressors are correctly estimated to first order. If a proportional hazard model is fit, but relevant explanatory variables, independent of included variables, are left out the remaining coefficients are biased toward zero, with the size of the bias depending on the effect of the omitted regressor. These results are asymptotic. Of course, if the hazard function is calculated from the estimated coefficients, the estimated hazard can be expected to be systematically biased.[19]

E. *Informal Methods of Specification Checking*

Careful residual analysis, as in the linear regression model, is the key to assessing a specification. Residual plots can reveal surprising departures from a hypothesized model, and can sometimes suggest directions in which to improve the specification. Perhaps the easiest procedure is to examine estimated values of the integrated hazard $\Lambda(t_k,x,\hat{\theta})$. If the specification is correct, these values should look in the absence of censoring like a sample from the unit exponential distribution. Let

$$\epsilon_k = \Lambda(t_k,x,\hat{\theta}).$$

[18] Details of this local approach to asymptotic specification analysis are given by Nicholas M. Kiefer and Gary Skoog (1984).

[19] Geert Ridder and Wim Verbakel (forthcoming) obtain similar results on the proportional hazard model using different methods.

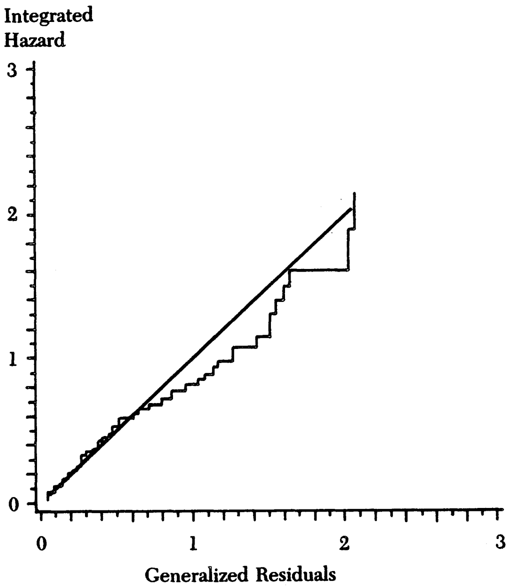


Figure 9. Integrated Hazard for Generalized Residuals from Exponential Regression (integrated hazard for standard exponential included)

To find the distribution of ϵ,

$$\begin{aligned} Pr(\epsilon < E) &= Pr(\Lambda(t_k,x,\hat{\theta}) < E) \\ &= Pr[t_k < \Lambda^{-1}(E,x,\hat{\theta})] \\ &= 1 - \exp(-E) \end{aligned}$$

using the fact that $F(t) = 1 - \exp[-\Lambda(t,x,\theta)]$ and assuming that $\hat{\theta} = \theta$. The estimated integrated hazard function for ϵ can be compared with the 45 degree line, the actual integrated hazard for the standard exponential distribution. The ϵ_k are known as *generalized residuals*. They are analogous to the ordinary residuals $\hat{\epsilon} = y - x\hat{\beta}$ in the linear models.[20]

Figure 9 shows the integrated hazard for the ϵ_k from the strike data and the exponential specification $\lambda(t,x,\theta) = \exp(\beta_0 + \beta_1 x)$. The solid line is the actual integrated hazard function for the standard exponential. The agreement appears good—especially when we recall that the values of the sample integrated hazard at longer durations have higher sampling variances.

Another method of assessing specification is to split the sample, perhaps into groups based on values of the explanatory variables, and fit the model to the groups separately. If the specification is correct, the estimated parameters should agree (up to estimation error). If semiparametric methods of estimating the hazards are used, these should provide similar estimates across groups. As in the linear regression model, different coefficients on explanatory variables across groups may indicate a need for respecification, perhaps involving nonlinearities (or additional nonlinearities).

F. *Tests and Diagnostics*

Tests of hypotheses about the coefficient vector can typically be constructed using the asymptotic distribution of maximum-likelihood estimators and the Wald, score (Lagrange multiplier), or likelihood ratio method. Choice among the three asymptotically equivalent methods depends primarily on convenience. If construction of confidence intervals is particularly important the likelihood ratio method is probably to be preferred, because it allows the data to determine the shape of multivariate confidence regions while direct use of the asymptotic distribution leads to elliptical regions.

