Semiparametric Estimation of Hazard Models[1]

Bruce D. Meyer

Department of Economics and Center for Urban Affairs and Policy Research,
Northwestern University, and NBER

October 1995

Abstract

Starting from the Prentice and Gloeckler [1978] approach to discrete or discretely observed continuous hazard models, this paper provides several new results on the interpretation, identification, efficiency and estimation of hazard models. The approach allows an unknown form for the baseline hazard and discrete observations of the form usually found in economics. It is shown that the approach has a high efficiency relative to fully parametric models. The model is extended to allow for unobserved heterogeneity and using the Kiefer and Wolfowitz [1956] theorem, consistency is shown for the case where both the baseline hazard and the heterogeneity distribution are unknown. An approach to estimation based on the results of Turnbull [1974, 1976] is proposed. Some simulations support the analytical results on efficiency and show the results of misspecifying the shape of the baseline hazard.

[1]This paper is a revision of Chapter 1 of my 1987 MIT Ph.D. Thesis. I am indebted to Aaron Han, Daniel McFadden, Kevin M. Murphy, Burton Singer, Glenn Sueyoshi, and especially Jerry Hausman, Bo Honoré and Hidehiko Ichimura for insightful discussions, and to the Alfred P. Sloan Foundation and NSF for financial support.

1. Introduction

In work on hazard models, economists have emphasized the need to allow for unobserved individual characteristics or heterogeneity. On the other hand, statisticians have argued for the nonparametric estimation of the part of the hazard common to all individuals (the baseline hazard). This paper combines the two approaches. The proposed technique leaves the baseline hazard unspecified, but also introduces unobserved heterogeneity either parametrically or nonparametrically. This allows consistent estimation in many situations where currently used estimators are inconsistent.

The econometrics literature has emphasized the importance of unobserved heterogeneity in hazard models. Unobservable attributes can be expected to affect failure times. Lancaster [1979] and Lancaster and Nickell [1980] provide early discussions of the biases caused by ignoring heterogeneity. One solution is to assume that individual attributes are draws from a distribution with a known shape. Lancaster assumes that the heterogeneity in the population has a gamma distribution; others have assumed a lognormal distribution or a discrete distribution.

Heckman and Singer [1982, 1984a, 1986] have argued that this solution is inadequate. They contend that parameter estimates depend greatly on the assumed shape of the heterogeneity distribution. They propose a nonparametric technique for the estimation of the heterogeneity distribution. However, both Lancaster, and Heckman and Singer assume a simple parametric form for the baseline hazard. The Weibull hazard is often used despite a lack of theoretical support for any particular shape. Furthermore, Trussell and Richards [1985] show that the Heckman and Singer approach is very sensitive to the assumed baseline hazard. This is not surprising since a misspecified baseline hazard causes all parameter estimates to be inconsistent.

The nonparametric estimation of the baseline hazard is important for two additional reasons. First, the shape of the baseline hazard is often irregular and unlikely to be well approximated by a simple parametric form. For example, studies of unemployment durations have found irregular spikes in the hazard at about 24 and 36 weeks.[2] Second, the shape of the baseline hazard is often central to tests of economic hypotheses. Heckman and Singer [1984a], Katz [1986] and Katz and Meyer [1990] use unemployment duration baseline hazard estimates as support for search theory. Blank [1989] uses baseline hazard estimates to examine hypotheses about the effects of AFDC. This suggests that imposing a shape for the baseline hazard may lead to incorrect conclusions about economic hypotheses.

Statisticians have emphasized the nonparametric estimation of the baseline hazard. This approach is often associated with Kaplan and Meier [1958] and Cox [1972, 1975]. Cox's partial likelihood technique allows consistent estimation of the part of the hazard which varies over individuals even when the baseline hazard is unknown. The baseline hazard then can be nonparametrically estimated. Unfortunately, partial likelihood has two important drawbacks. First, the likelihood becomes intractable when many failures occur at the same time.[3] The problem is particularly severe in data usually encountered in economics, which have many tied failure times. Second, even without ties, multiple integrals of the dimension of the sample are required to integrate over the heterogeneity distribution. This alone would lead to an intractable likelihood function.

---

[2]See Moffitt [1985], Katz [1986], Meyer [1990], and Katz and Meyer [1990]. The spikes appear to be caused by unemployment insurance.

[3]The same is true of marginal likelihood estimation. The two approaches differ when tied failure times are present. Let the number of individuals in the risk set at t (those alive and uncensored just prior to t) be R_t and let the number of failures occurring at t be d_t. Then the partial likelihood contains a term for every subset of the risk set at time t which contains exactly d_t elements. The number of such terms is $R_t!/((R_t-d_t)!(d_t)!)$. The marginal likelihood contains a term for every permutation of the d_t failures. The number of such terms is $d_t!$. In applications, the number of terms is likely to be extremely large in both cases. There are approximations to these likelihoods, but their biases are unclear (see Kalbfleisch and Prentice [1980] and Cox and Oakes [1984] for a discussion of these issues).

Two other methods allow an unknown baseline hazard.[4] Both methods however, need modification for unobserved heterogeneity. First, Moffitt [1985], Green and Shoven [1986], and van den Berg and van Ours [1994] use a special discrete time model. The probability of an individual's spell ending in a period, given survival to the beginning of the period, is a period specific constant times a function of the exogenous variables for the individual. This approach is similar in spirit to the method discussed here, but it has several difficulties. The introduction of unobserved heterogeneity is difficult, and the functional form does not guarantee that failure probabilities are between zero and one. Also, this model cannot be derived from a continuous time process. Since there usually is no natural unit of time in economic problems, it is often easier to interpret coefficients and relate them to economic theory if the modeling is done in continuous time. Also, a discrete time model defined for one time interval usually has a different functional form when applied to another time unit.[5]

A second method which also allows an unknown baseline hazard is introduced in Prentice and Gloeckler [1978].[6] The present paper proves several properties of this model and extends it to allow unobserved heterogeneity. The Prentice and Gloeckler approach has several advantages. First, it avoids inconsistency caused by misspecification of the baseline hazard. Second, individual heterogeneity is easily added to the model. Gamma heterogeneity gives a closed form solution for the likelihood and avoids numerical integration. Other distributions can be used with some loss in computational

---

[4]Several papers fit simple hazard models into the single index or transformation model framework. See Dabrowska and Doksum [1985], or Stoker [1986], for example. This approach is attractive, but it cannot accommodate time dependent covariates and common forms of censoring. A different approach is taken by Honoré (1990), who allows arbitrary unobserved heterogeneity, but parameterizes the baseline hazard.

[5]Arguments for continuous time models are given in Singer and Spillerman [1976], Heckman and Borjas [1980], and Heckman and Singer [1986].

[6]Kalbfleisch and Prentice [1980] also discuss some aspects of this model. Han and Hausman [1986] extend the Prentice and Gloeckler model to introduce heterogeneity and competing risks.

simplicity. By combining the Prentice and Gloeckler approach with the Heckman and Singer approach one can obtain an estimator which nonparametrically estimates both the baseline hazard and the heterogeneity distribution. Third, the method uses data of the form usually found in economics. Specifically, the underlying process is assumed to be continuous, but the process is observed only at discrete times. Several kinds of censoring are allowed, including censoring which is conditional on the covariates.

Nonparametric estimation may result in an efficiency loss when the technique does not employ all available a priori information.[7] Fortunately, nonparametric estimation of the baseline hazard usually causes only a small efficiency loss. Furthermore, with large economic samples we are willing to use a possibly inefficient technique if it has a smaller bias. Section 2 describes the basic model, while Section 3 extensively treats the efficiency issue. The efficiency results are similar to panel data results from the linear model. The close relationship between the Prentice and Gloeckler approach and partial likelihood is also discussed. Section 4 describes how unobserved heterogeneity of a known form can be added. Section 5 proves, using the Kiefer and Wolfowitz conditions, that the parameters of the model can be consistently estimated even when the shape of both the baseline hazard and the heterogeneity distribution are unknown. Section 6 discusses a reformulation of the model and a new estimation technique. Section 7 discusses testing. Section 8 describes the simulations which examine efficiency and bias, while Section 9 concludes.

---

[7]In special circumstances, a second disadvantage may arise. Covariates which are the same for all subjects do not have identifiable coefficients. This problem is even less important than it initially seems. For example, if there is a policy change at a given calendar time (which is reflected in one of the covariates), it will generally occur at a different point in the subject's durations. Thus, if the subjects began their spells at different calendar times, the covariates will differ.

2. The Basic Model

Economic surveys rarely monitor individuals continuously. A much more common situation is panel data which provides information on individuals at discrete times. Also, retrospective surveys often give data which are discrete. For example, the length of unemployment spells is usually given in weeks. The following model accounts for the usual discreteness of economic data, but the economic process is allowed to take place continuously. As in most other continuous time models, the hazard is parameterized and estimated. Continuous time estimates are often easier to interpret and relate to economic theory than estimates from discrete time models.

Typically, the end of some durations is not observed because an individual leaves the survey. This occurrence is called right censoring and is explicitly allowed for in the model.

Following Prentice and Gloeckler, assume that the underlying process for observation i, where $i=1, \ldots, N$, consists of the positive failure time T_i, the positive censoring time C_i, and the time path of a p dimensional vector of covariates $z_i(t)$.

Assume that the censoring variable is integer valued and has a maximum of T. Censoring has this form when individuals are only observed at discrete times and there is a maximum length of observation. The censoring time for individual i is also assumed to be independent of T_i conditional on the path of the covariates. Let $t_i=\min(T_i, C_i)$ and let $\delta_i=1$ if individual i is observed to fail, i.e., $T_i < C_i$, and $\delta_i=0$ otherwise.

Suppose we only observe if t_i is in the interval I_t, where $I_t \equiv [t, t+1)$ for $t=0,1,\ldots,T-1$, and $I_T \equiv [T, \infty)$. This will occur if we only have information on individuals at discrete times. If $t_i \in I_t$, then let $k_i = t$. Assume that $z_i(\cdot)$ is constant on I_t for $t=0,1,\ldots,k_i+\delta_i-1$. Generally, economic data and computational limitations require this assumption in duration models. Often in econometrics z_i is taken to be independent of t.[8] When z_i is not constant over an interval, it has usually been approximated by its within interval mean.[9] Thus, the data available to the econometrician are (k_i, δ_i, $\{z_i(t)\}$), where k_i equals the integer part of min(T_i, C_i), and $\{z_i(t)\}$ denotes the path of the covariates, i. e. the values of the covariates when $t=0,1,...,k_i+\delta_i-1$.

Let the hazard for individual i be of the Cox [1972] proportional hazards form[10] with baseline hazard $\lambda_o(t)$,

(2.1) $$\lambda_i(t) = \lambda_o(t)\exp\{z_i(t)'\beta\}$$

(2.2) $$\lim_{h \to 0^+} \frac{prob\left[t + h > T_i \geq t \mid T_i \geq t\right]}{h} = \lambda_o(t)\exp\left\{z_i(t)'\beta\right\} \quad ,$$

where $0 \leq t \leq T < \infty$, and $\lambda_o(t)$ and β are unknown.

The probability that a duration will last until t+1 given that it has lasted until t is

(2.3) $$\begin{aligned} P\left[T_i \geq t+1 \mid T_i \geq t\right] &= \exp\left[-\int_t^{t+1} \lambda_i(u)du\right] \\ &= \exp\left[-\left[\int_t^{t+1} \lambda_o(u)du\right]\left[\exp\left\{z_i(t)'\beta\right\}\right]\right] \\ &= \exp\left\{-\exp\left[\gamma(t) + z_i(t)'\beta\right]\right\} \quad \text{, where} \end{aligned}$$

---

[8]See, for example, Lancaster [1979] or Kiefer and Neumann [1980].

[9]See Diamond and Hausman [1984], and Hausman and Wise [1985].

[10]Most of the following results are changed only slightly by replacing $\exp\{z_i(t)'\beta\}$ with $\varpi\{z_i(t)'\beta\}$, where $\varpi(\cdot)$ is positive and increasing.

(2.4) $$\gamma(t) = \ln\left\{\int_t^{t+1} \lambda_o(u)du\right\} \quad .$$

The likelihood of the data is

(2.5) $$L(\gamma,\beta) = \prod_{i=1}^{N}\left(\left[1 - \exp\left\{- \exp\left[\gamma(k_i) + z_i(k_i)'\beta\right]\right\}\right]^{\delta_i} \prod_{t=0}^{k_i-1} \exp\left\{- \exp\left[\gamma(t) + z_i(t)'\beta\right]\right\}\right) \quad ,$$

where $\gamma = [\gamma(0)\ \gamma(1)\ ...\ \gamma(T\text{-}1)]'$.

