SEARCH MODELS AND DURATION DATA

By

George R. Neumann*

Contents

---

* Professor of Economics, University of Iowa, Iowa City, Iowa 52242. This work was supported by NSF grant SES-9122253. I have benefitted from valuable comments made by Audra Bowlus, Tue Gørgens, Joel Horowitz, Nick Kiefer, Dan McFadden, Paul Ruud, and participants in the NSF Symposium on Duration/Hazrd Models held August 2-9, 1995 at the University of California, Berkeley.

1. Introduction.

Many topics of interest to economists involve the passage of time. How long does a typical spell of unemployment last? Has the time between births increased in recent years as women's relative earnings increased? Answers to these questions are desired both for the purpose of sorting out various theories of social behavior and to answer questions of public policy. For example, the effects of Unemployment Insurance benefits on the duration of unemployment remains a subject on continued interest and controversy (Atkinson et al. [1984]), as does the effect of public sector dispute resolvement procedures on strike durations in collective bargaining (Butler and Ehrenberg, 1981); Schnell and Gramm, 1987). Analysis of duration data is complicated by the fact that duration data typically come in incomplete form; that is, some observations will, at the time of survey, be unfinished. Sometimes the data will consist of complete and incomplete spells, as in follow up studies of medical treatments. In other cases, such as the Occupational Mobility and Job Tenure survey that the United States Bureau of Labor Statistics conducts, employment spell lengths are incomplete of necessity: individuals are asked for the "length of time working for the present employer." The set of special methods that have been developed to deal with data with this structure are known as failure time models, that is, they deal with data that measure the time until an event occurs. Although the methods originated in industrial engineering and the biomedical sciences, where they were used to study machine breakdown and the effects of medical procedures, they had natural application to diverse social science phenomena from wars to strikes to unemployment spells, and they were rapidly adopted in economics, demography, and sociology.

Duration modelling has a long and established history in biostatistics and engineering, and it is not surprising that early uses in social sciences borrowed liberally from models used in other scientific contexts. Social science analysis of duration data raises new issues concerning sampling structures, the completeness of models, and testing. For example, randomization in experimental design allows biostatisticians to be harmlessly ignorant of "the true model" when estimating the treatment effect of a particular drug. In contrast, econometricians have become painfully aware of the modest role that measured covariates play in economic data and consequently have been much more receptive to models where unmeasured heterogeneity -- the term given to everything else that matters-- plays a significant role. In a similar vein, many test statistics used in duration analysis are applicable only for discrete covariates, or for orthogonal factors. In some areas discreteness and/or orthogonality can be generated in the design of the sample, but such options usually are not available to applied econometricians. This chapter surveys and summarizes tools and techniques that have proven useful for duration modelling, with an emphasis on those that have become useful in applied econometric work. The plan of the chapter is as follows. The second section describes the fundamental tools of duration models. The hazard function is defined and methods of estimation, both parametric and non-parametric, are described. In the third section covariates are introduced, as are tests of model specification. Two broad classes of models --proportional hazards and accelerated failure time models -- are introduced and illustrated using a popular data set. The fourth section of the chapter reviews applications of these tools in the search economics literature. Section five summarizes the results.

2. Duration data and distributions.

2.1 Survivor, Density and Hazard Functions

To fix ideas we consider the case of a homogeneous population; the extension to account for heterogeneity --both observed and unobserved -- is pursued in section 3. Consider then a non-negative random variable, T, which describes the length of time until an event of interest occurs. In most areas of econometrics it is customary to describe T by its **cumulative distribution function**

$$F(t) = \text{Prob}(T < t), \quad 0 < t < \infty, \qquad [1]$$

which specifies the probability that the random variable T is less than some value t. For continuous random variables description by the **probability density function**

$$f(t) = \lim_{dt \to 0} \frac{Prob(t \le T < t + dt)}{dt} = \frac{\partial F(t)}{\partial t} \qquad [2]$$

provides an equivalent view. This view of the data is the unconditional approach: F(t) specifies, say, the probability that a spell of unemployment will last no longer than t weeks. In applications it is frequently more convenient to reason using conditional probabilities: "if unemployment has lasted 10 weeks already, than the probability a worker will become employed next week is now 10%." Define the **survivor function**

$$S(t) = Prob(T \ge t) = 1 - F(t) \qquad [3]$$

which give the probability that a spell will last t periods or longer. The **hazard function** specifies the instantaneous rate of failure at T=t, conditional upon survival to time t and it is defined as

$$h(t) = \lim_{dt \to 0} \frac{Prob(t \le T < t + dt \mid T \ge t)}{dt} = \frac{f(t)}{S(t)} \qquad [4]$$

Note that this representation of the random variable contains the same information as the more familiar characterization given by the CDF. To see this observe that $h(t) = - d \log(S(t))/dt$. Upon integration and using that $S(0) \equiv 1$ we have

$$\mathbf{S(t) = \exp\left(- \int_0^t h(s)\, ds\right) = \exp(-\Lambda(t))} \quad [5]$$

The term

$$\mathbf{\Lambda(t) = \int_0^t h(s)\, ds}$$

is called the **integrated hazard function** and appears frequently in diagnostic tests. Usually we require that $\lim_{t\to\infty} \Lambda(t) = \infty$ in order to impose a non-defective distribution of failure times, i.e., $S(\infty) = 0$[2]. The shape of the hazard function provides a characterization of the underlying stochastic process. If $\partial h(t)/\partial t > 0$ the process is said to exhibit positive duration dependence, while if the sign of the derivative is negative the process is said to exhibit negative duration dependence (Heckman and Borjas,1980). Positive duration dependence means that the chances of failing are increasing over time (or whatever metric that t represents) while negative duration dependence means that

---

[2] In some models of job matching an individual learns about a match sufficiently rapidly that a job separation occurs with probability zero; consequently, defective distributions, those for which $\int f(s)ds < 1$, are sometimes of use in applied work, for example, in mover-stayer problems.

the chance of failure falls with time. The condition that $\partial h(t)/\partial t \equiv 0 \ \forall \ t$ defines a memoryless system, which uniquely defines the exponential distribution. Characterizing duration distributions as having hazard rates that are monotonically increasing or decreasing in t is purely for convenience. Real world data are rarely so nice, and in the reliability literature engineers routinely encounter data characterized by "bathtub" - shaped hazard rates: initially h(t) declines due to processes like infant mortality; h(t) is constant during the "useful" life" phase; finally, failure rates rise during the "wearout" phase (Barlow and Proschan, 1981).

2.2 Censoring Mechanisms

An important aspect of duration data is that some observations may be incomplete; such an observation is termed **censored**. Let T^* be the random stopping time of the event of interest in the absence of censoring, and let C be the censoring time. Let δ be a variable that records the value 1 if the observation is censored, and zero otherwise. The random variable observed is $T = \min\{ T^*, C\}$. If the censoring times have survivor function G(C), with associated density function g(c), and if censoring and failure times are independent, the density function evaluated at T is

$$\mathbf{k(T) = G(T) f(T) + g(T)S(T)} \qquad [6]$$

The first term on the left hand side of [6] is the probability that an observation fails at time T, that is,the joint event {t=T, C>T} occurs, while the second terms gives the probability that T is a censored time.

While the distribution of the stopping times frequently is unknown, the distribution of the censoring times is (partially) under the control of the sample designer. In the case of **Type I Censoring** the sample is observed for a period of length Z, and all spells not completed at z are censored. In this case $G(C) = 1 \ \forall\ c \leq Z$, and $G(C)= 0 \ \forall\ c > Z$. Alternative censoring mechanisms include **Type II Censoring**, where sampling continues until the r-th smallest failure time is observed, and **Progressive Type II Censoring**, where a given fraction of the sample may be censored at several ordered failure times. There are, of course, numerous other ways that a sample could be designed for censoring. The important point will be whether the censoring mechanism is informative about the stochastic process under observation. For example, studies of income (Hausman and Wise, 1977) show that sample attrition is income related; hence inference about the income determination process must allow for cross effects with attrition.

2.3 Nonparametric Estimation.

Graphical plots of the survivor function , S(t), are a simple way to describe duration data. Ignoring censoring, for a sample of size N from a homogeneous population the empirical survivor function (ESF) is

$$\bar{S}(t) = \frac{\#\text{ of } T \geq t}{N} \qquad [7]$$

For continuous data the ESF is a step function with steps (of size 1/N) at each failure time; if there are ties in the data, [7] implies that the ESF is a step function but that the size of the step is proportional to the number of failures at each distinct failure time. To handle censored data, it is assumed that if T = t and δ=1, censoring happens immediately after time t. With this convention let $t_1 < t_2 < .. < t_k$ be the observed failure times in a sample of N. Let d_i be the number of spells that end at time t_i, and let m_i be the number of spells censored between t_i and t_{i+1}. The **risk set**, the set of spells that are eligible to fail at time t_i, is defined as

$$n_i = \sum_{j \geq i} (m_j + d_j) \qquad [8]$$

Since h(t)dt is the probability of completing a spell in the interval t+dt given that the spell lasts at least to t, a natural estimator of h(t) is

$$\bar{h}(t_i) = d_i \,/\, n_i \qquad [9]$$

The corresponding estimator of the survivor function is

$$\bar{S}(t_i) = \prod_{j=1}^{i} (1 - \bar{h}_j)$$

$$= \prod_{j=1}^{i} (n_j - d_j)/ n_i \qquad [10]$$

This is the Kaplan-Meier, or product-limit, nonparametric maximum likelihood estimator of the survivor function.[3] Greenwood's formula

$$\mathrm{var}[\bar{S}(t_i)] = \bar{S}^2(t_i) \sum_{j \le i} d_i/(n_j(n_j - d_j)) \qquad [11]$$

can be used to estimate the asymptotic variance of S.

To illustrate these methods Table 1 shows data on employment

---

[3] Kaplan and Meier [1958] introduced this estimator; subsequent work by Johansen [1978] provides interpretations of it as the maximum likelihood estimator in a wide variety of sampling schemes.

durations of college graduates in their first job[4]. The first column contains the times when jobs were observed to end or were censored. Observation were recorded at 3-month intervals for seven years. Column 2 reports the number of individuals known to be still employed at their first job as of the beginning of the time interval. It is the number employed at the start of the last period minus those who failed or were censored in that interval. Column 3 of the table contains the risk set - n_i in [8] - while columns 4 and 5 show the number censored and failed, respectively, in the time interval. In column 6 we show the estimated hazard function given by [9], in this case stated as a monthly rate,i.e, $h(t_i) = (d_i/n_i)*(1/3)$. The last column contains the estimated survivor function, calculated according to [5]. Graphs of the hazard and the survivor functions are shown in figure 1.

[4] The data are taken from table 1.1 of Lancaster [1990], p. 15.

Table 1
Employment Durations of College Grads in First Jobs

Duration of Job (months)	Number Still Employed at T	Risk Set	Number censored in interval	Number failed in interval	(Monthly) Hazard Rate	Survivor Function
0	703	703	0	3	0.0014	1.000
3	700	700	0	17	0.0081	0.996
6	683	683	0	31	0.0151	0.972
9	652	651	1	33	0.0169	0.929
12	618	618	0	60	0.0324	0.883
15	558	558	0	22	0.0131	0.801
18	536	536	0	29	0.0180	0.770
21	507	505	2	33	0.0218	0.730
24	472	471	1	40	0.0283	0.683
27	431	430	1	17	0.0132	0.628
30	413	408	5	28	0.0229	0.603
33	380	376	4	22	0.0195	0.563
36	354	350	4	16	0.0152	0.531
39	334	333	1	10	0.0100	0.508
42	323	318	5	13	0.0136	0.493
45	305	295	10	9	0.0102	0.473
48	286	280	6	8	0.0095	0.459
51	272	269	3	8	0.0099	0.446
54	261	258	3	10	0.0129	0.433
57	248	247	1	4	0.0054	0.416
60	243	230	13	7	0.0101	0.410
63	223	221	2	1	0.0015	0.397
66	220	216	4	2	0.0031	0.396
69	214	203	11	2	0.0033	0.392
72	201	191	10	4	0.0070	0.388
75	187	177	10	1	0.0019	0.380
78	176	161	15	1	0.0021	0.378
81	160	130	30	1	0.0026	0.376
84	129	129	--	-	--	0.373

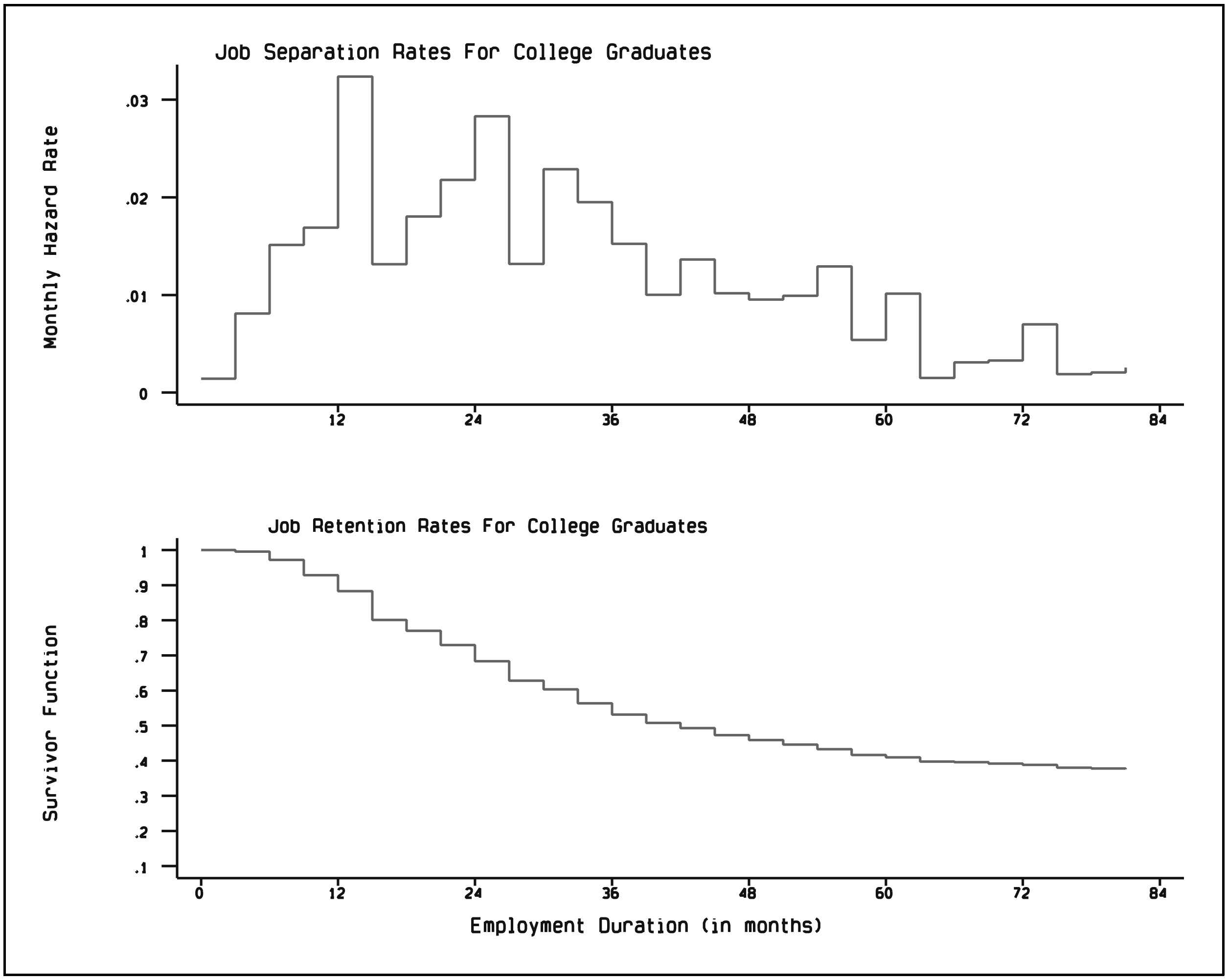

Figure 1: Hazard and Survivor Functions For Lancaster's Data

These duration data are typical of the types encountered in applied work, save only that there are no covariates. The hazard function gives some indication of non-monotonicity: it appears to rise initially and then to fall. However, the roughness of the data does not rule out multi-modality of the hazard. The noisiness of non-parametric estimates of the hazard function suggests that some smoothing operation be applied to the estimated hazard function. By smoothing I mean that the estimate of the hazard rate at point t_* is affected by the actual failure rate at nearby points. This can be done in several ways --kernel estimation and explicit Bayesian methods are two obvious methods-- but the most popular method for smoothing is the assumption of an explicit functional form for the density or the hazard rate.