How can the fit be assessed? In nonlinear models with explanatory variables it is useful to report a chi-square test for the joint hypothesis that all coefficients apart from a constant are equal to zero. This statistic is analogous to the F-statistic in a linear regression. It provides the reader with an indication of how much the explanatory variables jointly contribute to the fit of the model. Note, how-

[20] The ϵ are generalized residuals in the sense of Cox and Snell (1968). Residual analysis for economic duration data is discussed by Lancaster and Andrew Chesher (1985). See also J. Crowley and M. Hu (1977).

ever, that a high value of this statistic does not necessarily mean that the model is satisfactory. Residual plots and possibly calculation of diagnostics based on residuals are in order.

A simple and frequently used diagnostic statistic is based on the fact that the generalized residuals $\Lambda(t,x,\hat{\theta})$ are distributed approximately as standard exponential under the null hypothesis that the specification is correct. The standard exponential has rth moment equal to $r!$, so it is natural to base a test on these moment restrictions. It is easy to check that, if the specification of a parametric model includes a constant term, the generalized residuals sum to one, so there is no point in testing the restriction on the first moment. The second moment is not given identically by the estimating equations, so one might look at the statistic

$$\rho = n^{-1} \sum_{i=1}^{n} \Lambda(t_i,x_i,\hat{\theta})^2 - 2$$

which clearly goes to zero under the null hypothesis of current specification (the second moment of a standard exponential variable is equal to two). Upon dividing ρ by its asymptotic standard error we obtain an asymptotically normally distributed test statistic. This statistic can be given a variety of interpretations.[21]

Chi-square tests to fit can be obtained based on grouped data. Alternatively, tests can be obtained based on the estimated survivor function. These methods are treated by Lawless (1982, ch. 9).

[21] It is the numerator of a score test for heterogeneity when the variance of the heterogeneity term is small. It is also the numerator of an information matrix test, as shown by Chesher (1984). Tests of higher-order moment restriction can be interpreted as tests of the significance of higher-order terms in a Laguerre expansion of the actual distribution of Λ about the exponential as in Kiefer (1985). Sunil Sharma (1987) gives a generalization. Monte Carlo evidence on the performance of these tests is mixed. Peter Jensen (1986) and John Kennan and George R. Neumann (1987) provide some evidence.

G. *Estimation in the Presence of Heterogeneity*

A natural approach to accommodate heterogeneity is to base inference on the mixed distribution resulting from the presence of heterogeneity. Consider the example of Section VIB in which the population is made up of two subgroups, with duration distributions $f_1(t)$ and $f_2(t)$. The distribution for the observations is $f(t) = pf_1(t) + (1 - p)f_2(t)$. The distribution $f(t)$ can be used to form a likelihood function and the parameters can be estimated on the basis of that likelihood function. More generally, the individual densities can be written conditionally on a heterogeneity term v, as $f(t|v)$ and inference can be based on the distribution of observed durations

$$f(t) = \int f(t|v)p(v)dv.$$

The distribution $p(v)$ is typically unknown, and economic theory gives little or no guidance on its form. Consequently, p may be specified only to be a member of a parametric family and these parameters may be estimated along with others. The term v can be modeled as having a discrete distribution and the integral replaced by a summation as in our example. In fact, this is a case of practical importance. Suppose the distribution is given by $Pr(V = v_i) = p_i$, $i = 1, \ldots, I$, and let the parameter vector $\delta = (v_1, \ldots, v_I, p_1, \ldots, p_I)$. These parameters and I, the number of points in the discrete distribution, can be estimated by maximum likelihood.[22]

The question of identification arises immediately. A simple example illustrates the difficulty. Consider the survi-

[22] This estimator is the nonparametric maximum-likelihood estimator suggested by Heckman and Burton Singer (1984). Calculation of the asymptotic distribution of this estimator is problematic. Examples suggest that a small number of points, three to five, is typically adequate as a practical matter.

vor function $p\exp(-\gamma_1 t) + (1 - p)\exp(-\gamma 2t)$. This function, perhaps estimated from data on individuals, could represent the survivor function for each individual if there is no heterogeneity, or it could represent a population in which the fraction p of individuals has survivor function $\exp(-\gamma t)$ and $1 - p$ has survivor function $\exp(-\gamma 2t)$. Sorting out these possibilities in the absence of explanatory variables is analogous to the standard problem of sorting out demand and supply parameters in a simultaneous equations model. The problem arises in practice. Kennan, for example, found that heterogeneity appeared to be important for some specifications of the hazard and not for others. Elbers and Ridder (1982) studied identification in the proportional hazard setting and found essentially that the presence of explanatory variables in the individual distributions is sufficient to assure identification.[23]

