Econometric Modelling of Duration Analysis:

Its Methodology and Problems

Masako KUROSAWA

1. Introduction

On many occasions, econometric analysis involves an investigation of the duration that a particular subject occupies a state under uncertainty. These states may represent a condition of the subject in question, or they may simply distinguish the period before or after the subject makes a particular decision. For example, economists may wish to know factors that influence spells of unemployment, workers' decisions to retire, the length of strikes, or the intervals between successive bargaining. The termination of a spell is called failure, and the duration of the spell until failure is called the failure time. The process that determines the failure time is comprised of a sequence of choices by the subject on whether to continue or to exit his or her current state. For example, the duration of unemployment is determined by a sequence of choices regarding whether the subject takes a job offer and becomes employed or continues searching for another job. Hence, by investigating the effects of potentially interesting factors on the probability that a choice will occur, we can study their influence on the duration of time spent in a certain state, that is, on the failure time.

The purpose of this paper is to review the basic concepts and problems associated with the economic applications of duration analysis, which have mostly been confined to the field of labor economics, and to present ways of alleviating such problems.

First, the paper discusses a basic concept of the hazard function—in particular, its relation with the underlying duration distribution. The simple proportional hazards model and how such a model is applied to a simple job-search model is also presented. This is followed by a discussion of parametric maximum likelihood estimation, which is a conventional estimation method for the proportional hazards model. Section 4 extends the basic model to incorporate random heterogeneity and time-varying explanatory variables. While the generalization of the model is needed to avoid omitted variables bias in the estimated parameters, they also create difficulties in determining the parametric forms for both the heterogeneity and conditional duration distributions. This is important because economic theory rarely provides any rationale for them, nor can the observed data uniquely determine the functional forms for such distributions. Yet, incorrectly assumed distributions lead not only to inefficiency but also to inconsistent estimates. In Section 5, I introduce four different ways to combat such inference problems by semi-parametric methods that free the analysis from making unnecessary distributional assumptions. The paper concludes that the adoption of semi-parametric hazard formulation is the best remedy in alleviating the inference problems. Determining which method best suits the particular problem, however, depends on the type of data available and the nature of the problem.

2. The Hazard

In this section, I introduce a definition of the hazard rate in a continuous time setting.[1] Unless otherwise indicated, it is assumed the start of a failure time is exogenous, so that all distributions are conditioned on such an event. This is appropriate since the same conditioning often applies to the nature of sampling in practice. When studying the duration of unemployment, for instance, samples are extracted from the pool of unemployed, which means that the sampling is conditioned on the event that these workers are already unemployed.

2.1 Hazard Definition

Let t be a time since the start of a spell, and δ be a random variable that represents the length of failure time (i.e., the length of a spell until it terminates). It is important to bear in mind that δ represents a period of time and not a point in time. Instead, t represents points in time, measured from the beginning of the state. Thus, for duration observations that have different starting points, the same value of t corresponds to different calendar dates.

A study of the behavior of duration leads to studying the distribution of a random variable, δ, which is characterized by the following cumulative distribution function, $F(\cdot)$:

$$Pr(\delta \leq \Delta) = F(\Delta | x; \theta) \tag{1}$$

where x is a vector of explanatory variables and θ is a vector of parameters that characterizes the distribution that requires estimation. We assume, for the moment, that x is time invariant. Complement to the above c.d.f., $1 - F(\Delta | x; \theta)$, is called the survivor function, $S(\Delta | x; \theta)$, which represents the probability that the duration of a spell lasts at least as long as Δ. Analogously, a density function for δ, $f(\cdot)$, can be obtained by differentiating $F(\cdot)$.

In practice, economic theory does not usually lead directly to the duration distributions, however. Instead, a distribution conditioned on the past, called the hazard function, is constructed first, from which the underlying probability distributions, $f(\cdot)$ or $F(\cdot)$, are derived. They are then inserted into the likelihood function to estimate a vector of unknown parameters, θ. Most economic models tend to specify the hazard rather than the distribution function itself since the problems of economic choice are often conditioned on the past, making the formulation of the hazard more intuitive. The hazard function measures the probability of exiting the current state at t, provided that the spell has lasted at least as long as t. It is related to the duration distributions as follows:

$$h(t,x;\theta)=\lim_{dt\to 0}Pr(\delta\in(t,t+dt)|\delta\geq t,x)=\lim_{dt\to 0}\frac{Pr(t\leq\delta\leq t+dt|t\leq\delta,x)}{dt}$$

$$=\lim_{dt\to 0}\frac{F(t+dt|x;\theta)-F(t|x;\theta)}{dt}\quad\frac{1}{1-F(t|x;\theta)}$$

$$=\frac{f(t|x;\theta)}{1-F(t|x;\theta)} \tag{2}$$

The hazard rate is also called the local exit rate, or the escape rate. Using the relation (2), it is also possible to specify the corresponding distribution functions in terms of $h(\cdot)$.

First, differentiate the log of survivor function with respect to t:

$$\frac{d\log S(t|x;\theta)}{dt}=\frac{d\log(1-F(t|x;\theta))}{dt}$$

$$=\frac{-f(t|x;\theta)}{1-F(t|x;\theta)}=-h(t,x;\theta) \tag{3}$$

Integrating both sides over the range, $(0, \Delta)$, for example, gives:

$$\log(1-F(\Delta|x;\theta))=-\int_0^{\Delta}h(t,x;\theta)dt \tag{4}$$

Given that $F(0|x)=0$, by definition, it follows that the cumulative distribution function of a duration is expressed in terms of $h(\cdot)$ as follows:

$$F(\Delta|x;\theta)=1-\exp[-\int_0^{\Delta}h(t,x;\theta)dt] \tag{5}$$

Given (5), the corresponding density function can be obtained.

$$f(\delta|x;\theta)=h(\delta,x;\theta)\exp[-\int_0^{\delta}h(t,x;\theta)dt]$$

$$=h(\delta,x;\theta)[1-F(\delta|x;\theta)] \tag{6}$$

Denoting the integrated hazard, $\int_0^{\Delta}h(t,x;\theta)dt$, as $\Lambda(\Delta,x;\theta)$, the survivor function can be written as:

$$S(\Delta|x;\theta)=\exp(-\Lambda(\Delta,x;\theta)) \tag{7}$$

Non-defectiveness is required for the duration to have a proper distribution so that the density integrates to 1. This is satisfied when $S(\infty)=0$, in other words, $\lim_{\Delta\to\infty}\int_0^{\Delta}h(t,x;\theta)dt=\infty$, which is called the admissibility condition. In practice, however, this condition is not necessarily met. Some people may stay unemployed forever. In such cases, defective duration distribution arises, since $F(\infty)<1$.

2.2 Forms of the Hazard

How the hazard is affected by the length of the elapsed spell is called a duration dependence, and its form depends crucially on the parametric specifications assumed for the hazard function, which in turn determines the underlying duration distribution. If the conditional probability to exit the spell does not change no matter how long the spell has endured, then there is no duration dependence (i.e., $\partial h(t)/\partial t=0$), and its process is called stationary. The hazard rate stays constant over time, and the corresponding duration distribution will be exponential. Usually, however, it is expected that the hazard will either increase or decrease with the length of the elapsed spell, and it is necessary to allow the hazard to have a specification that is flexible enough to accommodate either case.

In practice, the most frequently used formulation for the hazard is the proportional form:

$$h(t, x;\theta)=h_1(x;\beta)h_0(t;\alpha) \qquad (8)$$

where $h_1(x;\beta)$ depends only on the explanatory variables and not on time, while $h_0(t;\alpha)$ depends only on the elapsed duration. The latter is known as the baseline hazard, which is the value of the hazard when $h_1(x;\beta)=1$. β and α constitute the unknown parameter vector, θ. In this formulation, the duration dependence is determined entirely by $h_0(t;\alpha)$. Hence, the explanatory variables can influence the scale of the hazard but not the form of its dependence on time, which makes it easy to discriminate the effects of time and other explanatory variables on the hazard. The explanatory variables, x, are usually assumed to affect the hazard exponentially so that $h_1(x;\beta)=\exp(\beta' x)$, although linear polynomials or logistics can also be used. Exponential form is flexible and it naturally satisfies the basic requirement that the term $h_1(x;\beta)$ is always non-negative.

