ECONOMIC RESEARCH REPORTS

ESTIMATION OF THE SURVIVOR MODEL BY NONPARAMETRIC MAXIMUM LIKELIHOOD, MAXIMUM PENALIZED LIKELIHOOD AND SIMULATION BASED ESTIMATION

by

Keun Huh
and
Robin C. Sickles

RR #91-16 February, 1991

C. V. STARR CENTER
FOR APPLIED ECONOMICS

NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, N.Y. 10003

ESTIMATION OF THE SURVIVOR MODEL BY NONPARAMETRIC MAXIMUM LIKELIHOOD, MAXIMUM PENALIZED LIKELIHOOD AND SIMULATION BASED ESTIMATION*

Keun Huh
Economic Research Institute
Samsung Industries
Seoul

Robin C. Sickles
Department of Economics
Rice University
Houston

Revised February, 1991

*Research support was provided by National Institute on Aging Grant 1–R01–AG–05384–01. Computer resources for the NEC SX–2 supercomputer used in our analyses were made available through a grant from the Houston Area Research Center. Earlier versions and drafts of the paper were presented at the 1987 and 1989 Winter Meetings of the Econometric Society, the Third Conference on Panel Data, E.N.S.A.E., Paris, June, 1990, and the 6th World Conference of the Econometric Society, Barcelona, August, 1990. The authors would like to thank Siu Fai Leung, Bryan W. Brown, and participants in the microeconometrics seminar at the University of Georgia and the University of Rochester for useful comments. The usual caveat applies. The authors acknowledge support from the C. V. Starr Center for Applied Economics at New York University.

ABSTRACT

Standard treatments of heterogeneity components in typical longitudinal analyses can result in an incorrect parameterization of the survivor model. As a consequence, estimation bias is not limited to duration dependence but extends to the structural parameters as well. One approach to dealing with the heterogeneity components is to use a nonparametric mass point estimator to specify the marginal likelihood. We propose two additional methods to deal with this issue: maximum penalized likelihood estimation and simulation based estimation. Maximum penalized likelihood estimates the mixed joint density while smoothing the influences of unobserved heterogeneity and maximizing goodness of fit. Simulation based estimation maximizes the Pearson correlation between the simulated and observed frequencies of duration times based on axioms that describe the data generating process. It is computationally efficient for large panel data analyses with arbitrary forms of heterogeneity because it avoids closed–form expressions for the likelihood function. Monte Carlo experimental results indicate that these methods are computationally feasible and may provide attractive alternatives to the mass point estimator.

Keywords: survival model; unobserved heterogeneity; semi–nonparametric estimation; simulation based estimation.

Journal of Economic Literature classification number(s): 211, 841.

1.INTRODUCTION

The presence of an unobservable individual factor in modeling survivor hazards is a problematic confounding factor when the underlying hazard exhibits duration dependence [Lancaster and Nickell, 1980; Lancaster, 1979; Neyman and Scott, 1948]. Theoretical treatments [Simar, 1976; Laird, 1978; Lindsay, 1983a,b; Heckman and Singer, 1984; Manton, Stallard, and Vaupel, 1986] provided ways to control for unobserved heterogeneity in the context of a mixture probability density. With the exception of Manton et al., these authors assumed that unobserved heterogeneity was drawn from an unknown distribution which was independent of the observed variables. Two estimation approaches have distinguished themselves in the literature: the nonparametric approach of Heckman and Singer [1984] and the sufficient statistic method of Andersen [1970]. Heckman and Singer adopted a nonparametric method to identify an unobserved distribution from a mixed distribution assuming random effects, while Andersen used sufficient statistics to avoid the incidental parameters problem in assuming fixed effects. Andersen's application of conditional likelihood, which estimates structural parameters conditional on sufficient statistics for unobserved fixed effects, has had limited appeal due to the difficulty in finding the sufficient statistics for particular applications.

In this paper, we propose two new estimators for the survivor model with heterogeneity: Maximum penalized likelihood estimation (MPLE) and simulation based estimation (SIMEST). These and other alternatives recently have been used by Behrman, Sickles, and Taubman [1988, 1990], Behrman, Sickles, Taubman, and Yazbeck [1990], and Behrman, Huh, Sickles, and Taubman [1990] in studies of mortality that extended the morbidity studies outlined in Sickles and Taubman [1986] and Sickles [1989]. MPLE is a semi–nonparametric method based on the conditional likelihood method which has intuitive appeal and is relatively easy to compute. SIMEST is a computer intensive method based on the axioms which presumably govern the system's stochastic behavior and minimizes distance between observed and simulated sample frequencies.

The focus of our research is on estimators of compound processes. However, a number of authors have pointed out that focusing attention on the mixing heterogeneity distribution at the expense of a richly parameterized baseline duration distribution may have serious spefication error consequences [Behrman, Sickles and Taubman, 1989; Han and Hausman, 1990; Kiefer, 1988; Newman and McCullock, 1984; Ridder, 1986; Trussell and Richards, 1985]. The trade–offs between these sources of possible mispecification is a fertile research topic not addressed in our paper.

The paper is organized as follows. Section 2 presents the generic properties of the

survivor model in terms of a finite mixture continuous time stochastic process. Because of its ubiquitous use in applied work and because it is the only baseline distribution that can be viewed as either proportional hazard or accelerated time to failure model, we use the Weibull hazard model to motivate our theoretical discussion. Section 3 briefly outlines the mass point method and section 4 its extension to the finite mixture model considered by Heckman and Singer. Section 5 presents our alternative maximum penalized likelihood estimator and outlines algorithms that can be used to implement MPLE in survivor modeling. Section 6 shows how a simulation based estimator can be used to handle mixture densities based on axioms that govern the behavior of the mixture densities. Variants of this estimator, introduced by Lerman and Manski [1981] and first discussed in the survivor model context by Thompson, Atkinson, and Brown [1987], have recently been analyzed in depth by Gourieroux and Monfort [1989], McFadden [1989] and Pakes and Pollard [1989]. In section 7, we present the data generation design and results from a set of Monte Carlo experiments that assess the relative performance of the nonparametric maximum likelihood estimator (NPMLE), MPLE, and SIMEST estimators for the duration model with heterogeneity. Section 8 concludes.

2. THE WEIBULL PROPORTIONAL HAZARD FUNCTION

We begin with a discussion of the Weibull proportional hazard model incorporating heterogeneity. The functional form for the conditional hazard that we examine in this paper is a special case of Heckman and Singer's generalized Box–Cox form and is

$$(2\text{–}1) \qquad h(t_i \mid \underline{X}_i, \theta_i) = \exp(\gamma \ln t_i) \exp(\underline{X}_i \underline{\beta} + \theta_i)$$

for individual i=1,..,N with the log hazard function given by

$$(2\text{–}2) \qquad \ln h(t_i \mid \underline{X}_i, \theta_i) = \gamma \ln t_i + \underline{X}_i \underline{\beta} + \theta_i,$$

where t_i is the absolutely continuous time of a completed spell, $\underline{X}_i$ is an m–vector of strictly exogeneous and (possibly) time varying covariates, and where unobserved scalar heterogeneity is θ_i. Censored observations are given by

$$T_i = \min(t_i, t_c)$$
$$d_i = I(t_i < t_c),$$

where t_c is the censored time of an incomplete spell and I is a indicator function: $d_i = 1$ if $t_i <$

t_c and $d_i = 0$ otherwise. Assuming independence over individual duration spells, the joint likelihood of duration times and unobserved heterogeneity can be written as:

$$(2\text{–}3) \qquad L = \prod_{i=1}^{N} f(t_i,\theta_i|\underline{X}_i),$$

where

$$(2\text{–}4) \qquad f(t_i,\theta_i|\underline{X}_i) = \begin{cases} h(t_i,\theta_i|\underline{X}_i)\exp\left(-\int_0^t h(s_i,\theta_i|\underline{X}_i)ds\right), & \text{if } d_i = 1 \\ \exp\left\{-\int_0^t h(s_i\theta_i|\underline{X}_i)ds\right\}, & \text{if } d_i = 0. \end{cases}$$

The joint density is

$$(2\text{–}5) \qquad f(t_i,\theta_i|\underline{X}_i) = g(t_i|\ \underline{X}_i,\theta_i)\ \mu(\theta_i)$$

and the marginal likelihood of duration times, $f(t_i|\underline{X}_i)$, is

$$(2\text{–}6) \qquad L = \prod_{i=1}^{N} \int_{\Theta} g(t_i|\ \underline{X}_i,\theta_i)\ d\mu(\theta),$$

where $\theta \in \Theta$ in $\mathbb{R}$ and the conditional density, $g(\cdot)$ is the probability density function corresponding to (2–1) conditional on the distribution $\mu(\theta)$. The likelihood function (2–6) is a typical form of the statistical mixture model. The problem is how to control for the unobserved mixing distribution $\mu(\theta)$ [Lancaster, 1979; Lancaster and Nickell, 1980; Heckman and Singer, 1982,1984]. Standard approaches to the estimation of (2–6) require a parametric distribution on θ. However, if the density function $\mu(\theta)$ is specified parametrically, then estimation bias due to an incorrect parameterization of $\mu(\theta)$ is not limited to duration dependence effects, but extends to the structural parameters of included variables as well. Moreover, Heckman and Singer [1984] show that the problem of overparameterization can lead to the observational equivalence of two different sets of distributions.

A class of nonparametric estimators which can avoid the ad hoc specification of the mixing distribution $\mu(\theta)$ in (2–6) is the nonparametric MLE [Robbins, 1964; Laird, 1978; Lindsay, 1983a,b; Heckman and Singer, 1982, 1984]. The following assumptions will define

the estimation problem of (2–6) in terms of the estimation of a general finite mixture density over the closed interval [a,b].

Assumption 1. The marginal density of duration times, $f(\cdot)$, is bounded and continuous almost everywhere.

Assumption 2. The countable set of continuous functions $g_j(\cdot)$, j=1,..,k, is a linearly independent set over $\mathbb{R}$.

Assumption 3 The functional form of $g_j(\cdot)$ is known and is the same for j=1,..,k. Each $\theta_1,...,\theta_k$ is an element of the same parameter space Θ.

Assumption 1 allows the function f to be semicontinuous while Assumptions 2 and 3 are essential for the identifiability of the mixture distribution.

Definition 1. If $\mu(\cdot)$ denotes the probability measure over Θ defined by $\underline{p} = (p_1,...,p_k)$, then a finite mixture density function f is defined as

$$(2\text{–}7) \qquad f(\cdot) = \sum_{j=1}^{k} p_j g_j(t|\underline{X},\theta_j),$$

where $p_j = Pr(\theta = \theta_j) \geq 0$, $j = 1,...,k$, $\Sigma_j p_j = 1$, $\theta \in \Theta$.

3. THE MASS POINT METHOD

Suppose there are n independent sets of survival times $t_1,....,t_n$ with densities $g(t_i|\theta_1),...,g(t_i|\theta_n)$, i=1,...,n, where $\theta_1,...,\theta_n$ are unknown and regarded as a random sample from the prior $\mu(\theta)$. Then the marginal density of t_i, i=1,...,n, is (2–7). Robbins used $(t_1,...,t_{n-1})$ to estimate $\mu(\theta)$, and then used $\hat{\mu}(\theta)$ to estimate the posterior density $g(\cdot)$ of θ_n given t_n. Simar [1976] proved the existence and uniqueness of the maximum likelihood estimator of the mixed Poisson distribution. Laird [1978] extended Simar's and Robbins' methods to handle a general stochastic process model and proved the existence of a nonparametric representation under a self–consistency property. Lindsay [1983a,b] subsequently extended and generalized this to any mixture density and proved the existence of

a consistent estimate based on the optimality of the support of the mixture density.

The general maximization problem of the mixture density has a straightforward solution. From Definition 1, the maximum likelihood estimates become

$$(3\text{–}1) \qquad \max_{p} \ln L(\mu) = \sum_{i=1}^{n} \log \left[\sum_{j=1}^{k} p_j \, g(t_i | \theta_j) \right].$$

If $\hat{\theta}_k \in \Theta$ maximizes $g(\cdot)$, then (3–1) can be maximized by having all its mass at $\hat{\theta}_k$. Suppose that $\hat{\theta}_k$ has a point mass p_k with a mixture density $g(\cdot)$ and that $\theta^* = \{\theta_j,\ j=1,..,n,\ j \neq k\}$ has a probability mass $(1-p_k)$ with density $g^*(\cdot)$. Then the estimates are the solution to

$$(3\text{–}2) \qquad \max_{p_k} \ln L(\mu) = \sum_{i=1}^{n} \log \left[p_k \, g(t_i | \hat{\theta}_k) + (1-p_k) g^*(t_i | \theta^*) \right].$$

However, since the solution to (3–2) may not satisfy the condition $0 \leq p \leq 1$ [Hasselblad, 1966][1], Laird [1978] and Lindsay [1983a] proposed the mass point method by data grouping. Lindsay's characterization can be summarized by the following theorem.