2.4 Parametric Estimation.

Suppose that a family of duration distributions, F, has been specified up to a finite parameter vector θ, and the censoring mechanism, G, also has been specified up to a finite parameter vector ν. The data consist of N observations on T and a value of the indicator variable δ, where δ=1 indicates that T is censored. Conditional on δ, the contribution to the likelihood function by the i-th observation is

$$\begin{aligned} L_i(\theta,\nu|T_i,\delta) &= G(T_i,\nu)\ f(T_i,\theta) \quad \text{if } \delta_i = 1 \\ &= g(T_i,\nu)\ S(T_i,\theta) \quad \text{if } \delta_i = 0. \end{aligned} \quad [12]$$

With independent censoring ν is not informative for θ and the likelihood function constructed from [12] factors into terms involving the censoring mechanism, G, and terms involving the failure distribution, F. Under these circumstances θ can be estimated without regard to the precise form by which censoring takes place. The log likelihood function, up to a constant, for the sample is

$$\pounds(\theta|\delta) = \sum_{i=1}^{N} \delta_i \ln h(T_i,\theta) - \sum_{i=1}^{N} \Lambda(T_i,\theta) \quad [13]$$

where [13] uses the relation between density functions and survivor functions shown in [4]. The representation of the likelihood function given in [13] stresses the hazard function approach, and, indeed, in applied work most characterization of duration distributions is done in terms of the hazard function. Of course, in cases where there is no closed form solution for the hazard rate the representation in terms of hazard rates and integrated hazard functions offers no advantage. Subject to standard regularity conditions the estimator, $\hat{\theta}$ obtained by maximizing [13] is consistent, and $\sqrt{N}(\hat{\theta}-\theta)$ is asymptotically normal with mean zero and a variance that can be estimated by N x the inverse of the negative of the matrix of second derivatives. Note that a consequence of not being specific about the censoring mechanism is that Fisher's information matrix is not available for use.

Table 2 lists hazard and integrated hazard functions for four specifications popular in applied work. Some hazard functions have monotone hazard rates,

Table 2

Parametric Models of Duration Distributions		
Model	Hazard	Integrated Hazard
1.) Exponential	λ	λt
2.) Weibull	$\lambda\rho(t)^{\rho-1}$	λt^{ρ}
3.) Log-logistic	$\{\lambda\rho(\lambda t)^{\rho-1}\}/\{1+(\lambda t)^{\rho}\}$	$\ln(1+(\lambda t)^{\rho})$
4.) Gompertz	$\lambda\exp(\gamma t)$	$(\lambda/\gamma)[\exp(\gamma t)-1]$

like the Weibull, but these can be made non-monotone by adding higher order powers in t to the specification. Lundberg et al. [1985] provides a description of such extensions, and an application to unemployment data. Note that the choice of an inappropriate functional form for the hazard function may torture the data beyond recognition. For example, fitting an exponential model to data that have an increasing hazard will produce a pattern of misfitting similar to ignoring a trend in time series data. Similarly, a data that have a hazard that increases over some of its range and decrease on the remainder will be ill-treated by a monotonic hazard function such as the Weibull. Diagnostic methods for analyzing goodness-of-fit are discussed in section 3.

Applied work in econometrics has emphasized maximum likelihood methods in fitting parametric models, but this is not the only approach. In the exponential specification a straightforward analysis of $Y = \log(T)$ yields the probability density function

$$f(Y) = \exp(Y - \alpha - \exp(Y-\alpha)),\ -\infty < Y < \infty. \qquad [14]$$

with $\alpha = -\ln(\lambda)$. Write the regression as $Y = \alpha + W$, with the density of W given, from [14], by $\exp(W - e^W)$, which is an extreme value distribution.[5] The variate W does not have mean zero; its expectation is -.5722 (the negative of Euler's constant) and its variance is $\pi^2/6 \approx 1.645$. Thus with uncensored data one could fit least squares to logged duration data and expect to recover the parameters of interest, perhaps after transforming the OLS estimates. These estimates would, of course, be inefficient relative to maximum likelihood when the model is true. Also, there is a hint of how one might test whether the model is adequate by comparing the variance of the residuals to their expected value. Of course, since the defining feature of failure time models is the presence of censoring, and because regression models for censored data are of comparatively recent origin in applied economics, few examples of this approach are available. Nonetheless, the idea provides a useful way of distinguishing among families of duration models.

2.5 Sampling Issues

So far we have treated observations on durations as coming from a

[5] Kalbfleisch and Prentice [1980] provide details for the exponential and other cases.

random sample of all spells that occur. Thus $T \sim f(T)$ refers to the population distribution, and samples of this sort are referred to as **flow samples** (Chesher and Lancaster, (1984), Ridder (1984)). An alternative source of duration data involves **stock sampling**, that is drawing observations from the stock of objects in a particular state. For example, locating a survey unit near the entry door to the unemployment office and surveying those who come in would be sampling the flow of unemployment[6]. In contrast, if a labor force survey were taken and the length of time unemployed recorded for those who were unemployed at the time of the survey we would be sampling from the stock. In general the distribution of an observed time, T, will be different depending upon whether the sample is from the flow or the stock. The distribution of completed spells from the flow is F(t), with density f(t), as we described above. We will use k to describe the density, and K the distribution function, of spells sampled from the stock. The r-th uncentered moment is denoted by μ_r Suppose that a survey of the stock is taken at time t_0. At that time an interrupted spell of length T_b is observed. This is usually called the backward recurrence time, or elapsed duration. The forward recurrence time , T_f, is the time remaining until the spell ends. Completed durations satisfy $T_c = T_b + T_f$. To examine the relation between the two sampling methods we first derive the

---

[6] Assuming, of course, that all unemployed workers must register with the unemployment office.

distribution of elapsed duration and then extend it to the distribution of completed spells. Denote the flow into the state at time t as p(t). The probability that an individual who entered the state at t is still in it at time t_0 is $1-F[t_0-t]$. The density of elapsed durations, t_b, at date t_0 is the ratio of the stock that entered at time s ($=t_0 - t_b$) and remains, divided by the total number of spells, or

$$k(t_b) = \frac{p(t_0-t_b)\,[1 - F(t_b)]}{\int_{-\infty}^{t_0} p(s)[1 - F(t_0 - s)]ds} \qquad [15a]$$

This density clearly depends upon the entire previous history of the process, as the denominator of [15a] makes plain. However, the ergodic property for regular Markov and semi-Markov processes (Karlin and Taylor [1975]) implies that p(s) converges to a constant, $\tilde{p}$ and therefore that [15a] converges to

$$k(t_b) = \frac{1 - F(t_b)}{\mu} \qquad [15b]$$

where we have used the fact that the integrated survivor function, $\int [1-F(u)]du$, equals the mean, μ, of the flow or population data.

The density of completed spell durations sampled from the stock is found by the following argument. In the flow data the conditional density of t_c, given that $t_b < t_c$ is

$$k(t_c | t_b) = \frac{f(t_c)}{1 - F(t_b)} \qquad [16]$$

Multiplying by the marginal density of t_b, integrating out t_b, recalling that $t_b < t_c$, gives the marginal density for a completed spell

$$\begin{aligned} k(t_c) &= \int_0^{t_c} k(t_c | t_b) k(t_b) dt_b \\ &= \int_0^{t_c} (f(t_c)/\mu)\, dt_b \\ &= \frac{t_c\, f(t_c)}{\mu} \end{aligned} \qquad [17]$$

This is the first moment distribution corresponding to f(t). Its moments are (Lancaster, 1990)

$$E(T^r) = \int_0^{\infty} \frac{t^{r+1}\, f(t) dt}{\mu} \qquad [18]$$

For r=1 we have $E(t) = \mu_2/\mu = (\sigma^2 + \mu^2)/\mu = \mu + \sigma^2/\mu > \mu$. Thus the expected duration of unemployment for individuals who are randomly selected from the stock of unemployed is greater than the expected duration of the newly unemployed. In the important special case where unemployment spells are exponentially distributed so that $\sigma^2 = \mu^2$ we have that the completed spells of unemployment sampled from the stock will be twice as long as those sampled from the flow. This puzzling result is due to the fact that "random" samples of the unemployed are not random samples of the unemployment process. This is the problem of length biased samples[7]. If the probability of being included in the sample is proportional to the length of the spell longer spells will be disproportionately represented, which is what leads to the longer durations in stock sampled data.

3. **Econometric Models for Durations**

Social science data rarely can be regarded as drawings from a homogeneous population. Some adjustment must be made for heterogeneity among the observations, perhaps along the lines used in general linear models. Unlike the linear model case there is no natural starting point where covariates shift the mean around but leave other moments unchanged. Consequently the "coefficients" on covariates in duration models do not usually have simple interpretations as partial derivatives. Econometric approaches to modeling durations have focused on parametric or semiparametric specification of the hazard function in the presence of covariates. Few attempts have been made to distinguish one particular structure from another, although tests

---

[7] Kaitz (1970) introduced the length bias problem into labor economics. Chesher and Lancaster (1983), Heckman and Singer (1984b), and Ridder (1984) provide extensive treatment.

to do so are becoming available. For exposition purposes in this section I assume that covariates are constant across time for a given individual and observable. Extensions to time varying covariates and to the problem of unobserved heterogeneity are deferred to sections 3.3 and 3.4.

3.1 Parametric Models.

Standard approaches to parametric modeling assume a specific form for the hazard function, $h(t| X,\beta) = \phi(X,\beta,t)$. For example, in the exponential case the specification frequently used is $h(t| X, \beta) = \exp(X\beta)$. Other forms of the ϕ function that do not depend on t are possible, but the exponential specification has the advantages of simplicity and of imposing the non-negativity constraint on the hazard rate. The log likelihood function with N observations, with $\delta_i=1$ indicating that the i-th observation is censored, is:

$$\mathcal{L}(\beta) = \sum_{i=1}^{N} (1-\delta_i)X_i\beta - \exp(X_i\beta)T_i), \qquad [19]$$

which is globally concave in β. Commonly the parameters of [19] are estimated by maximum likelihood methods, but other options are available in certain cases. If all stopping times are uncensored, the linear model given in [14] above can be used[8]. In this case one regresses ln(t) against X, remembering that the least squares constant has to be adjusted because of the non-zero mean of the error term in [14], but in other respects the model is a simple linear one. Tests of the model can be based on the known distribution of the residuals -- $\varepsilon \{\equiv (\ln(t) - X\beta)\} \sim \exp(-\exp(\varepsilon))$, a Type I extreme value distribution.

A second option is to note that with censoring the model in [14] becomes a particular case of the censored regression model. To date, applied econometricians have found it more convenient to maximize the likelihood function directly rather than to use the tools of censored regression, mainly because software implementing general censored regression has not been widely available. An exception to this is the case of normally distributed durations. Programs are widely available for the censored normal regression model and, because the normal hazard function does not have a closed form solution, the censored regression approach is more frequently used.[9]

---

[8] Least squares estimates will be inefficient relative to maximum likelihood estimates. Lancaster (1990) shows that the ratio of variances obtained by least squares and ML is 1.6449 when the model is correct.

[9] The hallmarks of duration data are skewness to the right and a standard deviation approximately the size of the mean. Given these features, the normal distribution is seldom used to fit duration data, although the log normal is.

A. Specification Tests

Formal tests of a parametric duration model against a parametric alternative that nests it can be carried out in the usual way by Likelihood Ratio, Lagrange Multiplier or Wald tests. Alternatives that are commonly used in practice (exponential, weibull, gompertz, generalized gamma) can be nested in Box-Cox transformations -- $t^* = (t^\theta - 1)/\theta$ if $\theta \neq 0$, $= \ln(t)$ if $\theta=0$ -- and a test of $\theta = \theta_0$, where θ_0 is the transform implied by the model specified under the null, can indicate whether the assumed model adequately characterizes the data. For example, in the exponential model estimated by maximum likelihood the test would examine whether $\theta = 1$.

As in linear regression models, alternatives to the specification $\ln(\phi(X,\beta,t)) = X\beta$ are possible. Adding powers and cross-products to $X\beta$ permits tests of functional form as in standard applications of the RESET test (Ramsey and Schmidt, 1976).

To illustrate the parametric approach I follow tradition (Kiefer, 1988a; Horowitz and Neumann, 1989a,b; Green, 1993) and apply these methods to Kennan's strike data, (Kennan, 1985). The data consist of 566 observations on durations of contract strikes involving 1000 or more workers. Observations were taken from

January, 1968 to December, 1976.[10] Kennan's focus is on the effect of business cycles on strike durations, with cyclical effects measured by the Index of Industrial Production (INDP). Strike data also exhibit some seasonal fluctuations as well, so monthly or quarterly indicators are frequently included. Horowitz and Neumann (1989a) found that seasonal effects were confined to the first quarter of the year, or to the month of February if monthly detail is given. Thus the data consist of the starting year and month of the strike, the value of INDP at the start of the strike, and the length of the strike in calendar days.

All of the strikes reported in Kennan (1985) are complete; there is no censoring. To illustrate the use of specification tests in the presence of censoring the data have been censored according to:

$$\mathbf{T = \min (T^{*},C)} \qquad [20]$$

where T = observed duration, T^* = uncensored duration, C = censoring time, and δ = 1 if the observation is censored. Typically in economics censoring of duration times occurs because events of interest have not ended when data gathering ends.

---

[10] The data are reproduced in Kennan (1985), Table 1, pages 14-16.

Thus longer spells are more likely to be censored.[11] To generate this type of censoring I followed Horowitz and Neumann [1989a] and defined $C = 40 + 6u$, where u is a uniform [0,1] random variable. This produced a censoring rate of 36%.

The parametric model considered is an exponential model, with density

$$\mathbf{T^* \sim \exp(X\beta)\exp(-\exp(X\beta)T^*)} \qquad [21]$$

where X = (1, INDP, FEB).

[11] Random censorship could be handled simply by fitting the model to the uncensored observations, although this would be inefficient.

Table 3

Maximum Likelihood Estimates
of Kennan's Strike Duration Data

Variable	Model			
	Exponential	Extended Exponential	Weibull	Extended Weibull
INDP	-3.3035	-3.7652	-3.2797	-3.7348
	(1.078)	(1.129)	(1.077)	(1.116)
FEB	0.5676	0.5215	0.5633	0.5170
	(0.256)	(0.259)	(0.253)	(0.256)
INDPxFEB	--	7.0655	--	6.9999
		(4.142)		(4.094)
$INDP^2$	--	6.1589	--	6.0892
		(18.79)		(18.58)
Intercept	3.7161	3.7065	3.7139	3.7041
	(0.055)	(0.072)	(0.055)	(0.071)
ρ	1.0	1.0	0.990	0.989
	--	--	(0.45)	(0.45)
$\mathcal{L}$	-800.568	-799.185	-800.546	-799.158

Column (2) of table 3 contains the results of fitting an exponential model to Kennan's data. If this specification is deemed adequate, Kennan's conclusions would be upheld. The presence of cyclical and seasonal effects are indicated by both INDP and FEB being statistically significant. One check on the adequacy of the exponential specification is to embed the model in a more general specification. In this case the power series transformation leads directly to the Weibull model, estimates of which are shown in column (4) of the table. The shape parameter, ρ, which is fixed at 1 in the exponential model, changes hardly at all when left unconstrained, and neither do the other parameters.

One source of possible misspecification in parametric regression models is the mean function, $E(t^*) = \exp(-X\beta)$, for the exponential. To test for this possibility the model was re-estimated with powers and cross-products of the variables. Note that because FEB*FEB=FEB the extended model adds only the $INDP^2$ and INDPxFEB terms. Redundancies of this sort arise especially when indicator and/or trend variables are included. The specification test is implemented as a likelihood ratio test, and comparison of columns (2) and (3) yield a χ^2(2 d.f.) value of (-2*(-800.548 + 800.546) = .044. From this I would conclude that the Weibull model is not favored over the exponential specification.

It is possible to perform these tests jointly. Column (5) of table 3 reports estimates for the extended Weibull specification which, when compared with the exponential specification in column (2), yields an χ^2 (3 d.f.) test statistic of 2.82. This result also does not suggest significant departure from an exponential model.

An alternative test of the exponential specification can be obtained using White's Information Matrix test (White, 1982). The IM identity in this case is

$$\mathbf{H + G'G = N^{-1}} \sum_{i=1}^{N} \{\varepsilon_i^2 + \varepsilon_i - (1-\delta_i)\}\ X_i{}'X_i, \qquad [22]$$

where

$$\varepsilon_i = (1-\delta_i) - \exp(X_i\beta)T_i. \qquad [23]$$

Evidently, the IM test considers whether the variance of the censored exponential is correctly specified and whether it is correlated with the regressors or their squares. Several authors (Chesher and Spady (1991), Kennan and Neumann (1988), Orme (1990)) indicate that critical values obtained from the χ^2 asymptotic distribution of the IM test are seriously inaccurate, and recommend caution in applying the test. Horowitz (1994), srguing that the IM test is asymptotically pivotal, shows that the bootstrap can control the size problem, but this approach has not yet seen wide use in applied work.

The exponential model does not lend itself to testing for other forms of misspecification, because even with regressors it still is a "one-parameter" model. Two-parameter models such as the weibull, with shape parameter ρ, and the normal distribution, with standard deviation σ, afford the opportunity to test if these parameters vary with regressors x. Horowitz and Neumann (1989a) illustrate these tests in the censored regression framework, but the methods also work using likelihoods.

As in linear models, visual analysis of the data is useful for uncovering model inadequacies that may not be detected by formal tests. Chesher, Lancaster, and Irish (1985) provide a diagnostic based on the integrated hazard function. For the exponential model this test is very simple. Let $\upsilon_i = \ln(T_i) - X_i\beta$. Because β must be estimated define the residuals $v_i = \ln(T_i) - X_i b$, where b is a consistent estimate of β. The empirical distribution function (EDF) of v is the Kaplan-Meier (1958) estimate. The test consists of plotting $\log[-\log[S(v_i)]]$ against v_i. If the model is correctly specified the plot should reveal a scatter around a 45^o line. Figure 2 shows this graph for Kennan's strike data.