In order to see how the covariates, x, are determined in the economic applications of duration analysis, let us briefly consider a simplified model of a job search. Assume that a person who is unemployed can exit this state only through obtaining employment and that the process of becoming employed consists of two steps. First, one has to receive a job offer. Given a certain job offer, a worker will accept it only if the offered wage is at least as high as his or her reservation wage (i.e., the minimum level of wage at which a worker is willing to work). Suppose that this worker has been unemployed for a period τ, and a job offer for such a person arises at rate $l(\tau)$. Let the distribution of a wage offer be $F_\omega(w)$ and the reservation wage, w^r, be a function of the elapsed unemployment spell, $w^r(\tau)$. The variable w^r is derived by optimizing the expected present value of the worker's future income stream. Then, this instantaneous probability of getting a job is a product of the probabilities that an offer will arise and that the offered wage will be higher than the reservation wage.

Pr(accepting a job after τ periods of unemployment spell)

$=Pr(\text{offered wage} > w^r(\tau))\,l(\tau)$

$$=(1-F_\omega(w^r(\tau)))l(\tau) \tag{9}$$

In other words, this equation represents the probability of exiting a spell of unemployment after τ periods of unemployment, given that the worker has been unemployed for at least a period of τ. Hence, it clearly is analogous to the hazard rate.

The next step is to designate, with the help of economic theory, a set of potentially influential variables—in this case, any variables that are likely to determine the reservation wages, offer rates, or offer wages—as explanatory variables in the hazard function formulation. For instance, in the pioneering application of duration analysis on unemployment spells by Lancaster (1979), the time invariant covariates included age, an individual's unemployment history, and a replacement ratio (the ratio between income during unemployment to that of the last job) for unskilled unemployed workers. Even though the hazard rate in this example is made up of two separate components, it is practically impossible to identify the structural parameters without further information. Notable exception is the analysis of job search by Lancaster and Chesher (1983). There, they were able to make use of survey data on a pair of random variables associated with the elapsed unemployment spell: what workers expect to earn in a new job, and reservation wages. Using this information, they were able to deduce, rather than to estimate, the sample average of structural parameters (elasticities).[2]

Apart from its simplified theoretical basis, this model is heavily restricted in two important respects: there is a lack of both a random heterogeneity component and duration dependence. As examined also in Lancaster's 1979 paper, since observable variables can never completely explain all the heterogeneities across observations, a need for a random heterogeneity term arises if inconsistencies in the estimated parameters are to be avoided. Also, in the current specification, the hazard is independent of the length of the elapsed spell, exhibiting a stationary structure. In practice, there are many cases where non-stationarity may enter the model: the hazard may be affected purely by the passage of time or indirectly via the explanatory variables, which continuously vary throughout the spell. In the latter case, the hazard would depend on the entire time path of the explanatory variables since the start of the spell, rather than on its value at a particular point in time. In general, as in (8), it is customary to add the baseline hazard, $h_0(t;\alpha)$, multiplicatively to the specification (9), where its underlying distribution is chosen rather arbitrarily out of analytical convenience. Examples of duration distributions that underlie $h_0(t)$ are listed in the appendix (see Cox and Oakes (1984) or Kalbfleisch and Prentice (1980) for more details).

In Section 4 below, these restrictions are relaxed and the model is generalized.

Before embarking on the generalization, however, it is important to first discuss the parametric maximum likelihood estimation method, which is most commonly used to estimate the fully parameterized duration models.

3. Maximum Likelihood Estimation

Once the parametric form of the hazard is defined uniquely in terms of a vector of unknown parameters θ, the corresponding duration distribution can also be parametrically specified using the relations (5)–(6). In general, it is then a simple matter to form a likelihood function to estimate the parameters of interests. Below, a formation of the likelihood function is explained in the presence of various types of duration data.

Depending on the nature of the sampling, data may contain uncompleted spells in addition to completed spell observations. Recorded uncompleted spells may have an ambiguous starting point (i.e., left censored) or ambiguous ending (i.e., right censored). Left censoring occurs for spells that have already started at the time of sampling, making the length of a spell prior to the sampling date unknown. Right censoring, on the other hand, occurs when some observed spells have not yet terminated at the time of sampling.

In general, the likelihood function to be maximized is:

$$L=\prod_{i\in A} f(\delta_i|x_i;\theta)\prod_{j\in B}[1-F(\Delta_j|x_j;\theta)]\prod_{k\in C}\frac{f(\Delta_k+s_k|x_k;\theta)}{1-F(\Delta_k|x_k;\theta)}\prod_{l\in D}\frac{1-F(s_l+\Delta_l|x_l;\theta)}{1-F(\Delta_l|x_l;\theta)} \tag{10}$$

Set A and B contain observations whose likelihood functions are the unconditional probabilities of observed durations. In particular, set A contains observations on completed durations, δ_i, and set B contains duration observations that are right-censored. Hence, the only information available on the j-th observation in set B, for instance, is that a spell has lasted at least as long as Δ_j and that such a probability arises with $Pr(\delta>\Delta_j)=1-F(\Delta_j|x;\theta)$.

Set C and D both contain observations which are conditioned on the spell duration prior to the sampling date. Consider an interview survey conducted at two different points in time. If the spell is known to have started Δ_k prior to the time of the first survey, then this particular observation could never have a value less than Δ_k. Hence, the corresponding likelihood for this observation has to be conditioned on the event, $\delta>\Delta_k$, which occurs with probability, $1-\mathrm{F}(\Delta_k|x;\theta)$. Set C, in particular, contains observations on spell durations whose complete spell lengths are known. This applies to cases in which the termination of the event took place prior to the second survey. If the k-th spell has lasted for s_k since the first survey, its total failure time is $s_k+\Delta_k$. Hence, the k-th likelihood component becomes $f(\Delta_k+s_k|x_k;\theta)/(1-F(\Delta_k|x;\theta))$. Set D, on the other hand, contains observations which are censored from the right with a known duration prior to the sampling date. Conditioned on the event, $\delta>\Delta_l$, all one knows is that the spell has lasted at least for another s_l periods, making the l-th spell duration at least as long as $s_l+\Delta_l$. This is the case if the spell has not ended by the second interview date. The l-th likelihood component is $(1-F(s_l+\Delta_l|x;\theta))/(1-F(\Delta_l|x;\theta))$.

Note here that there are cases in which it is better to consider the duration under study as a discrete random variable, having values such as τ, $\tau+1$, $\tau+2$, and so on. For example, a duration distribution for a strike can be considered in units of days if a union's decision to continue a strike is made daily. Obviously, most common approach is to model such a process in discrete time. Alternatively, the data can be considered as incomplete observations out of a continuous process. For instance, if the observed duration is τ-days, we can assume that the spell has in fact lasted somewhere between τ and $\tau+1$ days. Hence, its corresponding likelihood component is the probability that the completed duration falls in the interval, $(\tau, \tau+1)$, which is $(F(\tau+1|x;\theta)-F(\tau|x;\theta))$.

Left-censored observations are associated with problems known as length-biased sampling if entry to the spell is not exogenous (Kiefer 1988). In the study of unemployment spells, for instance, the probability of being sampled out of a pool of unemployed workers is higher for people who stay unemployed longer. In this respect, the issue of length-biased sampling also applies to observations in set C and D in (10), although they will not cause sample selection bias because they are conditioned on the pre-interview durations. A problem arises when entry to a spell is endogenous and its timing is unknown (see Amemiya 1985, chapter 11, section 11.2.6, or Pudney 1989, chapter 6, section 6.4.2). Suppose that the spell has lasted for period d since the sampling date, where the probability to enter the unemployment state is time-invariant (constant entry rate). An individual with the vector of explanatory variables x has a duration density, $f(\delta|x;\theta)$. In order to compute the likelihood contribution of the partially observed duration with unknown starting point, it is necessary to derive a joint density that an individual with x is sampled with completed duration, δ, and that such an observation is sampled d periods prior to the end of the spell. Since this individual is at risk of being sampled only while the spell is in progress, the probability density of sampling this individual with a completed duration δ is $\delta f(\delta|x;\theta)/E(\delta|x;\theta)$.[3] The denominator is inserted in order to lend a proper density distribution that integrates to 1. Given that the completed duration is δ, the probability of observing the post-sampling duration of length $d\le\delta$ is $1/\delta$ since it is the same for any d over the range of $(0, \delta)$. From these two probabilities, joint density between δ and d is calculated to be $f(\delta|x;\theta)/E(\delta|x;\theta)$. Hence, the likelihood contribution of the partially observed duration is the marginal density of d, which is this joint density with δ integrated out over the range (d, ∞):

$$f(d|x;\theta)=(1-F(d|x;\theta))/E(\delta|x;\theta) \tag{11}$$

As can be seen, (11) involves an evaluation of the expectation term, which will be numerically complicated even for a distribution with a simple closed form. Moreover, the constant entry rate assumed to derive the relation (11) is too strict in practice. If possible, the best strategy may be to avoid using these types of data.