The sensitivity of estimates of θ to specification error in $p(v)$ is an important consideration in modeling. This question has been considered recently by a number of authors. James Heckman and Burton Singer (1984) report the results of fitting a Weibull model with several different assumed distributions for the heterogeneity term to a subset of the data from the Kiefer and Neumann (1981) study. They found that the coefficients of explanatory variables were sensitive to the assumption on the functional form of $p(v)$. Other work, however, suggests that the specification of the mixing distribution is not so important as long as the individual distribution is correctly specified (or not too badly misspecified). Thus, the Heckman and Singer (1984) finding of sensitivity to specification of $p(v)$ may be due to the inappropriateness of the Weibull assumption. It appears, based on examples and experiments appearing in the literature, that estimates are typically more sensitive to specification of the survivor function for each individual or group than to that of the mixing distribution (see Kenneth G. Manton, Eric Stallard, and James W. Vaupel 1986, Charles E. McCulloch and John L. Newman 1983, Newman and McCulloch 1984, Ridder 1986, and James Trussell and Toni Richards 1985). A general theorem to this effect has not yet appeared.

[23] Tony Lancaster and Stephen Nickell discuss identification and given examples. The Elbers and Ridder result can probably be generalized to say that a regressor affecting either the conditional distribution or the mixing distribution but not both is sufficient.

VII. *Other Topics*

Competing risks occur when a spell can end in several different ways. For example, a spell of unemployment could end at employment or at withdrawal from the labor force. In this event, transition rates from unemployment to employment and from unemployment to outside the labor force are defined; the hazard rate from unemployment is the sum of the two. The destination-specific transition rates, for example from unemployment to employment, are often of interest and are typically identified. A model of this type is estimated by Burdett et al. (1984). Lawrence Katz (1986) argues that workers on temporary layoff face competing risks: finding new jobs or being recalled.

Multivariate duration data occur when several spells are observed for each individual in the sample. In this case it is possible to look into possible dependence across spells for the same individual. The topic is discussed by Kalbfleisch and Prentice (1980, section 7.3). It is sometimes possible to handle heterogeneity as fixed effects; see Chamberlain (1985). A variety of notions of dependence, on duration and across spells, can be considered (see James Heckman and George Borjas 1980 for a survey).

Bayesian methods can be used to analyze duration data. In the proportional hazard framework the baseline hazard function λ_0 can be regarded as a nuisance parameter. Prior information about the shape of λ_0 can be incorporated formally, as well as information about the effects of explanatory variables. A family of prior distributions for λ_0 is suggested by Kalbfleisch (1978). An alternative Bayesian approach is suggested by Wen-Chen Chen et al. (1985).

Grouped-data methods are currently being developed and applied to economic data. These methods account for the facts that economic data are typically grouped into intervals, for example unemployment into weeks, and typically there are many observations at each value of duration. Note that grouped data and discrete data are not the same. With many observations at 5 weeks, say, it is possible to estimate the conditional probability of concluding a spell at 5 weeks by dividing the number of observations finishing at 5 weeks by the number lasing more than 4 weeks. This procedure estimates each conditional probability separately and therefore does not require an assumption on the functional form of the hazard. This is essentially the approach in the economic literature on semiparametric hazard estimation.[24] Little experience with these methods has been accumulated, but the approach seems appropriate and promising. The approach was adopted by Robert Prentice and L. Gloeckler (1978) and by Moffitt (1985). Aaron Han and Jerry Hausman (1986) and Bruce Meyer (1986) propose improved semiparametric methods adapted to be suitable for economic data. An alternative semiparametric approach is suggested by Joel Horowitz and George Neumann (1987).

Asymptotic distribution theory for a variety of parametric and nonparametric estimators and test statistics can be derived elegantly using results from the theory of counting processes. See Odd Aalen (1978), Andersen and Gill (1982), and Martin Jacobsen (1982). The counting process approach may be useful in developing techniques particularly useful for economic data.