Theorem 1. [Lindsay 1983a] Suppose that $g_\theta = (g(\tau_1 | \theta),...,g(\tau_k | \theta))$ where $\tau_1,...,\tau_k$, $k \leq n$, $\theta \in \Theta$, are the set of distinct values among sample data $t_1,...,t_n$, let $G = \{ g_\theta \}$, and let $c(G)$ denote the convex combinations of at most $(k+1)$ members of G. Then for any μ, (3–1) is a strictly concave function of $f(\tau_j | \mu)$, $j=1,..,k$. Furthermore, if G and $c(G)$ are compact, (3–1) has a unique maximum at $\hat{g}_\theta$ on the boundary of $c(G)$ and $\hat{g}_\theta$ can be estimated by k points of support.

The necessary and sufficient conditions for maximizing $\hat{g}_\theta$ are provided by the Gateaux variation.

Theorem 2. [Necessary and Sufficient Conditions; Lindsay 1983a] The estimate $\hat{\mu}(\theta_n)$ maximizes (3–1) if and only if $D[\hat{\mu}(\theta_n), \mu(\zeta)] \leq 0$ for all $\theta \in \Theta$ and therefore, θ_n is in the support of $\hat{\mu}(\theta_n)$ only if $D[\hat{\mu}(\theta_n), \mu(\zeta)] = 0$ where $D[\mu(\theta_1), \mu(\theta_2)] = \lim \varepsilon^{-1} \{\ln L[(1-\varepsilon)\mu(\theta_1) +$

$\varepsilon\ \mu(\theta_2)] - \ln L[\mu(\theta_1)]\}$ for some ε, and where $\mu(\zeta)$ denotes the degenerate probability measure which assigns unit mass to θ.

The EM algorithm is useful in solving the nested maximization problems involved in the solution to (3–1). From Bayes' theorem, the probability that θ_i comes from the l^{th} point of support is

$$(3\text{–}3) \qquad P_l(\theta_i \in \mu(\theta_l)) = \frac{g(t_i|\theta_l)\,\mu(\theta_l)}{\sum_{j=1}^{k} \mu(\theta_j)\ g(t_i|\theta_j)}, \quad l=1,....,k.$$

The MLE of $\mu(\theta)$, which is the expected value of $P_l(\theta_i \in \mu(\theta_l))$, is thus

$$(3\text{–}4) \qquad \hat{\mu}(\theta_l) = n^{-1} \sum_{i=1}^{n} P_l(\theta_i \in \mu(\theta_l)),\ l = 1,...,k.$$

Equation (3–4) is the E step. The maximization of the log likelihood (3–1) using the expected value from (3–4) yields the M step [Dempster, et al., 1977]:

$$(3\text{–}5) \qquad \sum_{i=1}^{n} P_l(\theta_i \in \mu(\theta_l)) \frac{\partial \log g_{\underline{\beta}}(t_i|\theta_l)}{\partial \underline{\beta}} = 0.$$

Then, given identifiability and under regularity conditions given in Lindsay [1983a], $\hat{\mu}(\theta_n)$ is a consistent estimate of $\mu(\theta)$.

4. HECKMAN AND SINGER'S NONPARAMETRIC ESTIMATION (NPMLE)

Heckman and Singer [1984] proposed a nonparametric method for the duration model to evaluate a mixture defined as (2–7), and, therefore, the integral in the mixture (2–6). Their approach is semi–nonparametric because the need to specify a parametric functional form for the mixing heterogeneity distribution $\mu(\cdot)$ can be avoided. Their approach was based in part on work by Kiefer and Wolfowitz [1956], Laird [1978] and Lindsay [1983a,b]. Lindsay's results are used to estimate a mixture density which is the marginal of t_i. The nonparametric

characterization of the mixture density $f(t_i|\underline{X}_i)$ takes the form :

$$(4\text{–}1) \qquad f(t_i|\underline{X}_i) = \sum_{j=1}^{k} g(t_i|\underline{X}_i,\theta_j)\, P_j,$$

where $\Sigma P_j = 1$, $P_j \geq 0$, j=1,...,k, k is the number of point of supports, P_j is a probability mass point and θ_j, j=1,...,k, is a locator of P_j such that $P_j = \text{Prob}(\theta=\theta_j)$. The loglikelihood of the marginal density (4–1) is given by

$$(4\text{–}2) \qquad \ln L = \sum_{i=1}^{N} \ln \sum_{j=1}^{k} g(t_i|\underline{X}_i,\theta_j)P_j.$$

For a given number of points of support $k \geq 1$, the likelihood function may have multiple local maxima unless $\mu(\theta)$ has a unimodal distribution. Theorem 2 provides conditions for a global solution to the maximization of (4–2). The Gateaux derivative, $D(\theta,\mu)$, of the log–likelihood function (4–2) with respect to θ is defined as

$$(4\text{–}3) \qquad D(\theta,\mu) = \sum_{i=1}^{N} \left[\frac{g(t_i|\underline{X}, \theta_j)}{f(t_i|\underline{X}_i)} - 1 \right].$$

The log likelihood function is maximized iff $D(\theta,\mu) \leq 0$, for all $\theta_i \in \Theta \subset \mathbb{R}$.

Suppose that for each individual i, the function $g(t_i|\underline{X}_i,\theta_i)$ has a unique mode $\theta_i \in \Theta$. Let θ_{min} be the minimum of $\{\theta_i\}_n$ and θ_{max} be the maximum of $\{\theta_i\}_n$. Then $g(t_i|\underline{Z}_i,\theta_i)$ should have its support in the domain $[\theta_{min},\theta_{max}]$. In order to find out what value of θ_i maximizes the joint densities of marginal survival times $f(t_i|\underline{X})$ for the duration model based on the conditional hazard (2–1), let $t_i^* = \int_0^{t_i} s^{\gamma}\exp(\underline{X}_i\beta)ds$ and $\exp(\theta_i) = \upsilon_i^*$. Rewrite the conditional survival function with respect to t_i^* as:

$$(4\text{–}4) \qquad S(t_i^*|\underline{X}_i,\theta_i) = \exp(-t_i^*\upsilon_i^*)$$

and thus the conditional density function of t_i^* will be

$$(4\text{–}5) \qquad f(t_i^* | \underline{X}_i, \theta_i) = \begin{cases} \upsilon_i^* \exp(- t_i \upsilon_i^*), & \text{if } d_i = 1 \\ \exp(- t_i^* \upsilon_i^*), & \text{if } d_i = 0. \end{cases}$$

This implies that $\upsilon_i^* = 1$ if the i^{th} observation is censored and thus if the observation is censored at the end of the sample period then $\theta_i = 0$ and $\theta_{min} = 0$. On the other hand, for an uncensored observation, the maximum occurs at

$$dg(t_i^* | \underline{X}_i, \theta_i)/d\upsilon_i^* = 0$$

and

$$(4\text{–}6) \qquad \exp(-t_i^* \upsilon_i^*) - \upsilon_i^* t_i^* \exp(-t_i^* \upsilon_i^*) = 0.$$

By solving (4–6) we have

$$\upsilon_i^* = \frac{1}{t_i^*}$$

where t_i^* is positive and bounded. Thus,

$$(4\text{–}7) \qquad \theta_i = -\ln [1 / (\int_0^{t_i} s^{\gamma} \exp(\underline{X}_i \underline{\beta}) ds)],$$

where

$$\theta_i < 0 \text{ if } \int_0^{t_i} s^{\gamma} \exp(\underline{X}_i \underline{\beta}) ds > 1,$$

$$\theta_i \geq 0 \text{ if } \int_0^{t_i} s^{\gamma} \exp(\underline{X}_i \underline{\beta}) ds \leq 1.$$

Therefore, the range of θ is defined on $[-\infty, \infty]$. However, the distributional form is unknown. Choosing the largest and smallest value of υ_i^* from the uncensored observations gives θ_{max} and θ_{min}, which are the upper and lower limits on θ. The more censored are the data we use, the more problematic may be identification of the tail distribution since the small number of

mass points around clustered observations cannot trace a possibly long–tailed heterogeneity distribution.

Having defined the interval $[\theta_{min},\theta_{max}]$, the EM algorithm [Dempster, et.al, 1977] used on (3–5) and (3–6) provides a method for the maximum likelihood estimation of the parameters in the mixture models. The algorithm is computationally slow. However, it is attractive in situations where methods for maximizing the likelihood over all parameters jointly are inadequate. Application to the heterogeneity model is achieved by treating the sequence of unobservables $\{\theta_i\}$ as missing data.[3] Consistency of the Heckman–Singer estimators rests on the assumption that the mixing distribution must be characterized by a finite number of mass points. As a practical matter the number of these must be small enough for their identification to be empirically feasible.

5. MAXIMUM PENALIZED LIKELIHOOD ESTIMATION (MPLE)

Maximum penalized likelihood estimation (MPLE) was introduced by Good and Gaskins [1971] and developed by de Montricher, Tapia and Thompson [1975], and Silverman [1982]. They consider the piecewise smooth estimation of an unknown density function by adding a penalty term to the likelihood which penalizes unsmooth estimates. The general form of a penalized loglikelihood under random sampling is given by

$$\log L = \sum_{i=1}^{N} \log f(x_i) - \alpha R\{f(x)\} \tag{5–1}$$

where $f(x)$ is an unknown density, $R\{f(x)\} < \infty$, R is a functional, and where α is the smoothing parameter. The choice of α controls the balance between smoothness and goodness–of–fit, while the choice of penalty functional R determines the type of behavior in the estimated density considered undesirable.[4] In this section we will consider the application of maximum penalized likelihood estimation to the hazard function with a mixing heterogeneity distribution defined in (2–5).[5] We point out that the penalty functional in our problem has both a classical and Bayesian motivation.

Assumption 4. Suppose that the random variables $(\theta_1,...,\theta_m)$ are a subset of $\theta=(\theta_1,..,\theta_n)$, with conditional density for duration given as $g(t_i|\underline{X}_i,\theta_j)$, i=1,...,n, j=1,..,m, m < n, where the density of the random variable $\mu(\theta)$ is unknown. Let $f(\cdot)$ be the empirical mixture density,

and let the functional form for the conditional density $g(\cdot)$ be completed by a finite vector of parameters. The empirical density $f(\cdot)$ is a mixture of a conditional density $g(\cdot)$ and a contaminating density $\mu(\theta)$.

Definition 2. Let the empirical distribution $f(\cdot)$ be semicontinuous (by Assumption 1) and the mixing distribution $\mu(\theta)$ be a dirac delta function. The modified mixture model is defined as

$$(5\text{–}2) \qquad f(t_i,\underline{X}_i,\theta_{j_i}) = g(t_i \mid \underline{X}_i,\theta_{j_i})\mu(\theta_{j_i}),\ i=1,..,n,\ j_i=1,..,m,$$

where $m < n$.

Next consider the joint likelihood of the conditional densities given by

$$(5\text{–}3a) \qquad L = \prod_{i=1}^{n} f(t_i,\underline{X}_i,\theta_{j_i})/\mu(\theta_{j_i}),\ i=1,...,n,\ j=1,..m,$$

with log likelihood function

$$(5\text{–}3b) \qquad \log L = \sum_{i=1}^{n} \log f(t_i,\underline{X}_i,\theta_{j_i}) - \sum_{i=1}^{n} \mu(\theta_{j_i}).$$

In order to estimate the parameters that characterize the joint likelihood of the conditional density $g(\cdot)$ represented in (5–3a) and (5–3b), prior knowledge on $\mu(\cdot)$ is necessary. The corresponding optimization problem using the log penalized likelihood function and a particular choice of the penalty function R, suggested by de Montricher, et al. [1975][6] is

$$(5\text{–}4) \qquad \log L = \sum_{i=1}^{n} \log f(t_i,\underline{X}_i,\theta_{j_i}) - \sum_{j=1}^{k} \alpha_j \| f_j^{(2)} \|^2$$

where the penalty function is the linear combination of the square of the norm of the second derivative with respect to the included covariates and which varies according to the correlation between the unobserved heterogeneity and the observed covariates. The functions (5–4) and (5–3b) converge as $\sum_{j=1}^{k} \alpha_j \| f_j^{(2)} \|^2$ becomes arbitrarily close to the density of heterogeneity

$\mu(\theta)$. Since the norm of the second derivative weighted by the smoothing parameter determines the roughness of densities while estimating the best fitting structural relationship by MLE, the prior density on unobserved heterogeneity $\mu(\theta_{j_i})$ in function (5–3a,b) can be approximated by the penalty term nonparametrically if some degree of roughness is considered unknown heterogeneity. Furthermore, the smoothing parameter α is a hyper–prior which sets the degree of roughness which can be interpreted as the prior density of heterogeneity. Thus MPLE can be regarded as quasi–Baysian estimation with hyper–prior α. In the rest of this section, we develop (5–4) to handle the problem of unobserved variables under different assumptions about temporal and cross–sectional sources of heterogeneity.