In a straightforward case when all observations are uncensored or right censored, the log of full likelihood can be written simply in terms of the hazard as follows:

$$\log L = \sum_{i \in A} \ln h(\delta_i, x_i; \theta) - [\sum_{i \in A} \Lambda(\delta_i, x_i; \theta) + \sum_{j \in B} \Lambda(\Delta_j, x_j; \theta)] \tag{12}$$

where $\Lambda(\Delta)$ is the integrated hazard over $[0, \Delta]$ interval, A is the set of uncensored observations, and B is the set of right censored observations.

Under the regularity conditions for $f(\cdot)$, $\hat{\theta}$ is consistent and $\sqrt{N}(\hat{\theta} - \theta)$ is distributed asymptotically normal with a mean of zero and the following variance:

$$-\left[E\left(\frac{1}{N}\frac{\partial^2 \log L}{\partial\theta\partial\theta'}\right)^{-1}\right] \tag{13}$$

which can be consistently estimated by:

$$-\left[\frac{1}{N}\frac{\partial^2 \log L}{\partial\theta\partial\theta'}\right]^{-1} \tag{14}$$

When the expression for the derivative is not in closed form, numerical maximization should be adopted to obtain the maximum likelihood estimates, in which case, care must be taken to ensure that the global maximum is attained.

As seen, the parametric maximum likelihood estimation can easily accommodate censored data. It can also incorporate time varying explanatory variables, as we shall see later, although it will involve numerical integration.

There are, however, fundamental problems associated with the parametric maximum likelihood estimation. Particular parametric forms have to be determined a priori for the hazard function to construct the likelihood function to be estimated. While economic theory rarely provides any justification on what distributional form the duration distribution should take, such an assumption is crucial to the estimation in that if it is mis-specified, the estimated parameters will be inconsistent. The introduction of a random heterogeneity term further complicates the problem, since this necessitates yet another parametric assumption for its density function in order to calculate the corresponding marginal distributions. It also raises an identification problem between the heterogeneity distribution and the duration distribution conditional on such heterogeneity. Moreover, a possibility of correlation between the explanatory variables and the heterogeneity term cannot be easily handled.

In the following section, the generalization of the model—in particular, an introduction of the random heterogeneity and time varying covariates—are discussed. Two interrelated problems, inference and identification, inherent in the generalized model and some novel approaches that alleviate such problems will be discussed in Section 5.

4. Generalization of the Model

4.1 Time Varying Covariates

In practice, observable explanatory variables often vary over time during the course of a spell. If so, duration dependence is not only generated directly by t via $h_0(t:\alpha)$, but also by the time path of the explanatory variables during the spell. For instance, if a worker's reservation wages decline as unemployment spells lengthen because unemployment benefits cease to be provided after several months, then the likelihood of exiting from the unemployment state would be expected to increase with the spell, causing the hazard to increase with time. The hazard is then written as $h(t, x(t);\theta)$. The duration distribution now depends not on the value of x at certain point in time but on the entire form of the time path, $\{x(t)\}$, since the beginning of the spell.[4] It incorporates a possibility for the value of $x(t)$ after the start of a spell to influence the hazard. For instance, the studies of Kennan (1985) and Harrison and Stewart (1987) both examined the impact of business cycles on strike frequencies and strike durations. Allowing measures of business cycles to vary during the spell, they were able to estimate the effect of a continuously changing economic environment on the conditional probability of ending the strike.

In order to estimate the impact of continuously varying $x(t)$ on the duration distribution, it is necessary to have continuous observations so that the integral terms within the likelihood function ((15)) can be computed. Since observations are generally available at discrete points, an arbitrary time path for the variable has to be assumed based on the given observation points.

$$F(\Delta|\{x(t)\};\theta)=1-\exp[-\int_0^{\Delta} h(t, x(t);\theta)dt] \tag{15}$$

Another difficulty associated with the introduction of time varying covariates arises from identification issues. This is best illustrated in the proportional hazards formulation, $h(t, x(t);\theta)=h_0(t;\alpha)h_1(x(t);\beta)$. The function h_1 (effect of time varying variables) is only identified from h_0 (baseline hazard) if $x(t)$ varies substantially across observations or if strong functional form is assumed for h_0, so that $ln(h_0)$ and $ln(h_1)$ are linearly independent. Otherwise, there would be multicollinearity.

4.2 Random Heterogeneity

Thus far, it has been assumed that all the heterogeneities across individual observations are explained by the observed explanatory variables. In general, however, unless it is known a priori that individuals are homogeneous, it is necessary to take into account the population variation by unobservable random factors because of incomplete control by the explanatory variables or mis-specification of the functional form. We represent such random heterogeneity factors by a disturbance term, v, and call its distribution function, f_v, a mixing distribution defined over the range R_v. The hazard and the underlying duration distribution will then be conditioned on v, so that they are now written as $h(t, x|v;\theta)$ and $f(\delta|x, v;\theta)$ respectively. In order to obtain the unconditional distribution of duration, we need to integrate out for v over the range R_v:

$$F(\Delta|x;\theta)=\int_{R_v} F(\Delta|x, v;\theta)f_v(v)dv$$

$$=1-\int_{R_v}\exp\left[-\int_0^{\Delta}h(t,x|v;\theta)dt\right]f_v(v)dv \tag{16}$$

$$f(\delta|x;\theta)=\int_{R_v}h(\delta,x|v;\theta)\left[1-F(\delta|x,v;\theta)\right]f_v(v)dv \tag{17}$$

As before, the unconditional hazard is given in terms of $F(\Delta|x;\theta)$ and $f(\delta|x;\theta)$.

How do we know which distribution is suitable for the mixing distribution? In practice, a pair of distributions, $f(\delta|x,v;\theta)$ and $f_v(v)$, are selected for analytical convenience, such that they generate a simple closed form for the unconditional duration distribution, $f(\delta|x,v;\theta)$. This is often called a reduced form method.

Lancaster (1979) assumed a Weibull hazard function and a gamma mixing distribution in a proportional hazards model:

$$h(t,x_i|v_i;\theta)=v_i\exp(x_i'\beta)\alpha t^{\alpha-1} \tag{18}$$

This is a common specification which allows for a monotonic duration dependence, depending on whether the value of α is greater or less than 1. Flinn and Heckman (1982) further generalized the conditional hazard specification to allow the heterogeneity term to be correlated across spells. Their hazard was flexible enough to accommodate various distributional forms and time varying covariates. Their hazard is called the Box-Cox conditional hazard, since it involves a Box-Cox transformation:

$$h(t,x_i(t)|v_i;\theta)=\exp\left[\beta' x_i(t)+cv_i+\mu_1\frac{t^{\gamma_1}-1}{\gamma_1}+\mu_2\frac{t^{\gamma_2}-1}{\gamma_2}\right] \tag{19}$$

The last two terms approach $\log(t)$ as $\gamma_1\to 0$ or $\gamma_2\to 0$, respectively. This model converges to a Weibull duration distribution if $\gamma_1=0$ and $\mu_2=0$, to a Gompertz if $\gamma_1=1$ and $\mu_2=0$, and finally, to an exponential if $\mu_1=\mu_2=0$. Let us also briefly review a commonly used combination for discrete duration data. In Kennan's (1985) study on strike durations, he used a beta mixing distribution and logistic continuation probability distribution (1 minus the discrete analogue of the hazard rate) for the following three reasons: (1) the resulting unconditional density has a closed form, (2) a parameter in the mixing distribution indicates variance of the heterogeneity, and (3) the shape and location of the hazard can vary flexibly with the explanatory variables (it can be uni-modal, U-shaped, J-shaped or uniform). Harrison and Stewart (1987), also in their study of strike durations, adopted the discrete proportional hazards model with a multiplicative heterogeneity term. They assumed the exponential of v to be distributed as gamma and $\log(-\log(1-\lambda_t))$, where λ_t is a discrete time analogue of the baseline hazard at spell duration t, to be represented by a third-order polynomial in t to allow for a flexibility in the form of duration dependence.