VIII. *Conclusion*

Econometric methods based on hazard functions provide a natural approach to analysis of economic data that can be modeled as generated by series of sequential decisions. The methods do not provide a "correct" approach to data analysis to be contrasted with an "incorrect" approach based on specification of a density function: Specification of a hazard function is an alternative to specification of a density function, providing a simple way to choose a specification that allows plausible behavior to be modeled.

Many of the economic applications of hazard function methods have to date been in labor economics. Here, empirical questions about employment and unemployment spells are often framed in the language of hazard functions. It appears, as a rough generalization, that employment spells exhibit negative duration dependence, at least after the first few months. That is, the longer a job is held, the less likely it is to be lost. There is some evidence that the hazard for unemployment spells may be slightly downward sloping, though this is still a very active research area. Kennan (1985) found some evidence of a U-shaped hazard in the full data set on strike durations. This literature is evidence on the useful-

[24] A better term would be *superparametric* in view of the typically large number of parameters introduced. Since Cox (1972), the term *semiparametric* is used when there are finitely many parameters of interest and infinitely many nuisance parameters. Thus, the partial-likelihood estimator is semiparametric. Methods for grouped data may or may not be semiparametric.

ness of the hazard function approach. Nevertheless, a confident assessment of the importance of hazard function methods in applied economics will have to await further applications. There are possibilities in many applied fields.

References

AALEN, ODD. "Nonparametric Inference for a Family of Counting Processes," *Annals of Statistics,* July/Aug. 1978, *6,* pp. 701–26.

ANDERSEN, PER K. AND GILL, RICHARD D. "Cox's Regression Model for Counting Processes," *Annals of Statistics,* Dec. 1982, *10,* pp. 1100–20.

ATKINSON, ANTHONY B. AND MICKLEWRIGHT, JOHN. *Unemployment benefits and unemployment duration: A study of men in the United Kingdom in the 1970s.* London: London School of Economics and Political Science, Suntory-Toyota International Centre for Economics and Related Disciples, 1985.

BURDETT, KENNETH ET AL. "Earnings, Employment and the Allocation of Time Over Time," *Rev. Econ. Stud.,* Oct. 1984, *51*(4) pp. 559–78.

BURDETT, KENNETH; KIEFER, NICHOLAS M. AND SHARMA, SUNIL. "Layoffs and Duration Dependence in a Model of Turnover," *J. Econometrics,* Apr. 1985, *28*(1), pp. 51–69.

CHAMBERLAIN, GARY. "Heterogeneity, Omitted Variable Bias, and Duration Dependence," *Longitudinal analysis of labor market data.* Eds.: JAMES J. HECKMAN AND BURTON SINGER. Cambridge U. Press, 1985, pp. 3–38.

CHEN, WEN-CHEN ET AL. "Bayesian Analysis of Survival Cures for Cancer Patients Following Treatment," in *Bayesian statistics 2.* Eds.: J. M. BERNARDO ET AL. Amsterdam: North-Holland, 1985, pp. 299–328.

CHESHER, ANDREW. "Testing for Neglected Heterogeneity," *Econometrica,* July 1984, *52*(4), pp. 865–72.

CHIANG, CHIN-LONG. *The life table and its applications.* Malabar, FL: Robert E. Krieger Publishing Co., 1984.

COX, DAVID R. "Regression Models and Life-Tables (with discussion)," *J. Roy. Statist. Soc.,* May/Aug. 1972, *B 34,* pp. 187–220.

______. "Partial Likelihood," *Biometrika,* May/Aug., *62*(2), pp. 269–76.

COX, DAVID R. AND LEWIS, P. A. W. *The statistical analysis of series of events.* London: Chapman & Hall, 1966.

COX, DAVID R. AND OAKES, DAVID. *Analysis of survival data.* London and NY: Chapman & Hall, 1985.

COX, DAVID R. AND SNELL, E. J. "A General Definition of Residuals (with discussion)," *J. Roy. Statist. Soc.,* May/Aug. 1968, *B 30,* pp. 248–75.

CROWLEY, JOHN AND HU, MARIE. "Covariance Analysis of Heart Transplant Survival Data," *J. Amer. Statist. Assoc.,* Mar. 1977, *77*(357), pp. 27–36.

DEVINE, THERESA J. AND KIEFER, NICHOLAS M. "Empirical Labor Economics in the Search Framework," manuscript, Cornell U., 1987.