The likelihood is now a function of a finite number of parameters and is easily maximized. The log-likelihood function can be written:

(2.6) $$L_1(\gamma,\beta) = \sum_{i=1}^{N}\left\{\delta_i \log\left[1 - \exp\left\{-\exp\left[\gamma(k_i) + z_i(k_i)'\beta\right]\right\}\right] - \sum_{t=0}^{k_i-1} \exp\left[\gamma(t) + z_i(t)'\beta\right]\right\} \quad .$$

Three more substitutions simplify things further. Define

$D_i(t) = 1$ if $k_i = t$ and $\delta_i = 1$, i.e., the end of the spell was observed during the t_{th} interval, and 0 otherwise,

$R_i(t) = 1$ if $k_i \geq t+1$, i.e., the spell lasted at least until the end of the t_{th} interval, and 0 otherwise, and

$h_i(t) = \exp\{\gamma(t) + z_i(t)'\beta\}$.

Note that

(2.7) $$h_i(t) = \int_t^{t+1} \lambda(u)\exp\left\{z_i(t)'\beta\right\}du \quad ,$$

the average hazard over the interval I_t. The final expression for the log-likelihood is

(2.8) $$L_1(\gamma,\beta) = \sum_{i=1}^{N} \sum_{t=0}^{T-1} \left\{ D_i(t)\log\left[1-\exp\left\{-h_i(t)\right\}\right] - R_i(t)h_i(t) \right\} \quad .$$

Maximization of $L_1(\gamma,\beta)$ allows consistent estimation of β and $\gamma(t) = 1n\int_t^{t+1}\lambda(u)du$, (t=0,1,...,T-1). Two baseline hazards, $\lambda_o(t)$ and $\lambda_o'(t)$, are observationally equivalent if γ(t) = γ'(t) for all t, where $\gamma'(t) = 1n\int_t^{t+1}\lambda_o'(u)du$. In effect, one identifies the equivalence class in which λ_o (t) lies; the equivalence classes being defined by unique sequences

of γ(t), (t=0, 1,...,T-1).

The first and second partial derivatives of the log-likelihood can now be written as

(2.9) $$\frac{\partial L_1}{\partial\gamma(t)} = \sum_{i=1}^{N} \left\{ D_i(t)\frac{h_i(t)\exp\left\{-h_i(t)\right\}}{1-\exp\left\{-h_i(t)\right\}} - R_i(t)h_i(t) \right\} \quad , \qquad (t=0,1,...,T-1)$$

(2.10) $$\frac{\partial L_1}{\partial\beta} = \sum_{i=1}^{N} \sum_{t=0}^{T-1} \left\{ D_i(t)z_i(t)\frac{h_i(t)\exp\left\{-h_i(t)\right\}}{1-\exp\left\{-h_i(t)\right\}} - R_i(t)z_i(t)h_i(t) \right\} \quad ,$$

(2.11) $$\frac{\partial^2 L_1}{\partial\gamma(t)^2} = -\sum_{i=1}^{N} \left\{ D_i(t)\frac{h_i(t)\exp\left\{-h_i(t)\right\}\left[\exp\left\{-h_i(t)\right\}+h_i(t)-1\right]}{\left[1-\exp\left\{-h_i(t)\right\}\right]^2} + R_i(t)h_i(t) \right\} \quad , \qquad (t=0,1,...,T-1)$$

(2.12) $$\frac{\partial^2 L_1}{\partial\gamma(t)\partial\gamma(\tau)} = 0 \quad \text{for all } t \neq \tau$$

(2.13) $$\frac{\partial^2 L_1}{\partial\beta\partial\beta'} = -\sum_{i=1}^{N} \sum_{t=0}^{T-1} \left\{ D_i(t)z_i(t)z_i(t)'\frac{h_i(t)\exp\left\{-h_i(t)\right\}\left[\exp\left\{-h_i(t)\right\}+h_i(t)-1\right]}{\left[1-\exp\left\{-h_i(t)\right\}\right]^2} + R_i(t)z_i(t)z_i(t)'h_i(t) \right\}$$

$$(2.14) \qquad \frac{\partial^2 L_1}{\partial\beta\partial\gamma(t)} = -\sum_{i=1}^{N} \left\{ D_i(t) z_i(t) \frac{h_i(t)\exp\{-h_i(t)\}\left[\exp\{-h_i(t)\} + h_i(t) - 1\right]}{\left[1 - \exp\{-h_i(t)\}\right]^2} + R_i(t) z_i(t) h_i(t) \right\} \quad .$$

The Interpretation of the First Order Conditions

The first order conditions for a maximum of the log-likelihood have a simple interpretation. $\exp\{-h_i(t)\}/[1-\exp\{-h_i(t)\}]$ is the odds ratio or the ratio of the survival probability to the probability of failing in the interval I_t. $\partial L_1/\partial\gamma(t)=0$ requires that the average hazard for those who survive the interval I_t equal the average hazard for those that fail times the estimated odds ratio. If $z_i(t)=z(t)$ for all i, this requires that the predicted fraction of individuals surviving an interval equal the actual fraction. If $z_i(t)$ varies over i, the predicted fraction equals the actual fraction after weighting by the hazards. This occurs because in the proportional hazards model the effect of a unit change in $\gamma(t)$ on the hazard is proportional to the hazard. $\partial L_1/\partial\beta=0$ requires that this same equality hold after weighting by each of the covariates and summing over all time periods. Again, the weighting by the hazard occurs because the effect of a unit change in $z_i(t)$ is proportional to the hazard.

3. Efficiency Comparisons

The main alternative to the Prentice and Gloeckler approach is maximum likelihood with an assumed shape for the baseline hazard. The advantage of maximum likelihood is a possible efficiency gain if the assumed shape is correct. This section compares the efficiency of the Prentice and Gloeckler approach to maximum likelihood when β is estimated and the shape of the baseline hazard is known.

Note that if the assumed baseline hazard is incorrect, only the Prentice and Gloeckler approach gives consistent estimates.

The first part of this section explains how the hazard model comparisons are analogous to familiar efficiency comparisons from panel data. The comparisons are then derived formally. The Prentice and Gloeckler approach is then compared to partial likelihood, which is very similar. All of these comparisons use the average asymptotic information matrix, since its inverse is the asymptotic variance. The gain from parameterizing the baseline hazard, rather than using the Prentice and Gloeckler approach, is shown to be small in many cases.

The comparison of the Prentice and Gloeckler approach to parametric maximum likelihood is very close to the comparison of fixed effects to OLS in panel data.[11] The Prentice and Gloeckler baseline hazard parameters are analogous to time dummies in panel data. Specifically, consider the linear model

$$(3.1) \qquad D_{it} = \gamma + Z_{it}^{\prime}B + \varepsilon_{it} \quad ,$$

where i=1,. . .,N_t indexes individuals, and t indexes time periods.

The version of the model with fixed time effects is

$$(3.2) \qquad \begin{aligned} \left(D_{it} - D_{.t}\right) &= \left(Z_{it} - Z_{.t}\right)^{\prime}B + \left(\varepsilon_{it} - \varepsilon_{.t}\right) \quad , \text{ or} \\ D_{it} &= \gamma_t + Z_{it}^{\prime}B + \varepsilon_{it} \quad , \end{aligned}$$

where $D_{.t} = \frac{1}{N_t}\sum_{i=1}^{N_t} D_{it}$, and similarly for $Z_{.t}$ and $\varepsilon_{.t}$. Let $Var\left(\varepsilon_{it}|Z_{it}\right) = \sigma_{\varepsilon}^{2}$.

Note that time means are subtracted in (3.2) rather than individual means.

The average asymptotic information matrix for $\hat{B}_{OLS}$ from (3.1) is

[11]See Wallace and Hussain [1969], or Maddala [1971] for panel data efficiency comparisons.

$Var(Z)\sigma_\varepsilon^{-2}$ while that of $\hat{B}_{FE}$ from (3.2) is $E[Var(Z|t)]\sigma_\varepsilon^{-2}$, for fixed T letting $N_t \rightarrow \infty$ for all t. The expectation here is over t, which will be nontrivial if the sample is unbalanced.

A useful equation is

Var(Z) = E[Var(Z|t)] + Var[E(Z|t)] .

E[Var(Z|t)] is the within variance of Z, and Var[E(Z|t)] is the between variance. Fixed effects uses only the within time period variation in Z_{it}. The time period dummy variables in (3.2) absorb the variation over time in the mean of Z_{it}. When OLS is consistent, it also allows use of the between variance.

Maximum likelihood with $\lambda_o(t)$ assumed constant is analogous to OLS, while the Prentice and Gloeckler approach is similar to fixed effects. The baseline hazard parameters absorb the variation over time in the mean of the covariates. In microeconomic data there is typically much greater variation in the explanatory variables across individuals at a point in time than in the average over time. Usually, fixed effects estimation removes individual means causing a large efficiency loss. Here, time means are removed. The much larger variance across individuals will in most cases cause maximum likelihood to have only a small efficiency gain over the Prentice and Gloeckler approach.

A revised equation (3.1) is probably closer to a typical hazard model. Suppose the effect of time is parameterized using a vector of explanatory variables Y_t, which is common to all individuals and includes a constant. Then, model (3.1) becomes

(3.3) $$D_{it} = Y_t'\zeta + Z_{it}'B + \varepsilon_{it} \quad .$$

The fixed effects equation is unchanged and has information equal to σ_ε^{-2} times the within variance in Z. The average asymptotic information for β using OLS in (3.3) is no longer $Var[Z]\sigma_\varepsilon^{-2}$ but $E[Var(Z|Y)]\sigma_\varepsilon^{-2}$

instead. E[Var(Z|Y)] is smaller than Var[Z] since it is the residual variance in Z after projection onto the Y space. This residual variance is the within variance plus a fraction of the between variance. The fraction will be small if linear combinations of the elements of Y closely approximate the time pattern in the mean of Z. Since the within variance is usually larger than the entire between variance, fixed effects will likely have a high efficiency. A similar result occurs when the Prentice and Gloeckler approach is compared to maximum likelihood with the baseline hazard parameterized.

Consider the flexible parameterization of the baseline hazard, $\lambda_o(t) = \exp\{y(t)'\zeta\}$, which implies that $\lambda_i(t) = \exp\{y(t)'\zeta + z_i(t)'\beta\}$. This general parameterization includes the Weibull; let $y(t)' = (1,\ln(t))$. By using the Prentice and Gloeckler approach rather than assuming a shape for the baseline hazard one again loses a fraction of the between variance in z. The fraction is again the part of the between variance that cannot be predicted using linear combinations of the y vector. The fraction is likely to be small relative to the within variance used by the Prentice and Gloeckler approach. This result is complicated by weights which enter the expectations and variances. This issue is discussed extensively later. The next several pages make the above intuitive arguments precise.

The variance of the estimators are compared by examining their information matrices. Inversion of the average asymptotic information matrices gives the asymptotic variances. Consider the observed information matrix S for the Prentice and Gloeckler approach:

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} -\dfrac{\partial^2 L_1}{\partial\gamma\partial\gamma'} & -\dfrac{\partial^2 L_1}{\partial\gamma\partial\beta'} \\ -\dfrac{\partial^2 L_1}{\partial\beta\partial\gamma'} & -\dfrac{\partial^2 L_1}{\partial\beta\partial\beta'} \end{bmatrix} .$$

The marginal information on β, $\left(S^{22}\right)^{-1}$, is given by

(3.4) $$\left(S^{22}\right)^{-1} = S_{22} - S_{21}S_{11}^{\ -1}S_{12} \quad ,$$

using the partitioned inverse formula. Equation (3.4) can be rewritten in a form which gives it a straightforward interpretation. It is useful to rewrite (2.13) as

(3.5)
$$\begin{aligned} S_{22} &= -\frac{\partial^2 L_1}{\partial\beta\partial\beta'} \\ &= \sum_{t=0}^{T-1}\sum_{i=1}^{N}\left\{D_i(t)z_i(t)z_i(t)'\frac{h_i(t)\exp\{-h_i(t)\}\left[\exp\{-h_i(t)\} + h_i(t) - 1\right]}{\left[1-\exp\{-h_i(t)\}\right]^2} + R_i(t)z_i(t)z_i(t)'h_i(t)\right\} \\ &= \sum_{t=0}^{T-1}\sum_{i=1}^{N} W_i(t)z_i(t)z_i(t)' \end{aligned}$$

where

(3.6) $$W_i(t) = D_i(t)\frac{h_i(t)\exp\{-h_i(t)\}\left[\exp\{-h_i(t)\} + h_i(t) - 1\right]}{\left[1-\exp\{-h_i(t)\}\right]^2} + R_i(t)h_i(t)$$

Note that $W_i(t) \geq 0$, for all i and t. Equations (2.11) and (2.14) can be rewritten in the same manner as (3.5). Equation (2.11) yields

(3.7)
$$\begin{aligned} \left(S_{11}\right)_t &= -\frac{\partial^2 L_1}{\partial\gamma(t)^2} \\ &= \sum_{i=1}^{N} W_i(t) \end{aligned} \quad ,$$

where $\left(S_{11}\right)_t$ denotes the (t+1)*st* diagonal element of S_{11}. The off-diagonal elements of S_{11} are all zero.