Case 1: Correction for Individual Specific Heterogeneity. We begin with a single covariate and heterogeneity that varies across individuals but stays constant through time. The maximum penalized likelihood estimates are then given by the solution to

(5–5)

$$\text{Max } L = \prod_{i=1}^{n} \left[h(t_i,x_i,\theta_{j_i}) \exp\left(-\int_0^{t_i} h(\tau,x_i,\theta_{j_i})d\tau\right)\right]^{d_i} \left[\exp\left(-\int_0^{t_i} h(\tau,x_i,\theta_{j_i})d\tau\right)\right]^{1-d_i} / \|h(t_i,x_i,\theta_{j_i})\|^2.$$

With the hazard function specified as (2–1), the loglikelihood function corresponding to (5–4) is

$$(5\text{–}6) \qquad J = \log L = \sum_{i=1}^{N} \left[d_i[\, \gamma \log t_i + x_i\beta + \theta_{j_i} \,] - \int_0^{t_i} \tau_i^{\gamma} \exp(x_i\beta + \theta_{j_i})\, d\tau \right]$$

$$- \alpha \sum_{j=1}^{m-1} \int_{x_j}^{x_{j+1}} \left[d^2 \{\, \tau_i^{\gamma} \exp(x_j\beta + \theta_j) \,\} / dx^2 \right]^2 d\tau,$$

where $j = 1,..,m-1$ denotes the order of bins evenly spaced in the interval $I = [x_{min}, x_{max}]$ and $m-1 \le N$. Equation (5–6) can be reexpressed as

$$(5\text{–}7) \qquad J = \log L = \sum_{i=1}^{N} d_i \log h(t,x,\theta) - \sum_{i=1}^{N} \int_0^{t_i} h(s_i,x_i,\theta_j)\, ds - \log \|h(t,x,\theta)\|^2.$$

Maximization of (5–7) is carried out subject to $h(x) \in S$, $h(x) \geq 0$, $\forall\, x \in (a,b)$, for a given set of points, $\{x_j\}$, j=1,..,m–1, where $S = \{v \mid v$ is continuous in (a,b), $v(a) = v(b) = 0$, $v^{(2)} \in L^2$, $v^{(2)}(a) = v^{(2)}(b)=0\}$.

The existence of a unique maximum of MPLE was proved by Tapia and Thompson [1978] in the Sobolev space. For our problem in (5–7), the following theorems are useful:

Lemma 1. The problem (5–7) has a solution if v is a subset of the Hilbert space.

Proof. Let $v_n = x^n$ on [0,1]. $v_n(1) = 1 \neq \|v\|_{L^2} = 0$. i.e. a sequence of $\{v_n\}$ converges in a particular norm but is not pointwise convergence since the point evaluation functional in L^2 is not continuous.□

Definition 3. A Hilbert space defined on (a,b) is a Reproducing Kernel Hilbert Space (RKHS) if for all $x \in (a,b)$ there exists M such that $|v(x)| \leq M \|v(x)\|$ for all $v \in H$.

Lemma 2. S is a RKHS.

Proof. $|v(x)| = |\int_a^x v'(x)dx| \leq \int_a^x |v'(x)|dx \leq \int_a^b |v'(x)|dx = \int_a^b 1\, |v'(x)|dx \leq \| 1 \| \|v'(x)\|_{L^2}$, by the Cauchy Schwartz inequality. Since $\| 1 \| \|v'(x)\|_{L^2} = (b-a)^{1/2} \| v \|_H$, then $|v(x)| \leq (b-a)^{1/2} \|v\|_H$.□

Lemma 3. Let **S** be a closed convex subset of a Hilbert space H^s. Let **J**:H→**R** be continuous and twice Gateaux differentiable in **S** and uniformly negative definite in **S**. Then J has a maximizer and is unique.

Proof. See Theorem 7 of Appendix I of Tapia and Thompson (1978).□

Theorem 3. The maximization of (5–7) has a unique solution.

Proof. From the definition of the norm induced by the inner product and by the linearity of the constraints, i.e., integration and the inequality, S is convex. Since integration is continuous, i.e. linear and bounded such that

$$|\int_a^b v(x)dx| = |x\, v(x)]_{x=a}^{x=b} - \int_a^b xv'(x)dx|$$
$$|\int_a^b xv'(x)dx|$$
$$\leq \|x\|_{L^2} \|v'\|_{L^2} = [(b^3-a^3)/3]^{.5} \|v\|_H,$$

and since S is RKHS, S is closed. J:H→R is continuous since point evaluation is continuous (by Lemma 2.) and the functions such as log, summation and the norm are continuous. Let $\Psi(t) = J(h+t\eta)$. Then

$$\Psi(t) = J(h+t\eta) = \log L(h+t\eta)$$
$$= \sum_{i=1}^{n} \log [h(x_i) + t\eta(x_i)] - \sum_{i=1}^{n} \int h(x_i)+t\eta(x_i)dx - \int_I \left[(h+t\eta)^{(2)}(x_i)\right]^2 dx.$$

The Gateaux derivatives of J at h(x) are given as

$$\Psi^{(1)}(t) = \frac{\sum_{i=1}^{n} \eta(x_i)}{h(x_i) + t\eta(x_i)} - \sum_{i=1}^{n} \int \eta(x_i)dx - \int_I 2\left[h^{(2)}(x_i)+t\eta^{(2)}(x_i)\right]\eta^{(2)}(x_i)\, dx \quad (5\text{–}8)$$

$$\Psi^{(2)}(t) = - \frac{\sum_{i=1}^{n} \eta^2(x_i)}{[\, h(x_i) + t\eta(x_i)]^2} - \int_I 2\left[\eta^{(2)}(x_i)\right]^2 dx$$

$$\Psi^{(2)}(0) = - \frac{\sum_{i=1}^{n} \eta^2(x_i)}{[\, h(x_i)\,]^2} - 2\| \eta \|^2 < - \| \eta \|^2, \text{ for all } \eta \neq 0 \text{ and are symmetric.}$$

Thus the second Gateaux variation is uniformly negative definite and therefore the maximizer exists and is unique.□

Theorem 4. The maximization of the norm can be solved by a natural cubic spline

interpolating $(x_1,h_1),....,(x_m,h_m)$, where m is the number of bins, subject to constraints such that

$$\text{Max } K(h) = \| h(t,x,\theta) \|^2 = \alpha \int_I \left[d^2 \{ h(t,x,\theta) \} / dx^2 \right]^2 d\tau$$

subject to

$$h \in C^2[x_{min},x_{max}]$$

$$h^{(2)}(x_{min})=h^{(2)}(x_{max})=0,$$

where $C^2[a,b]$ is the normed linear space which posses the second derivatives.

Proof. From (5–8) of Theorem 3, the first Gateaux derivative of K(h) at the direction η is given by

$$\begin{aligned} K^{(1)}(0)(\eta) &= -\int_I 2\, h^{(2)}(x_i)\eta^{(2)}(x_i)\, dx \\ &= -\sum_{i=1}^{m-1} \int_{x_i}^{x_{i+1}} 2\, h^{(2)}(x_i)\eta^{(2)}(x_i)\, dx. \end{aligned}$$

Integration by parts reveals that the right–hand–side is

$$= -\sum_{i=1}^{m-1} 2\, h^{(2)}(x_i)\eta^{(1)}(x_i) \Big|_{x_i}^{x_{i+1}} - \sum_{i=1}^{m-1} \int_{x_i}^{x_{i+1}} 2\, h^{(3)}(x_i)\eta^{(1)}(x_i)\, dx.$$

If $h^{(3)}(x_i)$ is constant, then the second term vanishes since

$$\begin{aligned} &\int_{x_i}^{x_{i+1}} \eta^{(1)}(x_i)\, dx = \eta(x_{i+1}) - \eta(x_i) = 0, \text{ and } \eta(x_i)=0 \ \forall\ i=1,...m-1 \\ &= -\sum_{i=1}^{m-1} 2\, h^{(2)}(x_i)\eta^{(1)}(x_i) \Big|_{x_i}^{x_{i+1}} \\ &= 2\, h^{(2)}(x_1)\eta^{(1)}(x_1) - 2\, h^{(2)}(x_m)\eta^{(1)}(x_m) \\ &= 0\ , \ (\text{since } h^{(2)}(x_1) = h^{(2)}(x_m) = 0). \end{aligned}$$

Therefore the solution must be a function where the third derivative is constant and satisfies the conditions of the problem.□

By Theorem 4, evaluation of the MPLE is computationally straightforward. For the norm of the second derivative, the function $f(\cdot)$ is in the class of generalized cubic splines. Therefore the interpolating problem is to fit a curve through the points $(x_i, f(x_i))$, i=1,...,n, in the plane. A mesh $\Lambda = \{\xi_1 (= x_1) < \xi < .. < \xi_m (= x_n)\ ;\ m \le n\ \}$ is chosen with the points ξ_j, j = 1,...,m, being the knots. For notational convenience, define the knots $\xi_j = x_j$, j = 1,...,m, such that x_j are evenly spaced in the interval (x_{min}, x_{max}) by m bins where t_j and θ_j, j=1,...,m, are the meshes coinciding with x_j for j=1,...,m. Then a function may be written as a cubic polynomial with mesh Λ in the interval (x_{min}, x_{max}) having certain discontinuities shown where the polynomials join, and the piecewise polynomial is chosen to minimize the mean square curvature. Consider now its evaluation in practice. Suppose the piecewise cubic polynomial interpolating $(x_j, h(x_j))$ and $(x_{j+1}, h(x_{j+1}))$, j=1,...,m, is given by

$$h(x_j) = a_j(x-x_j)^3 + b_j(x-x_j)^2 + c_j(x-x_j) + d_j \tag{5–9}$$

where x is a covariate with $\{x_j\}$, $x \in (a,b)$ and j=1,...,m–1. Using the methods in calculus of variation [Reinch, 1967] and by taking various order derivatives and evaluating at the knot points,

$$b_j = h^{(2)}(t_j,\theta_j,x_j) / 2, \tag{5–10}$$

$$a_j = \{\ h^{(2)}(t_{j+1},\theta_{j+1},x_{j+1}) - h^{(2)}(t_j,\theta_j,x_j)\ \} / \{\ 6\ (x_{j+1} - x_j)\}, \tag{5–11}$$

$$c_j = \frac{h(t_{j+1},\theta_{j+1},x_{j+1}) - h(t_j,\theta_j,x_j)}{x_{j+1} - x_j} - \frac{2(x_{j+1}-x_j)\ \{h^{(2)}(t_{j+1},\theta_{j+1},x_{j+1}) + h^{(2)}(t_j,\theta_j,x_j)\}}{6}, \tag{5–12}$$

$$d_j = h(t_j,\theta_j,x_j), \tag{5–13}$$

where $h^{(2)}(t_j,\theta_j,x_j)$ is the second derivative at knot points $(x_j, h(x_j))$. Thus the solution of

MPLE reduces to the evaluation of $h^{(2)}(t_j,\theta_j,x_j)$. With the relations (5–10) – (5–13), the equations relating $h^{(2)}(t_j,\theta_j,x_j)$ can be obtained based on the continuity of the first derivative of the spline. We have

$$(5\text{–}14)\qquad (x_j - x_{j-1})h^{(2)}(t_{j-1},\theta_{j-1},x_{j-1}) + 2\{(x_j - x_{j-1}) + (x_{j+1} - x_j)\}\, h^{(2)}(t_j,\theta_j,x_j) +$$

$$(x_{j+1} - x_j)h^{(2)}(t_{j+1},\theta_{j+1},x_{j+1}) =$$

$$\frac{h(t_{j+1},\theta_{j+1},x_{j+1}) - h(t_j,\theta_j,x_j)}{x_{j+1} - x_j} - \frac{h(t_j,\theta_j,x_j) - h(t_{j-1},\theta_{j-1},x_{j-1})}{x_j - x_{j-1}}.$$

The conditions, $h^{(2)}(x_1^-) = h^{(2)}(x_m^+) = 0$, lead the recurrence relation (5–14) to a tridiagonal system of linear equations for $h^{(2)}(x_1),\ldots,h^{(2)}(x_{m-1})$. The resulting system can be solved by Gaussian elimination. It follows from Theorem 3 and 4 that the MPLE (5–7) will have a unique solution.