EFRON, BRADLEY. "The Efficiency of Cox's Likelihood Function for Censored Data," *J. Amer. Statist. Assoc.,* Sept. 1977, *72*(359), 557–65.

ELANDT-JOHNSON, REGINA C. AND JOHNSON, NORMAN L. *Survival models and data analysis.* NY: Wiley, 1980.

ELBERS, CHRIS AND RIDDER, GEERT. "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model," *Rev. Econ. Stud.,* 1982, *49*(3), pp. 403–09.

HAN, AARON AND HAUSMAN, JERRY. "Semiparametric Estimation of Duration and Competing Risk Models," M.I.T., working paper, Nov. 1986.

HARRISON, ALAN AND STEWART, MARK. "Cyclical Variation in Strike-Settlement Probabilities," McMaster U., working paper, Aug. 1987.

HECKMAN, JAMES J. AND BORJAS, GEORGE J. "Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence," *Economica,* Aug. 1980, *47*(187), pp. 247–83.

HECKMAN, JAMES AND SINGER, BURTON. "A Method for Minimizing the Impact of Distribution Assumptions in Econometric Models for Duration Data," *Econometrica,* Mar. 1984, *52*(2), pp. 271–320.

HOROWITZ, JOEL AND NEUMANN, GEORGE R. "Semiparametric Estimation of Employment Duration Models," U. of Iowa, working paper, 1987.

JACOBSEN, MARTIN. *Statistical analysis of counting processes,* NY: Springer-Verlag, 1982.

JENSEN, PETER. "Testing for Unobserved Heterogeneity and Duration Dependence in Econometric Duration Models," U. of Aarhus, working paper, 1986.

JOHANSEN, SØREN. "The Product Limit Estimator as Maximum Likelihood Estimator," *Scand. J. Statist.,* 1978, *5*(2) pp. 195–99.

______. "An Extension of Cox's Regression Model," *International Statistical Review,* Aug. 1983, *51*(2), pp. 165–74.

JOHNSON, NORMAN L. AND KOTZ, SAMUEL. *Distributions in statistics: Continuous univariate distributions,* Vol. 1. NY: Wiley, 1970.

JOVANOVIC, BOYAN. "Wages and Turnover: A Parametrization of the Job-Matching Model," in *Studies in labor market dynamics.* Eds.: GEORGE R. NEUMANN AND NIELS WESTERGAARD-NIELSEN. Berlin: Springer-Verlag, 1984, pp. 158–67.

KALBFLEISCH, JOHN D. "Non-Parametric Bayesian Analysis of Survival Time Data," *J. Roy. Statist. Soc.,* May/Aug. 1978, *B 40,* pp. 214–21.

KALBFLEISCH, JOHN D. AND PRENTICE, ROBERT L. *The statistical analysis of failure time data.* NY: Wiley, 1980.

KATZ, LAWRENCE F. "Layoffs, Recall and the Duration of Unemployment," U. of California, Berkeley working paper, Jan. 1986.

KENNAN, JOHN. "The Duration of Contract Strikes in U.S. Manufacturing," *J. Econometrics,* Apr 1985, *28*(1), pp. 5–28.

KENNAN, JOHN AND NEUMANN, GEORGE R. "On the

Size of Information Matrix Tests," U. of Iowa working paper, 1987.

KIEFER, NICHOLAS M. "Specification Diagnostics Based on Laguerre Alternatives for Econometric Models of Duration," *J. Econometrics*, Apr. 1985, *28*(1), pp. 135–54.

KIEFER, NICHOLAS M.; LUNDBERG, SHELLEY AND NEUMANN, GEORGE R. "How Long Is a Spell of Unemployment?" *J. Bus. Econ. Statist.*, Apr. 1985, 3(2), pp. 118–28.

KIEFER, NICHOLAS M. AND NEUMANN, GEORGE R. "An Empirical Job Search Model, with a Test of the Constant Reservation-Wage Hypothesis," *J. Polit. Econ.*, Feb. 1979, *87*(1), pp. 89–107.

______. "Individual Effects in a Nonlinear Model: Explicit Treatment of Heterogeneity in the Empirical Job-Search Model," *Econometrica*, July 1981, *49*(4), pp. 965–79.