Equation (2.14) yields

$$(3.8)\qquad \begin{aligned}\left(S_{21}\right)_t &= -\frac{\partial^2 L_1}{\partial\beta\partial\gamma(t)} \\ &= \sum_{i=1}^{N} z_i(t)W_i(t)\end{aligned} \quad ,$$

where $\left(S_{21}\right)_t$ denotes the (t+1)*st* column of S.

The observed marginal information for β, equation (3.4), can now be written as

$$(3.9)\qquad \begin{aligned}\left(S^{22}\right)^{-1} &= \sum_{t=0}^{T-1}\sum_{i=1}^{N} z_i(t)z_i(t)'W_i(t) \\ &\quad -\sum_{t=0}^{T-1}\left\{\left[\sum_{i=1}^{N} z_i(t)W_i(t)\right]\left[\sum_{i=1}^{N} W_i(t)\right]^{-1}\left[\sum_{i=1}^{N} z_i(t)'W_i(t)\right]\right\}\end{aligned} \quad .$$

The average information is obtained by dividing by the sample size. After some manipulation, it can be written as

$$(3.10)\qquad \frac{1}{N}\left(S^{22}\right)^{-1} = \left\{\sum_{t=0}^{T-1}\frac{1}{N}\sum_{i=1}^{N} W_i(t)\right\}\sum_{t=0}^{T-1}\left[\frac{\frac{1}{N}\sum_{i=1}^{N} W_i(t)}{\sum_{t=0}^{T-1}\frac{1}{N}\sum_{i=1}^{N} W_i(t)}\left[\frac{\frac{1}{N}\sum_{i=1}^{N} z_i(t)z_i(t)'W_i(t)}{\frac{1}{N}\sum_{i=1}^{N} W_i(t)} - \frac{\frac{1}{N}\sum_{i=1}^{N} z_i(t)W_i(t)}{\frac{1}{N}\sum_{i=1}^{N} W_i(t)}\;\frac{\frac{1}{N}\sum_{i=1}^{N} z_i(t)'W_i(t)}{\frac{1}{N}\sum_{i=1}^{N} W_i(t)}\right]\right] \quad .$$

This information matrix for the Prentice and Gloeckler approach will be compared to the fully parametric maximum likelihood information matrix. A flexible parameterization of the baseline hazard is $\lambda_o(t) = \exp\{y(t)'\zeta\}$, which implies that $\lambda_i(t) = \exp\{y(t)'\zeta + z_i(t)'\beta\}$. y(t) is a q dimensional vector constant on each interval (one component of y(t) can vary within intervals). The vector y(t) is also assumed to include the constant 1. This general parameterization includes the Weibull; let q=2 and y(t)'=(1, ln(t)). The marginal observed information matrix for β using maximum likelihood is now

$$\tilde{S} = \begin{bmatrix} \tilde{S}_{11} & \tilde{S}_{12} \\ \tilde{S}_{21} & \tilde{S}_{22} \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2 L_2}{\partial\zeta\partial\zeta'} & -\frac{\partial^2 L_2}{\partial\zeta\partial\beta'} \\ -\frac{\partial^2 L_2}{\partial\beta\partial\zeta'} & -\frac{\partial^2 L_2}{\partial\beta\partial\beta'} \end{bmatrix} .$$

The submatrices are

(3.11) $$\tilde{S}_{11} = \sum_{t=0}^{T-1} \sum_{i=1}^{N} y(t)y(t)'W_i(t)$$

(3.12) $$\tilde{S}_{21} = \sum_{t=0}^{T-1} \sum_{i=1}^{N} z_i(t)y(t)'W_i(t) \text{ and}$$

(3.13) $$\tilde{S}_{22} = S_{22} \quad .$$

The marginal information on β when maximum likelihood is used is then

(3.14) $$\begin{aligned} \left(\tilde{S}^{22}\right)^{-1} &= \tilde{S}_{22} - \tilde{S}_{21}\tilde{S}_{11}^{-1}\tilde{S}_{12} \\ &= \sum_{t=0}^{T-1} \sum_{i=1}^{N} z_i(t)z_i(t)'W_i(t) \\ &\quad - \sum_{t=0}^{T-1} \sum_{i=1}^{N} z_i(t)y(t)'W_i(t)\left[\sum_{t=0}^{T-1} \sum_{i=1}^{N} y(t)y(t)'W_i(t)\right]^{-1} \sum_{t=0}^{T-1} \sum_{i=1}^{N} y(t)z_i(t)'W_i(t) \quad . \end{aligned}$$

The average information can be written as

(3.15) $$\frac{1}{N}\left(\tilde{S}^{22}\right)^{-1} = \left\{\frac{1}{N}\sum_{t=0}^{T-1} \sum_{i=1}^{N} W_i(t)\right\}\left[\frac{\frac{1}{N}\sum_{t=0}^{T-1} \sum_{i=1}^{N} z_i(t)z_i(t)'W_i(t)}{\frac{1}{N}\sum_{t=0}^{T-1} \sum_{i=1}^{N} W_i(t)} - \frac{\frac{1}{N}\sum_{t=0}^{T-1} \sum_{i=1}^{N} z_i(t)y(t)'W_i(t)}{\frac{1}{N}\sum_{t=0}^{T-1} \sum_{i=1}^{N} W_i(t)}\left[\frac{\frac{1}{N}\sum_{t=0}^{T-1} \sum_{i=1}^{N} y(t)y(t)'W_i(t)}{\frac{1}{N}\sum_{t=0}^{T-1} \sum_{i=1}^{N} W_i(t)}\right]^{-1} \frac{\frac{1}{N}\sum_{t=0}^{T-1} \sum_{i=1}^{N} y(t)z_i(t)'W_i(t)}{\frac{1}{N}\sum_{t=0}^{T-1} \sum_{i=1}^{N} W_i(t)}\right] .$$

Under regularity conditions, the average information matrices $\frac{1}{N}\left(S^{22}\right)^{-1}$ and $\frac{1}{N}\left(\tilde{S}^{22}\right)^{-1}$ converge. This result is stated as a proposition.

PROPOSITION 1 (Convergence of Information Matrices):

If (T_i, C_i, {$z_i(t)$, t=0,1,...,T-1}), (i=1,...,N), are independent, identically distributed random vectors, {$z_i(t)$, t=0,1,...,T-1} has bounded second moments, and y(t), t=0, . . .,T-1 is nonstochastic and finite, then as $N \to \infty$

(3.16)
$$\frac{1}{N}\left(S^{22}\right)^{-1} \underset{p}{\to} W E_w\left\{E_w\left[zz'|t\right] - E_w\left[z|t\right]E_w\left[z'|t\right]\right\}$$
$$= W\left\{E_w[zz'] - E_w\left[E_w[z|t]E_w[z'|t]\right]\right\} = W E_w\left\{Var_w[z|t]\right\}$$
, and

(3.17)
$$\frac{1}{N}\left(\tilde{S}^{22}\right)^{-1} \underset{p}{\to} W\left\{E_w\left[zz'\right] - E_w\left[zy'\right]E_w\left[yy'\right]^{-1}E_w\left[yz'\right]\right\}$$

where $W = \underset{n\to\infty}{plim} \sum_{t=0}^{t-1} \frac{1}{N}\sum_{i=1}^{N} W_i(t)$ and $E_w[\quad]$, $Var_w[\quad]$ denote the weighted expectation and variance, respectively, weighting by the $W_i(t)$.

PROOF:

[This will be replaced by conditions for the consistency and asympotical normality of the estimators]

The WLLN implies that

$$plim \frac{1}{N}\sum_{i=1}^{N} W_i(t) = E[W(t)] \quad ,$$

$$W = \sum_{t=0}^{T-1} E[W(t)] \quad ,$$

$$plim \frac{1}{N}\sum_{i=1}^{N} z_i(t)W_i = E[z(t)W(t)] \quad \text{, and}$$

$$plim\frac{1}{N}\sum_{i=1}^{N} z_i(t)z_i(t)'W_i(t) \;=\; E[z(t)z(t)'W(t)] \quad ,$$

provided that the expectations on the right exist. Remember that

$$W_i(t) \;=\; D_i(t)\frac{h_i(t)\exp\{-h_i(t)\}\left[\exp\{-h_i(t)\} + h_i(t) \;-\; 1\right]}{\left[1 - \exp\{-h_i(t)\}\right]^2} \;+\; R_i(t)h_i(t) \quad .$$

The first term of $W_i(t)$ is bounded, while the second is not since $h_i(t)$ may be infinite. However, there exists a Δ such that $E[R_i(t)h_i(t)|z_i(t)] \leq \exp\{-h_i(t)\}h_i(t) < \Delta < \infty$. The law of iterated expectations then gives the existence of the required moments. The above results imply that

$$\frac{1}{N}\left(S^{22}\right)^{-1} \underset{p}{\rightarrow} W\sum_{t=0}^{T-1}\frac{E[W(t)]}{W}\left[\frac{E[z(t)z(t)'W(t)]}{E[W(t)]} \;-\; \frac{E[z(t)W(t)]}{E[W(t)]}\;\frac{E[z(t)'W(t)]}{E[W(t)]}\right]$$

$$\frac{1}{N}\left(\tilde{S}^{22}\right)^{-1} \underset{p}{\rightarrow} W\left[\frac{\sum_{t=0}^{T-1} E[z(t)z(t)'W(t)]}{\sum_{t=0}^{T-1} E[W(t)]} \;-\; \frac{\sum_{t=0}^{T-1} E[z(t)y(t)'W(t)]}{\sum_{t=0}^{T-1} E[W(t)]}\left[\frac{\sum_{t=0}^{T-1} y(t)y(t)'E[w(t)]}{\sum_{t=0}^{T-1} E[W(t)]}\right]^{-1}\frac{\sum_{t=0}^{T-1} E[y(t)z(t)'W(t)]}{\sum_{t=0}^{T-1} E[W(t)]}\right] \quad .$$

Rewriting the above expressions as weighted expectations completes the proof. Q.E.D.

Equation (3.16) indicates that the information from the Prentice and Gloeckler approach is W times the within variance in z. To interpret the information matrix for maximum likelihood with a paremeterized baseline hazard, rewrite the right-hand side of (3.17) as

$$W\left\{E_w[zz'] \;-\; E_w[zy']E_w[yy']^{-1}E_w[yz']\right\} \;=\; W\left\{E_w[zz'] \;-\; E_w\left(E_w[z|t]E_w[z'|t]\right)\right\}$$
$$+\; W\left\{E_w\left(E_w[z|t]E_w[z'|t]\right) \;-\; E_w[zy']E_w[yy']^{-1}E_w[yz']\right\} \quad . \tag{3.18}$$

The second term on the right-hand side above is the efficiency gain from using maximum likelihood rather than the Prentice and Gloeckler approach. This term is a fraction of the between variance in z equal to the variance of $E_w[z|t]$ minus its projection onto the y space. The Prentice and Gloeckler approach is fully efficient when $E_w[z|t] = Ay(t)$ for some p by q matrix A.

In microeconomic data there is typically much greater variation across individuals than variation in the population mean over time. In hazard models, many of the explanatory variables are constant over time. The much larger variance across individuals will likely cause the Prentice and Gloeckler approach to have a relatively small efficiency loss for most elements of β. If a time varying explanatory variable is the same for almost all individuals, its coefficient may be imprecisely estimated unless the baseline hazard is known a priori.