What happens if we ignore such random heterogeneity altogether? Lancaster (1979) discovered that this will bias the estimate of duration dependence downwards. His maximum likelihood estimate of duration dependence parameter increased as he included additional explanatory variables that were significant in the model. It increased even further upon incorporating a gamma mixing distribution. Without the generalization of the model by introducing a random heterogeneity, short spell observations would contribute to raise the hazard above its true value, and long spell observations would lower the hazard below its true level, leading to a spurious negative duration dependence. Consider, for instance, a study of a strike duration. Suppose v_i measures a degree of militancy of the i-th union. Then, it is natural to assume that the higher the v, the longer the strike is likely to last (i.e., the lower the hazard). If we fail to take v into account, the estimated hazard will be higher than the true value for short strikes due to the presence of groups with a low level of v, and will be lower than the true value of the hazard for longer strikes due to a high proportion of remaining groups with a high level of v. On the whole, the effect of omitting v alone would make it look as though the hazard declines with the length of the strike.

Hence, the entire purpose of introducing random heterogeneity is to generalize the model in order to avoid biases in the estimated effect of variables of interest, duration dependence in particular. However, values of such estimates vary significantly depending on the mixing distribution assumed for a given specification of the hazard (Heckman and Singer 1984), even though estimates are typically more sensitive to the specification of the duration distribution than to that of the mixing distribution. This makes, the correctness of the assumed functional forms for both the hazard and the mixing distribution crucial. The problem arises because economic theory rarely gives any guidance concerning the nature of their functional forms.

Is the making of such ad hoc functional forms really necessary for identification? In fact, given only the data on durations, there exists more than one pair of specifications for $f(\delta|x, v;\theta)$ and $f_v(v)$ that yield the same unconditional duration distribution, $F(\delta|x;\theta)$. This is why the methods outlined above have specified the functional forms for both $f(\delta|x, v;\theta)$ and $f_v(v)$ in controlling for random heterogeneity. Only in some cases where $f(\delta|x, v;\theta)$ is known and certain conditions are satisfied, is it possible to uniquely solve for $f_v(v)$ for a given $f(\delta|x, v;\theta)$. In such cases, the functional forms for both $f(\delta|x, v;\theta)$ and $f_v(v)$ do not have to be assumed, and hence, can avoid the risk of producing inaccurate estimates of the unknown parameters. Specifically, Elbers and Ridder (1982) have shown that if the hazard was known to be of the proportional form with a multiplicative random heterogeneity term:

$$h(t|x, v;\theta)=h_1(x;\beta)h_0(t;\alpha)v \tag{20}$$

with at least one exogenous variable taking values along the real line and $E_v(v)<\infty$, then, both the conditional hazard and the mixing distribution will be identified. The identification condition given by Heckman and Singer (1984), on the other hand, allows $E_v(v)=\infty$, which permits a wider range of distributions to be candidates for $f_v(v)$. But it has a heavier restriction on the form of the baseline hazard. It requires the existence of a known constant, c, where the integrated hazard, $\Lambda_0(\Delta')=c$ for a certain Δ' for all admissible Λ_0. Hence, for most widely used parametric hazard models (a class of Box-Cox hazard or a non-monotonic log logistic model), identification can be achieved without any regressors by specifying a functional form of the hazard up to a finite number of parameters and placing some restrictions on the moments of the admissible mixing distribution. In particular, provided that there is a parameterization for $h(t|x,v;\theta)$, non-parametric identification of $f_v(v)$ is achieved without any regressors.

This directs us to consider a totally non-parametric method or a semi-parametric method that removes part of the necessity for specifying functional forms as ways to alleviate such problems.

5. Ways to Avoid Inference Problems: Semi-parametric Methods

The previous section discusses the identification problem in which the mixing distribution and conditional duration distribution are not uniquely determined by the knowledge of duration observation alone. At the same time, mis-specifications of distributional forms not only lead to inefficiency but also to inconsistent estimates. By estimating at least some component of the hazard non-parametrically, it is less necessary to remain dependent on the correctness of the underlying distributional assumptions that are too often arbitrary.

This section describes several methods of estimation and formulations of the hazard function that attempt to alleviate the inference and identification problems inherent in the parametric maximum likelihood estimation by not assuming parametric forms for both $f(\delta|x,v;\theta)$ and $f_v(v)$.[5] In particular, the partial likelihood and regression methods are robust in terms of functional assumptions for the baseline hazard and mixing distribution, respectively. A semi-parametric maximum likelihood method estimates either the baseline hazard or $f_v(v)$ parametrically, while the rest is estimated non-parametrically. A formulation of the hazard entirely by non-parametric means is called an actuarial approach and is introduced last.

5.1 Partial Likelihood

Under the proportional hazards formulation ((8)), it is possible to draw an inference about the effects of explanatory variables without any knowledge of the functional form of the baseline hazard, $h_0(t;\alpha)$. In this method, likelihood is constructed by utilizing ranks rather than values of the observed durations so that the parametric form does not have to be assumed for the baseline hazard.

First, order all observations of spell duration such that $\delta_1<\delta_2<\delta_3<\ldots\delta_N$. If the j-th spell is of length δ_i, it implies that the j-th spell lasted for δ_i, which was the i-th shortest amongst all the duration observations. The j-th observation has a vector of explanatory variables x_j. Then, the conditional probability that the j-th spell is of length δ_i is:

$$h(\delta_i,x_j;\theta)/\sum_{k\in R_i}h(\delta_i,x_k;\theta) \tag{21}$$

where $h(\delta_i,x_j;\theta)=h_0(\delta_i;\alpha)h_1(x_j;\beta)$. R_i represents the risk set prior to δ_i, which includes observations on durations that are longer than or equal to δ_i but not yet terminated at the spell duration δ_i. (21) represents a probability that the j-th spell ends at δ_i, given that one of the spells in the risk set R_i ends at δ_i. This probability is conditional on the entire history of failures and censoring prior to the elapsed duration δ_i. The baseline hazard $h_0(\cdot)$ cancels out because of the multiplicative form assumed by the proportional specification. Hence, a contribution to the likelihood of the j-th duration observation that is the i-th shortest is, $h_1(x_j;\beta)/\sum_{k\in R_i} h_1(x_k;\beta)$. The joint distribution of the entire set of observed failure times, δ_1, δ_2, $\ldots\delta_N$, can be obtained using the chain rules. Hence, the log of partial likelihood to be maximized is:

$$\log L=\sum_{j=1}^{N}[\ln h_1(x_j;\beta)-\ln(\sum_{k\in R_i} h_1(x_k;\beta))] \tag{22}$$

The intuition behind the fact that only the rank information is sufficient to make an inference about β can be seen by considering any one-to-one strictly increasing transformation. Under such a transformation, $h_0(t;\alpha)$ will not be identified, but the order of the durations will remain unchanged and can be used to identify β.