The expression (5–7) deals with a single explanatory variable but can be extended to the multivariate case. In the multivariate case, (5–7) becomes

$$(5\text{–}15)\qquad J = \log L = \sum_{i=1}^{N} \left[d_i[\, \gamma \log t_i + x_i^T\beta + \theta_{j_i} \,] - \int_0^{t_i} \tau^\gamma \exp(x_i^T\beta + \theta_{j_i})\, d\tau \right]$$

$$- \sum_{l=1}^{k} \alpha_l \int_{X^l_{min}}^{X^l_{max}} \left[d^2 \{ \tau^\gamma \exp(x_i^l\beta + \theta_j) \} / d(x^l)^2 \right]^2 d\tau$$

where $x_i^T = (x_{1,i},\ldots,x_{k,i})$. The maximum exists and is unique for the multivariate case since Theorem 3 and 4 can be applied to a linear combination of separated penalty functions.

Case 2: Correction of Time Specific Heterogeneity[7]. Suppose that the omitted variable varies through time but remains constant across individuals. For i=1,...,n, $x \in R$ and $t \in (0, t_i)$, the estimates are solutions to

$$(5\text{–}16) \qquad \text{Max } J = \sum_{i=1}^{N} d_i \log h(t_i,x_i,\theta_t) - \sum_{i=1}^{N} \int_0^{t_i} h(\tau,x_i,\theta_t)\, d\tau \; - \| h(t,x,\theta) \|^2,$$

subject to $h \in S = \{ h \in H, h \geq 0, h^{(2)} \in L^2(0,T), h^{(2)}(0)=h^{(2)}(T)=0 \}$ or, with the hazard function specified as (2–1),

$$(5\text{–}17) \qquad \text{Max } J = \sum_{i=1}^{N} \left[d_i[\, \gamma \log t_i + x_i\beta + \theta_t \,] - \int_0^{t_i} \tau^{\gamma} \exp(x_i\beta + \theta_\tau)\, d\tau \right]$$
$$- \alpha \int_0^{T} \left[d^2 \{ \tau_i^{\gamma} \exp(x_i\beta + \theta_\tau) \} / d\tau^2 \right]^2 d\tau,$$

where $T = \max\{ t_i \}$. The penalty function uses the second derivative of the hazard function with respect to t. The second Gateaux variation at h in the direction of η is given by

$$(5\text{–}18) \qquad \Psi^{(2)}(0) = - \frac{\sum_{n} \eta^2(t_i)}{[\, h(t_i) \,]^2} - 2\| \eta \|^2 < - \| \eta \|^2, \text{ for all } \eta \neq 0.$$

The second Gateaux variation is negative definite and therefore a unique maximum exists for this problem.□ According to Theorem 4, (5–17) is the solution of the cubic spline function of time t.

Case 3: Correction of Individual and Time Specific Heterogeneity. Suppose that the omitted variable varies through time as well as across individuals. The application of MPLE in this case has the same intuition as Generalized Least Square estimation does in the linear model. where i=1,..,n, $t \in (0,T)$, $x \in (a,b)$. In this case the MPLE estimates are solutions to

$$(5\text{–}19) \qquad \text{Max } J = \log L = \sum_{i=1}^{N} d_i \log h(t_i,x_i,\theta_{it}) - \sum_{i=1}^{N} \int_0^{t_i} h(\tau_i,x_i,\theta_{it})\, d\tau - \| h(t,x,\theta) \|^2,$$

subject to $h \in S = \{ h \in H, h \geq 0, h_x^{(2)} \in L^2(a,b), h_x^{(2)}(a)=h_x^{(2)}(b)=0, h_t^{(2)} \in L^2(0,T),$ $h_t^{(2)}(0)=h_t^{(2)}(T)=0\}$. With the hazard function specified as (2–1), the MPLE are solutions to

$$(5\text{–}20)\qquad \text{Max } J = \sum_{i=1}^{N} \left[d_i[\, \gamma \log t_i + x_i\beta + \theta_{it} \,] - \int_0^{t_i} \tau_i^{\gamma} \exp(x_i\beta + \theta_{i\tau})\, d\tau \right]$$
$$- \alpha_1 \sum_{j=1}^{m-1} \int_{x_i}^{x_{i+1}} \left[d^2 \{ \tau_i^{\gamma} \exp(x_j\beta + \theta_{j\tau}) \} / dx^2 \right]^2 d\tau$$
$$- \alpha_2 \int_0^{T} \left[d^2 \{ \tau_i^{\gamma} \exp(x_i\beta + \theta_{i\tau}) \} / d\tau^2 \right]^2 d\tau,$$

where $T = \max\{ t_i \}$, $x \in (x_{min}, x_{max})$, $m-1 \le n$. The penalty function uses the second derivatives of the hazard function with respect to t and x.

Theorem 6. The problem (5–19) has a unique maximum.

The second Gateaux variation at h in the direction of (η_1,η_2) is given by

$$(5\text{–}21)\qquad \Psi^{(2)}(0)(\eta_1,\eta_2) = - \frac{\sum^{n} \eta_1^2(x_i)}{[\, h(x_i) \,]^2} - \frac{\sum^{n} \eta_2^2(t_i)}{[\, h(t_i) \,]^2} - 2\|\eta_1\|^2 - 2\|\eta_2\|^2 < - \|\eta_1,\eta_2\|^2 \text{ for}$$

all $\eta \neq 0$. Then the unique maximum exists.□

For these three classes of models with unobserved heterogeneity, MPLE avoids the problem of overparameterization of heterogeneity associated with parametric approaches because a parametric specification of the heterogeneity distribution is not needed. It also avoids the computational problem of alternatives such as the finite point of support probability estimator (mass point estimator) employed by Heckman and Singer because the heterogeneity distribution under MPLE is defined by the norm of a certain function. MPLE does not require a finite mass point. The basic idea in MPLE is to smooth heterogeneity from the likelihood by including penalty terms which take into account the degree of roughness or local variability which has not been controlled for by the covariates. MPLE is a versatile method for our purpose because the functional form of the penalty term R can be chosen according to various assumptions about the covariance structure of the unobserved heterogeneity whose parametric distribution is not specified. There has been very little work on the asymptotic properties of MPLE. However, it is clear that MPLE for our **Case 1** problem considered in the Monte Carlo experiments below is consistent as $\alpha/\sqrt{n} \to 0$ for bounded α, if the mixing distribution can be characterized by a finite number of supports and if the joint density in (5–4) is characterized

by (4–1). The reason is that the NPMLE and the MPLE converge to the same function for large N since the penalty term becomes negligible as estimates of unobserved heterogeneity become less and less rough.

For our numerical experiments, we have used the discrete minimization routine (ZXCGR) in the IMSL library since a step function approximation was employed in the interval (x_{min},x_{max}) divided by (m–1). The time interval was divided by a natural time unit for Case 2 and 3. When the subjective choice method for the hyper–prior α was introduced, we increased or decreased α until there was no significant pattern in $\{\theta_j\}$ and/or $\{\theta_t\}$ since if roughness in the function is not smoothed out, estimates of θ will exhibit wide fluctuations.

6. SIMULATION BASED ESTIMATION (SIMEST)

Monte Carlo approaches to probability calculations are well known in the area of computer simulation and have received recent interest in econometrics [Gourieroux and Monfort, 1989; McFadden, 1989; Pakes and Pollard, 1989]. As computing technology advances to handle bigger inputs with shorter processing time, computer intensive statistical methods have been introduced and developed to solve more complicated problems in stochastic process modeling. Simulating methods have many potential values but have not been widely used in econometric applications since approaches are based on frequency or density simulation [Lerman and Manski, 1981; Diggle and Gratton, 1984]. For example, the sequence of observations $\{x\}$ is used to construct an estimate of the true density, $\hat{f}$, and then as many independent realizations as required can be drawn from $\hat{f}$. However, the construction of $\hat{f}$ is not a easy task, especially with stochastic process models. Thus it may be desirable to simulate not from $\hat{f}$ itself but from the underlying true structure of the observed data.

In this section we will apply simulation based estimation (SIMEST) to our heterogeneity problem. It is a computationally efficient method to estimate stochastic processes since it does not require direct calculation of the probability densities. The probability density function is essential for any stochastic process model and its specification is problematic when there is no closed–form solution for the density functional. Thus it is a common practice to use ad hoc models though more formal stochastic process models are appropriate. SIMEST is based on axioms which presumably govern the data generating process and does not require closed form expressions for the likelihood. The concepts of the simulation based estimation method used herein were introduced by Atkinson, et al. [1983], Diggle and Gratton [1984] and Thompson, et al. [1987] and are summarized in Thompson [1989]. Here we will review the work by

Thompson, et al. and extend their procedure to a multivariate framework.

6.1. UNIVARIATE SIMULATION BASED ESTIMATION FOR THE WEIBULL PROPORTIONAL HAZARD MODEL

Suppose that we wish to estimate only duration dependence (γ) without covariates. The hazard function is

$$\lambda = t_i^{\gamma} \exp(\theta_i),$$

where γ is the duration dependence parameter, θ_i, i=1,...,n, is an unknown heterogeneity component and t_i, i=1,...,n, is time until failure.

Suppose the transition of states follows the Poisson process. According to the Poisson axioms, the probability that failure can occur in the time interval $[\ 0,t_i)$ is

$$\Pr[x(t+\Delta t)=1] = \Pr[x(t)=1]\ \Pr[x(\Delta t)=0] + \Pr[x(t)=0]\ \Pr[x(\Delta t)=1] + o(\Delta t).$$

Let the probability that one failure takes place in [t, (t+Δt)) be $\lambda\Delta t$ for every t in [0, t) and the probability that more than one failure happens in [t,(t+Δt)] is given by o(Δt) where $\lim_{\Delta t \to 0}$ o(Δt)/Δt = 0. Then

$$\Pr[x(t+\Delta t)=1] = \Pr[x(t)=1]\ (1-\lambda\Delta t) + \Pr[x(t)=0]\ \lambda\Delta t + o(\Delta t),$$

and

$$\frac{\Pr[x(t+\Delta t)=1\ ] - \Pr[x(t)=1]}{\Delta t} = \lambda\ \{\ \Pr[x(t)=0] - \Pr[x(t)=1]\ \} + o\Delta t/\Delta t.$$

Let $\Delta t \to 0$. Then

$$d\Pr[x(t)=1] / dt = \lambda\ \{\ \Pr[x(t)=0] - \Pr[x(t)=1]\ \}$$

and thus

$$\Pr[x(t)=1] = \lambda t \exp(-\lambda t) \qquad (6\text{–}1)$$

(6–2) $\Pr[x(t)=0] = \exp(-\lambda t)$.

The cumulative distribution function for at least one failure on or before t becomes

(6–3) $F(t) = 1 - \Pr[x(t)=0] = 1 - \exp(-\lambda t)$.

A common practice is to use maximum likelihood with a parametric specification for the heterogeneity distribution and the probability density function of failure, $f(\cdot)$. An alternative approach is nonparametric specification of the heterogeneity distribution but with the form of the density function $f(\cdot)$ required. We wish to estimate the parameter γ without specifying the probability density function.