A complete discussion of the components of the variance of z needs to cover the weights which alter the weighted distribution of z over time. The weights, which were given earlier in equation (3.6), are

$$W_i(t) = D_i(t)\frac{h_i(t)\exp\{-h_i(t)\}[\exp\{-h_i(t)\} + h_i(t) - 1]}{[1-\exp\{-h_i(t)\}]^2} + R_i(t)h_i(t) \quad .$$

For small values of $h_i(t)$ ($h_i(t) < .5$), a close approximation which aids the interpretation of the weights is

$$W_i(t) \approx (1/2)D_i(t)h_i(t) + R_i(t)h_i(t) \quad . \tag{3.19}$$

The weights depend on the covariates, the coefficients β, the baseline hazard, and the effect z has on censoring. These effects do not necessarily work in the same direction and may cancel each other out. In most applications the $R_i(t)$ term is the most important; $D_i(t)=0$ for most t, and the baseline hazard enters $h_i(t)$ multiplicatively so it will not affect $E_w[z|t]$ except through $R_i(t)$. Since $R_i(t)$ is an indicator variable for survival until t+1, it will tend to equal one for observations with covariates associated with long durations. This will likely induce at least one monotonic function of t, and much of this trend is likely to be absent after projection on y(t). In general, one doesn't expect the weights to alter dramatically the earlier conclusions of high efficiency of the Prentice and Gloeckler approach.

These efficiency comparisons are almost identical to those comparing partial likelihood and maximum likelihood in data with the exact failure times.[12] A large number of papers have been written on this subject and they generally conclude that partial likelihood has a high relative efficiency.[13] The evidence includes several Monte Carlo studies. The Prentice and Gloeckler approach is like partial likelihood in many respects. Both estimators make no assumption about the baseline hazard. Both use only variation in the explanatory variables across individuals; the over time variation is not used. In fact, Bailey [1984] shows that in a special case partial likelihood is asymptotically equivalent to the present model. This result however, requires knowledge of the exact times of failure, no tied failure times, and assumes that a separate baseline hazard parameter is estimated between each time of failure.

4. Parametric Unobserved Heterogeneity

It is reasonable to assume that the econometrician will not observe all of the determinants of an individual's hazard. Numerous authors have emphasized the biases caused by ignoring this variation in the hazard.[14] One reasonable approach, which is used by all of the cited authors, assumes that the heterogeneity enters the hazard multiplicatively.[15] In the present model the hazard then takes the form

$$(4.1) \qquad \lambda_i(t) = \theta_i \lambda_o(t) \exp\{z_i(t)'\beta\} \quad ,$$

---

[12]Oakes [1977], Efron [1977], and Cox and Oakes [1984] are closest to this paper.

[13]See Kalbfleisch and Prentice [1980] for a summary of several studies.

[14]A partial list includes Lancaster [1979], Lancaster and Nickell [1980], Heckman and Borjas [1980], Heckman and Singer [1984a], Harris [1982], Diamond and Hausman [1984], and Hausman and Wise [1985].

[15]Multiplicative heterogeneity more closely corresponds to the idea of omitted variables. Additive heterogeneity is easily handled in this approach. Let $\lambda_i(t)=\theta_i\lambda_o(t)\exp\{z_i(t)'\beta\} + \kappa_i$, where κ_i is independent of θ_i, the covariates, and censoring. The new likelihood involves T additional parameters for $E[\exp\{-t\kappa\}]$, $(t=0,1,...,T-1)$.

where θ_i is a draw from a nonnegative distribution $\mu(\theta)$, which is independent of the covariates and censoring. A log-likelihood is obtained by conditioning on the unobserved θ and then integrating over its distribution. For the Prentice and Gloeckler approach one obtains

$$(4.2)\quad L(\gamma,\beta,\mu) \;=\; \sum_{i=1}^{N} \log\left\{ \int \exp\left[-\theta \sum_{t=0}^{k_i-1} \exp\left\{\gamma(t) + z_i(t)'\beta\right\}\right] d\mu(\theta) - \delta_i \int \exp\left[-\theta \sum_{t=0}^{k_i} \exp\left\{\gamma(t) + z_i(t)'\beta\right\}\right] d\mu(\theta) \right\} \quad .$$

To implement this approach one typically assumes a shape for the distribution $\mu(\theta)$. Estimation without this assumption is discussed in the next section. It is plausible that much of the parameter instability found by Heckman and Singer is due to the assumption of a Weibull baseline hazard. When the baseline hazard is nonparametrically estimated, which heterogeneity distribution is chosen may be unimportant.

A convenient and commonly used distribution for θ is the gamma. The gamma distribution gives a closed form expression for the likelihood, avoiding numerical integration. The gamma distribution is used in the following discussion, although other distributions could be used following the same approach. If θ is distributed gamma with mean one (a normalization) and variance σ^2, then the log-likelihood becomes

$$(4.3)\quad \begin{aligned} L_3\left(\gamma,\beta,\sigma^{-2}\right) &= \sum_{i=1}^{N} \log\left\{\left[1 + \sigma^2 \sum_{t=0}^{k_i-1} \exp\left\{\gamma(t) + z_i(t)'\beta\right\}\right]^{-\sigma^{-2}} - \delta_i\left[1 + \sigma^2 \sum_{t=0}^{k_i} \exp\left\{\gamma(t) + z_i(t)'\beta\right\}\right]^{-\sigma^{-2}}\right\} \\ &= \sum_{i=1}^{N} \log\left\{\left[1 + \sigma^2 \sum_{t=0}^{k_i-1} h_i(t)\right]^{-\sigma^{-2}} - \delta_i\left[1 + \sigma^2 \sum_{t=0}^{k_i} h_i(t)\right]^{-\sigma^{-2}}\right\} \quad . \end{aligned}$$

where $h_i(t) = \exp\{\gamma(t) + z_i(t)'\beta\}$.

5. Nonparametric Estimation of the Heterogeneity

In a series of papers, Heckman and Singer [1982, 1984a, 1986] argue that the distribution of the population heterogeneity should be nonparametrically estimated. They show the sensitivity of parameter estimates to the assumed distribution for θ, using the Kiefer and Neumann [1981] data. However, Trussell and Richards [1985] show that the Heckman and Singer estimator is very sensitive to the assumed shape of the baseline hazard. A solution to these problems is obtained by combining the Prentice and Gloeckler approach and the Heckman and Singer approach. This section proposes an estimator which does not require knowledge of the shape of the baseline hazard or the heterogeneity distribution. The results of this section also extend Heckman and Singer [1984a] to interval data and censoring of the kind typically found in economics.

Heckman and Singer [1984a] is based on the Kiefer and Wolfowitz [1956] theorem on maximum likelihood estimation in the presence of infinitely many incidental parameters.[16] Under various conditions Heckman and Singer verify the required assumptions for the theorem. Heckman and Singer assume that the data contain the exact times of failure. They also assume that the distribution of the censoring variable C_i is known and independent of the covariates $\{z_i(t)\}$.

In most economic applications these assumption do not hold. Failures are usually grouped in intervals because the data come from panel surveys or are imprecisely measured. Furthermore, the distribution of the censoring variable is usually unknown and likely to depend on the covariates. For example, in the Moffitt [1985] study of unemployment durations, C_i is the maximum potential duration of unemployment benefits. C_i will depend on an individual's work history and the state unemployment rate. The following proof allows interval data and censoring of an unknown form which may depend on the covariates.

Some notation is helpful before giving the proof of consistency of maximum likelihood applied to equation (4.2). Let $\pi \equiv (\beta,\gamma) \in P$, where P is the Cartesian product of (p+T) finite closed subintervals of the real line. Let θ denote a value of a random variable Θ taking values in the positive real line and define $M \equiv \{\mu\} \equiv$ the set of probability distributions with values on the nonnegative real line. Let a point in the product space Ψ = P x M be denoted by $\psi = (\pi,\mu)$. Denote the true values of π and μ by $\psi_0=(\pi_0,\mu_0)$. Let $m(\{z(t)\})$ denote the frequency function of the covariates $\{z(t)\}$, where $(t = 0,1,...,(k-1+\delta))$. $m(\{z(t)\})$ is assumed to be a bounded density with respect to Lebesgue measure on the continuous coordinates of $\{z(t)\}$ and a distribution function on the discrete

[16]Cosslett [1983] also relies on the Kiefer and Wolfowitz theorem. He nonparametrically estimates the distribution of an additive error term in the binary choice model.

coordinates. Lastly, denote the probability distribution of δ given {z(t)} by c(δ|{z(t)}). It is assumed that both m(·) and c(·) do not depend on ψ.

The random vector (k,δ,{z(t)}) has frequency function f_ψ(k,δ,{z(t)}) which can now be written as

$$(5.1)\qquad f_\psi(k,\delta,\{z(t)\}) = m(\{z(t)\})c(\delta\,|\,\{z(t)\})\int_0^\infty g_\pi(k\,|\,\delta,\{z(t)\},\theta)\,d\mu(\theta) \quad ,$$

where

$$(5.2)\qquad g_\pi(k\,|\,\delta,\{z(t)\},\theta) = \begin{cases} 1 - \exp\{-\theta\, h(0)\} & \textit{if } \delta = 1 \text{ and } k = 0 \quad , \\ \exp\left\{-\theta\sum_{t=0}^{k} h(t)\right\} & \textit{if } \delta = 0 \quad , \\ \exp\left\{-\theta\sum_{t=0}^{k=1} h(t)\right\} - \exp\left\{-\theta\sum_{t=0}^{k} h(t)\right\} & \textit{if } \delta = 1 \text{ and } k \geq 1 \quad , \end{cases}$$

and h(t) = exp {γ(t) + z(t)'β} as before.

Kiefer and Wolfowitz use the metric

$$d(\psi_1,\psi) \equiv d((\pi_1,\mu_1),(\pi_2,\mu_2)) = \sum_{j=1}^{T+p} \left|\arctan\pi_{1j} - \arctan\pi_{2j}\right| + \int_0^\infty \left|\mu_1(\theta) - \mu_2(\theta)\right| \exp\{-\theta\}d\theta \quad ,$$

where π_{ij} denotes the j_{th} element of π_i.

Note that $d((\pi_i\mu_i),(\pi_0,\pi_0)) \rightarrow 0$ implies that $\pi_i \rightarrow \pi_0$ and $\mu_i(\theta) \rightarrow \mu_0(\theta)$ at all points of continuity of $\mu_0(\theta)$. Under assumptions which will be stated shortly, ψ is complete, i.e., all Cauchy sequences in ψ converge to an element of ψ. The metric d(·,·) is useful because it makes the space ψ totally bounded. Since a complete and totally bounded metric space is compact, the parameter space ψ is compact.

One comment is needed to clarify how the present model fits in the Kiefer and Wolfowitz framework. Kiefer and Wolfowitz assume that m(·), c(·), and $g_\pi(\cdot)$ are known up to a finite set of parameters. However, the maximization of $f_\psi(\cdot)$ over ψ does not depend on m(·) and c(·), so one can treat the problem as if these functions were known.

Given independent and identically distributed observations on the random vector (k,δ,{z(t)}), Kiefer and Wolfowitz give five conditions which imply the consistency of maximum likelihood. More precisely, the estimate of π converges almost surely to π_0, and the estimate of μ(θ) converges almost surely to $\mu_0(\theta)$ at all points of continuity of $\mu_0(\theta)$. With these preliminaries, the main result can be stated and proved.

PROPOSITION 2 (Nonparametric Estimation of Heterogeneity and Baseline Hazard):

If (A) M is restricted to a class of uniformly integrable[17] distribution functions on [0,∞),

(B) There exists at least one component of the covariates $\{z^j(t)\}$, $j \in \{1,...,p\}$, with $\beta_j \neq 0$ and with distribution that has an everywhere positive Lebesgue density for t = 0,1 conditional on some realization of the other covariates $\{z^1(t),..., z^{j-1}(t), z^{j+1}(t),...,z^p(t)\}$,

(C) $$E_{\psi_0}\left[\ln \int_0^\infty g_{\pi_0}(k \mid \delta, \{z(t)\}, \theta)\, d\mu(\theta)\right] > -\infty \quad ,$$

then the maximum likelihood of estimator of ψ in (4.2) gives an estimate of π which converges almost surely to π_0 and an estimate of μ(θ) which converges almost surely to $\mu_0(\theta)$ at all points of continuity of $\mu_0(\theta)$. A discussion of the assumptions follows the proof.

PROOF:

The five Kiefer and Wolfowitz conditions are verified:

[17]A family of random variables $\{\theta_j\}$, j∈J, is uniformly integrable iff $\lim_{\Delta \to \infty} \int_{|\theta_j| > \Delta} \theta_j \, dP = 0$, uniformly in j∈J.