Censored observations can easily be incorporated into this partial likelihood framework. However, if an observation is censored between duration δ_i and δ_{i+1}, its contribution to the likelihood will be confined to the risk set. Ties (when more than one duration observation with the same spell length exist) will reduce the efficiency of the estimator, but can still be dealt with in a similar manner. Likelihood contributions of tie observations will have common denominators with different numerators associated with a different set of explanatory variables. This method of estimation is also suitable for the hazard that involves time varying explanatory variables. Conditional probability that the j-th spell terminated at δ_i (again, the i-th shortest of all observations) becomes $h_1(x_j(\delta_i);\beta)/\sum_{k\in R_i} h_1(x_k(\delta_i);\beta)$, where $x_k(\delta_i)$ denotes the value of explanatory variables facing the k-th spell at duration δ_i. When the covariates do not depend on time, the product of conditional probability can be interpreted as the "marginal" likelihood of ranks. With continuously varying variables, the likelihood of marginal or conditional probability is no longer inferable, but corresponds to the full likelihood without the term that reflects information on the gaps between the successive failure times. Cox and Oaks (1984) show that the estimator derived from this method will be consistent and asymptotically normally distributed. The information matrix of the full and partial maximum likelihood estimator converge to var $(x(t))$ and $E(\mathrm{var}(x(t)|t))$, respectively, for groups defined by the failure time, t. Then, the relation:

$$\mathrm{Var}(x(t))=E\{\mathrm{Var}(x(t)|t)\}+\mathrm{Var}\{E(x(t)|t)\} \tag{23}$$

shows that the asymptotic efficiency of the partial likelihood relative to the full likelihood will be high if the ratio of the between-spells component of var$(x(t))$ (i.e.,Var $\{E(x(t)|t)\}$) to the within-spells component (i.e.,$E\{\mathrm{var}(x(t)|t)\}$) is small. It depends on how useful the information provided by the gaps between the successive failures would be in determining the coefficients on x. This condition applies unless the coefficients on x are far from zero, unless censoring depends strongly on $x(t)$, or unless there are strong time trends in x. As discussed in Section 4.1, this implies that the parameters of explanatory variables will not be identified if they are closely correlated with the elapsed duration. Also, a loss in precision is greater for the partial likelihood for a small sample.

A multiplicative random heterogeneity term can also be incorporated, although it will involve multiple integrations of order R_i, which is a dimension of the risk set; however, such a computation will be tremendously messy. Lastly, if one is interested in the duration dependence rather than the effect of time varying covariates, this estimation method will not be useful since parameters of the baseline hazard cannot be estimated.

5.2 Regression Method

An alternative approach to estimating the proportional hazards model is to transform it into a regression format with the log of integrated baseline hazard as a dependent variable. In general, for the hazard function, $h(t, x; \theta) = h_0(t;\alpha)h_1(x;\beta)$, the underlying duration density is:

$$f(\delta | x;\theta) = h(\delta, x;\theta)\exp(-\Lambda) \tag{24}$$

where Λ denotes the integrated hazard, $\int_0^\delta h(t, x;\theta)dt$. The transformation of a variable from δ to Λ will give the density of Λ as $f_\Lambda(i|x) = \exp(-i)$, since the Jacobian of transformation is $|d\delta/d\Lambda| = h(\cdot)$. Hence, Λ has a unit exponential distribution. Further transformation from Λ to $\varepsilon = -\log\Lambda$ would give:

$$f_\varepsilon(\varepsilon|x) = e^{-\varepsilon}\exp(-e^{-\varepsilon}) = \exp(-\varepsilon - e^{-\varepsilon}) \tag{25}$$

This transformed variable, ε, is the log of an exponentially distributed random variable, and it has a type I extreme value distribution. For the proportional hazards model with a multiplicative random heterogeneity, $h(t, x;\theta) = \exp(\beta' x)h_0(t;\alpha)v$, an expression for $-\log\Lambda$ is as follows:

$$-\log\Lambda = \beta' x + \log v + \log\int_0^\delta h_0(t;\alpha)dt \tag{26}$$

Denoting the last term as $\ln\Lambda_0(\delta)$ and replacing $-\log\Lambda$ by ε, this can be rewritten as:

$$-\ln\Lambda_0(\delta_i) = \beta' x_i + \log v_i + \varepsilon_i \tag{27}$$

for the i-th observation. This is a regression equation with a fixed effect, $\log v_i$, conditional on v_i. Unconditionally, it has a composite error term $\varepsilon_i + \log v_i$. The variable ε_i is distributed as a type I extreme value, and hence has a constant mean $\Psi(1)$ and variance $\Psi'(1)$ where $\Psi(\cdot)$ is a digamma function. Its mean can be absorbed into the constant term, and its variance can be compared with the variance of the least square residuals to test the existence of omitted regressors. As long as x is not correlated with the composite error term and all the duration observations are complete, the regression of $-\ln \Lambda_0(\delta_i)$ on x_i would provide consistent estimates of β and $\mathrm{var}(v_i)$ without assuming any functional forms for the distribution of v.

A special case in which the dependent variable simply reduces to a log of duration is the case of a constant hazard, $h(t, x;\alpha, \beta)=\alpha e^{\beta' x}v$. Since $\ln \Lambda_0(\delta_i)=\ln \alpha+\ln(\delta_i)$,

$$\log \delta_i = -\beta' x_i - \log \alpha - \log v_i - \varepsilon_i \tag{28}$$

On the other hand, under Weibull distribution, $h(t, x;\alpha, \beta)=\tau \alpha t^{\alpha-1} e^{\beta' x}$:

$$\log \delta_i = a_0 + a_1' x_i + \varepsilon_i \tag{29}$$

where $a_0=-\log \tau/\alpha$ and $a_1=-\beta/\alpha$ and ε_i again has a type I extreme value distribution. Note, however, that it is not possible to identify β from α under the Weibull baseline hazard formulation.

The advantage of this least square method is that it is simple to compute and is distribution free, since no assumption is required for the distribution of the error term (i.e., mixing distribution). Moreover, the correlation between the error term and the explanatory variables can be easily handled by an instrumental variables technique. In addition to the examples above, log normal or log logistic distribution for the baseline hazard with time invariant explanatory variables can also be transformed into a simple regression of the log duration. For instance, in a study of unemployment spells by Stephen Jones (1988), he transformed a proportional hazards model with no duration dependence into a log-linear function in the elapsed duration. Even though his data consisted mainly of censored unemployment spells, reservation wages at the time of each censoring were available, enabling him to study the relation between the elapsed unemployment spell and reservation wages. He assumed an offer probability to be distributed as Pareto, from which, he derived the expected value of the elapsed unemployment spell to be a log linear function of reservation wages and a random error term. A possible endogeneity of the reservation wage was easily dealt with by instrumenting. The instruments were chosen so that they affect the reservation wage but not the exit rate (i.e., search cost) such as the level of benefits. They found that the results from this IV estimation were more significant and also economically sensible than the simple OLS estimates.

This regression method can also incorporate explanatory variables that are time-varying, although the dependent variable becomes a non-linear function of durations and parameters of the baseline hazard, α. For instance, Kurosawa and Pudney (1993) examined the impact of time varying covariates—in particular, incomes policy dummies—on contract duration. It was important to allow the impact of incomes policies not only at the time of the previous bargain, but also at any time after the bargaining is struck, so that any new policy that came into effect after the last bargain may exert some impact on the probability of striking another bargain. In this way, the model enabled them to distinguish the two separate effects of incomes policies on wage changes: delay and moderation. Their hazard specification involved time-varying explanatory variables, time invariant observables unique to each group, and the fixed effect:

$$h(t|c_n, z_{ns}, x_n(t), u_n, v_{ns};\theta)=\exp[\beta' x_n(t)+\gamma_0+\beta_1' c_n+\gamma_2' z_{ns}+u_n+v_{ns}]h_0(t;\alpha) \qquad (30)$$

where suffix n indicates groups, and s indicates spells. Their data contained more than one observation by the same group (multiple spell data), hence some variables were specific to spells as well as to groups (z_{ns}), while others were only group-specific (c_n). Using the same transformation process as in (27) and assuming a constant baseline hazard, this hazard formulation was transformed to the following regression format.

$$\iota_{ns}=\gamma_0^* \ +\gamma_1' c_n+\gamma_2' z_{ns}+u_n+\varepsilon_{ns} \qquad (31)$$

where $\gamma_0^*=\gamma_0+\Psi(1)$ and where the dependent variable involved a numerical integration:

$$\iota_{ns}=-\ln[\int_0^{\delta_{ns}}\exp[\beta' x_n(t)]h_0(t;\alpha)dt] \qquad (32)$$

for each ns-th observation. Their model was estimated by the General Method of Moment which easily handled the possibility of correlation between the covariates and the error term by instrumenting.

There are a few drawbacks associated with this method, however. First, there is no simple way of handling the severely censored data. A method used to build the Tobit model can be applied based on the extreme value distribution of ε, though the resulting non-linear computation of the likelihood will no longer be simple. In this respect, the regression method is particularly suitable in a case such as Kurosawa and Pudney's, where the study is concerned not only with the process of failure times but also with another random variable called mark—which, in their study, was wages—that is observed as a result of a failure, since marks are necessarily associated only with completed spells. Also, as can be seen from the example of Weibull distribution, it is not possible to identify β and α since a constant term absorbs α. In the presence of time varying covariates, their effects—in particular, those of trending variables—would be difficult to identify from α, unless there are sufficient variations in the covariates across individuals.