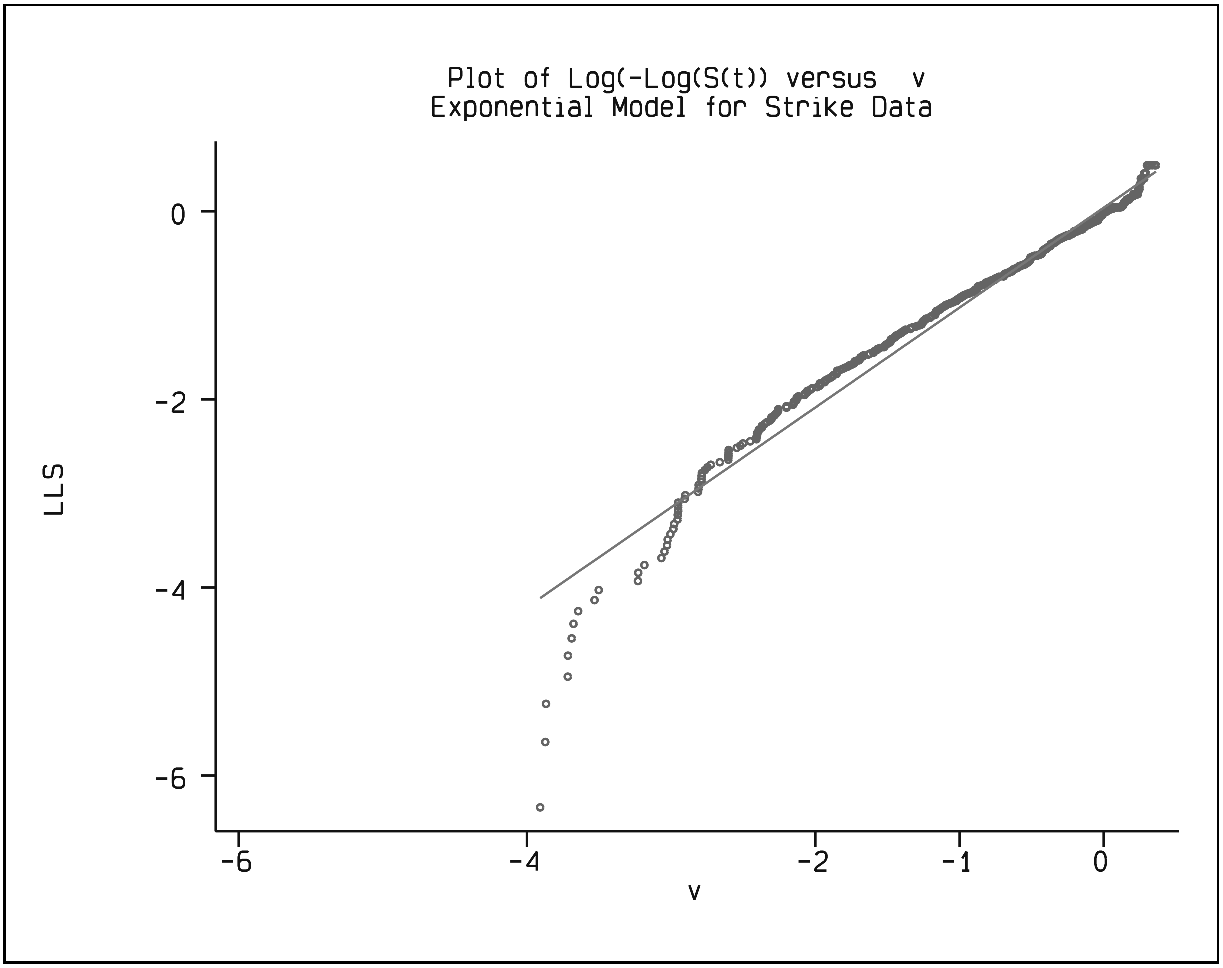

Figure 2

The agreement of the strike data with the model is evident at higher levels of v (longer strikes) but something clearly is amiss at the lower end. There is evidence of a shortage, so to speak, of short strikes. Short strikes in Kennan's data correspond to strikes lasting less than 2 weeks ($v_i < -3$). Thus the data suggest that strike durations are not completely described by an exponential model, but that departures from this model are concentrated in the lower tail of the error term. A natural question to ask is whether the departures from the maintained model shown in figure 2 affect the inferences one would draw about the relation between strike length and business cycle conditions. To ask this question we need to examine a broader class of models.

3.2 Semiparametric Models.

The exponential model described in the previous section was convenient for estimation and for use but, as the graphical test indicated, it proved to be too restrictive. Although several formal tests did not reject the exponential formulation, the graphical test gave evidence of model failure but did not suggest an alternative specification. A variety of semiparametric models that nest the exponential model as a special case are available. The term "semiparametric" has come to have several meanings in the area of duration models, some of which seem to conflict. The standard description of semiparametric has been, following Cox [1972], the case where the parameters of interest are of finite, and fixed, dimension and the nuisance parameters are of infinite dimension. I continue to use the term semiparametric to describe models that do not fit this definition but which have become popular in applied research. Although this tortures the language, I use it because it captures the spirit of the approach, even where it involves estimating lots of parameters.

3.2.a Proportional Hazard Models

Recall that the hazard function $h(t|X)$ in principle allows a wide range of interaction among durations, t, and regressors, X. To obtain generality a model has to restrict the possible range of interactions. Cox (1972) introduced the continuous proportional hazard model by specifying:

$$\lambda(t|X) = \lambda_0(t)\exp(X\beta), \qquad [24]$$

that is, by restricting the interaction of t and X to be multiplicative through the **baseline hazard rate**, $\lambda_0(t)$, and the regressors embodied in the link function, $\exp(X\beta)$.[12] Cox (1972, 1975) showed that the parameter β could be estimated without specifying the form or family for λ_0. Specifically, order the durations from smallest to greatest - $t_1 < t_2 \ldots < t_N$. The conditional probability that observation 1 ended at t_1, given that any of N observations could have ended, is (ignoring censoring)

$$\frac{\lambda(t_1, X_1, \beta)}{\sum_{i=1}^{N} \lambda(t_1, X_i, \beta)} \qquad [25]$$

[12] Here, as in other models, nothing requires the link function to be exp(XB). For the reasons noted above this is the popular specification in applied work, even though alternatives such as $\sinh(X\beta)$, $\mathrm{abs}(X\beta)$, etc. exist.

which, given the proportionality assumption, reduces to

$$\frac{\exp(X_1\beta)}{\sum_{i=1}^{N} \exp(X_i\beta)} \qquad [26]$$

Similarly, the conditional probability of the j-th shortest duration is the ratio of hazard for the individual completing a spell at time t_j to the sum of the hazards for individuals whose spells were in progress just prior to t_j.
To accomodate censoring, a spell that is censored between durations t_j and t_{j+1} appears in the summation in the denominator of [26] for observations 1 through j, but not in any others, and never enters the numerator. In this manner we use the information that the spell was in progress, and thus could have failed, up to a certain date, and thereafter we have no further information about the spell.

Cox (1972), Kalbfleisch and Prentice (1973), and Breslow (1974) show how to estimate the baseline survivor function, $S_0(t) = \exp(-\int_0^t \lambda_0(s)ds)$, for the proportional hazards model. The baseline survivor function is rarely used by itself, although as we shall see, it plays a significant role in diagnostic tests.

Kiefer (1988a) notes that the proportional hazard specification has a linear model interpretation; specifically that it satisfies

$$-\ln(\Lambda_0(t)) = G(t) = X\beta + v, \qquad [27]$$

where G(•) is a monotone function and v is a random variable distributed as a unit extreme value; i.e., $F(v) = 1 - \exp(-\exp(v))$, $-\infty < v < \infty$. This feature of the proportional hazard model facilitates development of diagnostic tests.

Cox's proportional hazard model is a continuous time specification, which leads to difficulties in developing formal tests and to problems in implementation when duration data are discrete. In the continuous case, one has to find moments or distributions of empirical processes, and this can be complicated. Similarly, ties are a minor nuisance if they are rarely present, but a major inconvenience if they are. In such cases there is an incentive to using a discrete data version of Cox's model. Han and Hausman (1990) and Kiefer (1988b) provide extensions to Prentice and Gloeckler's (1978) grouped data treatment of the PH model that address the discreteness issue. The approach is to specify a specific form of the link function $\phi(X,\beta)$ and to estimate the parameters of a flexible duration model.[13] For example, treating X as fixed, using $\phi(X,\beta) = \exp(X\beta)$ as the link function, and assuming that individual durations are observed at distinct points $t_1 < t_2 < \ldots < t_K$, the probability that

---

[13] Kiefer [1988a, n. 24] terms this approach *superparametric* because of the many parameters that are estimated.

an observation with characteristics X survives the i-th interval, given survival through the (i-1)-th interval is

$$\mathbf{Prob}\ (T \geq t_i | T > t_{i-1}) = \exp[-\int_{t_{i-1}}^{t_i} h_0(s)ds\ \exp(X\beta)$$

$$= \Theta_i(X, \beta, \gamma) \qquad [28]$$

The conditional probability of a spell ending in interval i is 1 - $\Theta_i(\bullet)$. Using the complete PH specification yields

$$\Theta_i = \exp(-\exp(X\beta + \gamma_i)) \qquad [29]$$

where

$$\gamma_i = \log(\int_{t_{i-1}}^{t_i} h_0(s)\ ds) \qquad [30]$$

The contribution to log likelihood made by the n-th observation, which fails or is censored in the k_n-th interval, is

$$£_n(\beta, \gamma) = \ln(1-\Theta_k(X_n, \beta, \gamma))*(1-d_n) + \sum_{j=1}^{k_n} \ln(\Theta_j(X_n, \beta, \gamma)) \qquad [31]$$

An alternative way of writing the contribution to log likelihood for observation n that fails in the i-th interval is

$$\pounds_n(\beta,\gamma) = \ln \int_{\gamma_{i-1} - X\beta}^{\gamma_i - X\beta} f(\varepsilon)\, d\varepsilon \qquad [32]$$

where $f(\varepsilon)$ is the density function of an extreme value random variable under the PH specification. If observation n is censored, replace the upper limit of the integral with $+\infty$. Equation [32] shows that the discrete PH structure is closely related to standard dichotomous choice models; indeed, the discrete duration density is like an ordered logit under the PH model where $F(\varepsilon)$ is the extreme value distribution. Reinterpreting the data in the dichotomous choice framework, each individual contributes k_n pseudo-observations to the likelihood function. The first k_n-1 pseudo observations for an individual consist of a piece of the survivor function; the last piece is $(1-\Theta_k)$, which appears only if the observation is known to have failed. Hypotheses about specific forms of β and γ are the basis for specification tests, which are discussed below.

Differences between the continuous and discrete versions of the proportional hazards model are subtle. In the continuous case the baseline hazard is treated as a nuisance function, and the ordered failure (and censoring) ranks are used to condition the baseline hazard out of the estimation procedure. In the discrete case the baseline hazard is treated as constant within intervals, and a parameter for each interval is estimated. Thus the scale of the failure time data is explicitly estimated in the discrete

PH model. Because only a finite number of intervals are considered the discrete version of the PH model potentially can be misspecified due to aggregation bias (Bergström and Edin, 1992; Sueyoshi, 1992).

3.2.b Accelerated Failure Time models.

An alternative family of models with restrictions of strength similar to the proportional hazard specification is the class of Accelerated Failure Time models. In this case the hazard function is

$$\lambda(t|X) = \lambda_0(t\phi(X,\beta))\phi(X,\beta). \qquad [33]$$

Note that in contrast with the PH model, the effect of regressors in accelerated failure time models is to rescale time. That is, here a covariate accelerates (or decelerates) the time to failure while in the proportional hazard model a covariate changes the hazard rate. For the case where $\phi(\bullet) = \exp(X\beta)$ the AFT model also has a linear models interpretation:

$$-\ln t = X\beta + v, \quad v \sim F(v). \qquad [34]$$

The distribution function, $F(v)$, is continuous, but is otherwise unconstrained. Thus, AFT and PH models make assumptions about different features of duration data. PH models allow some general transform of duration time to be linearly related to $X\beta$, but completely specifies the distribution of the error term. AFT models, on the other hand, completely restrict the transform of duration time, but allow arbitrary structure on the error term. It can be shown that for the link function $\phi(X,\beta) = \exp(X\beta)$ the weibull (including the exponential as a subcase) specification is the only member of both the proportional hazard and accelerated failure time families.

Parametric implementation of an AFT model requires a choice, up to a finite vector of parameters, of the distribution of v, F(v). Semiparametric implementation requires that F be estimated along with β. If censoring were not present the semiparametric tool of choice would be least squares, with the distribution of the residuals, $\bar{F}_n(\varepsilon)$, being obtained as the empirical distribution function. If censoring is present, the regression structure of ln(t) suggests that censored regression methods be used. Many estimators have been proposed (see Horowitz and Neumann (1988, 1989a,b) for discussion) but three methods have attained significant usage -- quantile estimators (Powell, 1986a), symmetrically censored least squares (Powell, 1986b), and semiparametric M estimators (Horowitz, 1986, 1988). Horowitz and Neumann (1988, 1989a) survey their application.

Ridder (1990) has proposed a model --the Generalized Accelerated Failure Time (GAFT) -- that nests both AFT and PH specifications. Indeed, the GAFT model has the linear structure of the PH model in equation [27] but the error term is not required to be extreme value. Horowitz (1996) provides $\sqrt{N}$ consistent, asymptotically normal estimators of G and F, given a consistent estimate of β for the case where there is no censoring. As yet the GAFT approach cannot handle censored data and so it has seen limited use in duration models.

Because economic theories about durations typically have implications about the behavior of the hazard rate, an estimate of the hazard function is usually desired. This requires extra effort in the semiparametric case. Both PH and AFT models specifications deliver a consistent estimate of the integrated hazard function at the observed failure times. To obtain the hazard function the log survivor function must be differentiated. Typically the result is very noisy, and some smoothing procedure must be applied. Watson and Ledbetter [1964] examine this problem; recent descriptions of kernel estimation of the hazard function are given in Silverman [1986], and Scott [1992]. Ramlau-Hansen [1983] shows, in the no covariate case, that kernel estimates of the hazard function are normally distributed as the sample size gets large. Wells [1990] extends this approach to include covariates. Thus in principle it is possible to test whether the hazard function is increasing, decreasing, or constant over an interval.

3.2.c Diagnostic tests

Semiparametric models are popular because they provide protection against some forms of model misspecification. Thus the exponential model and the weibull model, to which it was compared in table 2 using Kennan's strike data, are both members of the PH family; consequently, estimates of the effect of covariates obtained from Cox's model should be the same, up to a scalar, as those obtained from the parametric specification.[14] Of course, one never knows if the data generating process really has proportional hazards, or accelerated failure times, so there is a need for formal methods of testing. Methods for testing specifications are relatively thoroughly worked out for the proportional hazard model, in both discrete and continuous forms, but less so for censored regression models or for accelerated failure time models.

The workhorse for developing test statistics has been the integrated hazard function, $\varepsilon_t = \int_0^t h(s,x,\beta)ds$. It is well known that ε_t is distributed extreme value with mean 1, and this

---

[14] If the weibull specification is correct, the PH slope coefficients are β/α, where α is the shape coefficient in the weibull. Setting α =1 is the exponential specification.

fact allows tests based on the integrated hazard, or on functions of it. By analogy with residual testing in linear models, it is convenient to work with centered versions of the integrated hazard. Lancaster (1985, 1990), following Cox and Snell [1969] calls 1 - ε_t a generalized error, and defines generalized residuals as the errors with maximum likelihood estimates of unknown parameters inserted, i.e., $e_t = \int_0^t h(s,X,b)ds$, where b is the mle for β. On the null hypothesis that the model is correctly specified a plot of the logarithm of the sample survivor function against e_t should produce a straight line with a 45° slope. For Cox's proportional hazard model this is Kay's [1977] graphical test.

Graphical tests are useful for detecting departures from a specified model, but a formal test is often needed. One approach is the two-sample test of Breslow, Elder, and Berger [1984], Wei [1984], and Gill and Schumacher [1987]. Consider a binary explanatory variable, Z_{2i}, which takes on the value 1 if the i-th individual has characteristic A, and zero otherwise. The link function $\phi(X_i\beta) = \exp(Z_{1i}\beta_1 + Z_{2i}\beta_2)$ forms the basis for the test: reject the null hypothesis that there is one population (H_0: $\beta_2 \equiv 0$) if $\beta_2/\sqrt{Var(\beta_2)} > t_{N\alpha}$, where $t_{N\alpha}$ is a predetermined critical value of the student t-distribution. The extension to a k-sample test is immediate.

Two-sample tests are of limited use in econometric applications where covariates typically are continuous. One approach to testing the PH assumption with continuous covariates involves introducing quadratic or higher order terms in X into the link function. This is simply the RESET procedure, which in this context tests the adequacy of the link function specification. An alternative approach tests the proportionality hypothesis by testing the constancy of β in [24]. Schoenfeld [1980] proposes partitioning the time axis into L_1 sets, and the regressor space into L_2 sets, and testing failure rates within the L_1 x L_2 cells. If the PH hypothesis does not hold, then over a certain interval of time the hazard will be greater than in other periods. The test that Schoenfeld proposes is a χ^2 test with $L_1 x L_2$ degrees of freedom. Moreau et al. [1985] consider a slightly restricted version of the Schoenfeld test. They discretize the time interval into r units $((0,b_1),(b_1,b_2)...(b_r,\infty))$ and define the hazard as:

$$\mathbf{h(t|X(t)) = h_0(t)exp((\beta+\gamma_j)X(t)), \quad b_{j-1} \leq t < b_j,} \qquad \mathbf{[35]}$$

where β and $\gamma_j = (\gamma_{1j}, \gamma_{2j}, ..., \gamma_{pj})$ are (px1) dimensional vectors. The Moreau et al. test is a score test of the null hypothesis that $\gamma_j = 0$, j= 2,..,p. Its asymptotic distribution is χ^2 with (r-1)*p degrees of freedom. Compared to the Schoenfeld test there is some economy in that only one degree of freedom is used up for each of the p covariates, but the partition of T must be chosen. As is well known, the outcomes of the test can be sensitive to the choice of the partition, but methods for making choices that assure high power have not been developed.