For the univariate case, we can assume that time to failure for each individual is recorded as $t = (t_1 \leq t_2 \leq .....\leq t_n)$. Using this data we can divide the time axis into k bins, the m^{th} of which contains n_m observations. Having an initial value for the parameter γ, the simulation mechanism is employed to generate a large number of simulated failure times $s = (s_1 \leq s_2 \leq .....\leq s_N)$, where $N > n$. The simulation mechanism here is the cumulative distribution function (6–3). From (6–3), we can assume the probability of failure for a certain duration of time lies within [0,1] by the nature of the cumulative distribution. Then a random number u_i, i=1,...,N is generated from the uniform distribution. Using the generated numbers, the simulated time to failure s_i can be generated by equating u_i to (6–3). Let the number of simulated observations which fall into the m^{th} bin be $\hat{\nu}_{km}$. Then the simulated probability of the m^{th} bin becomes $\hat{P}_{km}(\gamma_0) = \hat{\nu}_{km} / N$. If the probability of the data at the corresponding bin is $P_m = n_m /N$, the natural criterion function is to minimize the distance between $\hat{P}_{km}(\gamma_0)$ and P_m. This turns out to be Pearson's goodness of fit. Thompson et al. suggest three possible criterion functions under the equal binning scheme.[8] However, Pearson's function is the only criterion that remains unchanged when, e.g. two cells are combined into a single cell. The goodness of fit is defined as

(6–4) $$S(\gamma_0) = \sum_{j=1}^{k} \frac{(\hat{P}_{kj}(\gamma_0) - P_j)}{P_j},$$

where k is the number of bins, $\hat{P}_{kj}(\gamma_0)$, $k \geq j$ is the simulated probability of j^{th} bin with estimated parameter γ_0. The function (6–4) is minimized when $\hat{P}_{kj}(\hat{\gamma}) = P_j$, j=1,...,k. Once the

criterion function (6–4) converges to a value $\hat{\gamma}$, confidence intervals for the true value of γ can be derived using a modified bootstrap method. Let $\hat{\gamma}$ be the value γ that minimizes $S(\gamma)$. The estimate $\hat{\gamma}$ is used to generate M simulated data sets of size n_m. Having simulated data $s_{ij}(\hat{\gamma})$, i=1,..,n, j=1,..,M, we then calculate the criterion function (6–4) for each simulated data set. Let $S_j(\hat{\gamma},s_n)$ be the value of (6–4) for the j^{th} simulated data set. The bootstrap mean and variance can be determined by

$$E(S(\hat{\gamma},s_n)) = \frac{1}{M}\sum_{i=1}^{M} S_j(\hat{\gamma},s_n)$$

$$Var(S(\hat{\gamma},s_n)) = \frac{1}{M}\sum_{i=1}^{M} \{ S_j(\hat{\gamma},s_n) - E(S(\hat{\gamma},s_n)) \}.$$

A 95% confidence interval for $S_j(\hat{\gamma},s_n)$ is given by

$$S_j(\hat{\gamma},s_n) = E(S(\hat{\gamma},s_n)) \pm \frac{2}{\sqrt{M}} Var(S(\hat{\gamma},s_n)).$$

Using $\hat{\gamma}$ as the center of a rotatable design, $S_j(\hat{\gamma},s_n)$ can be expressed as the quadratic curve such that $S_j(\hat{\gamma},s_n) = A + B\gamma + C\gamma\gamma$. Then the 95% confidence can be approximated by

(6–5)
$$E(S(\hat{\gamma},s_n)) - \frac{2}{\sqrt{M}} Var(S(\hat{\gamma},s_n)) \le A + B\gamma + C\gamma\gamma$$
$$\le E(S(\hat{\gamma},s_n)) + \frac{2}{\sqrt{M}} Var(S(\hat{\gamma},s_n)).$$

SIMEST can be an attractive method for the multivariate stochastic model. In Thompson's algorithm, the equal binning procedure is essential to establish asymptotic properties and is a necessary condition for minimizing the criterion function (6–4) without defining properties regarding the choice of bin width (k) as well as the number of simulated observations (N). His equal binning scheme sets the number of observations to be equal in each bin such that $P_j = 1/k$, j=1,...,k. This equal probability binning is particularly important to dampen the sensitivity of estimates to small perturbations of $\hat{\gamma}$ from the true value γ. When

equal probability binning is not possible, we need to define both the bin width and number of simulated observations in the light of Min Max var$\{P_j\}$, j=1,...,k.

6.2 MULTIVARIATE SIMULATION BASED ESTIMATION FOR THE WEIBULL PROPORTIONAL HAZARD MODEL

Next, suppose that the probability of failure follows the Poisson axioms and is conditional on a set of exogenous variables and duration time. Then the parameter λ of the Poisson process is given by

$$\lambda = t_i^{\gamma} \exp(\underline{X}_i\underline{\beta}_i + \theta_i), \qquad i= 1,...,n. \tag{6–6}$$

We wish to estimate the parameters $\delta=(\underline{\beta},\gamma)$ by SIMEST. This multivariate situation raises some complexity in simulation procedures. Due to the presence of heterogeneity, it is possible that λ in (6–6) is not monotone in $\underline{X}$ and t. The nonmonotonicity prevents us from using the equal probability binning procedure. If we cannot use the equal probability binning procedure, some modification is needed. First, it is necessary to introduce a robust procedure to minimize the variance of each bin probability ($\hat{P}_{kj}$, j=1,..,k) due to small changes in the value of estimates of δ. At the same time, we need enough bins to secure the identifiability of the SIMEST procedure. Second, we need a procedure to avoid empty bins which make the criterion function uninformative because $S(\delta) \to \infty$ if $P_j=0$. Alternatively, the criterion function ($S(\delta)$) has to be modified. Third, it is important to set a criterion to decide how many simulated observations, more precisely, how many replications for each fixed exogenous observation, are required. Because of the presence of fixed values of the exogenous variables, we need to decide the number of replications of simulated failure times (for each set of exogenous observations) by which the population distribution can be reflected. As a result, since the first and second problems are essentially the problem of bin width, multivariate SIMEST has to be considered in a multi–dimensional space with an appropriate bin width and a proper set of simulated observations.

Suppose we know the proper number of bins in terms of minimizing Mean Square Error (MSE) and have also decided on a number of replications to reflect the underlying population distribution. Without loss of generality, we will consider the three dimensional case (i.e. one covariate($\underline{X}$) and a duration time(t)). Let $t = \{t_i(\underline{X}_i)\}$, i=1,...,n, be failure time data conditional on exogenous variable $\underline{X}_i$, i=1,...,n, and let k_1 and k_2 be the number of bins dividing the time axis and the covariate axis, respectively. Then the set of vectors $Z = \{ (t_1,x_1),$

$(t_2,x_2),....,(t_n,x_n)\}$ lies on the real space R^2. Let m be the number of repeated simulations. Then simulated time to failure in the time axis will be $0 \le s_{11}(X_1), s_{12}(X_2), \ldots, s_{1n}(X_n)$, $s_{21}(X_1)$, $s_{22}(X_2)$, ..., $s_{2n}(X_n)$, ..., $s_{m1}(X_1)$, $s_{m2}(X_2)$, ..., $s_{mn}(X_n)$ with the corresponding value of exogenous variable $X = \{X_i\}$ in the covariate axis. The number of these simulated times and values of a covariate which falls into $(\ell_1,\ell_2)^{th}$ bin will be denoted by $V_{\ell_1\ell_2}$, where $\ell_1 = 1,...,k_1$, $\ell_2 = 1,...,k_2$. If δ is close to the true value, then the simulated bin probability

$$\hat{P}_{\ell_1\ell_2}(\delta) = \frac{V_{\ell_1\ell_2}}{m \times n} \qquad (6\text{–}7)$$

should approximate the corresponding portion of data (time and a covariate) in the same bin,

$$P_{\ell_1\ell_2} = \frac{n_{\ell_1\ell_2}}{n}. \qquad (6\text{–}8)$$

Therefore $P_{\ell_1\ell_2} = \hat{P}_{\ell_1\ell_2}(\delta)$, $\ell_1 = 1,...,k_1$, $\ell_2 = 1,...,k_2$ if δ is close to the true value. However, the asymptotic value of $\hat{P}_{\ell_1\ell_2}(\delta)$ can only be achieved by increasing the number of observations, n. Since $V_{\ell_1\ell_2} = m(n_{\ell_1\ell_2})$ if δ is close to the true value, where $\ell_1 = 1,...,k_1$, $\ell_2 = 1,...,k_2$, then

$$\hat{P}_{\ell_1\ell_2}(\delta) = \frac{V_{\ell_1\ell_2}}{l(n)} = \frac{l(n_{\ell_1\ell_2})}{l(n)} = \frac{n_{\ell_1\ell_2}}{n}. \qquad (6\text{–}9)$$

Equation (6–9) indicates that the number of replications has no effect on the asymptotic value of $\hat{P}_{\ell_1\ell_2}(\delta)$. Rather $\hat{P}_{\ell_1\ell_2}(\delta)$ has a limiting value as the number of observations n goes to infinity (McFadden, 1989).

Another unsolved question is how to decide the number of bins, k_1 and k_2, to prevent outlier sensitivity. Although a number of data–based density estimators exist, the histogram method is picked here because of its wide–spread use and its asymptotic efficiency. Scott [1979] proposed an optimal histogram method in terms of minimizing integrated mean square error (IMSE). Following Scott's approach, the sum of integrated variance(IV) and integrated bias(IB) is defined as integrated mean square error (IMSE)

$$IMSE = \int_{I(x)} \{Var\, f_k(x) + E(f_k(x)) - f_k(x)\}\, dx$$

where $x \in I(x)$, k is the number of bins, $f_k(x)$ is the probability in the k th bin. IV can be approximated by

$$IV = \sum_k \frac{P_k(1 - P_k)}{nh_k}$$

$$= \frac{1}{nh_k} - \sum_k \frac{P_k^2}{nh_k}$$

since P_k is the probability of the k th bin such as $P_k = \int_{I_k(x)} 1\, f(s)\, ds$ and $\Sigma P_k = 1$, h_k is the k^{th} bin width. By the Taylor series expansion of P_k,

$$IV = \frac{1}{nh_k} - \frac{\int f_k^2(x)\, dx}{n} + o(n^{-1})$$

and Integrated Bias is expressed by

$$IB = \sum_k \{\frac{P_k}{h_k} - f(x_k)\}.$$

By the generalized mean value theorem, if f is Lipschitz continuous, the integrated bias becomes

$$IB = (1/12)\, h^2 \int (d\hat{f}(x)/dx)^2 dx.$$

Then the minimization of IMSE gives us the optimal bin width, $h^* = [\frac{6}{n\, R(df/dx)}]^{(1/3)}$, where R(.) is the roughness of f. In multidimensional spaces, h^* is approximated by

$$h_i^* = 3.5\, s_i\, n^{-(1/(2+d))},$$

where i denotes each axis in the multidimensional space, s_i is the standard deviation, n is the number of observations. Scott and Thompson [1983] suggest to choosing $h_i^* = 2\, s_i\, n^{-(1/(4+d))}$, which has the convergency rate $o(n^{-2/3})$, by minimizing the frequency polygon.

Finally we need a minor modification of the criterion function since the presence of

empty bins will make Pearson's goodness of fit criterion (6–4) uninformative. To prevent this, the modified Pearson goodness of fit using the results of (6–8) and (6–9) is given by

$$S_m(\delta) = \begin{cases} \sum_{\ell_1=1}^{k_1}\sum_{\ell_2=1}^{k_2} \dfrac{(\hat{P}_{\ell_1\ell_2}(\delta) - P_{\ell_1\ell_2})^2}{P_{\ell_1\ell_2}}, & \text{if } \hat{P}_{\ell_1\ell_2}(\delta), P_{\ell_1\ell_2} \neq 0 \\ \sum_{\ell_1=1}^{k_1}\sum_{\ell_2=1}^{k_2} \dfrac{(\hat{P}_{\ell_1\ell_2}(\delta) - P_{\ell_1\ell_2})^2}{\hat{P}_{\ell_1\ell_2}(\delta)}, & \text{if } \hat{P}_{\ell_1\ell_2}(\delta) \neq 0, P_{\ell_1\ell_2} = 0 \\ 0, & \text{otherwise.} \end{cases}$$

The modified minimization criterion substitutes the observation probability with the simulated probability when the observed probability of a certain bin is zero. This may be possible since the simulated probability should approximate the observation probability if the estimate of parameter δ is close to the true value. The criterion is also minimized when $P_{\ell_1\ell_2} = \hat{P}_{\ell_1\ell_2}(\delta)$, $\ell_1 = 1,...,k_1$, $\ell_2 = 1,...,k_2$. Once the parameter $\hat{\delta}$ estimated, confidence intervals for the true value of the parameter δ can be derived as in (6–5). Consistency and asymptotic normality of the simulation based estimator for large N and M are proven in Lerman and Manski [1981]. McFadden (1989) and Pakes and Pollard (1989) prove similar results for alternative simulation estimators when the number of simulations (M) is finite.