1. The conditional frequency function

$$f_\pi(k,\delta,\{z(t)\}|\theta) \equiv m(\{z(t)\})c(\delta|\{z(t)\})g_\pi(k|\delta,\{z(t)\},\theta)$$

is a density with respect to a σ-finite measure ξ defined on

$$\{(0,1),(T,0),(i,j);\ i = 1,...,T\text{-}1, j = 0,1\} \text{ x } R_{pT}\ .$$

$d\xi = dk\ d\delta\ dz_1(0)...dz_p(0)...dz_1(T\text{-}1)...dz_p(T\text{-}1)$, where $dk, d\delta, dz_1(t),...,dz_r(t); t=0,1,...,T\text{-}1$ are counting measure (the first r coordinates of z(t) are taken to be discrete) and $dz_{r+1}(t),...dz_p(t)$; $t = 0,1,...T\text{-}1$ are Lebesgue measure on the real line.

2. (Identification) If $\psi_1 \in \Psi$ and $\psi_1 \neq \psi_0$ then identification requires that for at least one region A

$$\int_A f_{\psi_1}(k,\delta,\{z(t)\})d\xi(k,\delta\{z(t)\}) \neq \int_A f_{\psi_0}(k,\delta,\{z(t)\})d\xi(k,\delta,\{z(t)\}) \quad .$$

Suppose $\psi_1 \equiv (\pi_1, \mu_1) \neq (\pi_0, \mu_0) \equiv \psi_0$, then, since c(•) and m(•) do not depend on ψ, it is necessary that

$$(*) \qquad \int_0^\infty g_{\psi_1}(k|\delta,\{z(t)\},\theta)d\mu_1(\theta) \neq \int_0^\infty g_{\psi_0}(k|\delta,\{z(t)\},\theta)d\mu_0(\theta)$$

for some (k,δ,{z(t)}) such that $f_{\psi_0}(k,\delta,\{z(t)\})d\xi(k,\delta,\{z(t)\}) > 0$.

We prove this proposition by contradiction. Suppose that (*) holds for k=0,1, i.e.,

$$(5.3) \qquad \int_0^\infty g_{\pi_1}(k = 0|\delta,\{z(t)\},\theta)d\mu_1(\theta) = \int_0^\infty g_{\pi_0}(k = 0|\delta,\{z(t)\},\theta)d\mu_0(\theta) \text{ and}$$

(5.4) $$\int_0^\infty g_{\pi_1}(k = 1|\delta,\{z(t)\},\theta)d\mu_1(\theta) = \int_0^\infty g_{\pi_0}(k = 1|\delta,\{z(t)\},\theta)d\mu_0(\theta) \quad ,$$

for all {z(t)} such that $f_{\psi_0}(k,\delta,\{z(t)\})d\xi(k,\delta,\{z(t)\}) > 0$. One can show that this supposition leads to a contradiction.

First, let z(0) = z(1) = $(0,...,0,z^j(0),0,...,0)$, and let $z \equiv z^j(0)$. If the value of the covariates $(z^1(t),...,z^{j-1}(t),z^{j+1}(0),...,z^p(t))$ in Assumption (B) does not equal zero for t=0,1, one can translate the coordinates so that it does equal zero. By assumption (B) $f_{\psi_0}(\,) > 0$ for k = 0,1 and all real z.

Next, one can normalize E[θ] = 1 because θ belongs to a uniformly integrable class so it has finite mean. Using the definition of $g_\pi(\cdot)$, it is easy to see that the two equalities (5.3) and (5.4) are equivalent to

(5.5) $$\int_0^\infty \exp\left[-\theta \exp\{\gamma_1(0) + z\beta_1\}\right]d\mu_1(\theta) = \int_0^\infty \exp\left[-\theta \exp\{\gamma_0(0) + z\beta_0\}\right]d\mu_0(\theta) \text{ and}$$

(5.6) $$\int_0^\infty \exp\left[-\theta\left(\exp\{\gamma_1(0)\} + \exp\{\gamma_1(1)\}\right)\exp\{z\beta_1\}\right]d\mu_1(\theta) = \int_0^\infty \exp\left[-\theta\left(\exp\{\gamma_0(0)\} + \exp\{\gamma_0(1)\}\right)\exp\{z\beta_0\}\right]d\mu_0(\theta) \quad .$$

Now, by judiciously making substitutions, equations (5.5) and (5.6) can be put in the form of (A.2) in Elbers and Ridder [1982]. One can then follow their proof. Define

$$r = \exp\{z\} \quad ,$$

$$a_i(r) = \left[\exp\{\gamma_i(0)\} + \exp\{\gamma_i(1)\}\right] r^{\beta_{ij}} \quad \text{and}$$

$$b_i = \frac{\exp\{\gamma_i(0)\}}{\exp\{\gamma_i(0)\} + \exp\{\gamma_i(1)\}} \qquad (i = 0, 1) \quad .$$

where β_{ij} is the jth component of β_i .

Let $\varphi_{\mu_i}(s) \equiv \int_0^\infty \exp[-\theta s] d\mu_i(\theta)$ denote the Laplace transform of μ_i.

Note that the support of r is (0,∞) by Assumption (B). Now equations (5.5) and (5.6) can be rewritten as

$$\varphi_{\mu_1}\big(b_1 a_1(r)\big) = \varphi_{\mu_0}\big(b_0 a_0(r)\big) \text{ and}$$

$$\varphi_{\mu_1}\big(a_1(r)\big) = \varphi_{\mu_0}\big(a_0(r)\big) \text{ for all } r > 0 \quad .$$

These last two equations imply

(5.7) $$b_0 a_0(r) = \varphi_{\mu_0}^{-1}\Big(\varphi_{\mu_1}\big(b_1 a_1(r)\big)\Big) = b_0 \varphi_{\mu_0}^{-1}\Big(\varphi_{\mu_1}\big(a_1(r)\big)\Big) \text{ for all } r > 0 \quad .$$

Define $F = \varphi_{\mu_0}^{-1} \varphi_{\mu_1}$. Then $\lim_{s \to 0^+} F(s) = 0$.

Also, since $\int_0^\infty \theta d\mu_1(\theta) = \int_0^\infty \theta d\mu_0(\theta) = 1$, one has $\lim_{s \to 0^+} F'(s) = \lim_{s \to 0^+} \frac{\varphi'_{\mu_1}(s)}{\varphi'_{\mu_0}\Big(\varphi_{\mu_0}^{-1}\big(\varphi_{\mu_1}(s)\big)\Big)} = 1$ by the properties of the Laplace transform.[18] From (5.7) one has

[18]See Feller [1971, chapter XIII].

(5.8) $F(b_1 s) = b_0 F(s)$ for all $s > 0$.

Let $s = b_1 s'$ in (5.8), then $F((b_1)^2 s') = (b_0)^2 F(s')$ for all $s' > 0$. Repeating this substitution gives $F((b_1)^n s) = (b_0)^n F(s)$ for all s > 0, and all positive integers n. Differentiating with respect to s and rearranging terms yields

$$(5.9) \qquad F'(s) = \left[\frac{b_1}{b_0}\right]^n F'((b_1)^n s) \text{ for all } n \quad .$$

Now let $n \to \infty$ in (5.9) and use the fact that $0 < b_1 < 1$ to obtain

$$F'(s) = F'(0^+) \lim_{n\to\infty} \left[\frac{b_1}{b_0}\right]^n = \lim_{n\to\infty} \left[\frac{b_1}{b_0}\right]^n \text{ for all } s > 0 \quad .$$

This last equation implies $b_1 = b_0$ and $F'(s) = 1$ for all s > 0. Together with $F(0^+) = 0$ one now has $\varphi_{\mu_1}^{-1}(\varphi_{\mu_1}(s)) = F(s) = s$ for all $s > 0$. Therefore, $\varphi_{\mu_1} = \varphi_{\mu_0}$ and $a_1(r) = a_0(r)$ for all r > 0. Finally, $\mu_1 = \mu_0$ by the uniqueness of the Laplace transform.

3. (Continuity) One can show that for any $\{\psi_i\}$ and ψ_* in Ψ, if $\psi_i \to \psi_*$ as $i \to \infty$, then $f_{\psi_i}(k,\delta,\{z(t)\}) \to f_{\psi_*}(k,\delta,\{z(t)\})$. First, note that with the given assumptions $\Psi \equiv P \times M$ is complete. P is complete because the intervals in which the components of π are located are closed and finite. M is complete because uniform integrability rules out limit measures which have total measure or mean not equal to one.

Assume that $\psi_i \to \psi_*$, i.e., the limit as $i \to \infty$ *of* $d(\psi_i,\psi_*) = 0$, where $\psi_i = (\pi_i,\mu_i)$. Now examine

$$| f_{\psi_i}(k,\delta,\{z(t)\}) - f_{\psi_*}(k,\delta,\{z(t)\}) |$$

$$= m(\{z(t)\})c(\delta \mid \{z(t)\}) \left| \int_0^\infty g_{\pi_i}(k \mid \delta,\{z(t)\}) d\mu_i(\theta) - \int_0^\infty g_{\pi_*}(k \mid \delta,\{z(t)\}) d\mu_*(\theta) \right|$$

(5.9)
$$\leq m(\{z(t)\})c(\delta \mid \{z(t)\}) \int_0^\infty \left| g_{\pi_i}(k \mid \delta,\{z(t)\}) - g_{\pi_*}(k \mid \delta,\{z(t)\}) \right| d\mu_i(\theta)$$

$$+ m(\{z(t)\})c(\delta \mid \{z(t)\}) \left| \int_0^\infty g_{\pi_*}(k \mid \delta,\{z(t)\}) d\mu_i(\theta) - \int_0^\infty g_{\pi_*}(k \mid \delta,\{z(t)\}) d\mu_*(\theta) \right| \quad .$$

Since the components of β and γ lie in finite intervals, one can show that $g_\pi(k|\delta,\{z(t)\})$ is jointly uniformly continuous in β, γ, and θ. Therefore, for each $\epsilon > 0$, there exists an $I(k,\{z(t)\})$ such that if $i > I(k,\{z(t)\})$, then

$$| g_{\pi_i}(k \mid \delta,\{z(t)\}) - g_{\pi_*}(k \mid \delta,\{z(t)\}) | < \epsilon \quad .$$

Consequently, the first term on the right-hand side of (5.9) is bounded above by $\epsilon m(\{z(t)\})c(\delta|\{z(t)\})$ for large i. The second term also vanishes as $i \to \infty$, by a fundamental theorem of measure theory[19] which states that if $\mu_i \to \mu_*$ weakly and g is bounded and continuous, then $\int g d\mu_i \to \int g d\mu_*$.

4. (Measurability) For any $\psi \in P \times M = \Psi$ and any $\rho > 0$, $w_\psi(k,\delta,\{z(t)\};\rho)$ must be a measurable function of $(k,\delta,\{z(t)\})$ where

$$w(k,\delta,\{z(t)\};\rho) = \sup_{\psi':d(\psi,\psi')<\rho} f_{\psi'}(k,\delta,\{z(t)\}) \quad .$$

N is separable implying that there exists a countable dense subset Ψ' of Ψ.

Now

$$\sup_{\psi':d(\psi,\psi')<\rho} f_{\psi'}(k,\delta,\{z(t)\}) = \sup_{\substack{\psi'':d(\psi,\psi'')<\rho,\\ \psi''\varepsilon\Psi'}} f_{\psi''}(k,\delta,\{z(t)\}) \quad ,$$

so that $w_\psi(k,\delta,\{z(t)\};\rho)$ is the supremum of a countable sequence of measurable functions and must also be measurable. The countable dense subset Ψ' can be taken to be $P' \times M'$ where P' is all (T+p) tuples with rational coordinates. M' is all distributions with finitely many points of increase with these points and the values of the distribution function taking rational values.