Horowitz and Neumann [1992] present a method of moments test that can be used with continuous covariates and which does not require partitioning of the time and/or covariate axis. The test exploits the integrated hazard ($U = \int_0^t h_0(s)\exp(\beta X)ds$) function for the proportional hazard model, which is distributed as a (possibly censored) unit exponential variate under the null hypothesis that the model is correctly specified. A possibly vector-valued function, W(U), is chosen such that $E_P W(U,X,\delta) = 0$ when the model is true, and $E_P W(U,X,\delta) \neq 0$ when U does not have a unit exponential distribution. Expectations are taken over the joint distribution, P, of U, X, and δ, the censoring variable.

Tests of discrete proportional hazard models, because they do not have to rely on dependencies created by non-parametric estimation of the baseline hazard function, are the standard Wald, LM, and LR tests. For example, if the intervals on which the discrete PH model is defined are the same length, then a test of the exponential model for durations can be constructed by noting that

$$\gamma_i = \ln h_0 + \ln(t_i - t_{i-1}). \qquad [36]$$

Thus a test of the exponential model can be constructed as a test of the restriction $\gamma_1 = \gamma_2 ... = \gamma_K$. A test of the proportionality hypothesis itself can be made in the discrete case. The specification given in [29] reveals that one restriction of the PH model is that the link function, $\exp(X\beta)$ have constant β's. Assuming that there is enough variation in the X's, the specification in [29] can be relaxed to

$$\Theta_i = \exp(-\exp(X\beta_i + \gamma_i)) , i = 1,..,K \qquad [37]$$

and a test of $\beta_i = \beta$, i = 1,..K can be implemented using LR, LM, or Wald tests (Kiefer (1990); Han and Hausman (1990); Sueyoshi, (1991); McCall (1994)).

3.3 Semiparametric Estimation of Kennan's Strike data

To illustrate these semiparametric techniques I apply them to Kennan's strike data, which was described earlier. The estimates are shown in table 4. The second column contains estimates based on the maximizing the partial likelihood function suggested by Cox. The third column contains estimates based on maximizing the conditional likelihood function for discrete data, the sum over the 566 observations of equation [31]. To implement the discrete model I divided the time axis into 10 intervals - (0,5], (6-10], (10-15], (15-20], (20-25], (25-30], (30-35], (35-40], (40-45], (45 - ∞) - the first 9 of which are of equal length. The fourth and fifth columns contain censored regression estimates of ln (duration) models, which are semiparametric alternatives to the proportional hazard specification. The column labelled "Quantile" contains parameter estimates obtained using Powell's [1986] censored regression quantile method for the .3 quantile, while the column labelled "SGLS" reports estimates based on Horowitz's [1986,1988] semiparametric generalized least square approach, using the .3 Quantile estimates as a starting point. Note that because the coefficients in the Proportional hazard model are identified only up to scale, they are not directly comparable to the censored regression coefficients. Only for the exponential distribution are the coefficients directly comparable.

Table 4

Semiparametric Estimates
of Kennan's Strike Duration Data

	Model			
Variable	Pro. Haz. (Cont.)	Pro. Haz. (Discrete)	Quantile (θ=.3)	SGLS
INDP	3.2453	3.2224	2.1521	1.9054
	(1.08)	(0.93)	(1.76)	(1.33)
FEB	-0.5671	-0.5626	-0.6123	-0.6453
	(0.266)	(3.89)	(0.37)	(0.28)
Intercept	--	--	-2.6094	
			(0.09)	
$\mathcal{L}$	-2125.14	-1100.47		

Both versions of the proportional hazard model are in agreement that the level of industrial production relative to its trend (INDP) significantly affects strike duration, but the existence of a February (FEB) effect is not clearly evident in the discrete PH model. In contrast, estimates of the effect of INDP using either AFT (quantile or SGLS) model would not indicate a statistically significant role for industrial production. The quantile estimate has an asymptotic t-value of 1.22, while the t-statistic for the SGLS estimate is 1.43. Thus the inferences that one would draw will depend upon whether the proportional hazard or accelerated failure time model is correct.

As discussed in section 3.2a above, it is possible to test features of the proportional hazards assumption.[15] To get a visual impression of the adequacy of the PH assumption I used Kay's graphical test based on the estimates of the continuous model in column (2) of table 4. Specifically, denoting by $\hat{S}$ the conditional survivor function, define LLS $\equiv$ log[-log[$\hat{S}$]], and let ESF be the Kaplan-Meier estimate of the survivor function of LLS. If the proportional hazard model is correct the plot of log[-

[15] Graphical tests are based on the idea that large sample properties of residuals can be relied upon in finite situations with estimated parameters. This idea underlies Kay's test. Baltazar-Aban and Peña [1995] discuss the limitations of this assumption, especially with regard to the proportional hazards model

log[ESF]] against LLS will be approximately a 45° line. Figure 3 shows the plot obtained from the strike data with a 45° line superimposed. In all graphical tests one is left with the question of how close is close, but certainly nothing in figure 3 would indicate a significant departure from proportional hazards.

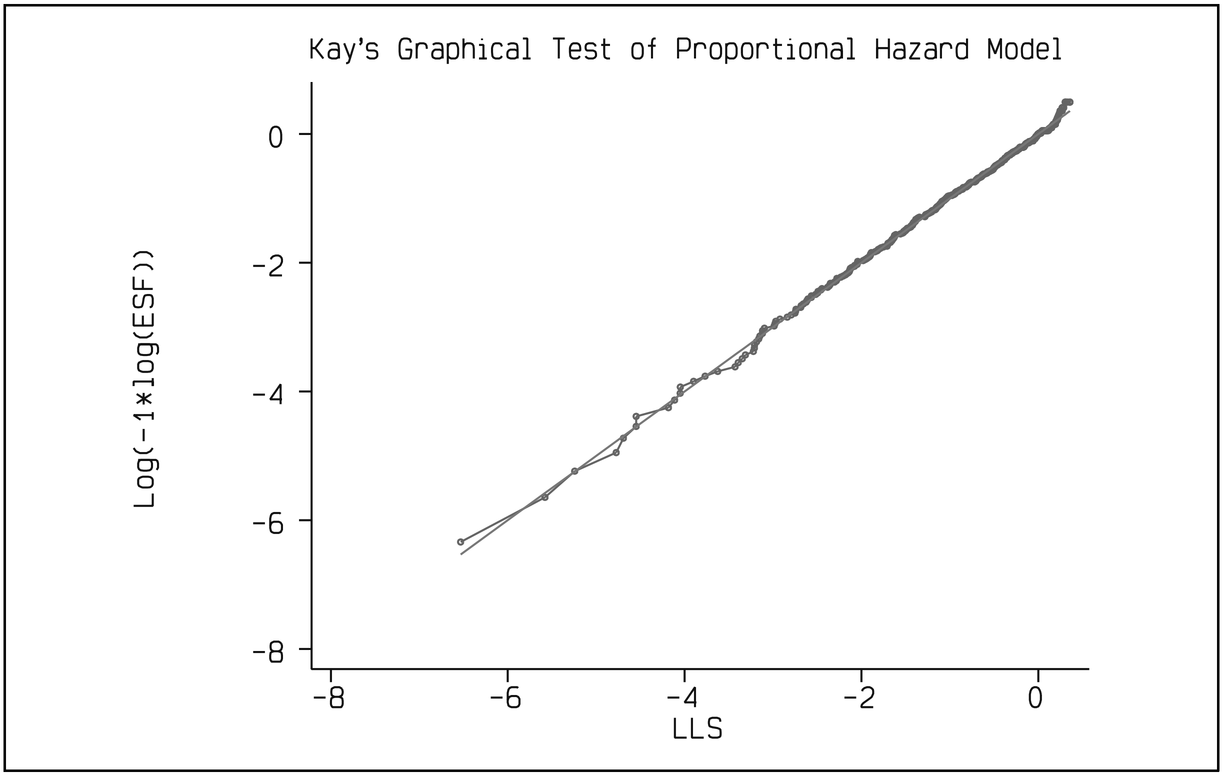

Figure 3

To formally test the proportional hazards hypothesis with continuous covariates two tests are available. For the continuous model the Horowitz-Neumann test calculates the moments of $W(U,X,\delta)= (1+\delta)\exp(-U) -1$, where U is the empirical integrated hazard function computed from the coefficients estimated by Cox's partial likelihood method. Under the null hypothesis that the

Proportional Hazard model is correct W has expected value 0 and is distributed asymptotically normal. Computing the test statistic yields a value of 0.001. If the proportional hazards model is correct the test statistic is distributed as χ^2 with 1 degree of freedom, so this test accepts the proportional hazard model (p=.92).

For the discrete PH model the test of the proportionality assumption is a test for constancy of the elements of β. As the time axis has been partitioned into r bins to implement the discrete model, one could potentially estimate (r-1) coefficient vectors, each of dimension p. For the strike data p=2 and r=10, so a full test would involve testing 18 restrictions. This is unlikely to have much power. Instead of testing all possible variations in β, I considered the case where $\beta = \beta_1$ for t ε (0,23], and $\beta = \beta_2$ for t $\varepsilon(23,\infty]$. The test is implemented as an LR test. The value of the maximized log likelihood function is -1100.3057. Therefore the test statistic -2*ΔLogL = -2*(-1100.3057 +1100.4735)= 0.33563. If the proportional hazard model is correct this statistic is distributed as χ^2 with 2 degrees of freedom, so this test accepts the proportional hazard model also (p=.85).

Equivalent tests of the accelerated failure time model in the presence of censoring have not yet appeared. In particular there is not, to my knowledge, a test that compares proportional versus accelerated hazards. However, it is likely that the GAFT model (Ridder [1990]; Horowitz [1996]) will be able to distinguish between the non-parametrically estimated function G(t), given in [27], and the deterministic function ln(t).

Of course it is possible that neither the proportional hazards specification nor the accelerated failure specification is correct. Each limits the interaction of duration time and covariates in forming the hazard function in a specific way, and a particular problem may require a more flexible interaction between covariates and duration. For example, using British data Atkinson et al. [1984] and Narendranathan and Stewart [1993] find that unemployment benefits have different effects on the hazard from unemployment as the spell lengthens. Mortensen [1977] presents a search theoretic argument of why this should be so based on a maximum benefit receipt period. In cases such as this, hazard rate specifications like $\ln h(t|X) = \beta_0 X + \beta_1 I(t < t_*) + \beta_2 X^* I(t \geq t_*)$ are sensible specifications, where I(X) is the indicator function taking the value 1 if event X is true and zero otherwise, and t_* is an exogenously determined change point of the process. Other forms that have "bathtub"-like shapes or that change at discrete points are, of course, possible. In such cases the regression approach described above is likely to have little utility.

3.4 Time Varying Covariates.

Our development so far has treated X, the set of covariates, as fixed. Yet it frequently is the case that elements of X will change over the length of a spell. For example, unemployment insurance payments typically are made at a certain level for some period of time, and subsequently are paid at a different, usually lower, level thereafter. Similarly, the probability of discharge in an industrial setting typically varies with a worker's disciplinary history, which itself may change over time. Finally, in the context of job matching models, (e.g., Jovanovic, [1979, 1984]), the duration of a job depends upon the wage received, which typically is modelled as stochastic.

Time dependent covariates can be put into two classifications: external and internal. **External** covariates are the outcome of a stochastic process whose marginal distribution does not involve the parameters of the duration model. Fixed covariates are one example of an external covariate, where the stochastic process is degenerate; predetermined processes such as an age indicator, which changes at a fixed, predetermined time, are another example. Local labor market conditions, as represented by the monthly unemployment rate, would be a typical example of a process which is stochastic, external to individual unemployment durations, but not defined in advance. **Internal** covariates are the output of a stochastic process that is generated by the individual under study and is observed as long as the duration has not ended (or been censored).   Lancaster [1990] notes that all external covariates are exogenous, in the sense of Engle, Hendry, and Richard [1983], and the conditionality principle would suggest conditioning on their entire observed path. Thus, in a study of the effect of UI payments on the duration of unemployment it would make sense to consider the effect of the level of next period's UI payments on today's hazard, or to introduce a stochastic variable like "time to subsequent benefit exhaustion" (Meyer, 1990) into today's hazard when the covariate is exogenous. For internal covariates such conditioning is not possible, and depending upon the definition of the covariate, proper treatment may require modelling the joint probability of T and X(t). Lancaster [1990] provides a useful discussion of this issue.

Assuming that the time varying covariate affects the hazard only through its current value the integrated conditional hazard may be written as

$$\Lambda(t|X(t)) = \int_0^t h(s,x(s),\beta)\, ds \qquad [38]$$

where X(t) is the history of X up to time t. The density of a completed spell is $h(t,x(t)) \exp(-\Lambda(t|X(t))$, so the only difficulty in implementation is the integration in [38]. If X changes infrequently the integral can be simplified into a sum of a few terms; in other cases numerical integration may be required.

Note that identification of the effects of a time varying covariate from duration dependence is not trivial. As an example, suppose that the divorce rate (the hazard from marriage) is affected by the aggregate unemployment rate, $h_M(t|X, U(t)) = h_o(t)\exp(X\beta + U(t)\gamma)$. By construction this model is a proportional hazard model, but Cox's partial likelihood methods will not separately identify h_0 and γ in a panel of newly weds (a flow sample). Because the current unemployment rate affects all persons married at time t in the same manner, its effect cannot be distinguished from duration dependence in $h_o(t)$. In general, substantial cross-individual variation in the paths of time varying covariates will be needed to reliably identify their effects.

3.5 Unmeasured Heterogeneity.

Covariates are incorporated into the hazard function control for measured heterogeneity across individuals, but it would be a rare case where a researcher had data that measured all relevant variables, both T and X, without error. If data are incomplete in this fashion then, much as in the linear case, standard methods of estimation may lead to biased and inconsistent estimates of parameters. To illustrate the issue write the conditional hazard as $h(t|X,\nu)$ where X is the set of measured, time-invariant covariates, and ν is a scalar representing unmeasured heterogeneity.[16] Here υ is an unobserved, time-invariant effect. If υ varied with t it could be absorbed into the process that produces randomness in t, and thereafter ignored. Let ν have distribution function $K(\nu|X)$. The observations that we see on (t,X) are draws from the unconditional distribution of t, given X, p_m:

$$\begin{aligned} \mathbf{p_m(t|X)} &= \int \mathbf{p(t|X,\nu)\ dK(\nu|X).} \qquad \mathbf{[39]} \\ &= \int \mathbf{\{h(t|X,\nu)\ exp(-\textstyle\int_0^t h(s|X,\nu)\ ds)\}\ dK(\nu|X)} \end{aligned}$$

The unconditional distribution is then the average, taken with

[16] It is possible to interpret ν as measurement error in t, as measurement error in X, or as the combined effect of a left-out set of regressors Z (Lancaster, (1985)).

respect to the mixing distribution, K. The survivor function is similarly defined as

$$S_m(t|X) = \int \exp(-\int_0^t h(s|X,\nu)\, ds)\}\, dK(\nu|X), \qquad [40]$$

and the mixture hazard function can be found by differentiating minus the logarithm of [40] to get

$$h_m(t|X) = \frac{\int h(t|X,\nu) S(t|X)\, dK(\nu|X)}{\int S(t|X,\nu)\, dK(\nu|X)} \qquad [41]$$

Thus the mixture hazard function is a weighted average of the conditional hazard functions, where the weight function is

$$d\, w_t(\nu|X) = \frac{S(t|X)\, dK(\nu|X)}{\int S(t|X,\nu)\, dK(\nu|X)} \qquad [42]$$

The averaging of ν is with respect to the distribution of v over the survivors at date t. Note that since large ν implies a large hazard, the mixture hazard eventually looks like the conditional hazard evaluated at inf ν. It can be shown (Heckman and Singer, (1984a), Proposition 1) that the slope of $h_m(t|X)$ is always less than the slope of $h(t|X,\nu)$.