7. MONTE CARLO RESULTS

7.1 DESIGN OF EXPERIMENTS AND DATA GENERATION

We consider the Weibull proportional hazard model defined in (2–1) in which observed data are generated as realizations of the stochastic process

$$h_i = t_i^{\gamma} \exp(\beta_0 + \underline{X}_i\underline{\beta} + \theta_i), \quad i = 1,...,n,$$

where $t_i \geq 0$, $\underline{X}_i=(x_{1i}, x_{2i})$, $\underline{\beta} = (\beta_0,\beta_1,\beta_2)$ and where θ_i is an unobserved stochastic process defined on a complete probability space. Heterogeneity, θ_i, need not be i.i.d., but for our experiments we assume that it is. The artificial samples are generated by the following

procedures. First, we draw a uniform random variable u_i in the interval [0, 1] and generate heterogeneity θ_i according to the implicit function, μ , where

$$\theta_i = \mu^{-1}(u_i), \quad i = 1,...,n$$

and where μ^{-1} is inverse of an appropriate cumulative probability function. Next, we draw values of two exogenous variables $\underline{X}_i=(x_{1i}, x_{2i})$ from a standard normal random number generator. Another uniform random number in the interval [0, 1] is drawn for the survival function $S_i = (1-F(.))$. We then solve for the implied duration t_i from the survival function with given values of parameters, $\underline{\beta}$ and γ. Thus

$$(7-1) \qquad t_i = \exp\left[\{\ln(-\ln S_i) + \ln(\gamma+1) - (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \theta_i)\} \frac{1}{\gamma+1}\right].$$

Different mixing distributions for the heterogeneity are drawn to compare the performances of the different estimators. We use the standard normal as our unimodal contamination. In addition, a bivariate normal distribution representing a bimodal heterogeneity distribution, and a multinomial distribution representing multimodal distribution are also employed. Given the duration t_i from (7–1) with true parameter value $\beta_0=0.1$, $\beta_1=\beta_2=\gamma=1$, the right censored times T_r are set to ensure that about fifteen percent of observations are censored. We increase the censoring rate to twenty percent. When left censoring is allowed, censoring times T_l are arbitrarily set to be 1 time–unit since the mean duration of t_i is 3.43 time units. As a result, about 25% were censored. Samples of 100, 500 and 1000 are used. These are in the range of sample sizes for the bulk of empirical duration studies.

Computing algorithms were developed in Fortran77. In addition to the computing source codes, STEPIT of Chandler [1969] is employed as the maximization method for SIMEST. STEPIT is useful for the SIMEST procedure because when the steps oscillate it detects the fashion of zigzags and shortcuts the optimizations. The minimization routine ZXCGR in the IMSL library was used for MPLE. We also use the computer code of CTM documented by Yi, et al. [1987] for Heckman and Singer's NPMLE, which is based on the EM algorithm of Dempster, et al. [1977].

The basic logic of the simulation based estimator in this context is rather straightforward. Suppose that sample observations for 2 covariates and duration variable t are given by the data generation procedure described above. The simulation algorithm in each replication is as follows:

Input initial values for the parameters β_0, β_1, β_2, γ,

Repeat until $t_i^* > 0$, where t_i^* is a simulated time,

Generate θ_i from U(0,1),

Generate a simulated time t_i^* through the survival function,

if $t_i^* > 0$, then discard,

End repeat,

Return t_i^*.

The simulation algorithm can be easily adapted for more complicated models. Finally, the method employed to choose the smoothing parameter α of MPLE is the subjective choice method [Bartoszynski, et al., 1981]. We attempted to adapt the cross–validation method by minimizing

$$CV(\alpha) = n^{-1} \sum_{i=1}^{n} (t_i - t_{\delta\alpha})^2,$$

where $t_{\delta\alpha}$ is the inverse function of the hazard function, h, such as $F^{-1}(\underline{x}_i,h ; \delta, \alpha)$ and where $\underline{x} = (x_1,x_2)$, $\delta = (\beta_1, \beta_2, \gamma)$. However, the evaluation of $CV(\alpha)$ is too computationally burdensome even for pseudodata sets of size 100 because to find a maximum it requires no less than: the number of function evaluations x the number of observations x the number of function evaluations with new α. In a typical case, about 14700 iterations were needed for pseudodata sets of size 100. The adaptation of cross–validatory methods to our model merits further investigation.

7.2 COMPARISONS AMONG DIFFERENT ESTIMATORS

Typical outcomes of our Monte Carlo experiments are shown in Tables 1–13. These results are suggestive of some possible discrepancies among the different estimators in different cases but also suggest substantial comparability between them when the underlying stochastic process is not too complicated and has been correctly modeled. Table 1 presents results based on the three estimators when there is no censoring and heterogeneity is drawn from a standardized normal distribution. Both the duration parameter and structural parameters are estimated well for all three estimators. We began to estimate NPMLE using 2 points of support to identify the heterogeneity distribution. Since the standard normal distribution has a mass point at 0.0, one point of support locates at –3.0 with cumulative

probability 0.0 and the other point is set to locate at 0.0 with the expected cumulative probability 0.5. However, the sampling distribution of the NPMLE appears to be much less biased when 4 points of support are used [$(\theta,\mu(\theta))$ = { (.012,.276), (.232,.343), (1.22,.760) }][9]. Standard deviations from SIMEST are calculated by the bootstrap method with 30 replications. The bin width for SIMEST is based on the expression introduced by Scott [1979] which chooses an optimal bin width by minimizing the integrating mean square error (IMSE) of the multidimensional histogram. In this case with the sample size of 500, the number of bins is six[10]. Samples of size of 1000 and 100 require seven bins and four bins for each dimension, respectively. However, SIMEST continually converges to a local optimum when starting values quite different from the true values are used. For example, when starting values are (β_1,β_2,γ) = (.4,.4,.4), our estimates are (.532, .613, .276). Results for SIMEST in Table 1 are based on starting values of (.8,.8,.8). For MPLE, the smoothing parameters, α_i, i=1,2, are chosen by the subjective choice method [Bartoszynski, et al., 1981]. We start from α_i=10 for i=1,2. For the purpose of comparison, $\| h^{(2)}(x)\|$, the norm of the second derivative of hazard function with respect to $\underline{X}$, was used as the penalty function[11]. After searching seven times for the optimal value of α^* using the starting value of α = 10, we found the α^* for which the $\{\theta_j\}$, j=1,..,m, do not exhibit significant fluctuations[12]. When α was chosen between .6 and .4, there were no significant differences in both estimates and values of $\{\theta_j\}$. It is possible to choose the different values for each smoothing parameter but the same value (.5) was chosen since we generated the pseudodata for $\underline{X}$ from the same distribution. Finally, the number of bins to calculate derivatives was 10 in the interval $[\underline{x}_{min},\underline{x}_{max}]$.

We next assume that there is right censoring after 5 time–units, which censors about 15% of the sample observations. As seen in Table 2, NPMLE and MPLE slightly underestimate the true values. However, the degrees of underestimation for the structural parameters is greater with NPMLE than MPLE. On the other hand, SIMEST overestimates the duration dependence parameter, but estimates the structural parameters very well.

The principal findings of our experiments are reported in Tables 3–13. First, both MPLE and NPMLE perform poorly in small samples while SIMEST performs relatively well. As the number of observations increases to 500 and more, both MPLE and NPMLE begin to track the underlying stochastic model, in contrast to SIMEST whose stochastic axioms are at variance with the data generation process and thus should not be expected to perform well asymptotically (see Tables 1,3,4).

Second, as we increase the proportion of censored observations NPMLE looses any advantage over MPLE. Of the three methods, SIMEST appears to be quite robust. However,

when left and right censoring coexist, SIMEST also becomes unstable (see Table 2, 5 and 6). Unfortunately since our version of the CTM program -- the computing source codes for the NPMLE – does not appear able to handle left censored data, comparisons with NPMLE are not available.

Third, Table 7 reports how the choice of the smoothing parameter affects parameter estimates from MPLE. Two smoothing parameters are chosen subjectively and are used to estimate MPLE. One is chosen to be 0.2, which is smaller than α =0.5. With this choice of α, estimates tend to be biased downward. The result is expected because the estimates of MPLE should be the same as those of QMLE if α = 0. Furthermore, when we select α to be 1.0, which is greater than the best choice of α, estimates are also underestimated due to oversmoothing.

Fourth, the choice of the bin width for SIMEST is quite essential, especially for the multivariate nonlinear function–fitting problem, because the estimates become unstable as the chosen bin width differs from the optimal bin width (Table 8).

Fifth, Table 9 and 10 demonstrate the results when the heterogeneity distribution is drawn from a bimodal and multimodal distribution. MPLE and NPMLE performed well when actual heterogeneity is not unimodal. However, NPMLE has mass points at (.274, .783, .823) for a bimodal distribution and shows all negative directional derivatives. For the multimodal distribution, 4 points of support appear adequate. These results, as well as those with the unimodal distribution, suggest that the mass point method employed by NPMLE has difficulty reflecting the true distribution of heterogeneity and that the choice of optimal supporting points requires further research.

Sixth, we investigated the predictive power of NPMLE, MPLE and SIMEST with different true parameter values under the complete data with 500 observations. Tables 11, 12 and 13 summarize the results of three different cases. Evidence indicates that these three estimators have substantial predictive power.

Finally, we repeat the same experiment as many as 100 times for different seed values and then use the pseudodata set for each estimator under different censoring schemes and true values. The results are shown in Table 14 to 19. The Monte Carlo results are quite stable, except those for SIMEST, and support the outcomes of typical experiments shown in Tables 1–13.

8. CONCLUSIONS

This paper has investigated the inherent problems of the duration model in longitudinal

data analyses where the data are contaminated by individual specific heterogeneity. We have furthermore outlined and studied several methods well–suited to measure the mixture distribution. Until the work of Heckman and Singer (1984), few in the field of econometrics had paid attention to semi–nonparametric estimation methods for the identification of mixed unobservables. We have proposed two additional estimators which also address the existence of an unobserved mixing distribution in the sample density. MPLE smooths out roughness while maximizing goodness of fit. SIMEST is an estimator based on the axioms on which the structural relationships are based. Our Monte Carlo results suggest that NPMLE has some disadvantages relative to MPLE when censoring exists but that the two estimators often are quite comparable. SIMEST appears to out perform NPMLE and MPLE when the censoring rate grows. However, SIMEST has a serious problem of locating local maximum. It would appear that the application of SIMEST to the multidimensional model needs more investigation. Finally, although MPLE performs well in most cases, the choice of smoothing parameters is an open question when the multiterm of different norms is used for the penalty function.

FOOTNOTES

[1]For the mixture distribution, the likelihood function is often unbounded as $\partial L/\partial\theta=\infty$. See Hasselblad [1966].

[2]For the M–step, we can get the following first–order condition from (2–1):

$$\sum_{i=1}^{n} h(\theta_1) \frac{\partial \log g_{\underline{\beta}}(x_i|\theta_1)}{\partial \underline{\beta}} = 0 .$$

[3]Heckman and Singer suggested the steps to find a global maximum. For the first step, start to maximize with one point of support (k=1) with initial value for $\delta^{(m)}$. Let $\delta^{(m+1)}$ denote the estimated parameter. Divide interval $[\theta_{min}^{(1)}, \theta_{max}^{(1)}]$ into a representative mass of points of support (k=1) and find the points which have the Gateaux derivative of the loglikelihood function, $D(\theta,\mu_1) > 0$, for all $\theta \in \Theta$. If there is no point showing the positive Gateaux derivative, a global solution has been found. If $D(\theta,\mu_1) > 0$, add more points of support and divide the interval. Proceed to the subsequent step until there is termination by the criterion $D(\theta,\mu) \leq 0$, for all $\theta \in \Theta$. Trussell and Richards [1985] suggest that one more point of support be added until no improvement in the likelihood value is achieved.

[4]For example, if R is defined as the norm of the first derivative, then a penalty functional R will smooth the slope of the density f(x) which is semi–discontinuous. If R uses the norm of the second derivative, the curvature will be smoothed as well. Therefore, this smooth estimator is an application of the spline function.