Lastly, because of the explosive nature of a logarithm near 0, it is not suitable for the data if they contain observations on short durations which are contaminated with measurement errors, or if the study is aimed particularly at short durations. Still, for its great simplicity, the least square regression estimator is useful as a tool for preliminary analysis and a specification test.

5.3 Semi-parametric Maximum Likelihood

The semi-parametric maximum likelihood approach can generally take two forms: one is to assume a specific form for $f(\delta|x, v;\theta)$ and to estimate $f_v(v)$ non-parametrically, while another is to assume a specific form for $f_v(v)$ and estimate the baseline hazard non-parametrically. Either way, it is possible to derive the estimates of the interesting parameters by imposing fewer arbitrary specification assumptions, thereby avoid bearing the unnecessary risk of estimating inaccurately.

First, consider the case of an unknown mixing distribution. Heckman and Singer (1984) have proven that it is possible to use observed duration data to consistently estimate both $f_v(v)$ and the parameters in the conditional duration distribution, provided that we know the specification of $f(\delta|v, x;\theta)$. In particular, for the proportional hazards model in the presence of censoring and time-varying covariates, the non-parametric maximum likelihood estimation (NPMLE) yields consistent estimates when $h(t, x;\theta)$ is specified up to a finite number of parameters and $f_v(v)$ has a certain behavior in its tail distribution. Their proof verified the conditions stated in Kiefer and Wolfowitz (1956) that ensure the existence of a consistent estimator for both the mixing distribution and θ.

Their maximum likelihood estimator for the vector of parameters θ and the mixing distribution $f_v(v)$ are derived by solving the following problem:

$$\max_{f_v \in V, \theta \in \Theta} \sum_{i=1}^{N} \log \left(\int_{R_v} f(\delta_i | x_i, v;\theta) df_v(v) \right) \tag{33}$$

where Θ is the parameter space for θ, where $V=\{f_v : f_v(v) \geq 0$ is non-decreasing and right continuous, where $\int_v df_v(v)=1\}$, and where N is the total number of observations. They then showed that for a fixed θ, the non-parametric maximum likelihood estimator of $f_v(v)$ is a step function with, at most, N^* discrete points, where N^* is smaller than the total number of distinct values of (δ_i, x_i). Hence, the problem reduces to:

$$\theta, N^*, \max_{p_1, p_2, .. p_{N^*}, v_1, v_2, .. v_{N^*}}, \sum_{i=1}^{N} \log \left[\sum_{j=1}^{N^*} f(\delta_i | x_i, v_j;\theta) p_j \right] \tag{34}$$

subject to $\sum_{j=1}^{N^*} p_j = 1$ and $p_j \geq 0$ for all j. In the case of a general proportional hazards model with a multiplicative heterogeneity, $h(t|x, v;\theta) = h_1(x;\beta) h_0(t;\alpha) v$, the transformation of a variable from t to the integrated hazard, Λ, (i.e., $\Lambda_i = \int_0^{\delta_i} h_1(x;\beta) h_0(t;\alpha) dt$) simplifies the likelihood. The density function of Λ, conditional on v is exponential. Hence, $f(\Lambda|v) = v \exp(-\Lambda v)$. The optimization problem can be rewritten accordingly:

$$\theta, N^*, \max_{p_1, p_2, .. p_{N^*}, v_1, v_2, .. v_{N^*}}, \sum_{i=1}^{N} \log \left[\sum_{j=1}^{N^*} v_j \exp(-\Lambda_i v_j) p_j \right] \tag{35}$$

subject to $\sum_{j=1}^{N^*} p_j = 1$ and $p_j \geq 0$ for all j.[6] Even though asymptotic standard error cannot be computed from the Hessian of the likelihood since the dimension of the parameter space, N^*, varies with N, the resulting estimates are consistent as long as N^* is allowed to increase appropriately with N. Heckman and Singer suggested the use of an EM algorithm (Dempster, Laird, and Rubin 1977) to achieve convergence to a stationary point. This iterative process starts with estimating $f_v(v)$, using the above NPMLE for a given θ that determines $h_1(x;\beta)$ and $h_0(t;\alpha)$. Given a set of estimated parameters for $f_v(v)$, θ is estimated by the parametric maximum likelihood, which would yield a new value for Λ. This is inserted back into the above optimization problem to yield a new set of estimates for $f_v(v)$. The process is iterated in this way until convergence is achieved. Although only the global maximum ensures consistency of the estimator, this type of likelihood is plagued with multiple local maxima. This calls for experimentation with several alternative starting points, which makes this method computationally demanding. Monte Carlo experiments conducted in Heckman and Singer's other paper (Heckman and Singer 1982) showed that the NPMLE succeeded in estimating θ that predicts the sample well, although the mixing distribution was imprecisely estimated. Also, the NPMLE could not estimate more than four points of increase for $f_v(v)$. While mis-specification of the parametric $f_v(v)$ significantly biased the estimates of θ, allowing for a very flexible parametric form (with more than 2 parameters) for $f_v(v)$ seems to guard the estimates of θ from becoming heavily biased. In this sense, the NPMLE can be used as a test in determining the plausibility of the parametric MLE, although there are no formal test statistics to conclude its plausibility.

In Heckman and Singer (1984), they only estimated the random heterogeneity term non-parametrically, while the baseline hazard remained heavily parameterized. In light of the fact that the parametric form of the baseline hazard is often difficult to determine except for its overall shape, it would be convenient if we could also estimate it non-parametrically—in terms of a discrete step function. Ideally, a separate parameter should be assigned for every distinct duration observation in the sample. But in practice, a number of steps have to be consulted for the sake of efficiency. Meyer (1986) discusses efficiency comparisons between the semi-parametric hazard and the parametric maximum likelihood estimates. He states that in a situation where the explanatory variables differ more across observations than over time, or when the explanatory variables include a time trend, the semi-parametric method not only loses little efficiency but also ensures consistency. He found that the resulting estimates of the covariates' parameters are very similar to those obtained by Cox's partial likelihood.

Furthermore, Han and Hausman (1986) have found that, given a certain condition, both the baseline hazard and the mixing distribution can be identified from the observed data, and are hence estimable non-parametrically. They proved that the estimates of the covariates' parameters are asymptotically normal if at least one pre-determined variable is partly continuous. However, as Meyer (1990) found in his study on unemployment spells, the estimation involving a discrete distribution with many points are difficult in practice, particularly with respect to the mixing distribution.

Meyer tried to estimate the effects of time-varying covariates—in particular, that of unemployment benefits—while estimating both the baseline hazard and random heterogeneity non-parametrically. The data contained both completed and right censored observations. Since they are all recorded in term of weeks, observations could be regarded as being sampled out of a discrete time duration distribution. Instead, he regarded them as incomplete observations from continuous time distribution, so that the likelihood component of a spell that was reported to have lasted for δ_j weeks is $Pr(\delta_j \leq \text{spell} < \delta_j + 1)$:

$$\{1 - Pr(\text{spell} \geq \delta_j + 1 | \text{spell} \geq \delta_j)\} Pr(\text{spell} \geq \delta_j) \qquad (36)$$

where the latter probability can be written as a product of a series of conditional distributions:

$$\prod_{k=0}^{\delta_i \;\; 1} Pr(\text{spell} \geq k + 1 | \text{spell} \geq k) \qquad (37)$$

Even though weekly observations on the covariates are required for each unemployment spell from its start to its end, by representing the entire likelihood in terms of conditional probabilities of the form (36), and assuming all the time varying covariates to be constant within any unit interval, numerical integration was avoided altogether.

Different parameter values were then assigned to every distinct value of duration observations in a sample and then estimated together with the parameters of the explanatory variables. Theoretically, joint estimation of the mixing distribution is also possible non-parametrically. In practice, however, Meyer found it computationally very difficult, and resorted instead to estimation with a gamma mixing distribution and also with no mixing distribution. He found that the estimates of the parameters were hardly different for either formulation of the mixing distribution. In addition, the Weibull baseline hazard was found to yield significantly different parameter estimates from those based on the non-parametric baseline hazard. In sum, he concluded that non-parametric estimation of the baseline hazard avoids inconsistent estimation of β due to mis-specified baseline hazard, and may even make the mixing distribution irrelevant. Non-parametric estimation of the baseline hazard may sacrifice efficiency, but it assures consistency. Note, however, that his well-behaved baseline hazard may have been due to having as many parameters as distinct duration observations. This was only possible because the number of distinct spell observations in his data was not too large.