Figure 4 illustrates this with a mixture of exponential hazard rates. In the figure there are two types - $h_1 = \nu_1 = .1$ and $h_2 = \nu_2 = .05$ - with an initial distribution $G(\nu_1) = .5$. The mixture hazard rate, h_m, starts halfway between the two conditional

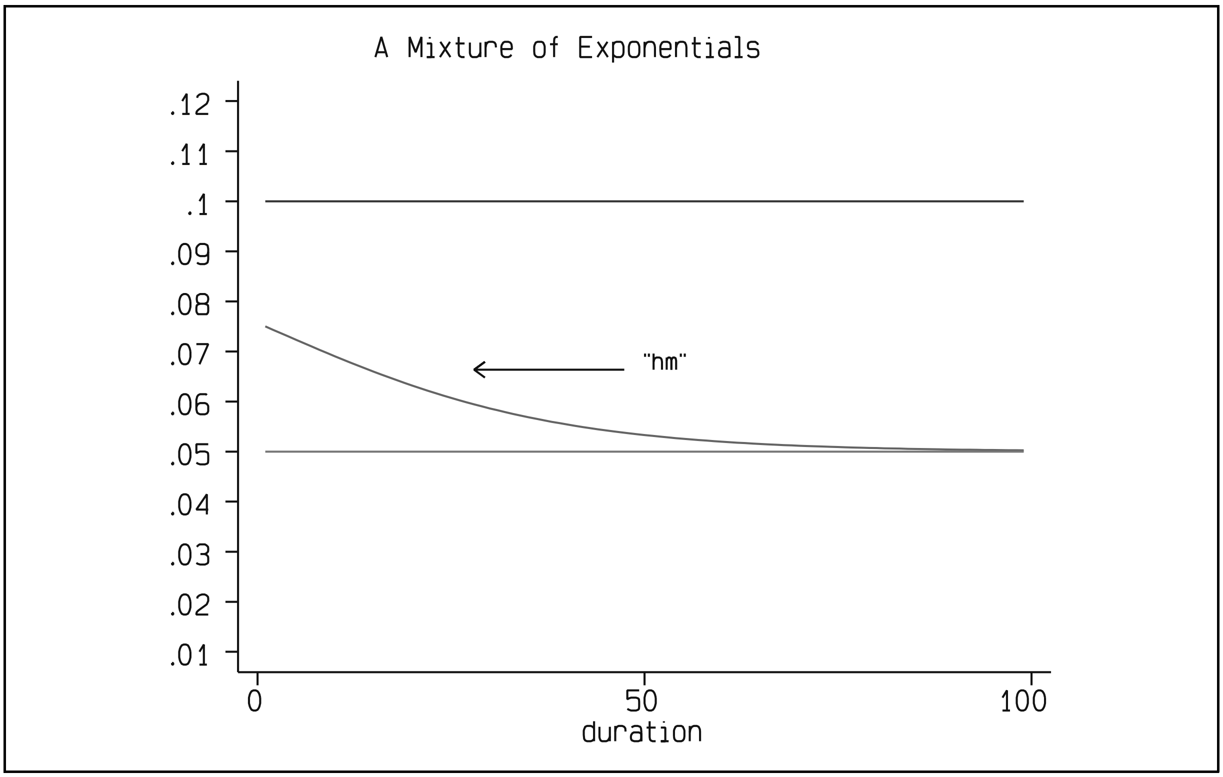

Figure 4

hazards and asymptotes to the lowest value. As the example shows, uncontrolled heterogeneity creates the appearance of duration dependence. Consequently, in cases where distinguishing duration dependence is important, methods must be used to control for unmeasured heterogeneity.

One approach to the problem of unmeasured heterogeneity is to assume a specific form for the mixing distribution, $G(v|X,\eta)$, where G is known up to a finite number of parameters, η. The mixture log likelihood function for N realizations of T and X is

$$\mathcal{L}(\beta,\eta) = \sum_{i=1}^{N} d_i(\log(p_m(T_i|X_i,\beta,\eta)) + (1-d_i)(\log(S_m(T_i|X_i,\beta,\eta) \quad [43]$$

and this function can be maximized using standard methods.

An example, due to Lancaster (1990), illustrates this approach. Assume that the heterogeneity acts multiplicatively on the hazard, i.e., $h(t|X,\nu) = \nu^* h(t|X)$. Suppose also that the conditional distribution of durations is exponential with parameter λ, that is, $p(T_i|X,\nu) = \exp(-\nu\lambda T_i)$. Let $dK(\nu|X)$ be a unit mean Gamma density.[17] The density function of the mixture is

$$p_m(T_i|X) = \int_0^\infty v^*\exp(-v^*\lambda^* T_i)\ v^{\beta-1}\exp(-\beta v)dv/\Gamma(\beta)$$

---

[17] The Gamma density is defined as $p(y)= \beta^\alpha y^{\alpha-1}\exp(-\beta y)/\Gamma(\alpha)$. Its mean is α/β, and its variance is α/β^2. The unit mean Gamma has $\alpha=\beta$, and variance = $1/\beta$.

$$= \{(\lambda\beta)/(\lambda T_i + \beta)\}^{1+\beta} \qquad [44]$$

where the assumption of unit mean ($\alpha=\beta$) has been used to arrive at [44]. The log likelihood for a sample of N independent observations is

$$\mathcal{L}(\lambda,\beta) = (1+\beta) \sum_{i=1}^{N} \{\ln(\lambda\beta) - \ln(\lambda T_i + \beta)\}. \qquad [45]$$

Maximization of $\mathcal{L}$ with respect to λ and β is straightforward, although solutions in the natural parameter space $\{ 0 < \beta \}$ may fail to exist. When this occurs it suggests that a mixture model is inappropriate. While some theoretical models deliver a specification of $p(t|X,\nu)$, to my knowledge none delivers a complete specification of the distribution of heterogeneity, observed or unobserved. Consequently, the choice of a mixing distribution, such as the Gamma in [44], is usually made on other grounds, such as computational ease.

Heckman and Singer [1984a,b] have criticized this approach, arguing that it over-parameterizes the model, and that a faulty choice of K leads to inconsistent estimates of the parameters of interest. They argue instead for the use of a non-parametric estimate of the mixing distribution based on the results of Lindsay [1983a, 1983b] and Laird [1978]. The maximum likelihood estimator of the mixing distribution is a discrete distribution with the number of points of support, s, being no greater than N, the sample size [Heckman and Singer 1984a, proposition 9]. In applications N is typically larger than 300 while s frequently turns out to be small, say 3 or 4. Why this is so is not well understood.

To see what the nonparametric approach achieves, suppose the number of points of support, s, and their locations, ν_i, i=1,..,k, are known. Dropping the explicit conditioning on X for notational convenience, notice that the survivor function $S_m(t)$ is observed at $N \geq k$ distinct points and consequently that the probability mass functions, $g(\nu_i)$ can be solved for in terms of the assumed conditional survivor functions and the known support points as in

$$S_m(t) \quad = \sum_{i=1}^{k} \exp(-\int_0^t h(s|\nu_i)\, ds)\}\, g(\nu_i). \qquad [46]$$

In general [46] is a set of N equations in the s unknowns, $g(\nu_i)$. Lindsay [1983a] discusses the issue of the existence of solutions for $g(\nu_i)$ and how to find the support points[18].

[18] Hu and Sickles [1994] provide two additional estimators for the mixing distribution, essentially by smoothing the discrete estimator of Heckman-Singer. The asymptotic distribution of these estimators is also unknown. Heckman [1990] provides a method of moments estimator with the same features.

The non-parametric mixing approach has not seen much use in applied econometrics for several reasons. First, it is computationaly demanding and multiple local maxima commonly appear. Second, no asymptotic distribution theory for the estimator has been produced. Finally, there is the belief, and some evidence, that a more flexible parameterization of the conditional hazard function coupled with a flexible parametric specification of the mixing distribution is sufficient to avoid substantial bias in estimating the structural parameters of interest. Monte carlo evidence in Sueyoshi [1991], Han and Hausman [1990], Ridder [1986] and Ridder and Verbakkel [1983] supports this interpretation, although no general theorem explaining the mechanics has yet been produced. While the point that flexible modelling of the conditional hazard function may avoid, or at least lessen, the need for the non-parametric maximum likelihood approach in reduced form models, structural models of the sort that arise in search models, which were the original target of Heckman and Singer [1984a], do not benefit from this because the theory tightly predicts the form of the conditional hazard function. In these cases, if heterogeneity is an important concern, non-parametric estimation of the mixing density may be needed.

Treatment of unobserved heterogeneity becomes more interesting with data on multiple spells. Chamberlain [1985] noted that common factors canceled from the proportional hazard model and proposed eliminating unmeasured heterogeneity by conditioning on the common factors. As an example, suppose that each individual in the sample has exactly two spells of unemployment whose hazards are given by

$$h_{ij}(T_{ij} | X_{ij}, \nu_i) = \lambda_0(T_{ij})\exp(X_{ij}\beta)\nu_i \quad , j = 1,2; \; i = 1,...,N. \qquad [47]$$

The within-person partial likelihood for the longer of the two spell-lengths is

$$L_i = (h_{i1}^d \; h_{i2}^{1-d})/(h_{i1} + h_{i2})$$

$$= \frac{\exp(X_{i1}\beta)^d \; \exp(X_{i2})^{1-d}}{\exp(X_{i1}\beta) + \exp(X_{i2}\beta)} \qquad [48]$$

where d = 1 if spell 1 was the longer spell. Note that both λ_0 and ν_i cancel out in [47]. The proposed partial likelihood function is

$$L(\beta) = \prod_{i=1}^{N} L_i \qquad [49]$$

Of course for β to be estimable there must be within-person variation in X. Thus the practical issues are the same as in linear models: eliminating fixed effects may generate a selection bias because $\Delta X_i = 0$ for some individuals. Note that although we have used the index i to refer to persons, it may well be that the common effects are due to some other grouping variable, for example family, or industry. Little research is available about conditioning on groupings other than person-specific conditioning.

4. **Search Theory and Duration Models.**

Duration models have played an important role in several areas of economics, including the analysis of strike lengths (Lancaster, 1972; Kennan (1985), Schnell and Gramm (1987)), timing and spacing of births (Newman and McCullough (1984); Wolpin (1984)), and optimal replacement or renewal policies (Rust, (1987); Pakes, (1986)). The largest application has been in the area of search economics applied to the labor market. Indeed, Devine and Kiefer [1991] survey over 600 empirical studies dealing with labor market search, which indicates the scope of application.

My goal in this section is not to be encyclopedic; instead, I present a stylized development of the search ideas and illustrate how they have been used in and developed by duration models. I start with the simple search model, the case of searching without recall for a job that will last forever. This model has been a workhorse in applications using single spell data. While the model is very simple, the insights gained from it translate easily into more complicated models with multiple, endogenously chosen labor market states, possibly allowing for duration dependence in durations and in transition probabilities. A central feature in this literature has been the delicate interplay between theory and econometric practice. Unlike other areas, in search applications the stochastic element that drives the model plays an integral part in the theory and the econometrics rather than being just tacked on as a residual. But even when great efforts have been made to interrelate the structure and the stochastic specification some compromises in modeling must be made if only for computational reasons.

4.1 Search Theory

In the simplest version of the search model individuals are modeled as infinite-lived, with discount rate r, who shop for jobs that pay wages = w, where w is a draw from a wage offer distribution with cumulative distribution function F[w]. Job availability matters in the sense that the offer arrival rate, λ, is less than one. Once a job is accepted employment lasts forever. While searching, individuals receive a benefit of b per period. The goal of the individual is to maximize wealth, defined as the present value of the sum of future earnings and payments received while searching. As Mortensen [1970, 1986] shows, under these circumstances a reservation wage policy is optimal and exists. A reservation wage policy is one where the decision is : accept a job paying w iff $w \geq w^r$; otherwise continue to search. The quantity w^r is called the reservation wage and in the simple case under consideration it is the solution to

$$w^r = b + (\lambda/r) \int_{w^r}^{\infty} (w - w^r)\, dF(w), \qquad [50]$$

an expression that appears in many guises in the search literature. In this notation the instantaneous probability of a job offer arriving is λ, and since search is assumed to be random, the probability of an acceptable job offer arriving, Π, is the product of the arrival rate and the likelihood that it is acceptable,

$$\Pi = \lambda (1-F[w^r]). \qquad [51]$$

The simplifying assumptions used in the theory --stationary wage distribution, constant arrival rate, infinite life -- deliver a strong result. Search durations are exponentially distributed with intensity parameter Π, i.e.,
$t \sim \Pi \exp(-\Pi t)$.

The distribution of accepted wages is the truncated part of the wage offer distribution,

$$g(w) = \frac{f(w)}{1-F[w^r]} \qquad [52]$$

If both search duration and wages are observed the joint distribution of duration and wage is

$$g(w,t) = \lambda \exp(-\lambda(1-F[w^r]t)f(w)I(w \geq w^r) \qquad [53]$$

where I(X) is the indicator that the event X occurs. The inclusion of $I(w \geq w^r)$ in [53] is essential for obtaining maximum likelihood estimates of w^r (Christensen and Kiefer (1991), Flinn and Heckman (1982)); omission of the indicator function results in a likelihood function that is monotone in w^r.

The assumptions used to generate a constant reservation wage model are strong and some have felt the need for relaxing them (Flinn and Heckman, 1982). It is not obvious how best to do so. One possibility could be that arrival rates change as spell lengths increase, i.e., $\lambda_t = \lambda_0^{-\rho t}$, perhaps because of a stigma effect. Alternatively, the benefit received while searching, b, might change due to Unemployment Insurance provisions, which will induce a changing reservation wage (Mortensen, (1977)). The assumption of a finite life also will eliminate a constant reservation wage (Gronau (1971)), but this is an aging effect, not a duration effect, and it seems hardly appropriate to introduce a finite lifespan of about 80 years as the explanation of a declining reservation wage among youths whose unemployment durations last on average 1 to 3 months. The point is that there are many ways to generate a non-constant reservation wage, although some might be more appropriate. In any event, the survivor function for the duration data is, in the case of a changing reservation wage,

$$S(t) = 1-G(t) = \exp(-\int_0^t \Pi(s)\, ds). \qquad [54]$$

with the density of duration times given by

$$g(t) = \Pi(s)\exp(-\int_0^t \Pi(s)\, ds). \qquad [55]$$

The joint density of durations and accepted wages, constructed in a manner similar to [53], is

$$g_t(w,t) = \lambda_t \exp(-\int_0^t \lambda_s (1-F_s[w_s^r] ds) f_t(w) I(w \geq w_t^r) \qquad [56]$$

Econometric models which analyze only duration data use [55] with a flexible functional form for $\Pi(s)$; this is called the reduced form approach. In contrast, several authors work jointly with accepted wages and durations and attempt to identify λ, F[w], and w^r; this approach is called structural.

The central feature of the search model is the reservation wage. In theory it is chosen so that the marginal gain to an additional search is equated to the marginal cost of search. Unfortunately, reservation wages are not usually observable. The reduced form approach, which has been the dominant approach for measuring public policy effects (see Meyer (1990); Anderson and Meyer (1993,1994)), does not require information about the behavior of reservation wages for implementation. The structural approach does, and its implementation requires a choice to: (i) work with reported answers to a question about lowest acceptable wage; (ii) adopt some approximation to the reservation wage in terms of observables; or (iii) use an exact (or even approximate) solution to an approximation of the value function. Each approach has some merit, and some costs. I discuss them in turn, starting first with the evidence on measured reservation wages.

One of the more enduring issues in labor market policy is the nature of the adjustments that unemployed workers make. In particular, it is often argued that reservation wages should fall with search duration, thereby hastening re-employment. In the language of duration models this is equivalent to positive duration dependence, the measurement of which I noted earlier is quite difficult. An early, pre-search theory, study by Kasper (1967)[19] examines the correlation of changes in reservation wages with unemployment duration. It is worth revisiting this early study to get a feel for the magnitudes involved.

Kasper used a sample of 3,000 workers who applied for Temporary Extended Unemployment Compensation in Minnesota during the period April, 1961 to September, 1961. The average duration of unemployment to date (i.e., these were interrupted spells) was

---

[19] Search theory is commonly dated in economics to start with Stigler's pioneering articles (Stigler, (1961; 1962)). Kasper's 1967 article is drawn from his 1963 unpublished Minnesota Ph.D. thesis, whose inspiration is Reder's published 1947 dissertation.

7.5 months; some spells were for over 2 years.[20] Included in the data were answers to the questions "What rate of pay did you receive from your last employer? (W_0)" and "What wage ... are you [currently] seeking? (W_1)". Kasper constructed the percentage change of asking wage to previous wages -- $Y = (W_0 - W_1)/W_0$ and fit the regression $Y = \beta_0 + \beta_1$ UDUR + ε, where UDUR is the duration of unemployment, measured in months. He obtained (standard errors in parentheses):

$$\hat{Y} = \underset{(0.857)}{-0.808} + \underset{(0.099)}{0.357}\ \text{UDUR} \qquad r = .068 \qquad [57]$$

From this he concluded that "...(1) the average asking wage of the unemployed is significantly *less* than their former wage, [and] (2) the average asking wage of the unemployed significantly *declines* over the duration of unemployment ...(emphasis in the original)" (Kasper (1967), p.165-6.) There are, of course, reasons to believe that this group of workers would have relatively high reservation wages, and that observed changes in their reservation wages would be small. After all, they had been unemployed for an average of 7.5 months at the time of the survey, so the sample is a highly selected one relative to the pool of all workers who were unemployed at this time. But even

[20] Kasper (1967), p.167.

if we ignore sample selection issues, there are two greater difficulties that restrain a conclusion that reservation wages actually declined.[21] First, the relation shown in [57] is completely consistent with a constant reservation wage but heterogeneous searchers. In the case where two, otherwise identical, workers face wage offer distributions that differ in their mean values, so that $\mu_1 > \mu_2$, worker 2 will set a lower reservation wage than worker 1 and, as Kiefer and Neumann [1979a] demonstrate, worker 2 also will have a longer duration of search. In this case if the reservation wage was observable a regression of it, or a monotone transform of it like [57], on completed spell lengths would produce a negative relation, although reservation wages were unchanging. This is just a particular example of the general difficulty of distinguishing duration dependence (changing reservation wages) from heterogeneity (differences in the cost of search), as Heckman has repeatedly argued.[22]

Second, there is the magnitude of the effect. The term "significantly declines" must refer to the implied t-value of 3.6

---

[21] Ignoring the obvious criticism that the asking wage question may not measure a worker's true reservation wage.