5. (Integrability) It is required that for any $\nu \in N$,

[19] See Billingsley [1979, p. 288].

$$\lim_{\rho \to 0^+} E_{\psi_0}\left[\ln\left[\frac{\underset{\psi':d(\psi,\psi')<\rho}{sup} f_{\psi'}(k,\delta,\{z(t)\})}{f_{\psi_0}(k,\delta,\{z(t)\})}\right]\right]^+ < \infty \quad .$$

But for some $(\tilde{\pi},\tilde{\mu}) \in \Psi$

$$E_{\psi_0}\left\{\ln \underset{\psi':d(\psi,\psi')<\rho}{sup} f_{\psi_0}(k,\delta,\{z(t)\}) - \ln f_{\psi_0}(k,\delta,\{z(t)\})\right\}^+$$

$$= E_{\psi_0}\left\{\ln\ m(z(t)) + \ln\ c(\delta\,|\,\{z(t)\}) + \ln\int_0^\infty g_{\tilde{\pi}}(k\,|\,\delta,\{z(t)\},\theta)d\tilde{\mu}(\theta)\right.$$

$$\left. - \ln\ m(\{z(t)\}) - \ln\ c(\delta\,|\,\{z(t)\}) - \ln\int_0^\infty g_{\pi_0}(k\,|\,\delta,\{z(t)\},\theta)d\mu_0(\theta)\right\}^+$$

$$\leq E_{\psi_0}\left\{-\ln\int_0^\infty g_{\pi_0}(k\,|\,\delta,\{z(t)\},\theta)d\mu_0(\theta)\right\} < \infty \text{ by assumption} \quad .$$

Comments on the Assumptions:

The theorem holds for a more general proportional hazards model. Let $\lambda_i(t) = \theta_i\lambda_o(t)\omega\{z_i(t)'\beta\}$

where ω is a known continuous and strictly increasing function and ω' is continuous. Also assume that $\omega(x) \to 0$ as $x \to -\infty$, and $\omega(x) \to \infty$ as $x \to \infty$. Then the conclusions of Proposition 2 hold.

Assumptions (A) and (B) of Proposition 2 are also used by Heckman and Singer. A simple example of a family of distributions satisfying (A) is one such that with probability one $0 < \in < \theta < \Delta < \infty$ uniformly for all elements of the family. Heckman and Singer also consider an alternative assumption to

(A) which restricts the tail behavior of $\ln\theta$. This approach requires quite lengthy proofs which are avoided here with some loss of generality. Cosslett [1983] and many others make assumptions analogous to (B). Assumption (B) is likely much stronger than needed. Intuitively, variation in the covariates besides the jth and for $k \geq 2$ is useful in identifying the parameters. Variation in the covariates over time also seems likely to be useful. Assumption (C) is the standard maximum likelihood regularity condition that the likelihood cannot equal negative infinity at the true parameter values.

6. A Reformulation of the Estimation Problem and a New Approach

A hazard model is equivalent to the nonlinear model

$$(6.1) \quad \int_0^{T_i} \lambda_i(u)du = \varepsilon_i$$

where ε_i is unit exponentially distributed.[20] Using the parameterization of the hazard given above, $\lambda_i(t) = \theta_i \lambda_0(t)\psi(z_i(t)'\beta)$, (6.1) can be rewritten as

$$(6.2) \quad \int_0^{T_i} \theta_i \lambda_0(u)\psi(z_i(u)'\beta)du = \varepsilon_i.$$

Dividing both sides by $\theta_i \lambda_0(0)$ and assuming $\theta_i > 0$, we have

$$(6.3) \quad \int_0^{T_i} \lambda_0'(u)\psi(z_i(u)'\beta)du = \eta_i, \text{ where}$$

[20]See Kalbfleisch and Prentice (1980), pp. 30-32.

$$\eta_i = \frac{\varepsilon_i}{\theta_i \lambda_0(1)}, \ \lambda_0'(u) = \lambda_0(1)^{-1}\lambda_0(u), \text{ and } \lambda_0'(1) = 1 .$$

Equation (3) can be used to understand several different estimation strategies. If the distribution of η is taken to be completely unrestricted, this model is even more general than a proportional hazards model. The approaches of Heckman and Singer (1984) and Section 5 can be thought of as estimating η restricting it to be a mixture of exponentials. If subjects are continuously observed so that exact failure times are known, then (3) may be estimable without imposing any distribution assumptions on η using the Kaplan-Meier product-limit estimator for the distribution of η in a way analogous to Horowitz (1987).

Alternatively, if one only observes subjects at discrete times, then following the approach used throughout the paper, one can write the probability of T_i being in a given time interval as

$$(6.4) \quad P[t \le T_i < t+1] = P\left[\int_0^t \lambda_0'(u)\psi(z_i(u)'\beta)du \le \eta_i < \int_0^{t+1} \lambda_0'(u)\psi(z_i(u)'\beta)du\right] .$$

The likelihood of the data can be written as a function of terms of the form (6.4) and terms of the form

$$(6.5) \quad P[t \le T_i] = P\left[\int_0^t \lambda_0'(u)\psi(z_i(u)'\beta)du \le \eta_i\right] ,$$

for the case where T_i is right censored. These expressions combined with the cumulative distribution function for η give a likelihood function that can be maximized as a function of β. One alternative would be to estimate the distribution of η using a flexible family of distributions. Here I describe the maximum likelihood estimate of the cumulative distribution function of η, $F(\eta)$, conditional on β. To estimate β will require iterating between estimates of F and β.

Turnbull (1974, 1976) provides an algorithm to obtain the maximum-likelihood estimate of the distribution function characterizing a random sample of observations indexed by i=1, . . . ,N for the case where we only observes the left and right endpoints, L_i and R_i, of the interval in which the ith observation lies. The algorithm uses the idea of self-consistency. Turnbull shows that the estimate satisfying the self-consistency criteria is the maximum-likelihood estimate.

In our case,

$$(6.6)\quad L_i = \begin{cases} \sum_{t=0}^{k_i-1} \alpha(t)\psi(z_i(t)'\beta) & \text{for } k_i \geq 1 \\ 0 & \text{for } k_i = 0 \end{cases}$$

and

$$(6.7)\quad R_i = \begin{cases} \sum_{t=0}^{k_i} \alpha(t)\psi(z_i(t)'\beta) & \text{for } \delta_i = 1 \\ \infty & \text{for } \delta_i = 0 \end{cases} ,$$

where $\alpha(t) = e^{\gamma(t)}$. To calculate the distribution function, define the set of disjoint intervals $[q_1, p_1]$, $[q_2, p_2]$, . . . ,$[q_m, p_m]$, where q_j is an element of the set of left endpoints $\{L_i\}$ and p_j is an element of the set of right endpoints $\{R_i\}$, and $[q_j, p_j]$ does not include any other element of $\{L_i\}$ or $\{R_i\}$ except at the endpoints. Also define $C = \bigcup_{j=1}^{m}[q_j, p_j]$. It can be shown that the maximum likelihood estimate of F, the cumulative distribution of η, is constant outside of C and is independent of the behavior of F within each interval $[q_j, p_j]$. Turnbull's algorithm consists of iterating between two steps until the estimates converge. Let s_j be the probability we assign to η lying in the interval $[q_j, p_j]$, and let $\alpha_{ij}=1$ if $[q_j, p_j]$ is contained in $[L_i, R_i]$, and 0 otherwise. The first step is to calculate the predicted probability that the ith observation lies in the jth interval $[q_j, p_j]$. This probability will be zero for all intervals which are not within $[L_i, R_i]$. The form of the predicted probability is just

$$(6.8)\quad \mu_{ij}(s) = \frac{\alpha_{ij}\, s_j}{\sum_{k=1}^{m} \alpha_{ik}\, s_k},$$

where $s=(s_1, s_2, \ldots, s_m)$.

The second step is to obtain an estimate of the probability mass in the jth interval by summing over all observations i the predictions of being in interval j from (6.8), i.e.

$$(6.9)\quad \pi_j(s) = \sum_{i=1}^{N} \mu_{ij}(s).$$

One iterates between these two steps until $s_j = \pi_j(s_1, s_2, \ldots, s_m)$. The initial estimates of s_j can be taken to be 1/m for all j.

To implement this approach, one needs to assign values for $F(\eta)$ over (p_j, q_j). A simple approach is to linearly interpolate between $F(p_j)$ and $F(q_j)$. This solution yields a likelihood function that is continuous and differentiable in β given F. It will also likely be more convenient to use the normalization $\alpha(0)=1$, rather than $\lambda_0(1)=1$.

[Simulations of this approach are underway.]

7. Testing

Coefficient estimates from the Prentice and Gloeckler approach can be used to test assumptions about the shape of the baseline hazard. Two approaches can be taken. One can use conventional nonlinear hypothesis testing on the estimated γ. Alternatively, one can test using $\pi \equiv (\beta,\gamma)$ following the specification test approach of Hausman [1978]. This second approach can be used in more situations.

The nonlinear hypothesis testing approach is easily described in the Weibull case. Let the null hypothesis be that the baseline hazard has the Weibull form, i.e., $\lambda_o(t) = \nu\varphi t^{\varphi-1}$. Now consider

$$R(t) = \frac{\log\left[\sum_{j=0}^{t+1}\exp\{\gamma_j\}\right] - \log\left[\sum_{j=0}^{t}\exp\{\gamma_j\}\right]}{\log(t+1) - \log(t)} \qquad t=1,\ldots,T-1$$

$$= \frac{\log\left[\nu(t+1)^{\varphi}\right] - \log\left[\nu t^{\varphi}\right]}{\log(t+1) - \log(t)}$$

$$= \varphi \text{ under } H_o \quad .$$

Let $$R = \begin{bmatrix} R(2) - R(1) \\ R(3) - R(2) \\ \vdots \\ R(T-1) - R(T-2) \end{bmatrix} .$$

Then $$\hat{R}' \left[\frac{\partial \hat{R}}{\partial \gamma'} Var(\hat{\gamma}) \frac{\partial \hat{R}}{\partial \gamma}\right]^{-1} \hat{R} \underset{a}{\sim} \chi^2_{T-2}$$

under the null hypothesis that the baseline hazard is Weibull.

Alternatively, the Hausman [1978] approach can be used to compare the estimates of π from the semiparametric method to those from maximum likelihood with the baseline hazard restricted to a given functional form. Under the null hypothesis that the functional form is correct, both estimators are consistent. Under the alternative, only the semiparametric estimates are consistent and the two sets of estimates will diverge. A specification test determines whether the differences between the two estimates are significant. Let $\hat{\pi}_N$ and $\hat{\pi}_M$ denote the estimates of π from the semiparametric approach and Maximum Likelihood, respectively. Then

$$\left(\hat{\pi}_N - \hat{\pi}_M\right)\left(Var\left(\hat{\pi}_N\right) - Var\left(\hat{\pi}_M\right)\right)^{-1}\left(\hat{\pi}_N - \hat{\pi}_M\right) \underset{a}{\sim} \chi^2_{p+T}$$

under the null hypothesis, where p is the dimension of β. This approach may be preferred because there may not be a simple set of restrictions available for the nonlinear hypothesis test. Furthermore, it allows one to examine if the differences between $\hat{\pi}_N$ and $\hat{\pi}_M$ are large in an economic sense.

A specification test can be performed using a subset of the parameters rather than $\pi \equiv (\gamma,\beta)$. β is often easier to use than the entire vector π, because $\hat{\beta}_M$ and $Var(\hat{\beta}_M)$ are directly calculated by maximization routines. On the other hand, $\hat{\gamma}_M \equiv \gamma(\hat{\zeta}_M)$, where $\hat{\zeta}_M$ are the parameters of the baseline hazard. $Var(\hat{\zeta}_M)$ must be calculated using a first order Taylor series approximation. While $\hat{\gamma}_M$ is more difficult to use, it may give a test with greater power than $\hat{\beta}_M$.

8. Simulations to Assess Bias and Relative Efficiency of Parametric Models

Simulations are useful to further assess the magnitude of the bias, but possible efficiency gain, from using a parametric baseline hazard. To study the results under a large number of alternative assumptions, I examine the limiting distribution of the estimator as the sample size grows. In particular, I focus on the probability limit of the estimator and the asymptotic standard error of the estimator under various assumptions. Since the discretely observed hazard model is just a multinomial model many aspects of this process are simplified. The limiting parameter estimates are obtained as the solution to the limiting first order conditions. The asymptotic standard errors are obtained from the limiting information matrix calculated as the limiting outer product of the gradients. These expressions can be written as functions of the T probabilities of being at risk at the beginning of an interval and the T probabilities of a spell ending during an interval.

The procedure is to draw N observations on the covariates from a specified distribution. In the case of time-varying covariates, N paths of the variables must be chosen. A shape must also be assumed for the true baseline hazard. Given this information and a censoring rule, the distribution of the observed data from this true model is completely characterized by $E[D_i(t)]$ and $E[R_i(t)]$, i=1,...,N. $E[D_i(t)]$ is the probability of a spell being observed ending in the tth interval, while $E[R_i(t)]$ is the probability of a spell lasting and being uncensored at the end of the tth interval. The index i indicates that these probabilities are conditional on the ith set of covariates.