KIEFER, NICHOLAS M. AND SKOOG, GARY R., "Local Asymptotic Specification Error Analysis," *Econometrica*, July 1984, *52*(4), pp. 873–85.

LANCASTER, TONY. "A Stochastic Model for the Duration of a Strike," *J. Roy. Statist. Soc.*, Apr./June 1972, *135*, pt. 2, pp. 257–71.

______. "Econometric Methods for the Duration of Unemployment," *Econometrica*, July 1979, *47*(4), pp. 939–56.

______. "Generalised Residuals and Heterogeneous Duration Models: With Applications to the Weibull Model," *J. Econometrics*, Apr. 1985, *28*(1), pp. 155–69.

LANCASTER, TONY AND CHESHER, ANDREW. "Residuals, Tests and Plots with a Job Matching Application," U. of Hull, Mar. 1985.

LANCASTER, TONY AND NICKELL, STEPHEN J. "The Analysis of Re-Employment Probabilities for the Unemployed (with discussion)," *J. Roy. Statist. Soc.*, Apr./June, 1980, *143*, pt. 2, pp. 141–65.

LAWLESS, JERALD F. *Statistical models and methods for lifetime data.* NY: Wiley, 1982.

MANTON, KENNETH G.; STALLARD, ERIC AND VAUPEL, JAMES W. "Alternative Models for the Heterogeneity of Mortality Risks Among the Aged," *J. Amer. Statist. Assoc.*, Sept. 1986, *81*(398), pp. 635–44.

MCCULLOCH, CHARLES E. AND NEWMAN, JOHN L. "Using Parametric Heterogeneity Specifications in Hazard Rate Models," manuscript, Tulane U., 1983.

MEYER, BRUCE. "Semiparametric Hazard Estimation," working paper, M.I.T., 1986.

MILLER, RUPERT G., JR. *Survival analysis.* NY: Wiley, 1981.

MOFFITT, ROBERT. "Unemployment Insurance and the Distribution of Unemployment Spells," *J. Econometrics*, Apr. 1985, *28*(1), pp. 85–101.

MORTENSEN, DALE. "Job Search and Labor Market Analysis," in *Handbook of labor economics.* Vol 2. Eds.: ORLEY ASHENFELTER AND RICHARD LAYARD. Amsterdam: North-Holland, 1986, pp. 849–919.

NARENDRANATHAN, WIJI AND NICKELL, STEPHEN. "Modelling the Process of Job Search," *J. Econometrics*, Apr. 1985, *28*(1), pp. 29–49.

NEWMAN, JOHN L. AND MCCULLOCH, CHARLES E. "A Hazard Rate Approach to the Timing of Births," *Econometrica*, July 1984, *52*(14), pp. 939–61.

NICKELL, STEPHEN J. "Estimating the Probability of Leaving Unemployment," *Econometrica*, Sept. 1979, *47*(5), pp. 1249–66.

PRENTICE, ROBERT L. AND GLOECKLER, L. A. "Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data," *Biometrics*, 1978, *34*, pp. 57–67.

RIDDER, GEERT. "The Sensitivity of Duration Models to Misspecified Unobserved Heterogeneity and Duration Dependence," manuscript, U. of Amsterdam, 1986.

RIDDER, GEERT AND VERBAKEL, WIM. "On the Estimation of the Proportional Hazards Model in the Presence of Unobserved Heterogeneity," *J. Amer. Statist. Assoc.* Forthcoming.

SHARMA, SUNIL. "Specification Diagnostics for Econometric Models of Durations," UCLA, working paper, 1987.

SIDER, HAL. "Unemployment Duration and Incidence: 1968–82," *Amer. Econ. Rev.*, June 1985, *75*(3), pp. 461–72.

STRUTHERS, C. A. AND KALBFLEISCH, JOHN D. "Misspecified Proportional Hazard Models," *Biometrika*, Aug. 1986, *73*(2), pp. 363–69.

TRUSSELL, JAMES AND RICHARDS, TONI. "Correcting for Unmeasured Heterogeneity in Hazard Models Using the Heckman-Singer Procedure," in *Sociological methodology.* Ed.: NANCY BRANDON TUMA. San Francisco: Jossey-Bass, 1985, pp. 242–76.

TURNBULL, BRUCE W. "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data," *J. Roy. Statist. Soc.*, Sept./Dec. 1976, *B 38*, pp. 290–95.