[22] Conversely, had the regression relation shown in [57] had a positive slope the data would be consistent with a constant reservation wage with heterogeneity in the cost of search, rather than, as here, in the benefits of search.

of the coefficient of UDUR, and therefore refer to statistical significance. What is amazing is that the effect is so small. An entire year of unemployment would, for this sample, lead to a reduction in asking wages of 4%, an amount that seems quite small when compared to the 25% decline in wages experienced by workers who lost jobs due to plant closings around this time. In short, the data of this study, if they were to be believed relevant for the entire populace, would seem rather supportive of the constant reservation wage model.

More recent evidence on reservation wages is reported by Feldstein and Poterba (1984). They use a special study of job search methods of the unemployed contained in the May 1976 Current Population Survey of the U.S. A total of 4,668 persons were classified as unemployed, and 3,238 completed a set of questions dealing with search behavior. One of the questions asked was "What is the lowest wage or salary you would accept (before deductions) for this type of work?" Interpreting this as an individual's reservation wage Feldstein and Poterba [1984, Table 1, p.148] provide a summary of how the ratio of the reservation wage to previous earnings varies by duration of unemployment. Figure 6 shows the relation between the ratio of reservation wages and previous earnings to unemployment duration for the Feldstein -Poterba data and for the Kasper data as well.[23]

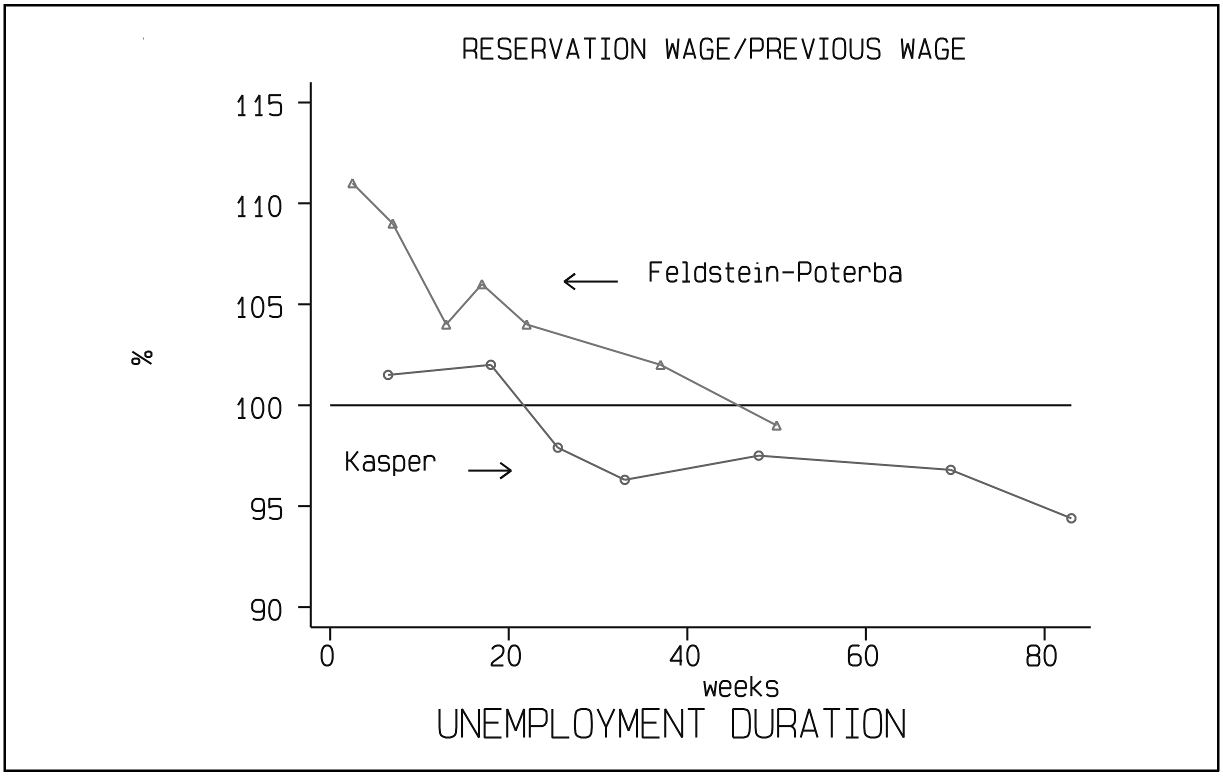

Figure 5

The relation between reservation wages and elapsed duration is remarkably similar in both studies. After 52 weeks of unemployment reservation wages had declined to 98% of previous earnings in Kasper's data, and to 97% of previous earnings in Feldstein-Poterba's data, levels that would not suggest much of a change in reservation wages. However, as figure 6 shows, the relation is steeper than the levels at 52 weeks of unemployment

---

[23] Data for Feldstein-Poterba are for all job losers and job leavers row 1 of their Table 1. Data for Kasper is computed from his figure 1.

would indicate because so many individuals state as reservation wages a number that exceeds, and often far exceeds, previous earnings. For example, of all unemployed workers in May, 1976 38% expressed a reservation wage that exceeded previous earnings and 27% expressed a reservation wage exactly equal to their previous wage. Perhaps even more curious is that 28% reported a reservation wage that was more than 110% of previous earnings. Conceivably workers on layoff might correctly report their reservation wage to equal previous earnings, and some workers might quit a job because it paid too lowly. As Feldstein and Poterba point out, for workers who were "job losers", i.e., not on layoff, not a quit or a labor force re-entrant, the fraction with a reservation wage exceeding previous earnings was 31%, and the fraction reporting a reservation wage in excess of 110% of earnings was 24%. Both numbers are inappreciably different from the total sample estimates.

It seems then that directly reported reservation wages are not a promising avenue for subsequent research. Given the state of the art in questionnaire design and interviewing techniques, individuals either can not or will not go through the decision calculus necessary to supply a meaningful answer about reservation wages. Accordingly, one will have to infer reservation wage patterns from other outcomes of the search process.

4.2 First Generation Search Models.

The motivation behind most early applications of search models was to examine the impact of public policy such as Unemployment Insurance (UI) on labor market outcomes. There are several ways that a program such as UI can alter search results, and these vary from program to program. For example, in the U.S. the individual states have different settings for the replacement rate -the ratio of Unemployment benefits to previous earnings- and for the maximum period for which benefits can be drawn. It is not clear a priori which policy lever would have the largest effects, and early research set out to measure this. Initial work, of which Burgess and Kingston (1971, 1977), Classen (1977), and Ehrenberg and Oaxaca (1976) is representative, proceeded to analyze models like

$$\begin{aligned} \mathbf{T_i} &= \mathbf{X_i}\, \beta_0 + \mathbf{UI_i}\, \beta_1 + \varepsilon_{1i} \\ \mathbf{Dln(W_i)} &= \mathbf{Z_i}\, \gamma_0 + \mathbf{UI_i}\, \gamma_1 + \varepsilon_{2i} \end{aligned} \qquad [58]$$

where UI is a vector of state-specific, possibly individual specific measures of the UI program, T_i is time spent unemployed, and Dln(w) is the difference in logarithms of post- and preunemployment wages. The standard method of analysis was to perform OLS on [58], possibly with some adjustment for heteroskedasticity. Many of the early studies used data from the employment service, so T_i refers to UI-compensated unemployment rather than total duration of unemployment.[24] In addition to the sample selection issue raised by eligibility for UI, this measurement strategy also occasions problems of censoring. Administrative records lose track of individuals when they exhaust their benefits, so the dependent variable in first part of [58] is censored, while that of the second part is completely unobserved. Note that this is in addition to the ordinary force of labor force withdrawal that is attendant to the job loss process. In the Burgess and Kingston data 11% of the Arizona UI claimants exhausted benefits, as did 41% of the recipients in San Francisco. In the same data, 23% of the recipients in Arizona report no earnings in the year following unemployment, as did 41% of those living in San Francisco. If these were independent processes (which they are not) somewhere between a 30% to 80% of the sample is being lost.

Not all early studies used administrative records as the source

[24] In the US new entrants and re-entrants to the labor market as well as in most states job quitters and those discharged for cause are ineligible for UI payments. Consequently, UI recipients are not a random sample of the unemployed; in 1992 UI recipients were 35% of all unemployed.

of unemployment spells. Ehrenberg and Oaxaca (1976) use data from the NLS survey for 1966-71 to measure durations and wage gains. This eliminates the censoring attendant to the use of administrative records in the US but it brings new problems. The NLS data, like most data sets, were not set up to record data by spell length, so only total weeks of unemployment divided by number of spells of unemployment can be used as a measure of search duration. Of course there is some gain in that data sets like the NLS have a richer set of controls, like age education, and family status that are absent in administrative data.

These early studies, although beset by technical problems of censoring and selectivity for which the econometric literature had yet to produce a treatment, were influential precisely because search ideas were in the air quite broadly, and many applied economists were looking for methods of dealing with duration data that were more tightly linked with the underlying economic theory.

4.3 Second Generation Studies.

Early studies of search outcomes can be criticized for applying inappropriate statistical methods to empirical models that were only loosely related to the economic theory; the hallmark of the second generation of search studies is an attention to specifying the stochastic process generating the data and then carefully developing the empirical model that is fit to the data. Lancaster (1979) and Nickell (1979) originally introduced the reduced form approach to analyzing unemployment durations. Attention to the data structure is particularly important. For example, Lancaster works with a sample of 479 unskilled workers who were found on the unemployment registers at date t, and interviewed some 5 weeks later. Had none of these workers found employment in the 5 week interregnum, the analysis would have been of completely censored data and could not have been informative. In an earlier incarnation these data might have been ignored because of their frailties, but Lancaster, and Nickell using similar data, exploited the hazard function approach to the data and teased out the inferences that could be made from partially censored duration data.

Lancaster [1979] introduces Cox's (1972) model {summarized in eq. [24] above} into the econometric literature, although he in fact does not estimate the general model. Instead, Lancaster first implements an exponential specification ($h(t|X) = \psi_1(X)\psi_2(t) = \exp(X\beta)$), and then allows for duration dependence, $\psi_2(t) = \alpha t^{\alpha-1}$. Finally, pointing out that unobserved heterogeneity can lead to misleading inferences about duration dependence, Lancaster deals with unobserved heterogeneity by assuming that ν affects the hazard multiplicatively, as in eq. [44], and that it has a Gamma distribution. Lancaster also points out that the likelihood function appropriate to the problem will depend upon how the data were gathered, in particular whether the data were sampled from the flow into unemployment or from the stock of unemployed.

Table 5

Summary of Lancaster's Duration Model Specifications

Specification	Replacement Rate	Duration Dependence
Exponential Hazard	-0.43 (0.21)	1.00 --
Weibull Hazard	-0.41 (0.21)	0.77 (0.09)
Weibull Hazard with Gamma Heterogeneity	-0.43 (0.26)	0.90 (0.22)

Source: Lancaster (1979), Tables 1, 2, and 5. Replacement rate is the log of the replacement rate; duration dependence is the coefficient, α, from the weibull specification. Standard errors are in parentheses.

The pattern of Lancaster's empirical results at this relatively early stage is interesting because it presages results that turn up over and over in the following decades. Focusing on the coefficient of the replacement rate (the ratio of unemployment payments to previous earnings, and on the duration dependence parameter, Lancaster's findings, shown in Table 5, indicate that adding regressors to the unemployment duration specification makes the model fit better, but even the best constant hazard rate specification is dominated by one that allows for duration dependence. Row 2 of table 5 shows that when the duration parameter, α, is not constrained to unity, as it is in the exponential specification, the MLE is significantly less than 1. However, note that the magnitude of the coefficient on the replacement rate is hardly changed. Row 3 of table 5 shows the results when both heterogeneity and duration dependence are allowed. As was suggested earlier, it is difficult to separate these competing explanations, and this difficulty manifests itself as a large standard error on α in column (3). In fact α now cannot be distinguished from 1, nor can the variance of ν be distinguished from zero. Thus the data are consistent with a pure heterogeneity explanation, a pure duration dependence explanation, or some mixture. But with 479 observations Lancaster cannot tell them apart. This pattern re-appears in many subsequent studies by a variety of authors. But note that, while disentangling duration dependence from heterogeneity is complicated, the coefficient (and its standard error) of the replacement rate are hardly affected, a result that was replicated by Nickell (1979) using British data and in Monte Carlo studies by Sueyoshi (1992) and Ridder (1986).

Nickell's (1979) approach is similar to Lancaster's, differing in three respects. First, and of lesser importance, Nickell uses a discrete time model as opposed to Lancaster's continuous-time formulation. Nothing important hinges on this specification, at least not at this level. Aggregation problems brought about by moving from weekly to quarterly, or even annual data are a different matter (see Bergstöm and Edin, (1992)), but these hardly play a role in this case - Nickell and Lancaster use duration data measured in weeks. The second difference is that Nickell adjusts for unobserved heterogeneity using a discrete distribution of heterogeneity rather than the more tightly paramaterized Gamma form used by Lancaster. Nickell uses $dK(\nu|X) = \nu_1$ with probability ϕ, and $dK(\nu|X) = \nu_2$ with probability $1-\phi$, where $<\nu_1, \nu_2, \phi>$ are parameters to be estimated. This discrete specification of the heterogeneity distribution anticipates subsequent development by Heckman and Singer (1984) of the non-parametric maximum likelihood estimator of the mixing distribution. Indeed, conditional on the correct choice of the number and location of points of support they are the same. Finally, Nickell allows for time variation in the effect of Unemployment Insurance and finds that the effect in his data are is confined to the first twenty weeks of unemployment. This sort of exploratory work, although it can lead to overfitting of the data, is very informative for the design of social welfare policies.

The elegant treatment of duration data by Lancaster and Nickell provided answers to half of the search economics research program. Kiefer and Neumann (1979a,b) provide an answer to the remaining part of the program. Using a discrete time model for job offers and treating λ, the offer arrival rate, as fixed at unity (which was the custom in theoretical search models at that time, cf. Mortensen (1970) and McCall (1970)), Kiefer and Neumann argued that the log of the joint density of a completed spell of unemployment, t, and a re-employment wage, w, was:

$$g(t,w) = \sum_{s=1}^{t-1} \log\{F(w^r(s))\} + \log\{ f(w)\, I(w \geq w^r(t))\} \qquad [59]$$

where F and f were the cdf and density of the wage offer distribution. Thus, in the constant reservation wage case the joint density was the product of a binomial term --the probability of getting t-1 failures followed by 1 success-- and the conditional density of wages given that the wage offer exceeded the reservation wage. Incomplete spells were handled as censored in the same manner as Lancaster (1979).

To implement the model Kiefer and Neumann assumed that offer wages and reservation wages could be written as

$$\ln(w_i^o) = X_i\beta + \varepsilon_i^o$$

$$\ln(w_i^r) = Z_i\gamma + \varepsilon_i^r \qquad [60]$$

with $[\varepsilon^o, \varepsilon^r]$ distributed $N(0,\Sigma)$. In this formulation w_i^r is regarded as a linear approximation to the solution to equation [50], composed of observable elements, Z, and unobservable (to the econometrician) elements, ε^r. The theory underlying equation [50] implies that all elements that affect wage offers, X, also affect reservation wages. Hence, $X \supset Z$. Identification requires that there be some elements in Z that are not in X.

Letting $Y_i = \ln(w_i^o) - \ln(w_i^r) = X_i\beta - Z_i\gamma + u_i$, where $u_i = \varepsilon_i^o - \varepsilon_i^r \sim N(0, \sigma_r^2 + \sigma_o^2 - 2\sigma_{r,o}^2)$, and $y_i = Y_i/\sigma_u$ the econometric problem can be cast, for the case of a constant reservation wage only, in the familiar framework of Heckman's (1976) two-step selectivity model. In this case a probit is fit to data on y_i, which yields consistent estimates of $(X\beta - Z\gamma)/\sigma_u$. These estimates are in turn used to construct the inverse Mills ratio, $\lambda_i = \phi(-y_i)/(1-\Phi(-y_i))$, which is needed to obtain consistent estimates of the wage equation regression:

$$E(\ln(w^o)) = X\beta + \rho\sigma_o \lambda_i \qquad [61]$$

where $\rho = (\sigma_o^2 - \sigma_{r,o})/(\sigma_o\sigma_u)$. Computing the uncensored mean as $X_i\beta$ and re-estimating the probit equation, now called the structural probit, yields an estimate of σ_u. Subject to the identification condition this yields estimates of the wage offer and of the reservation wage equation.

For the non-constant reservation wage case this regression shortcut is not available and estimation must make use of the full likelihood function. The issues are the same, with the addition of a time varying term involving elapsed duration, s. Kiefer and Neumann (1979b) use the linear specification

$$\ln(w_i^r(s)) = Z_i\gamma + s\delta + \varepsilon_i^r \qquad [62]$$

but it is clear that any function of elapsed duration, k(s), could be used instead. Like Lancaster (1979) they find substantial evidence of a declining reservation wage. In subsequent work (Kiefer and Neumann (1981, 1982)) they follow Lancaster and account for unobserved heterogeneity by specifying a distribution for it and integrating it out. Evidence in favor of a declining reservation wage remains but is substantially muted when heterogeneity is allowed. Integrating the heterogeneity out is difficult in this specification because the optimality constraint must be satisfied for values of the heterogeneity parameter. That is, writing the employment condition index as:

$$Y_i(s|\nu) = (X_i\beta - Z_i\gamma - s\delta + \nu)/\sigma \qquad [63]$$

it can be seen that not all values of ν are consistent with employment patterns. High values of ν are inconsistent with workers not finding employment, as are low values of ν for those who do become employed. Moreover, when working with the joint density of wages and durations it matters what the source of the unobserved heterogeneity is. Thus if ν affects only costs of search, or utility while unemployed, then heterogeneity enters the employment condition only through the reservation wage. However, if the heterogeneity enters through the wage offer distribution, then the heterogeneity must be consistent with both the employment rule and the observed pattern of wages. These restrictions make the problem of maximizing the likelihood function non-standard, and the usual methods of inference about the parameters β, γ do not apply directly. Methods for dealing with these issues are the subject of much current research. (See Christensen and Kiefer, (1994) for one approach.)