Now, an assumed and estimated, but not necessarily true, model can be characterized by it survivor function for a given set of covariates $S_i(t)$. The resulting limiting log-likelihood can be written as

$$(8.1)\quad E[L] = \sum_{i=1}^{N} \sum_{t=0}^{T} \left\{ \left(E\left[R_i(t-1)\right] - E\left[R_i(t)\right] - E\left[D_i(t)\right] \right) \log S_i(t-1) + E\left[D_i(t)\right] \log\left(S_i(t-1) - S_i(t)\right) \right\}$$

$$= \sum_{i=1}^{N} \sum_{t=0}^{T} \left\{ \left(E\left[R_i(t-1)\right] - E\left[R_i(t)\right] - E\left[D_i(t)\right] \right) L_{1it} + E\left[D_i(t)\right] L_{2it} \right\}$$

where $E\left[D_i(t)\right] = E\left[D_i\left(t \mid \{z_i(\tau)\}\right)\right]$, $E\left[R_i(t)\right] = E\left[R_i\left(t \mid \{z_i(\tau)\}\right)\right]$, $L_{1it} = \log S_i(t-1)$

$L_{2it} = \log\left(S_i(t-1) - S_i(t)\right)$, $R_i(-1) = 1$, $R_i(T) = 0$, $D_i(T) = 0$,

$L_{1i0} = 0$, $L_{2i0} = \log\left\{1 - S_i(0)\right\}$.

The log-likelihood function in (8.1) is the sum over each of the N covariate paths of the expected contribution to the likelihood conditional on that covariate path. The parameter estimates obtained by maximizing this function will be the probability limit of the estimator from a sequence of samples that consist of a growing number of replications of these N values of covariates. The information matrix is calculated as the expected value of the outer product of the gradients. The standard errors are the square roots of the diagonal elements of the inverse of this expected information matrix. For example the block of the expected information matrix corresponding to β has the form

$$(8.2)\quad E\left[\frac{\partial L}{\partial\beta}\frac{\partial L}{\partial\beta'}\right] = \sum_{i=1}^{N}\sum_{t=0}^{T-1}\left\{\left(E\left[R_i(t-1)\right] - E\left[R_i(t)\right] - E\left[D_i(t)\right]\right)\frac{\partial L_{1it}}{\partial\beta}\frac{\partial L_{1it}}{\partial\beta'} - E\left[D_i(t)\right]\frac{\partial L_{2it}}{\partial\beta}\frac{\partial L_{2it}}{\partial\beta'}\right\} .$$

The rest of the matrix is calculated analogously. Note that this matrix only gives the inverse of the asymptotic variances when the assumed model includes the true one as a special case.

In the case where we have no heterogeneity in the assumed model and

$$\lambda_i(u) = \lambda_0(u)\exp\{z_i(u)'\beta\}$$

the expected log-likelihood simplifies to

$$(8.3)\ E\left[L(\gamma,\beta)\right] = \sum_{i=1}^{N}\sum_{t=0}^{T-1}\left\{E\left[D_i(t)\right]\log\left[1-\exp\{-h_i(t)\}\right] - E\left[R_i(t)\right]h_i(t)\right\},$$

where $h_i(t) = \exp\left\{\gamma(t) + z_i(t)'\beta\right\}$.

The first order conditions for maximization of the expected log-likelihood function are

$$E\left[\frac{\partial L}{\partial \beta}\right] = \sum_{i=1}^{N} \sum_{t=0}^{T-1} \left\{ E\big[D_i(t)\big] z_i(t) \frac{h_i(t)\exp\{-h_i(t)\}}{1-\exp\{-h_i(t)\}} - E\big[R_i(t)\big] z_i(t) h_i(t) \right\} = 0 \quad .$$

$$E\left[\frac{\partial L}{\partial \gamma(t)}\right] = \sum_{i=1}^{N} \sum_{t=0}^{T-1} \left\{ E\big[D_i(t)\big] \frac{h_i(t)\exp\{-h_i(t)\}}{1-\exp\{-h_i(t)\}} - E\big[R_i(t)\big] h_i(t) \right\} = 0, \qquad (t=0,1,\ldots,T-1), \quad and,$$

where the derivatives for a parametric baseline can be written as function of γ.

The expected log-likelihood function and its derivatives are calculated in the same way when the *assumed* model allows gamma heterogeneity, but the expressions are more complicated. If the assumed model has gamma heterogeneity then

$$L_{1it} = \log\left\{ \left[1 + \sigma^2 \sum_{\tau=0}^{t-1} h_i(\tau) \right]^{-\sigma^{-2}} \right\}$$

$$L_{2it} = \log\left\{ \left[1 + \sigma^2 \sum_{\tau=0}^{t-1} h_i(\tau) \right]^{-\sigma^{-2}} - \left[1 + \sigma^2 \sum_{\tau=0}^{t} h_i(\tau) \right]^{-\sigma^{-2}} \right\}.$$

In the simulations with heterogeneity in the true model, I generate the expected survivor function for a large number M draws from a heterogeneity distribution for a given set of covariates. Average the survivor function over these draws. Repeat this for each of the N paths of the covariates.

When the true model has unobserved heterogeneity the log-likelihood functions, first order conditions and expected information matrices are as above except now

$$E\left[D_i(t)\right] = M^{-1}\sum_{j=1}^{M} E\left[D_i\left(t \mid \left\{z_i(\tau)\right\},\ \theta_j\right)\right]$$

$$E\left[R_i(t)\right] = M^{-1}\sum_{j=1}^{M} E\left[R_i\left(t \mid \left\{z_i(\tau)\right\},\ \theta_j\right)\right]$$

where $\theta_1,\ldots,\theta_M$ are independent draws from the true heterogeneity distribution.

RESULTS OF THE SIMULATIONS

Table 1 provides a summary of some of the details of the simulations to assess the bias or efficiency of different estimators under various assumptions. In all of the simulations I assume that observations are only observed for 20 periods, so that ongoing spells are censored at that point. I use 100 values of the 3 covariates and all true coefficients are set equal to one. When there is unobserved heterogeneity in the model I draw 1000 values from the heterogeneity distribution for each set of covariates. The results are reported in Table 2.

EFFICIENCY COMPARISONS

These simulations compare the limiting standard errors of a parametric true model to those from the Prentice-Gloeckler approach. I try simulations both with and without unobserved heterogeneity and I distinguish between the coefficients on time constant and on time-varying covariates. The simulations without unobserved heterogeneity are numbered 2, 5, 11, 14, 28 and 31. In all cases you should compare the standard errors from these simulations to those from the immediately preceding simulation. The simulations with unobserved heterogeneity are numbered 20, 37, 39 and 41. Again, the standard errors should be compared to the preceding simulation (except 20 should be compared with 16). These results that for the time-constant covariates the efficiency loss is never substantial. For the time-varying covariates, the efficiency loss is sometimes substantial, though never extreme. This difference between time-constant and time-varying covariates was predicted by the analytical results of Section 3.

BIAS

The simulations focus on bias due to misspecifying the baseline hazard or the heterogeneity distribution. The results distinguish between biases in coefficients and in derivatives and between the coefficients on time-constant and time-varying covariates. The simulations which illustrate the biases due to misspecifying the baseline hazard are numbered 3, 6, 8, 9, 12, 15, 29, 32, 34, and 35. A consistent estimator of these coefficients is provided in the preceding simulation, or in some cases, the simulation two earlier. The simulations show that misspecifying the baseline hazard can lead to large biases in coefficients and derivatives of mean spell length with respect to the covariates. The possibility of extreme bias appears to be much greater for time-varying covariates. Examples of such biases can be seen in simulations 3 and 9 and to a lesser extent 12. The time-varying covariates are not always more biased though (see simulation 15). While a likely greater bias in time-varying covariate coefficients was not shown analytically earlier, it is quite intuitive. If the time pattern of the hazard is misspecified, it seems more likely to bias coefficients on variables that may have a time trend themselves.

The biases in the derivatives of mean duration are usually qualitatively similar to the biases in the coefficients, though not always. When one parameter is badly biased the derivatives for others are sometimes not approximately biased by the same proportion as the coefficients.

It is worth noting, that the case of a Weibull baseline and Gamma heterogeneity is special. Lancaster (1985) shows that derivatives of the logarithm of duration in a continuous model with no censoring are unbiased when heterogeneity is ignored or misspecified. Simulation 17 shows that this result is not even approximately true under the somewhat different assumptions here. In these simulations we examine duration not its logarithm, as well as a discrete model with censoring. I should also note that a misspecified baseline hazard may cause one to think there is no unobserved heterogeneity when in fact it is present (simulation 18).

The simulations also give some evidence on the biases due to not allowing for unobserved heterogeneity or misspecifying its distribution. Simulations 17 and 21 show substantial biases when unobserved heterogeneity was ignored (compare to simulation 16). However, simulations 22, 25, and 26 suggest that the choice of heterogeneity distribution used in the estimated model may not be crucial. These simulations show that a gamma distribution does quite well when the true distribution is multinomial, uniform or an exponentiated normal.

9. Conclusions

This paper describes an estimator for the proportional hazards model which allows an unknown form for the baseline hazard. The estimator avoids inconsistent estimation due to misspecification of the baseline hazard. Several kinds of censoring are allowed and discrete data of the form usually found in economics is used. Analytical expression indicate that the approach appears has a high efficiency relative to fully parametric models in many cases. The model readily allows for parametric unobserved heterogeneity. The paper also shows that one can consistently estimate the parameters of the model even when both the baseline hazard and the heterogeneity distribution are unknown. The paper describes a reformulation of the estimation problem and a tractable algorithm for calculating the maximum likelihood estimates. Simulations support the analytical result of high efficiency. The simulations also indicate the biases that can occur due to misspecification of the baseline hazard.

References

Bailey, K. (1984): "Asymptotic Equivalence Between the Cox Estimator and the General Maximum Likelihood Estimator of Regression and Survival Parameters," *Annals of Statistics*, 12, 730-736.

Billingsley, P. (1979): *Probability and Measure*. New York: John Wiley & Sons.

Blank, Rebecca M. (1989): "Analyzing the Length of Welfare Spells," Journal of Public Economics, 39, 245-273.

Cosslett, S. R. (1983): "Distribution-Free Maximum Likelihood Estimator of the Binary Choice Model," *Econometrica*, 51, 765-782.

Cox, D. R. (1972): "Regression Models and Life-Tables," *Journal of the Royal Statistical Society B*, 34, 187-202.

——— (1975): "Partial Likelihood," *Biometrika*, 62, 269-276.

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

Dabrowska, D., and K. Doksum (1985): "Partial Likelihood in Transformation Models with Censored Data," unpublished.

Diamond, P., and J. Hausman (1984): "The Retirement and Unemployment Behavior of Older Men," in H. Aaron and G. Burtless (eds.), *Retirement and Economic Behavior*. Washington: Brookings Institution.

Efron, B. (1977): "The Efficiency of Cox's Likelihood Function for Censored Data," *Journal of the American Statistical Association*, 72, 557-565.

Elbers, C., and G. Ridder (1982): "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model," *Review of Economic Studies*, 49, 403-409.

Feller, W. (1971): *An Introduction to Probability Theory and Its Applications*, Second Edition. New York: John Wiley & Sons.

Green, Jerry and John B. Shoven (1986): JMCB.

Han, A., and J. Hausman (1986): "Flexible Parametric Estimation of Duration and Competing Risk Models," Journal of Applied Econometrics, 5, pp. 1-28.

Harris, J. (1982): "Prenatal Medical Care and Infant Mortality," in V. Fuchs (ed.), *Economic Aspects of Health*. Chicago: University of Chicago Press.

Hausman, J. (1978): "Specification Tests in Econometrics," *Econometrica*, 46, 1251-1271.

Hausman, J., and W. Taylor (1981): "Panel Data and Unobservable Individual Effects," *Econometrica*, 49, 1377-1398.

Hausman, J., and D. Wise (1985): "Social Security, Health Status, and Retirement," in D. Wise (ed.), *Pensions, Labor, and Individual Choice*. Chicago: National Bureau of Economic Research.

Heckman, J., and G. Borjas (1980): "Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence," *Economica*, 47, 247-283.

Heckman, J., and B. Singer (1982): "The Identification Problem in Econometric Models for Duration Data," in W. Hildenbrand (ed.), *Advances in Econometrics*. Cambridge: Cambridge University Press.

——— (1984a): "A Method for Minimizing the Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, 271-320.

——— (1984b): "Econometric Duration Analysis," *Journal of Econometrics*, 24, 63-132.

——— (1984c): "The Identifiability of the Proportional Hazard Model," *Review of Economic Studies*, 51, 231-241.