[5]The problems we pursue here are quite different from previous applications which were limited to univariate modeling under the assumption that the functional form for the density is unknown. For example, Bartoszynski, et al. [1981] applied this method to estimate Cox's [1982] proportional hazard function. However, their analysis concerned a smoothed pointwise estimate of an unknown hazard distribution which is characterized as a dirac delta function. Other writers who were concerned with the parametric curve–fitting problem discussed the possibility that the method could smooth out the random error component in the linear least squares regression model [Kimeldorf and Wahba, 1970a,b; Anselone and Laurent, 1968].

[6]Good and Gaskins [1971], the first authors who applied this method, based the penalty function on the first derivative $R(f)=\int(f')^{1/2}$. Since then, a different penalty functions have been introduced by de Montricher, et al. [1975] and Silverman [1982] such as $\{(d/dx)^3(\log(x))\}^2$, $\int(f')^2$.

[7]It should be noted that in the case of no individual specific heterogeneity, the time–specific heterogeneity and the baseline hazard could be factored out and partial likelihood could be

maximized [Cox, 1972], resulting in consistent, inefficient and asymptotically normal estimates. We would like to thank Siu Fai Leung for pointing this out to us.

[8]Two other criteria to assess the deviation of the simulated probabilities from the actual probabilities are defined by maximizing the multinomial likelihood such as

$$\text{Max } S_1(\gamma_0) = \sum_{j=1}^{k} n_j \ln \hat{P}_{kj}(\gamma_0) \text{ or Max } S_2(\gamma_0) = \sum_{j=1}^{k} \ln \hat{P}_{kj}(\gamma_0) \text{ depending on the binning scheme.}$$

[9]This estimated cumulative heterogeneity distribution could not produce the theoretical distribution as well as the observed distribution, a phenomenon that was also reported by Heckman and Singer [1984]. Therefore, for the rest of our study, we give an additional point of support until no directional directives show positive values and no improvement in the likelihood value is shown.

[10]Since the use of 5 bins produced less biased estimates than that of 6 bins, we use 5 bins for the 500 observation experiments.

[11]The second derivative with respect to $\underline{X}$ was applied because our heterogeneity problem was defined as Case 1 of Section 5.

[12]Other estimations took more than a seven-fold increase in search time.

REFERENCES

Atkinson, E. N., B. Brown, and J. R. Thompson, 1983, Simulation Techniques for Parameter Estimation in Tumor related Stochastic Processes, in Proceedings of the 1983 Computer Simulation Conference, North–Holland, New–York, pp. 754–757.

Andersen, E. B., 1970, Asymptotic Properties of Conditional Maximum Likelihood Estimators, Journal of Royal Statistical Society, Series B, 32:283–301.

Anselone, P. M. and P. J. Laurent, 1968, A General Method for the Construction of Interpolating or Smoothing Spline Functions, Numerische Mathematik 12:66–88.

Bartoszynski, R. B., B. Brown, C. McBride, and J. R. Thompson,1981, Some Nonparametric Techniques for Estimating the Intensity Function of a Cancer Related Nonstationary Poisson Process, The Annals of Statistics 9:1050–1060.

Behrman, J., R. C. Sickles, and P. Taubman, 1988, Age Specific Death Rates, in Issues in Contemporary Retirement, edited by E. Lazear and R. Ricardo–Campbell, Stanford: Hoover Institution Press, 162–190.

Behrman, J., R. C. Sickles, and P. Taubman, 1990, Survivor Functions with Covariates: Sensitivity to Sample Length, Functional Form, and Unobserved Frailty, Demography 27: 267–284.

Behrman, J., R. C. Sickles, P. Taubman and A. Yazbeck, Black–White Mortality Inequalities, 1990, forthcoming in Journal of Econometrics.

Behrman, J., K. Huh, R. C. Sickles, P. Taubman, Robust Procedures to Estimate Cohort Effects in the Adoption of Health Technology: An Analysis of Modified Tobacco Use and Survivor Hazards in the U.S., Rice University, 1990.

Chandler, J. P., 1969, STEPIT, Behavioral Science, 14:81.

Cox, D. R., 1972, Regression Models and Life–Tables (with discussion), Journal of Royal Statistical Society, Series B 34:187–202.

Dempster, A. P., N. Laird and D. B. Rubin, 1977, Maximum Likelihood from Incomplete Data via the EM algorithm, Journal of Royal Statistical Society, Series B 39:1–38.

Diggle, P. J., and R. J. Gratton, 1984, Monte Carlo Methods of Inference for implicit Statistical Models, Journal of Royal Statistical Society, Series B 46:193–227.

Good, I. J., and R. A. Gaskins, 1971, Nonparametric Roughness Penalties for Probability Densities, Biometrika 58:255–277.

Gourieroux, C., and A. Monfort, 1989, Simulation Based Inference in Models with Heterogeneity, INSEE Working Paper #8905.

Han, A., and J. Hausman, 1990, Semiparametric Estimation of Duration and Competing Risk Models, Journal of Applied Econometrics 5, 1–28.

Hasselblad, V., 1966, Estimation of Parameters for a Mixture of Normal Distributions, Technometrics 8:431–444.

Heckman, J.,and B. Singer, 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, 1980, Cambridge: Cambridge University Press.

————, 1984, A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data, Econometrica 52:271–320.

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., 1988, Economic Duration Data and Hazard Functions, Journal of Economic Literature 26, 646–679

Kimeldorf, G. S. and Wahba, G., 1970a, A Correspondence between Bayesian Estimation on Stochastic Processes and Smoothing by Splines, Annals of Mathematical Statistics 41:495–502.

————, 1970b, Spline Functions and Stochastic Process, Sankhya(A), 32:173–180.

Laird, N. , 1978, Nonparametric Maximum Likelihood Estimation of a Mixing Distribution, Journal of American Statistical Association 73:805–811.

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 Royal Statistical Society, Series A 143:141–165.

Lerman, S. R., and C. F. Manski, 1981, On the Use of Simulated Frequencies to Approximate Choice Probabilities, in C. F. Manski and D. McFadden, eds., Structural analysis of Discrete Data with Econometric Applications, pp 305–319, Cambridge, Mass.: MIT Press.

Lindsay, B. G. ,1983a, The Geometry of Mixture Likelihoods: a General Theory, The Annals of Statistics 11:86–94.

———————— ,1983b, The Geometry of Mixture Likelihoods Part II: The Exponential Family, The Annals of Statistics 11:783–792.

Manton, K. G., E. Stallard, and J.W. Vaupel, 1986, Alternative Models for the Heterogeneity of Mortality Risks Among the Aged, Journal of the American Statistical Society 81, 635–644.

McFadden, D., 1989, A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration, Econometrica 57:995–1026.

de Montricher, G. F., R. A. Tapia, and J. R. Thompson, 1975, Nonparametric Maximum Likelihood Estimation of Probability Densities by Penalty Function Methods, The Annals of Statistics 3:1329–1348.

Newman, J., and C. E. McCulloch, 1984, A Hazard Rate Approach to the Timing of Births, Econometrica 52, 939–961.

Neyman J. and E. L. Scott, 1948, Consistent Estimates Based On Partially Consistent Observations, Econometrica 16:1–32.

Pakes, A., and D. Pollard, 1989, Simulation and the Asymptotics of Optimization Estimators, Econometrica 57: 1027–1058.

Ridder, G., 1986, The Sensitivity of Duration Models to Misspecified Unobserved Heterogeneity and Duration Dependence, manuscript, U. of Amsterdam.

Reinch, C. H., 1967, Smoothing by Spline Functions, Numerische Mathmatika 10:177–183.

Robbins, H., 1964, The Empirical Bayes Approach to Statistical Decision Problems, Annals of Mathematical Statistics, 35:1– 20.

Scott, D. W., 1979, On Optimal and Data Based Histograms, Biometrika 66:605–610.

Scott, D. W., and J. R. Thompson, 1983, Probability Density Estimation in Higher Dimensions, in Computer Science and Statistics: Proceedings of the Fifteenth Symposium on the Interface, ed. by Gentle, J. E., Amsterdam, North Holland:173–179

Sickles, R. C. and P. Taubman, 1986, An Analysis of the Health and Retirement Status of the Elderly, Econometrica 54: 1339–1356.

Sickles, R. C., 1989, An Analysis of Simultaneous Limited Dependent Variable Models and Some Nonstandard Cases, in Advances in Statistical Analysis and Statistical Computing, Theory and Applications, Vol. 2, edited by R. Mariano, Greenwich: JAI Press, Inc., 85–122.

Silverman, B. W., 1982, On the Estimation of a Probability Density Function by the Maximum Penalized Likelihood Method, The Annals of Statistics 10:795–810.

Simar, L., 1976, Maximum Likelihood Estimation of a Compound Poisson Process, The Annals of Statistics 4, 1200–1209.

Tapia, R. A., and J. R. Thompson, 1978, Nonparametric Probability Density Estimation, Baltimore: Johns Hopkins University Press.

Thompson, J. R., 1989, Empirical Model Building, New York: John Wiley & Sons, 114–131.

Thompson, J. R., E. N. Atkinson, and B. Brown, 1987, SIMEST: An Algorithm for Simulation Based Estimation of Parameters Characterizing a Stochastic Process, in Cancer Modeling, edited by J. R, Thompson and B. Brown, New York: Marcel Dekker, 387–415.

Trussell, J. and T. Richards, 1985, Correcting for Unobserved Heterogeneity in the Hazard Models : An Application of the Heckman–Singer Procedure, in N. Tuma, Sociological Methodology, Jossey Bass.

Yi, K. M., B. Honore, and J. Walker, 1987, CTM: A Program for the Estimation and Testing of Continuous Time Multi–Spell Models, User's Manual, Program Version 50, Mimeograph, ERC/NORC and University of Chicago.

TABLE 1

(n=500, uncensored)

	MPLE[1]	SIMEST	NPMLE[3]
γ	.934	1.154	.967
	(.084)	(.016)	(.070)
β_0	3.243	2.877	2.656
	(.379)	(.125)	(.309)
β_1	.988	.983	.971
	(.075)	(.078)	(.041)
β_2	.956	.934	.961
	(.065)	(.024)	(.009)
$-\ln L$	685	.887[4]	702

NOTE: True parameter values are $\gamma = \beta_1 = \beta_2 = 1$.
Values in ( ) denote standard errors.
Heterogeneity is specified to be standardized normal.

[1]Smoothing parameters $\alpha_1=\alpha_2=.5$, bin width (BW) = 0.6.

[2]Number of simulated observations (SN) = 50000; number of bins for x_1, x_2, and t is five.

[3]Four points of support were used to identify heterogeneity.

[4]Value of the criterion function.

TABLE 2

(n=500, 15% right censored[1])

	MPLE[2]	SIMEST[3]	NPMLE[4]
γ	.902	1.136	.896
	(.126)	(.037)	(.138)
β_0	7.23	3.57	10.45
	(1.37)	(.173)	(1.98)
β_1	.912	1.021	.811
	(.105)	(.056)	(.267)
β_2	.899	1.001	.845
	(.192)	(.034)	(.293)
$-\ln L$	710	1.036	771

See note in Table 1.

[1]78 out of 500 observations are censored.

[2]The smoothing parameter $\alpha = 0.55$.

[3]SN=50000; Number of bins for x_1, x_2, and t is five.

[4]Four points of support were used to identify heterogeneity.

TABLE 3

(n=100, uncensored)

	MPLE[1]	SIMEST[2]	NPMLE[3]
γ	.656	.823	.753
	(.299)	(.107)	(.145)
β_0	3.653	2.111	3.997
	(.166)	(.198)	(.198)
β_1	.775	1.140	.798
	(.397)	(.101)	(.487)
β_2	.718	.877	.757
	(.425)	(.088)	(.396)
-ln L	224	.678	295

See note in Table 1.

[1]Smoothing parameters $\alpha_1 = \alpha_2 = 1.2$.

[2]SN = 10000; number of bins for x_1, x_2, and t is four.

[3]Four points of support were used to identify heterogeneity.

TABLE 4

(n=1000, uncensored)

	MPLE[1]	SIMEST[2]	NPMLE[3]
γ	.931	.752	1.057
	(.068)	(.087)	(.035)
β_0	8.271	3.997	9.221
	(.220)	(.363)	(.317)
β_1	.956	.812	.937
	(.043)	(.091)	(.062)
β_2	.965	.799	.958
	(.088)	(.071)	(.097)
-ln L	1377	2.07	1540

See note in Table 1.

[1]Smoothing parameters $\alpha_1 = \alpha_2 = 0.3$.