The Kiefer-Neumann approach to structural modelling of the joint distribution of wages and durations involves using a log-linear approximation to the solution of equation [50]. An alternative approach, employed by Narendranathan and Nickell (1985) and Wolpin (1984) is to find exact solutions to approximations of the value functions that underlie equation [50]. Wolpin (1987) goes even further and proposes to calculate equation [50] exactly.

Given knowledge of the wage distribution $F(w|\beta)$ up to a finite set of parameters β, the arrival rate of offers λ, the value of utility while unemployed, b, and the discount rate, r, one can iterate the contraction mapping implicit in [50] to solve exactly for w_i^r. Faster methods are available for the solution of [50] since it can be shown to be a Volterra equation, for which special solutions are available (Tricomi (1957); Linz (1985)). Wolpin (1987) considers the case of a declining reservation wage, in which case an interrelated series of equations like [50] must be solved. Specifically, Wolpin models the reservation wage to satisfy

$$w_s^r = b + (1/(1+r))\, w_{s+1}^r\,, \quad s < TMAX$$

$$w_{TMAX}^r = \sum_{s=TMAX}^{T+R} \{(1/(1+r))^{s-1} \prod_{j=1}^{s-1} [1-\lambda(j)]\ \}\{\lambda(s)E(w) + (1-\lambda(s))b\} \quad [64]$$

Arrival rates are modeled as coming from a distribution with drift, i.e., $\lambda(t) \sim \Phi(L_0 + L_1 t)$, a characterization that by itself would produce a changing reservation wage. The maximum search period TMAX, after which an individual will accept any job offer, the expected value of which is given in the last line in [64], and the length of life after T is reached are "free" parameters that can be chosen to calibrate the model to the data. While this approach is computationally more demanding than previous alternatives and in this sense more expensive, comparatively little is known about the benefits to using such models. Comparisons among different approximation strategies, which will depend upon the types of data available (for example, whether observations are available on b, the amount of net income received while employed), are clearly needed.

4.4 Multiple state duration models.

Economists have observed transitions among labor market states of employed, unemployed, and not in the labor force (see Troikka (1976); Marston (1976)) primarily with a view towards interpreting changes in the stocks in each labor market state. In this sense, treatment of the transition data was basically "unemployment accounting." In the 1980's a number of authors proposed behaviorally interesting models of labor market activity where arrival rates and offer distributions influence movement from state to state. All of the estimable models of this sort share the feature of being renewal processes. Tuma and Robbins (1980) were one of the earliest such empirical studies of a two-state model (Employed, Unemployed, or E,U) of labor force attachment. In their approach parameters of government income subsidy programs (i.e., Seattle and Denver Income Maintenance Experiments) influenced the lengths of employment and unemployment spells. More formal two-state models were articulated by Flinn and Heckman (1982), and Burdett, Kiefer, and Sharma (1985). Three state models were proposed and formally studied by Burdett et al. (1984 a,b), Flinn and Heckman (1983), and Mortensen and Neumann (1984). All of these models are cast as competing risk models where the transition from state i to state j is typically modeled as:

$$\Pi_{ij}(X_1,X_2) = \lambda_i(X_1)p_j(X_2) \qquad [65]$$

where λ_i is the arrival rate of some event ("news") in state i, which may depend on a vector of characteristics X_1, and p_j is the probability that state j will be chosen, which may depend on characteristics X_2, some of which might be specific to state i. The hazard out of state i is

$$\Pi_i(X) = \sum_{i \neq j} \Pi_{ij}(X) \qquad [66]$$

and the density of the state specific durations is:

$$g_{ij}(t|X) = \Pi_{ij} \exp(-\int_0^t \Pi_i(X)\, ds). \qquad [67]$$

No new issues arrive in estimating reduced form versions of these competing risk models. Correlations between spells can be allowed due to, say, individual specific components (Flinn and Heckman, (1983)), but this only serves to make the likelihood function to be non-separable across states. More difficult problems arise when there are three or more states. To generate a meaningful K-state model, there must be K-1 shocks. In the Burdett et al. model, for example, there are shocks to the wage process and shocks to the value of home time process that generate movements among the states of employment, unemployment, and non-participation. A reservation wage characterizes the simple search model shown above, and an analogous characterization of wages, w, and value of home time, v, partitions the sample space into sets A_i, i= E,U,N, where $\langle w,v \rangle \varepsilon A_i$ implies that state i is chosen. The problem is that the reservation wage is unknown, as before, but so is the entire process of changes in the value of home time. Consequently, empirical work with multi-state models has used the reduced form approach.

4.5 Equilibrium Search Models.

The prototypal search model provides a simple but elegant description of unemployment durations and accepted wages, conditional on the form of the wage distribution being known, but says nothing about the distribution itself. This is somewhat paradoxical, and places the theory of wage search as an explanation not of wages, but of durations. This paradox was pointed out by Diamond (1971), which generated a substantial literature deriving equilibrium wage or price distributions.[25] Subsequent developments by Albrecht and Axell (1984), Axel (1976), Burdett (1990), and Mortensen (1990) provide the theoretical basis for the empirical work that has ensued. To fix ideas, I briefly describe Mortensen's (1990) model for the case where workers and firms are each homogeneous.

Workers have a reservation wage, w^r, which solves the usual search problem for wealth maximization. Unemployed workers see jobs arrive at rate λ_0, and they accept the first job that offers more than their reservation wage. While employed at wage W a worker's reservation wage is also W. Job offers arrive at a rate λ_1 while employed and jobs "disappear" at the rate δ. Thus the steady-state reservation wage solves

$$w^r = b + (\kappa_0 - \kappa_1) \int_{w^r}^{\infty} \frac{(1-F(x)) \, dx}{[1 + \kappa_1(1-F(x))]} \qquad [68]$$

where $\kappa_i = \lambda_i/\delta$, $i = 0,1$.

---

[25] See, for example, Butters (1977), Reinganum (1979), and Burdett and Judd (1983).

Firms are identical with productivity level P, face constant returns to scale in production, and maximize profits by choosing the wage to pay. The balancing condition which equates supply and demand is that firms will offer higher wages if and only if they can expect to get an additional number of workers to cover the lower per worker profits. Higher wages attract more workers to a firm and allows firms to retain the workers longer. The unique equilibrium wage distribution implied by this process of wage and employment determination is:

$$F(w) = \frac{[1+\kappa_1]}{k_1} \left\{ 1 - \frac{(P-w)^{1/2}}{(P-w_r)^{1/2}} \right\} \qquad [69]$$

Here, the fundamental parameters are P, the productivity level, and the three arrival rates, λ_0, λ_1, and δ. Together, these three arrival rates determine the degree of competition in the labor market. As $\lambda_1 \rightarrow 0$, the Diamond's monopsony wage occurs: all workers are offered a wage = b; as $\lambda_1 \rightarrow \infty$, workers receive the competitive wage, P. Thus the income distribution in the model is intimately related to the performance of the labor market. Mortensen shows that the lowest wage offer will satisfy $w_L = w_r$, that is, workers will accept any job offered to them, and the highest wage offered will satisfy

$$w_H = P - \{1/(1+k_1)^2\}(P - w_L). \qquad [70]$$

The basic idea underlying the model is that workers prefer high wages to low, and that the process of job mobility will carry them through the wage distribution as the locate higher paying jobs. The exogenously given rate of job destruction, δ, prevents too much piling up at the top end. For firms, wage policy matters, unlike in the standard competitive model. Paying a low wage will attract some workers but turnover will be high. Since profits are $(P-w)\ell(w)$, low-wage firms make large profits per employee, but attract and retain few workers. Alternatively, a firm could pay a high wage, which will attract a larger labor force and retain it longer. Profit per worker will be lower, but the firm makes it up in volume. Because all firms are ex ante identical Nash equilibrium in this model requires that expected profits be equal, which implies that the equilibrium wage distribution traces out the equal profit constraint. And this means that the wage density must have an upward sloping density throughout its range, a result that seems at odds with evidence on wage distributions.

Obviously the assumption that workers and firms are homogeneous is an abstraction, and econometric models have attempted to relax this assumption in different ways. Eckstein and Wolpin (1990) estimate a related equilibrium search model due to Albrecht and Axell (1984). In the Albrecht-Axell model workers have differing values of non-market time, b, and there are K worker types. Firms may differ in their productivity, but because each firm knows that there are K types of workers, each firm specializes in offering a wage equal to the reservation wage of one of the K types of workers. Data on unemployment spells and wages received on the job are needed to estimate the parameters of this model. In this model the distribution function of wages is a step function with discrete jumps at each of the K reservation wages. Because the number of distinct wages seen in any data set would make K a very large number, Eckstein and Wolpin keep K manageable by assuming that there is measurement error in recorded wages. Applying their specification to NLSY data on wages and search durations produces moderately good fits to the duration data, but a terrible fit to the wage data, even though measurement error in wages is explicitly treated.

Kiefer and Neumann (1994) estimate the homogeneous equilibrium search model described above using panel data from the NLSY similar to that used by Eckstein and Wolpin. They look at the spell of unemployment between the end of formal education and the first job, the wage received on the first job, the duration of the job, and why the job ended. To produce homogeneous samples they stratify on sex, level of formal education, and race. Kiefer and Neumann (1994) propose the estimators

$$\mathbf{w^r = \min \{ w_i \}, \quad w^H = \max \{ w_i \}} \qquad [71]$$

$$\hat{\mathbf{p}} = \mathbf{[(1+}\eta\mathbf{)/} \eta\mathbf{] \, w^H - [1/} \eta\mathbf{] \, w^r} \qquad [72]$$

where

$$\eta = \mathbf{(1 +} \lambda_1\mathbf{/}\delta \mathbf{)^2 - 1 > 0} \qquad [73]$$

The estimators w^r and w^H are super-efficient estimators and the theory of local cuts (Christensen and Kiefer, 1994) justifies conditioning on these values to estimate the other parameters and allows development of the appropriate distribution theory for the parameter $\theta = (\lambda_0, \lambda_1, \delta)$ from the profile likelihood with w^r and w^H substituted in for the true values. As van den Berg and Ridder (1994) note, these estimators are sensitive to measurement error, but at least for the case of classical measurement error in small samples their performance actually is improved. In small samples the extremum estimators are biased - w^r is over-estimated and w^H is underestimated; this bias disappears asymptotically. However, classical measurement error adds noise to each observation so that the observed sample minimum is likely to have negative measurement error, and conversely for the sample maximum. This partially offsets the bias in the extremum estimator. Of course, "classical" measurement error may be the least likely type of measurement error to find in large panel data sets, so these comforting results about measurement error may be only a matter of curiosity. Applying this model to the NLSY data produces a reasonable fit to the duration data, but bad fit to the wage data, a result that echoes Eckstein and Wolpin (1990). Bowlus, Kiefer, and Neumann (1996) use the same data but introduce heterogeneity in market productivity with a discrete mixture of firm productivity types. They find that they can fit the wage data arbitrarily well with a small number of types, typically less than 8. A discrete mixture has the disadvantage that the likelihood function becomes non-differentiable in the support points and probability weights of the mixture, but this problem can be handled by an application of the EM algorithm.

Van den Berg and Ridder have examined aspects of the equilibrium search model using a consistent methodology in several papers (Van den Berg and Ridder (1993, 1994); Koning, Ridder, and van den Berg (1996)). In van den Berg and Ridder (1993) they treat both within and between market heterogeneity across worker types by allowing the fundamental parameters to vary with regressors, i.e.,

$$\lambda_0 = \exp(X_0\alpha),\ \lambda_1 = \exp(X1_1\beta),\ \text{and}\ \delta = \exp(X_2\gamma) \qquad [74]$$

and they treat wage offers as being measured with error. In subsequent work (Konig, Ridder, and van den Berg (1996)) using samples of the Organization for Strategic Labour Market Research panel data they allow for the fundamental parameters to vary with regressors as in [74] and also treat market productivity, P, as heterogenous across markets. In this latter case they use a log normal for the mixing distribution. Their results suggest that a two-parameter distribution like the log normal provides an acceptable fit to the data.

Of course, this work on empirical equilibrium models is too recent to have been subjected to a variety of comparisons to indicate which parts of the models are robust, and which are sensitive, to changes in assumptions and techniques. There is clearly room for comparative work among these approaches, particularly as the computational demands of the models increases. That said, it is equally clear that the equilibrium search framework has open new research areas in econometrics and in labor economics, which will, I hope, be filled in the near future.

5. Summary.

A considerable amount of work has been undertaken in bringing the analysis of dynamic event histories into the mainstream of applied econometric work in microeconomics, and the hazard approach to modelling these events is now standard. This has led to a careful modelling of the probabilistic structure of the data, including the sampling plan that generated the data. These considerations were rarely present in earlier work. The standard workhorse of applied econometrics -- the linear model-- retains some applications in duration models, but its use is limited, particularly where time-varying covariates are an essential element of the problem. The initial attraction of non-regression models was their facility at incorporating censored observations. Current work on the effects of turnover and Unemployment Insurance (e.g., Anderson and Meyer (1993,1994), Meyer (1990)) routinely uses flexibly parameterized proportional hazard models even when censoring is not an issue.

Diagnostic testing of empirical duration models translates well, although not perfectly, from standard biostatistical applications. Tests for functional form and for heteroskedasticity are familiar from linear model applications. Graphical tests remain useful, and in some circumstances observed departures from a hypothesized relation can be formally tested. Cox's (1972) proportional hazard model is in frequent use, and diagnostic tests based on it are available in both discrete and continuous varieties, with a choice between the two made presumably by weighing power and computation time.

Because duration models are typically non-linear there is heightened concern about the effects of errors in measurement or missing data (unobservables) on inference. A complete solution to this problem, in particular, disentangling duration dependence from unobserved heterogeneity, is not yet available. Methods in wide use emphasize the fitting of parametric mixing models, which can lead to misspecification bias. Elegant alternatives which emphasize the fitting of nonparametric mixing distributions are computationally awkward and lack an asymptotic distribution theory. Both approaches are avenues of current research in statistics and applied econometrics.

Finally, the theory of search has had a major impact on the way economists view temporal processes. Simple models of job search activity suggest precise channels by which labor market policy choices --such as the level of unemployment insurance or the maximum length of benefit receipt-- will work and these models provide direct guidance on how to incorporate theory and uncertainty. Applications of duration models in applied econometrics have been relatively successful in reduced form and structural modelling of discrete state models of small dimension. Extending these models to larger dimensions, for example, analyzing life-time labor supply decisions, will require both new insights from theory and further advances in computation in the next decade.

References

Albrecht, J. W., and B. Axell, [1984], "An equilibrium model of search unemployment," Journal of Political Economy,92:824-40.

Anderson, P. M., and B. D. Meyer, [1993], "Unemployment insurance in the United States: Layoff incentives and cross-subsidies," Journal of Labor Economics, 11:S70-S95.

Anderson, P. M., and B. D. Meyer, [1994], "The extent and consequences of job turnover," Brookings Papers on Economic Activity: Microeconomics, 177-248.

Atkinson, A., Gomulka, J., Mickelwright, J., and N. Rau, [1984], "Unemployment benefits, duration, and incentives in Britain: How robust is the evidence?" Journal of Public Economics, 23:3-26.

Axell, B., [1976], "Search market equilibrium," Scandinavian Journal of Economics, 79,1:20-40.

Baltazar-Aban, I, and E. Peña, [1995], "Properties of hazard-based residuals and implications in model diagnostics,", Journal of the American Statistical Association, 90, no. 429:185-97.

Barlow, R. E., and F. Proschan, [1981], Statistical Theory of Reliability and Life Testing, McArdle Press: Silver Spring, MD.

Bergström, R., and P.-A. Edin, [1992], "Time aggregation and the distributional shape of unemployment duration," Journal of Applied Econometrics, 7:5-30.

Bowlus, A., Kiefer, N. M., and G.R. Neumann, [1996], "Estimation of equilibrium wage distributions with heterogeneity," Journal of Applied Econometrics, forthcoming.

Breslow, N. E., [1974], "Covariance analysis of censored survival data," Biometrics, 30: 89-99.

Breslow, N. E., Elder, L., and R. D. Gill, [1984], "A two-sample censored-data rank test for acceleration," Biometrics, 40:1049-62.

Burdett, K., [1990], "Empirical wage distributions: A new framework for labor market policy analysis," in J. Hartog, G. Ridder, and J. Theeuwes, eds., Panel Data and Labor Market Studies, Amsterdam:North-Holland,297-312.

Burdett, K., and K. Judd, [1983], "Equilibrium price distributions," Econometrica, 51,4:955-970

Burdett, K., Kiefer, N. M., Mortensen, D.T., and G. R. Neumann, [1984a], "Earnings, unemployment, and the allocation of time over time," Review of Economic Studies, 51:559-78.