——— (1986): "Econometric Analysis of Longitudinal Data," in Z. Griliches and M. D. Intriligator (eds.), *Handbook of Econometrics,* Volume III. Amsterdam: North-Holland.

Honoré, B. (1990): "Simple Estimation of a Duration Model with Unobserved Heterogeneity," **Econometrica**, 58, 453-473.

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

Kaplan, E. and P. Meier (1958): "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, 53, 457-481.

Katz, L. (1986): "Layoffs, Recall and the Duration of Unemployment," NBER Working Paper Number 1825.

Katz, Lawrence F. and Bruce D. Meyer (1990): “Unemployment Insurance, Recall Expectations, and Unemployment Outcomes,” *Quarterly Journal of Economics*, 105, November 1990, 973-1002.

Kiefer, J., and J. Wolfowitz (1956): "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters," *Annals of Mathematical Statistics*, 27, 887-906.

Kiefer, N., and G. Neumann (1981): "Individual Effects in a Nonlinear Model: Explicit Treatment of Heterogeneity in the Empirical Job-Search Model," *Econometrica*, 49, 965-979.

Lancaster, T. (1979): "Econometric Methods for the Duration of Unemployment," *Econometrica*, 47, 939-956.

Lancaster, T., and S. Nickell (1980): "The Analysis of Re-employment Probabilities for the Unemployed," *Journal of the Royal Statistical Society A*, 143, 141-165.

Maddala, G. (1971): "The Use of Variance Components Models in Pooling Cross Section and Time Series Data," *Econometrica*, 39, 341-358.

Meyer, Bruce D. (1990): "Unemployment Insurance and Unemployment Spells," Econometrica, 58, 757-782.

Moffitt, R. (1985): "Unemployment Insurance and the Distribution of Unemployment Spells," *Journal of Econometrics*, 28, 85-101.

Oakes, D. (1977): "The Asymptotic Information in Censored Survival Data," *Biometrika*, 64, 441-448.

Prentice, R., and L. Gloeckler (1978): "Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data," *Biometrics*, 34, 57-67.

Ridder, G., and W. Verbakel (1983): "On the Estimation of the Proportional Hazards Model in the presence of Unobserved Heterogeneity," mimeo, University of Amsterdam.

Singer, B. and S. Spillerman (1976): "Some Methodological Issues in the Analysis of Longitudinal Surveys," *Annals of Economic and Social Measurement*, 5/4, 447-474.

Stoker, T. (1986): "Consistent Estimation of Scaled Coefficients," *Econometrica*, 54, pp. 1461-1481.

Trussell J., and T. Richards (1985): "Correcting for Unmeasured Heterogeneity in Hazard Models Using the Heckman-Singer Procedure," in N. Tuma (ed.), *Sociological Methodology*. San Francisco: Jossey Bass.

Turnbull, Bruce W. (1974), "Nonparametric Estimation of a Survivorship Function with Doubly Censored Data," Journal of the American Statistical Association, 69, 169-173.

Turnbull, Bruce W. (1976), "The Empirical Distribution Function with Arbitrary Grouped, Censored and Truncated data," Journal of the Royal Statistical Society, A, 290-295.

van den Berg, Gerard and Jan van Ours (1996): "Unemployment Dynamics and Duration Dependence", forthcoming, Journal of Labor Economics.

Wallace, T., and A. Hussain (1969): "The Use of Error Components Models in Combining Cross Section with Time Series Data," *Econometrica*, 37, 55-72.

Table 1
True and Estimated Models Used in Simulations

20 time periods, 100 paths of 3 covariates (2 time-constant, 1 time-varying)

Coefficient on z_i, β_i=1, i=1,2,3.

Heterogeneity simulated with 1000 values for each set of covariates

True models

Baseline Hazards (BL)

Weibull: weibull λ=.1 p=1.2, $\lambda(t)=\lambda p(\lambda t)^{p-1}$
Log-logistic: log-logistic λ=.12 p=2.5, $\lambda(t)=\lambda p(\lambda t)^{p-1}/(1+(\lambda t)^{p})$ (non-monotone)
Piece-wise: piece-wise constant $\lambda(t)$= .08 for t=1,2,...,7,
.16 for t=8,9,...,13,
.08 for t=14,15,...,20

Covariates

z_1=ln(Gamma σ^2=.2)
z_2=Normal σ^2=.250
z_3 Trending up: z_3(t)=ln(Uniform(.5,1.5)+t**.2), uniform iid over t
z_3 Trending down: z_3(t)=ln(Uniform(.5,1.5)+t**-.2), uniform iid over t
z_3 Trending up, same: z_3(t)=ln(Uniform(.5,1.5)+t**.2), uniform the same each t

Unobserved Heterogeneity (UH)

None: degenerate
Gamma: gamma σ^2=.2
Multinomial: support at .3, .5 and 1.4, with probabilities .2, .2, and .6 respectively
Uniform: uniform (.4,1.6)
Exp(Normal): exp(Normal σ^2=1/9)
Gamma σ^2=1/3: gamma σ^2=1/3
Exp(Normal σ^2=.25): exp(Normal σ^2=.25)

Table 2
Limiting Standard Errors and Solutions to First Order Conditions

	True model	Estimated model		Estimate	S.E.	Average derivative
1.	BL: Weibull	True model	β_1	1.000	.237	-4.596
	z_3 trending up		β_2	1.000	.241	-4.596
	UH: none		β_3	1.000	.503	-4.596
2.	Same	BL: Prentice-Gloeckler	β_1	1.000	.238	-4.596
			β_2	1.000	.242	-4.596
			β_3	1.000	.507	-4.596
3.	Same	BL: Log-logistic	β_1	.937		-4.457
			β_2	.941		-4.478
			β_3	1.422		-6.781
4.	BL: Log-logistic	True model	β_1	1.000	.236	-4.070
	z_3 trending up		β_2	1.000	.240	-4.070
	UH: none		β_3	1.000	.469	-4.070
5.	Same	BL: Prentice-Gloeckler	β_1	1.000	.238	-4.070
			β_2	1.000	.242	-4.070
			β_3	1.000	.547	-4.070
6.	Same	BL: Weibull	β_1	1.050		-4.231
			β_2	1.057		-4.258
			β_3	.971		-3.911
7.	BL: Piecewise	BL: Prentice-Gloeckler	β_1	1.000	.241	-4.580
	z_3 trending up		β_2	1.000	.244	-4.580
	UH: None		β_3	1.000	.506	-4.580
8.	Same	BL: Weibull	β_1	.978		-4.614
			β_2	.983		-4.638
			β_3	1.040		-4.908
9.	Same	BL: Log-logistic	β_1	.921		-4.492
			β_2	.930		-4.535
			β_3	1.457		-7.109
10.	BL: Weibull	True Model	β_1	1.000	.265	-5.171
	z_3 trending down		β_2	1.000	.267	-5.171
	UH: None		β_3	1.000	.302	-5.171
11.	Same	BL: Prentice-Gloeckler	β_1	1.000	.266	-5.171
			β_2	1.000	.268	-5.171
			β_3	1.000	.303	-5.171
12.	Same	BL: Log-logistic	β_1	.955		-5.039
			β_2	.957		-5.049
			β_3	.874		-4.614
13.	BL: Log-logistic	True Model	β_1	1.000	.265	-4.643
	z_3 trending down		β_2	1.000	.269	-4.643
	UH: None		β_3	1.000	.285	-4.643
14.	Same	BL: Prentice-Gloeckler	β_1	1.000	.266	-4.643
			β_2	1.000	.270	-4.643
			β_3	1.000	.294	-4.643
15.	Same	BL: Weibull	β_1	1.037		-4.746
			β_2	1.043		-4.775
			β_3	1.004		-4.598

(continued)

Table 2 continued

	True model	Estimated model		Estimate	S.E.	Average derivative
16.	BL: Weibull	True model	β_1	.999	.315	-4.506
	z_3 trending up		β_2	.999	.317	-4.506
	UH: Gamma		β_3	1.000	.521	-4.510
16.	BL: Weibull	True model	β_1	.999	.315	-4.506
	z_3 trending up		β_2	.999	.317	-4.506
	UH: Gamma		β_3	1.000	.521	-4.510
17.	Same	BL: Weibull	β_1	.885		-4.442
		UH: None	β_2	.887		-4.456
			β_3	.974		-4.889
18.	Same	BL: Log-logistic	would not converge,			
		UH: Gamma	tended towards σ^2=0			
19.	Same	BL: Log-logistic	β_1	.847		-4.331
		UH: None	β_2	.853		-4.360
			β_3	1.253		-6.406
20.	Same	BL: Prentice-Gloeckler	β_1	.998	.357	-4.505
		UH: Gamma	β_2	.998	.356	-4.506
			β_3	1.000	.524	-4.513
21.	Same	BL: Prentice-Gloeckler	β_1	.876		-4.411
		UH: None	β_2	.879		-4.426
			β_3	.996		-5.018
22.	BL: Weibull	UH: Gamma	β_1	.996		-4.493
	z_3 trending up		β_2	.996		-4.493
	UH: Multinomial		β_3	.999		-4.507
23.	Same	BL: Prentice-Gloeckler	β_1	.995		-4.491
		UH: Gamma	β_2	.995		-4.491
			β_3	1.000		-4.515
24.	Same	BL: Prentice-Gloeckler	β_1	.979		-4.473
		UH: None	β_2	.979		-4.474
			β_3	1.000		-4.570
25.	BL: Weibull	BL: Weibull	β_1	1.003		-4.512
	z_3 trending up	UH: Gamma	β_2	1.003		-4.511
	UH: Uniform		β_3	1.001		-4.504
26.	BL: Weibull	BL: Weibull	β_1	.995		-4.561
	z_3 trending up	UH: Gamma	β_2	.996		-4.562
	UH: Exp(Normal)		β_3	.999		-4.575
27.	BL: Weibull	True Model	β_1	1.000	.238	-4.621
	z_3 trending up, same		β_2	1.000	.252	-4.621
	UH: None		β_3	1.000	.505	-4.621
28.	Same	BL: Prentice-Gloeckler	β_1	1.000	.240	-4.621
			β_2	1.000	.252	-4.621
			β_3	1.000	.506	-4.621
29.	Same	BL: Log-logistic	β_1	.945		-4.480
			β_2	1.001		-4.746
			β_3	1.376		-6.519
30.	BL: Log-logistic	True Model	β_1	1.000	.238	-4.090
	z_3 trending up, same		β_2	1.000	.253	-4.090
	UH: None		β_3	1.000	.484	-4.090

(continued)

Table 2 continued

	True model	Estimated model		Estimate	S.E.	Average derivative
31.	Same	BL: Prentice-Gloeckler	β_1	1.000	.239	-4.090
			β_2	1.000	.253	-4.090
			β_3	1.000	.540	-4.090
32.	Same	BL: Weibull	β_1	1.061		-4.296
			β_2	1.054		-4.267
			β_3	1.008		-4.080
33.	BL: Piece-wise	BL: Prentice-Gloeckler	β_1	1.000		-5.177
	z_3 trending up, same		β_2	1.000		-5.177
	UH: None		β_3	1.000		-5.177
34.	Same	BL: Weibull	β_1	.983		-5.141
			β_2	.999		-5.228
			β_3	.999		-5.225
35.	Same	BL: Log-logistic	β_1	.947		-5.024
			β_2	.963		-5.113
			β_3	.887		-4.708
36.	BL: Weibull	True Model	β_1	1.003	.340	-4.384
	z_3 trending up		β_2	1.003	.343	-4.383
	UH: Gamma σ^2=1/3		β_3	1.001	.529	-4.374
37.	Same	BL: Prentice-Gloeckler	β_1	1.005	.387	-4.384
			β_2	1.005	.386	-4.383
			β_3	1.001	.532	-4.364
38.	BL: Weibull	True Model	β_1	1.002	.390	-4.682
	z_3 trending down		β_2	1.002	.389	-4.682
	UH: Gamma σ^2=1/3		β_3	1.000	.318	-4.672
39.	Same	BL: Prentice-Gloeckler	β_1	1.003	.420	-4.683
			β_2	1.003	.416	-4.682
			β_3	1.000	.319	-4.669
40.	BL: Weibull	True Model	β_1	1.004	.342	-4.405
	z_3 trending up, same		β_2	1.003	.361	-4.405
	UH: Gamma σ^2=1/3		β_3	1.004	.662	-4.410
41.	Same	BL: Prentice-Gloeckler	β_1	1.005	.390	-4.406
			β_2	1.005	.402	-4.405
			β_3	1.006	.677	-4.407