5.4 Empirical Hazard

The empirical hazard rate, or the empirical survivor function, are non-parametric estimates constructed from the data on elapsed and completed duration alone, and are called actuarial estimators. Because this method cannot estimate the impact of interesting covariates on the hazard, it is mainly used as a preliminary step to surmise the potential functional forms before embarking on the parametric or semi-parametric method.

Suppose that there is a set of N duration observations, among which $k(\leq N)$ observations are the completed durations. We can order observations on distinct completed durations so that $\tau_1 < \tau_2 < \tau_3 < \dots \tau_k$. Then, the number of spells neither completed nor censored before τ_j becomes:

$$n_j = \sum_{i=j}^{k} (m_i + g_i) \tag{38}$$

where m_i is the number of observations censored between τ_i and τ_{i+1}, and g_i is the number of spells completed at duration τ_i. In other words, n_j is the number of observations in a risk set at τ_j whose duration has lasted at least as long as τ_j. The corresponding hazard rate at τ_j is the probability of completing a spell at τ_j, conditioned on the event that a spell is at least as long as τ_j. There are n_j observations altogether with a duration at least as long as τ_j, out of which g_j observations have actually completed the spell at duration τ_j. Hence, the hazard rate can be estimated by: $\hat{h}(\tau_j) = g_j / n_j$. Also, applying the relation between a discrete time analogue of the hazard and duration distributions, $h(\tau_j) = f(\tau_j)/S(\tau_j)$, recursively, we obtain:

$$S(\tau_j) = \prod_{i=1}^{j-1} (1 - h(\tau_i)), \quad S(\tau_j + 0) = \prod_{i=1}^{j} (1 - h(\tau_i)) \tag{39}$$

and a probability of failure at τ_j is $S(\tau_j) - S(\tau_j + 0)$, where $S(\tau_j + 0) = \lim_{dt \to 0} S(\tau_j + dt)$.[7] Given (39), the corresponding estimator for the survivor function is:

$$\hat{S}(\tau_j) = \prod_{i=1}^{j-1} \frac{n_i - g_j}{n_j} \tag{40}$$

This is called the Kaplan-Meier, or product limit, estimator. The estimator for the integrated hazard is simply:

$$\hat{\Lambda}(\tau_j) = \sum_{i<j} \hat{h}(\tau_i) \tag{41}$$

If the number of distinct failure times $\tau_1, \tau_2, \dots \tau_k$ are fixed, and if the number of failures at each $\tau_j \, (j = 1..k)$ increases as the total sample size increases, then the standard asymptotic theory of inference on the maximum likelihood estimator can be applied.[8] An asymptotic variance estimator of log $(\hat{S}(\tau))$ is:

$$\begin{aligned} \mathrm{Var}(\log \hat{S}(\tau)) &= \sum_{j|\tau_j<\tau} (1 - \hat{h}_j)^{-2} \, \mathrm{Var}(1 - \hat{h}_j) \\ &= \sum_{j|\tau_j<\tau} \frac{g_j}{n_j(n_j - g_j)} \end{aligned} \tag{42}$$

Inductively, it is:

$$\mathrm{Var}(\hat{S}(\tau)) = \hat{S}^2(\tau) \sum_{\tau_j<\tau} \frac{g_j}{n_j(n_j - g_j)} \tag{43}$$

which is known as Greenwood's formula.

Another method of deriving the estimates of the empirical hazard or survivor function is to use a life table. Consider again the ordered observations of completed spells. Divide $[\tau_1, \tau_k]$ into some intervals (not necessarily equal) and build a life table —or, a listing of the number of censored and completed observations for each interval. Let those intervals be $I_1, I_2, \ldots I_k$. As before, let m_j be the number of censored data within the interval, I_j and let g_j be the number of completed durations in the interval I_j. The number of observations in the risk set at t_j is $n_j = \sum_{1 \geq j}(g_1 + m_1)$. Then the life table estimator of the hazard, h_j, gives the conditional probability of a failure during the j-th interval, I_j:

$$\hat{h}_i = \frac{g_j}{n_j - m_j/2} \tag{44}$$

for all n_j except for $n_j = 0$, in which case, $\hat{h}_j = 1$. In the denominator, m_j is divided by 2 by assuming that about half of the m_j observations are at risk throughout the interval, I_j. The corresponding life table estimator of the survivor function at the end of the interval I_j is:

$$\hat{S}(I_j) = \prod_{i=1}^{j}(1 - \hat{h}_i) \tag{45}$$

Variances for this estimator are given by replacing n_j by $n_j - m_j/2$ in the Greenwood's formula, (43).

This life table method is often used when the actual censoring times are unavailable, but g_j and m_j are still known for each j-th interval.

A plot of the empirical integrated hazard (41) or the survivor function (40) against the elapsed duration helps determine the parametric form of the hazard. For example, exponential duration distribution should generate a constant hazard and an integrated hazard that are linear in duration. From a practical point of view, plots of the integrated hazard are usually smoother and, hence, easier to interpret (Keifer 1988). The estimator is more accurate for the shorter durations, for which the number of observations in the risk set is high. In Kennan (1985), his decision to use the logit duration distribution was based on the finding that his empirical hazard was a U-shaped function of the elapsed duration. In Meyer (1990), he found several peaks in the empirical hazard that coincided with the timing when benefits entitlement changes. Upon incorporating the expected time until benefits lapse as covariates in the hazard, similar peaks disappeared from the baseline hazard that is estimated non-parametrically. Thus, comparison between the form of the empirical hazard and the non-parametrically estimated baseline hazard with time varying covariates helps to identify the effect of duration dependence and that of time varying covariates separately.

6. Conclusion

The first section of this paper discussed the basic concept and methodology of econometric modelling of duration analysis by referring to recent applications of such an analysis that have been mostly confined to the field of labor economics. In general, however, there are inference and identification problems inherent in the parametric approaches to the estimation of these models. Since the observed data, $f(\delta | x;\theta)$, cannot usually and uniquely identify the conditional duration distribution ($f(\delta|x, v;\theta)$) and the mixing distribution ($f_v(v)$), conventional studies have been specifying the parametric distributions for both of them even though the specification, particularly of $f_v(v)$, is rarely justified by economic theory. Worse still, incorrectly specified distributions lead not only to inefficient but also to inconsistent estimates of the parameters.

In light of these problems, we have introduced several semi- and non-parametric methods as solutions to combat them. In particular, this paper discussed four such methods: partial likelihood, regression, semi-parametric maximum likelihood, and empirical estimation. The first three are semi-parametric in the sense that they avoid making specific functional assumptions for either $f_v(v)$ or $f(\delta|x, v; \theta)$. The partial likelihood method is conditioned on the unknown form of the baseline hazard; hence, it avoids specification for the form of the baseline hazard. The regression approach is robust to the distributional forms of the mixing distribution by transforming the hazard model into a regression format. The semi-parametric maximum likelihood approach treats the baseline hazard and/or the heterogeneity distributions as step functions over the elapsed spell, whose discrete mass points are estimated together with the parameters of the explanatory variables. Lastly, the empirical approach formulates the hazard entirely by non-parametric means and is used primarily as a tool for preliminary analysis.

While the adoption of a semi-parametric method seems to be the best remedy for alleviating inference problems, the method that best suits the analysis depends on the circumstances of the particular economic problem involved and the type of data available. It should be emphasized that the design of the available data is one of the most important aspects in determining the appropriate method of estimation. If data are available not only on the duration but also on other variables inferable of the duration process, as in Lancaster and Chesher (1983), it should be possible to deduce rather than to estimate the values of parameters without being plagued by inference problems at all.