Burdett, K., Kiefer, N. M., Mortensen, D.T., and G. R. Neumann, [1984b], "Steady states as natural rates in a dynamic discrete choice model of labor supply, " in G.R. Neumann and N. Westergård-Nielsen, eds., Studies In Labor Market Dynamics, Heidelberg:Springer-Verlag,74-97.

Burdett, K., Kiefer, N. M., and S. Sharma, "Layoffs and duration dependence in a model of turnover," Journal of Econometrics, 28:51-69.

Burgess, P. L., and J. L. Kingston, [1971], "Unemployment insurance, job search, and the demand for leisure: Comment," Western Economic Journal, 9,4:447-50.

Burgess, P. L., and J. L. Kingston, [1977], "The effects of unemployment insurance benefits on reemployment success," Industrial and Labor Relations Review, 30,3:25-31.

Butler, R., and R. Ehrenberg, [1981], "Estimating the narcotic effect of public sector impasse procedures," Industrial and Labor Relations Review, 35,1:3-20.

Butters, G., (1977), "Equilibrium distribution of sales and advertising prices," Review of Economic Studies, 44:465-91.

Chamberlain, G., [1985], "Heterogeneity, omitted variable bias, and duration dependence," in J. Heckman and B. Singer, (eds.), Longitudinal Analysis of Labor Market Data, Cambridge: Cambridge University Press.

Chesher, A., Lancaster, T., and M. Irish, [1985], "On detecting the failure of distributional assumptions," Annales de l'Insee, 59/60:7-44.

Chesher, A., and R. Spady, [1991], "Asymptotic expansions of the information matrix test," Econometrica, 59, May:787-815.

Christensen, B. J. and N. M. Kiefer, [1994], "Local cuts and separate inference," Scandinavian Journal of Statistics, 20:1-13.

Christensen, B.J., and N. M. Kiefer, [1991], "The exact likelihood function for an empirical job search model," Econometric Theory, 7:464-86.

Classen, K., [1977], "The effect of unemployment insurance on the duration of unemployment and subsequent earnings," Industrial and Labor Relations Review, 30,4:438-44.

Cox, D. R., [1972], "Regression models and life tables," Journal of the Royal Statistical Society, series B, 34:187-220.

Cox, D. R., [1975], "Partial Likelihood," Biometrika, May/Aug, 62, 2: 269-76.

Cox, D. R., and E. J. Snell, [1969], "A general definition of residuals," Journal of the Royal statistical Society Series B, 20:215-32.

Devine, T. J., and N. M. Kiefer, [1991], Empirical Labor Economics: The Search Approach, New York:Oxford University Press.

Diamond, P., [1971], "A model of price adjustment," Journal of Economic Theory,3:156-68.

Eckstein, Z., and K. I. Wolpin, [1990], "Estimating a market equilibrium search model from panel data on individuals," Econometrica, 58:783-808.

Ehrenberg, R. G., and R. L. Oaxaca, [1976], "Unemployment insurance, duration of unemployment, and subsequent wage gain," American Economic Review, 66:754-66.

Engle, R.F., Hendry, D.F., and J.F. Richard, [1983], "Exogeneity," Econometrica, 51:277-304.

Feldstein, M., and J. Poterba, [1984], "Unemployment insurance and reservation wages," Journal of Public Economics, 23,1/2:141-67.

Flinn, C. J., and J. J. Heckman, [1982], "New methods for analyzing structural models of labor force dynamics," Journal of Econometrics, 18:115-168.

Flinn, C. J., and J. J. Heckman, [1983], "Are unemployment and out-of-the-labor force behaviorally distinct labor force states?," Journal of Labor Economics, 1:28-42.

Gill, R. D., and Schumacher, M., [1987], "A simple test for the proportional hazards assumption," Biometrika,74:289-300.

Greene, W.H., [1993], Econometric Analysis, 2nd ed., MacMillan:New York.

Gronau, R., [1971], "Information and frictional unemployment," American Economic Review, 61,3:290-301.

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

Hausman, J., and D. Wise, [1977], "Social experimentation, truncated distributions, and efficient estimation," Econometrica, 45,4:919-38.

Heckman, J. J., [1976], "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models," Annals of Economic and Social Measurement, 5:475-92.

Heckman, J. J., [1990], "A nonparametric method of moments estimator for the mixture of geometrics model," in J. Hartog, *et al.* (eds.), Panel Data and Labor Market Studies, Amsterdam, North-Holland:69-80-117.

Heckman, J. J., and B. Singer, [1984a], "A Method for minimizing the impact of distributional assumptions in econometric models for duration data," Econometrica, 52, 2:271-320.

Heckman, J. J., and B. Singer, [1984b], "Econometric Duration Analysis," Journal of Econometrics, 24:63-132.

Horowitz, J. L., [1986], "A distribution-free least squares estimator for censored linear regression models," Journal of Econometrics, 32:59-84.

Horowitz, J. L., [1996], "Semiparametric estimation of a regression model with an unknown transformation of the dependent variable", Econometrica, 64:103-38.

Horowitz, J. L., [1988], "The asymptotic efficiency of semiparametric estimators for censored linear regression models," **Studies in Empirical Econometrics: Semiparametric and nonparametric econometrics**, Physica, Heidelberg, 1-18.

Horowitz, J. L., [1994], "Bootstrap-based critical values for the information matrix test," Journal of Econometrics, 61:395-411.

Horowitz, J. L. and G. R. Neumann, [1988], "Semiparametric Estimation of Employment Duration Models," Econometric Reviews, 6:5-40.

Horowitz, J. L., and G. R. Neumann, [1989a], "Specification testing in censored regression models: parametric and semiparametric methods," Journal of Applied Econometrics, 4:S61-S86.

Horowitz, J. L., and G. R. Neumann, [1989b], "Computational and Statistical Efficiency of Semiparametric GLS Estimators of Censored Regression Models," Econometric Reviews, 8:223-225.

Horowitz, J. L., and G. R. Neumann, [1992], "A generalized moments specification test of the proportional hazards model," Journal of the American Statistical Association, 87(417):234-40.

Hu, K., and R. C. Sickles, [1994], "Estimation of the duration model by nonparametric maximum likelihood, maximum penalized likelihood, and probability simulators," The Review of Economics & Statistics, LXXVI,4:683-94.

Johansen, S., [1983], "An extension of Cox's regression model," International Statistical Review, 51:165-74.

Johansen, S., [1978], "The product limit estimator as maximum likelihood estimator," Scandinavian Journal of

Statistics, 5:195-99.

Jovanovic, B., [1979], "Job-matching and the theory of turnover," Journal of Political Economy, 87:972-90.

Jovanovic, B., [1984], "Matching, turnover, and unemployment," Journal of Political Economy,92:108-22.

Kaitz, H., [1970], "Analyzing the length of unemployment spells," Monthly Labor Review, 93:11-20.

Kalbfleisch, J. D., and R. L. Prentice, [1973], "Marginal likelihoods based on Cox's regression and life model," Biometrika, 60:267-78.

Kalbfleisch, J. D., and R. L. Prentice, [1980], The statistical Analysis of Failure Time Data, New York, John Wiley.

Kaplan, E.L., and Paul Meier, [1958], "Nonparametric estimation from incomplete observations," Journal of the American Statistical Association, 53:457-81.

Karlin, S., and H. M. Taylor, [1975], **A Fisrt Course in Stochastic Processes**, 2nd ed., Academic Press, New York.

Kasper, H., [1967], "The asking price of labor and the duration of unemployment," Review of Economics and Statistics, 49,2:165-72.

Kay, R., [1977], "Proportional hazard regression models and the analysis of censored survival data," Applied Statistics,26:187-220.

Kennan, J. F., [1985], "The duration of contract strikes in U.S. manufacturing, "Journal of Econometrics, 28:5-28.

Kennan, J. F., and G. R. Neumann, [1988], "Why does the information test reject too often?," University of Iowa Working paper, no. 88-4, January.

Kiefer, N. M., [1988a], "Economic duration data and hazard functions," Journal of Economic Literature, XXVI, 2, June:646-679.

Kiefer, N. M., [1988b], "Analysis of Grouped Duration Data," Contemporary Mathematics, 80:107-37.

Kiefer, N. M., [1990], "Econometric models for grouped duration data," in J. Hartog, *et al.* (eds.), Panel Data and Labor Market Studies, Amsterdam, North-Holland:97-117.

Kiefer, N. M., and G. R. Neumann, [1979a], "Estimation of wage offer distributions and reservation wages," in S. Lippman and J. McCall, eds., Studies in the Economics of Search, North-Holland:New York,171-90.

Kiefer, N. M., and G. R. Neumann, [1979b], "An empirical job search model with a test of the constant reservation wage hypothesis," Journal of Political Economy,87:89-107.

Kiefer, N. M., and G. R. Neumann, [1981], "Individual effects in a nonlinear model: Explicit treatment of heterogeneity in the empirical job search model," Econometrica, 49:965-79.

Kiefer, N. M., and G. R. Neumann, [1982], "Wages and the structure of unemployment rates," in M. Baily, ed., Workers, Jobs, and Inflation, The Brookings Institution: Washington, D.C.,325-57.

Kiefer, N. M., and G. R. Neumann, [1994], "Wage dispersion with homogeneity: The empirical equilibrium search model," in H. Bunzel, P. Jensen, and N. Westergård-Nielsen, eds., Panel Data and Labour Market Dynamics, Amsterdam:North-Holland, 57-74.

Koning, P., Ridder, G., and G. J. van den Berg, [1996], "Structural and frictional unemployment in an equilibrium search model with heterogeneous agents," Journal of Applied Econometrics, forthcoming.

Laird, N., [1978], "Nonparametric maximum likelihood estimation of a mixing distribution," Journal of the American Statistical Association, 73:805-11.

Lancaster, Tony, [1972], "A stochastic model for the duration of a strike," Journal of the Royal Statistical Society, 135,2:257-71.

Lancaster, T., [1979], "Econometric methods for the duration of unemployment," Econometrica, 47, 4: 939-56.

Lancaster, T., [1985], "Generalized residuals and heterogeneous duration models: With applications to the weibull model," Journal of Econometrics, 28,1:113-26.

Lancaster, T., [1990], The Econometric Analysis of Transition Data, Cambridge University Press.

Lindsay, B. G., [1983a], "The geometry of mixture likelihoods: A general theory," Annals of Statistics, 11, 1:86-94.

Lindsay, B. G., [1983b], "The geometry of mixture likelihoods, part II: The exponential family," Annals of Statistics, 11, 3:783-92.

Linz, P., [1985], Analytical and numerical techniques for Volterra Equations, Philadelphia: SIAM.

Marston, S. T. [1976], "Employment instability and high unemployment rates," Brookings Papers on Economic Activity, I, 169-210.

McCall, B. P., [1994], "Specification diagnostics for duration models: a martingale approach," Journal of Econometrics, 60:293-312.

McCall, J. J., [1970], "The economics of information and job search," Quarterly Journal of Economics, 84,1:113-26.

Meyer, B. D., [1990], "Unemployment Insurance and unemployment spells," Econometrica, 58,4:757-82.

Moreau, T., O'Quigley,J., and Mesbah, M., [1985], "A global goodness-of-fit statistic for the proportional hazards regression model," Applied Statistics, 34:212-18.

Mortensen, D. T., [1970], "Job search, the duration of unemployment, and the Phillips curve," American Economic Review, 60,5:505-17.

Mortensen, D.T., [1977], "Unemployment insurance and job search decisions," Industrial and Labor Relations Review, July, 30:505-17.

Mortensen, D. T., [1986], "Job Search and Labor Market Analysis," Chapter 15 in O. Ashenfelter and T. Layard, (eds.), Handbook of Labor Economics, North-Holland:Amsterdam, 849-919.

Mortensen, D. T., [1990], "Equilibrium wage distributions: A Synthesis," in J. Hartog, G. Ridder, and J. Theeuwes, eds., Panel Data and Labor Market Studies, Amsterdam:North-Holland, 279-296.

Mortensen, D. T., and G. R. Neumann, [1984], "Choice or Chance: A Structural Interpretation of Individual Labor Market Histories," in G.R. Neumann and N. Westergård-Nielsen, (eds.), **Studies in Labor Market Dynamics**, Springer-Verlag, New York and Heidelberg, 98-131.

Narendranathan, W., and S. Nickell, [1985], "Modelling the process of job search," Journal of Econometrics, 28:29-49.

Narendranathan, W., and M.W. Stewart, [1993], "How does the benefit effect vary as unemployment spells lengthen?," Journal of Applied Econometrics, 8:361-81.

Nickell, S., [1979], "Estimating the probability of leaving unemployment," Econometrica, 47,5:1249-66.

Newman, J. L., and C. E. McCullough, [1984], "A hazard rate approach to the timing of births," Econometrica, 52,14:939-61.

Orme, C., [1990], "The small sample performance of the information matrix test," Journal of Econometrics, 46, Dec: 309-31.

Pakes, A., [1986], "Patents as options: Some estimates of the value of holding European patent stocks," Econometrica, 54,4:755-84.

Powell, J. L., [1986a], "Censored regression quantiles," Journal of Econometrics, 32:143-55.

Powell, J. L., [1986b], "Symmetrically trimmed least squares estimation for tobit models," Econometrica, 54:1435-60.

Prentice, R. L., and L. A. Gloeckler, [1978], "Regression analysis of grouped survival data with applications to breast cancer data," Biometrics, 34:57-67.

Ramlau-Hansen, H., [1983], "Smoothing counting processes intensities by means of kernel functions," Annals of Statistics, 11:803-13.

Ramsey, J. B., and P. Schmidt, [1976], "Some further results on the use of OLS and BLUS residuals in specification error tests," Journal of the American Statistical Association, 71:389-90.

Reder, M., [1947], Studies in the Theory of Welfare Economics, New York: Columbia University Press.

Reinganum, J., [1979], "A simple model of equilibrium price dispersion," Journal of Political Economy, 87:851-58.

Ridder, G., [1984], "The distribution of single spell duration data," in G.R. Neumann and N. Westergaard-Nielsen, (eds.), Studies in Labor Market Dynamics, Heidelberg:Springer-Verlag.

Ridder, G. [1986], "The sensitivity of duration models to misspecified unobserved heterogeneity and duration dependence," University of Amsterdam mimeo.

Ridder, G., [1990], "The non-parametric identification of generalized accelerated failure-time models," Review of Economic Studies, 57:167-82.

Ridder, G., and W. Verbakel, [1983], "On the estimation of the proportional hazards model in the presence of heterogeneity," University of Amsterdam Actuarial Science and Econometrics mimeo.

Rust, J., [1987], "Optimal replacement of GMC bus engines: An empirical analysis of Harold Zurcher," Econometrica, 55,5:999-1035.

Schnell, J. and C. Gramm, [1987], "Learning by striking: Estimates of the teetotaler effect," Journal of Labor Economics, 5,2:221-41.

Schoenfeld, D., [1980], "Chi-squared goodness-of-fit tests for the proportional hazards regression model," Biometrika, 69:145-53.

Scott, D. W., [1992], Multivariate Density Estimation: Theory, Practice, and Visualization, New York: John Wiley.

Silverman, B. W., [1986], Density Estimation for Statistics and Data Analysis, London:Chapman & Hall.

Stigler, G. J., [1961], "The economics of information," Journal of Political Economy, 69:213-25.

Stigler, G. J., [1962], "Information in the labor market," Journal of Political Economy, 70: 94-105.

Sueyoshi, G. T., [1992], "Semiparametric proportional hazards estimation of competing risks models with time varying covariates," Journal of Econometrics, 51:25-58.

Sueyoshi, G. T., [1991], "Evaluating simple alternatives to the proportional hazard model: unemployment insurance receipt and the duration of unemployment," University of San Diego Working paper, February.

Therneau, T., Grambsh, P. M., and T.R. Fleming, [1990], "Martingale-based residuals for survival models," Biometrika, 77:147-60.

Toikka, R. S., [1976], "A markovian model of labor market decisions by workers," American Economic Review, 66: 821-34.

Tricomi, F. G., [1957], Integral Equations, Dover.

Tuma, N. B., and P. K. Robins, [1980], "A dynamic model of employment behavior: A application to Seattle and Denver income maintenance experiments," Econometrica, 48,4: 1031-52.

Van den Berg, G. J., and G. Ridder, [1993], "On the estimation of equilibrium search models from panel data," in J. Van Ours, ed., Labor Demand and Equilibrium Wage Formation, Amsterdam:North-Holland.

Van den Berg, G. J., and G. Ridder, [1994], "Estimating an equilibrium search model from wage data," in H. Bunzel, P. Jensen, and N. Westergård-Nielsen, eds., Panel Data and Labour Market Dynamics, Amsterdam:North-Holland, 43-55.

Watson, G. S., and Ledbetter, M. R., [1964], "Hazard Analysis I," Biometrika, 51:175-84.

Wei, L. J., [1984], "Testing goodness-of-fit for proportional hazards model with censored observations," Journal of the American Statistical Association, 79:649-52.

Wells, M.T., [1990], "On the estimation of hazard rates and their extrema from general randomly censored data," Cornell University CAE working paper 90-16.

White, H., [1982], "Maximum likelihood estimation of misspecified models," Econometrica, 50:1-26.

Wolpin, K. I., [1984], "An estimable dynamic stochastic model of fertility and child mortality," Journal of Political Economy, 92:852-74.

Wolpin, K. I., [1987], "Estimating a structural search model: The transition from school to work," Econometrica, 55,4:801-17.