Appendix:

(i) Exponential distribution

$$f(\delta|x;\theta)=h_1\exp(-h_1\delta)$$
$$h(t, x;\theta)=h_1(x;\beta)$$
$$h_0(t;\alpha)=1$$

When the underlying duration distribution takes the exponential form, the transition probability is governed by Poisson arrivals with rate h_1. In this case, the probability of a transition from the current state within a small interval between t and $t + dt$ is $h_1\,dt + o(dt)$, where the limit of $o(dt)/dt$ becomes zero as dt approaches zero. This indicates that the probability of making more than one transition within the interval dt is negligible. Hence, the probability of having n transitions (i.e., arrivals) in t periods is the familiar Poisson probability, $[(h_1t)^n/n!]\exp(-h_1t)$, where h_1 is the hazard rate that is constant over time.

(ii) Weibull distribution

$$f(\delta|x;\theta)=h_1\alpha\delta^{\alpha-1}\exp(-h_1\delta^{\alpha})$$
$$h(t,x;\theta)=h_1\alpha t^{\alpha-1}$$
$$h_0(t;\alpha)=\alpha t^{\alpha-1}$$

The Weibull distribution can be considered as an exponential distribution on a rescaled time axis, t^{α} with a parameter h_1 where such parameter depends on the explanatory variables, x. This distribution can bear either a negative or positive duration dependence depending on the value of α. If α is smaller than 1, the hazard decreases monotonically with t. When $\alpha = 1$, this hazard reduces to the constant exponential hazard. The admissibility condition is satisfied for $\alpha>0$.

(iii) Gamma distribution

$$f(\delta|x;\theta)=[\gamma(\gamma t)^{k-1}\exp(-\gamma t)]/\Gamma(k)$$

for $k>0$ and where $\gamma = h_1(x;\beta)$. This distribution reduces to the exponential distribution for $k=1$, and the log-normal distribution for $k\to\infty$. This distribution can be derived as the distribution of the waiting time to the k-th point from a Poisson process with rate γ. It can accommodate various forms of duration dependence, although its application is limited due to its survivor function that involves an incomplete gamma integral.

(iv) Gompertz distribution

$$S(\Delta|x;\theta)=\exp\{h_1/\alpha[\exp(\alpha\Delta)-1]\}$$
$$h(t,x;\theta)=h_1\exp(\alpha t)$$

The hazard is linear in the exponential of the failure time, making the hazard decrease or increase more rapidly compared to the Weibull hazard, for instance. The hazard will have a positive duration dependence for $\alpha>0$.

(v) Pareto distribution

This distribution can be interpreted as a distribution of exponential survival time whose rate of arrival differs across individuals. If so, the duration density conditional on the arrival rate, μ, is:

$$f(\delta|\mu;\theta)=\mu\exp(-\mu\delta)$$

The unconditional density of δ is then:

$$f(\delta;\theta)=\int_0^{\infty}u\exp(-u\delta)f_{\mu}(u)du$$

A convenient distribution for $f_{\mu}(\cdot)$ is the Gamma with mean and variance being h_1:

$$f_{\mu}(u)=\frac{u^{h_1-1}e^{-u}}{\Gamma(h_1)}$$

In this particular case, the unconditional density of duration becomes Pareto distribution with following functions:

$$S(\Delta|x;\theta)=(\Delta+1)^{-h_1},\ f(\delta|x;\theta)=h_1(\delta+1)^{-h_1-1}$$
$$h(t,x;\theta)=h_1(t+1)^{-1}$$

This distribution converges to the exponential as $\kappa\to\infty$, and has a very long tail for a small value of κ. It possesses a negative duration dependence.

Notes:

[1] A brief discussion on discrete duration distribution model is given in Section 4.4.

[2] This is also one possible means by which to avoid the inference problems discussed in Section 4.

[3] For more detailed discussion, see Pudney (1989) section 6.4.2.

[4] Precisely speaking, time varying explanatory variables can be broken down into two categories. One is called external covariates and the other is called internal covariates. The former refers to variables whose process may influence but is not influenced by the failure. Using these terminology, the hazard at the elapsed duration t_i is conditioned on the external covariates over the range $t(0,\infty)$ and the internal covariates over the range

$(0, t_i)$.

[5] An alternative approach to the inference problem is to consider the determination process of spells under study to follow some stochastic Markovian processes. Then, a task is to model a first passage time (i.e., duration since the start of a spell) to the absorbing barrier (a certain threshold which, when reached, is associated with the termination of the spell) of such process. Examples of this approach being applied in econometrics are Lancaster (1972) and Jovanovic (1979). In order to have a simple form for the barrier, this approach necessitates a strict distributional assumption for the first passage time process. Moreover, assumptions such as independent and stationary increments are too restrictive in practice.

[6] For a censored observation, the term inside the last summation becomes simply, exp $(-\Lambda_i' v_j)p_j$.

[7] For a discrete duration distribution, density and survivor functions are: $f(\tau_j)=Pr(\delta=\tau_j)$ and $S(\tau_j)=Pr(\delta\geq\tau_j)=\sum_{\tau_i\geq\tau_j}f(\tau_i)$. The discrete analogue of the hazard rate at duration τ_j is the probability of exit at $t=\tau_j$ given that the spell has not ended for the last τ_j period, which is, $h_j=Pr(\delta=\tau_j|\delta\geq\tau_j)=f(\tau_j)/S(\tau_j)$.

[8] Precisely speaking, Kaplan-Meier estimator can be only viewed as a maximum likelihood estimator under certain conditions. See Kalbfleisch and Prentice (1980, pp.11–14) for detail.

References:

Amemiya, T (1985), *Advanced Econometrics,* Oxford: Basil Blackwell.

Cox , D. R. and Oakes, D. (1984), *Analysis of Survival Data,* London: Chapman and Hall.

Dempster, A., Laird, N. and Rubin, D. (1977), "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society, Series B,* 39, 1–38.

Elbers, C. and Ridder, G. (1982), "True and spurious duration dependence: the identifiability of the proportional hazards model", *Review of Economic Studies,* 49,402–11.

Flinn, C. and Heckman, J. (1982), "Models for the analysis of labor force dynamics". In R. Basmann and G. Rhodes (eds), *Advances in Econometrics,* Greenwich, CT: JAI Press, vol. 1.

Han, A. and Hausman, J. (1986), "Semi-parametric estimation of duration and competing risk models", M. I. T., working paper.

Harrison, A. and Stewart, M. (1987), "Cyclical variation in strike-settlement probabilities", University of Warwick: mimeo.

Heckman, J. and Singer, B. (1982), "The identification problem in econometric models for duration data". In W. Hildenbrand (ed.), *Advances in Econometrics: Proceedings of World Meetings of the Econometric Society,* Cambridge; Cambridge University Press.

Heckman, J. and Singer, B. (1984), "A method for minimizing the impact of distributional assumptions in econometric models for duration data", *Econometrica,* 52, 271–320.

Jones, S. (1988), "The relationship between unemployment spells and reservation wages as a test of search theory", *Quarterly Journal of Economics,* vol.CIII, 741–766.

Jovanovic, B. (1979), "Job matching and the theory of turnover", *Journal of Political Economy,* 87, 972–90.

Kalbfleisch, J. and Prentice, R. (1980), *Statistical Analysis of Failure Time Data,* New York: Wiley.

Kennan, J. (1985), "The duration of contract strikes in US manufacturing", *Journal of Econometrics, Annals* 1985–1, 28, 1–24.

Kiefer, J. and Wolfowits, J. (1956), "Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters", *Annals of Mathematical Statistics,* 27, 887–906.

Kiefer, N. (1988), "Economic duration data and hazard functions", *Journal of Economic Literature,* vol.XXVI, 646–79.

Kurosawa, M. and Pudney, S. (1993), "A method for the analysis of the timing and magnitude of events in a continuous-time panel", *Journal of Econometrics,* 59, 161–85.

Lancaster, T. (1972), "A stochastic model for the duration of a strike", *Journal of Royal Sta-*

tistical Society, Series A, 135, 257–71.

Lancaster, T. (1979), "Econometric methods from the duration of unemployment", *Econometrica,* vol.47, 939–56.

Lancaster, T. and Chesher, A. (1983), "An econometric analysis of reservation wages", *Econometrica,* 51, 1661–76.

Meyer, B. (1986), "Semi-parametric hazard estimation", M. I. T., working paper.

Meyer, B. (1990), "Unemployment insurance and unemployment spells", *Econometrica,* 58, 757–82.

Pudney, S. (1989), *Modelling Individual Choice: The Econometrics of Corners, Kinks and Holes,* Oxford: Basil Blackwell.

(経博：Ph. D. in Economics. 助教授)