[2]SN=100000; number of bins for x_1, x_2, and t is seven.

[3]Four points of support were used to identify heterogeneity.

TABLE 5

(n=500, 20% censored[1])

	MPLE[2]	SIMEST[3]	NPMLE[4]
γ	.834	1.165	.842
	(.239)	(.043)	(.210)
β_0	8.611	4.55	12.97
	(1.86)	(.327)	(2.36)
β_1	.859	1.199	.731
	(.237)	(.042)	(.296)
β_2	.862	1.099	.720
	(.218)	(.037)	(.309)
$-\ln L$	743	1.15	806

See note on Table 1.

[1]102 of the 500 observations are censored

[2]Smoothing parameters are $\alpha_1 = \alpha_2 = 0.6$.

[3]SN= 50000; number of bins for x_1, x_2, and t is five.

[4]Four points of support were used to identify heterogeneity.

TABLE 6[1]

(n=500, Right and left censored[2])

	MPLE[3]	SIMEST[4]
γ	.697	1.223
	(.367)	(.134)
β_0	13.481	8.54
	(4.37)	(.56)
β_1	.766	1.205
	(.342)	(.097)
β_2	.733	1.118
	(.366)	(.110)
-ln L	987	1.46

See note in Table 1.

[1]NPMLE is not available

[2]15% right censored and left censored.

[3]Smoothing parameters are $\alpha_1 = \alpha_2 = 0.7$.

[4]SN = 50000; number of bins for x_1, x_2, and t is five.

TABLE 7

MPLE with Different Smoothing Parameters
(n=500, uncensored)

Smoothing parameter	$\alpha = .2$	$\alpha = .5$	$\alpha = 1.0$
γ	.123	.934	.355
	(.156)	(.084)	(.011)
β_0	12.005	3.243	5.811
	(3.899)	(.379)	(3.68)
β_1	.661	.988	.602
	(.423)	(.075)	(.365)
β_2	.612	.956	.623
	(.477)	(.065)	(.275)
-ln L	1079	685	776

See note in Table 1.

TABLE 8

SIMEST with Different Bin widths
(n=500, uncensored)

	Case 1[1]	Case 2[2]	Case 3[3]
γ	.435 (.176)	1.154 (.016)	.240 (.232)
β_0	8.243 (2.987)	2.87 (.125)	13.566 (2.87)
β_1	.234 (.175)	.983 (.078)	.399 (.198)
β_2	.431 (.099)	.934 (.024)	.356 (.101)

See note in Table 1.

[1]SN=50000; number of bins for x_1, x_2, and t is two.

[2]SN=50000; number of bins for x_1, x_2, and t is five.

[3]SN=50000; number of bins $x_1 = x_2 = t = 10$.

TABLE 9

Bimodal Heterogeneity[1]
(n=500, uncensored)

	MPLE[2]	SIMEST[3]	NPMLE[4]
γ	.921 (.179)	.965 (.046)	.954 (.122)
β_0	5.753 (.366)	4.547 (.047)	4.885 (.288)
β_1	.923 (.087)	.976 (.038)	.928 (.039)
β_2	.922 (.071)	.955 (.076)	.921 (.021)
$-\ln L$	743	.876	799

See note on Table 1.

[1]Heterogeneity is generated by

$d\mu(\theta)=p\ (2\pi\sigma_1^2)^{-1/2} \exp\{-(\theta^2/2\sigma_1^2)\} + (1-p)(2\pi\sigma_2^2)^{-1/2} \exp\{-(\theta^2/2\sigma_2^2)\}\ d\theta$ where $p = .5$, $\sigma_1 = 1$, $\sigma_2 = 2$.

[2]Smoothing parameter $\alpha = .65$

[3]SN=50000; number of bins for x_1, x_2, and t is five.

[4]Four points of support were used to identify heterogeneity.

TABLE 10

Multimodal Heterogeneity[1]
(n=500, uncensored)

	MPLE[2]	SIMEST[3]	NPMLE[4]
γ	.992	1.002	.986
	(.005)	(.021)	(.002)
β_0	3.43	2.14	3.16
	(.138)	(.015)	(.016)
β_1	.981	.977	.983
	(.012)	(.033)	(.001)
β_2	.966	.978	.968
	(.009)	(.019)	(.003)
-ln L	673	.904	676

See note in Table 1.

[1]Heterogeneity was generated from
$d\mu(\theta_i) = p_i$, for i = 1,..,7, where $p_1 = p_3 = p_5 = p_7 = .1$, $p_2 = p_4 = p_6 = 0.2$.

[2]Smoothing parameter $\alpha = 0.3$.

[3]SN=50000; $bin(x_1) = bin(x_2) = bin(t) = 5$.

[4]Four points of support were used to identify heterogeneity.

TABLE 11

The Predictive Power of Estimators
(n=500, $\gamma = 2$, $\beta_1 = \beta_2 = 1$)

	MPLE[1]	SIMEST[2]	NPMLE[3]
γ	1.876	2.019	1.941
	(.197)	(.034)	(.042)
β_0	7.453	2.866	5.456
	(.254)	(.033)	(.208)
β_1	.956	.945	.968
	(.016)	(.028)	(.007)
β_2	.947	.965	.976
	(.015)	(.024)	(.011)
-ln L	907	.998	887

See note in Table 1.

[1]Smoothing parameter $\alpha = 0.6$.

[2]SN=50000, $\text{bin}(x_1) = \text{bin}(x_2) = \text{bin}(t) = 5$.

[3]Four points of support were used to identify heterogeneity.

TABLE 12

The Predictive Power of Estimators
(n=500, $\gamma = 1$, $\beta_1 = 2$, $\beta_2 = 3$)

	MPLE[1]	SIMEST[2]	NPMLE[3]
γ	.975	.987	.992
	(.113)	(.047)	(.042)
β_0	6.215	3.664	5.757
	(.097)	(.012)	(.022)
β_1	1.831	1.883	1.929
	(.065)	(.020)	(.007)
β_2	2.772	2.688	2.977
	(.069)	(.018)	(.010)
-ln L	861	1.127	837

See note in Table 1.

[1]Smoothing parameter $\alpha = 0.45$.

[2]SN=50000, $\text{bin}(x_1) = \text{bin}(x_2) = \text{bin}(t) = 5$.

[3]Four points of support were used to identify heterogeneity.

TABLE 13

The Predictive Power of Estimators
(n=500, $\gamma = .5$, $\beta_1 = \beta_2 = 1$)

	MPLE[1]	SIMEST[2]	NPMLE[3]
γ	.483	.510	.491
	(.066)	(.015)	(.025)
β_0	6.13	2.443	3.545
	(.023)	(.018)	(.013)
β_1	.971	.982	.992
	(.012)	(.014)	(.003)
β_2	.973	.985	.995
	(.008)	(.012)	(.004)
-ln L	703	1.307	681

See note in Table 1.

[1]Smoothing parameter $\alpha = 0.6$.

[2]SN=50000, $\text{bin}(x_1) = \text{bin}(x_2) = \text{bin}(t) = 5$.

[3]Four points of support were used to identify heterogeneity.

TABLE 14

Comparison of Estimators with Uncensored Data
(Standard Normal Heterogeneity)

Sample Size		MPLE	SIMEST	NPMLE
n=100	γ	.661(.017)[1]	.818(.062)	.748(.011)
	β_1	.781(.015)	1.112(.061)	.799(.012)
	β_2	.721(.009)	.869(.070)	.760(.007)
n=500	γ	.933(.005)	1.127(.034)	.970(.004)
	β_1	.983(.010)	.941(.063)	.973(.007)
	β_2	.950(.009)	.932(.056)	.960(.009)
n=1000	γ	.936(.006)	.771(.045)	1.054(.004)
	β_1	.961(.008)	.833(.061)	.975(.008)
	β_2	.969(.005)	.907(.064)	.970(.004)

[1]Values in ( ) represent standard deviations.

TABLE 15

Comparison of Estimators with Censored Data
(Standard Normal Heterogeneity, 15% Right Censored)

Sample Size		MPLE	SIMEST	NPMLE
n=100	γ	.632(.018)[1]	.796(.043)	.622(.018)
	β_1	.766(.017)	.898(.043)	.689(.013)
	β_2	.707(.010)	.849(.052)	.667(.011)
n=500	γ	.901(.012)	1.158(.038)	.904(.006)
	β_1	.913(.011)	1.014(.039)	.812(.007)
	β_2	.902(.015)	.992(.031)	.851(.012)
n=1000	γ	.921(.005)	.886(.022)	.916(.003)
	β_1	.927(.009)	.805(.034)	.799(.004)
	β_2	.933(.007)	.847(.031)	.815(.005)

[1]Values in ( ) represent standard deviations.

TABLE 16

Comparison of Estimators with Censored Data
(Standard Normal Heterogeneity, 20% RIght Censored)

Sample Size		MPLE	SIMEST	NPMLE
n=100	γ	.612(.019)[1]	.788(.148)	.621(.013)
	β_1	.762(.016)	.877(.039)	.686(.015)
	β_2	.699(.016)	.845(.044)	.661(.016)
n=500	γ	.836(.013)	1.164(.038)	.844(.007)
	β_1	.856(.012)	1.194(.029)	.732(.007)
	β_2	.859(.017)	1.096(.028)	.722(.013)
n=1000	γ	.865(.006)	.891(.027)	.887(.005)
	β_1	.907(.008)	.815(.029)	.769(.005)
	β_2	.913(.008)	.837(.038)	.809(.007)

[1]Values in ( ) represent standard deviations.

TABLE 17

Comparison of Estimators with Uncensored Data
(Bimodal Heterogeneity)

Sample Size		MPLE	SIMEST	NPMLE
n=100	γ	.757(.009)[1]	.813(.025)	.788(.007)
	β_1	.843(.011)	.872(.031)	.854(.007)
	β_2	.838(.007)	.865(.029)	.853(.006)
n=500	γ	.922(.003)	.967(.039)	.953(.002)
	β_1	.921(.005)	.958(.021)	.929(.005)
	β_2	.921(.006)	.947(.024)	.921(.004)
n=1000	γ	.945(.004)	.878(.021)	.954(.003)
	β_1	.970(.004)	.874(.034)	.975(.002)
	β_2	.975(.006)	.905(.035)	.977(.003)

[1]Values in ( ) represent standard deviations.

TABLE 18

Comparison of Estimators with Uncensored Data
(Multimodal Heterogeneity)

Sample Size		MPLE	SIMEST	NPMLE
n=100	γ	.812(.007)[1]	.832(.019)	.828(.005)
	β_1	.839(.012)	.859(.025)	.847(.009)
	β_2	.842(.011)	.857(.030)	.836(.008)
n=500	γ	.992(.002)	.989(.017)	.985(.002)
	β_1	.980(.003)	.971(.020)	.985(.003)
	β_2	.964(.003)	.958(.028)	.969(.002)
n=1000	γ	.994(.002)	.877(.024)	.985(.002)
	β_1	.985(.002)	.892(.027)	.988(.002)
	β_2	.971(.002)	.912(.032)	.972(.002)

[1]Values in ( ) represent standard deviations.

TABLE 19

Comparison of Estimators with Uncensored Data
(Standard Normal heterogeneity, $\beta_1=\beta_2=1,\gamma=2$)

Sample Size		MPLE	SIMEST	NPMLE
n=100	γ	1.412(.013)	1.542(.037)	1.558(.014)
	β_1	.783(.012)	1.198(.043)	.810(.010)
	β_2	.771(.013)	.852(.042)	.801(.012)
n=500	γ	1.872(.005)	2.202(.027)	1.940(.002)
	β_1	.955(.004)	.939(.014)	.967(.002)
	β_2	.948(.006)	.956(.014)	.975(.003)
n=1000	γ	1.883(.003)	1.556(.031)	1.953(.002)
	β_1	.962(.003)	.843(.032)	.971(.002)
	β_2	.956(.002)	.889(.037)	.976(.002)

[1]Values in ( ) represent standard deviations.

TABLE 20

Comparison of Estimators With Uncensored Data
(Standard Normal Heterogeneity, $\beta_1=1,\beta_2=2,\gamma=3$)

Sample Size		MPLE	SIMEST	NPMLE
n=100	γ	.728(.018)[1]	.707(.033)	.734(.011)
	β_1	1.576(.021)	1.576(.049)	1.635(.010)
	β_2	2.213(.028)	2.478(.062)	2.502(.015)
n=500	γ	.973(.002)	.982(.019)	.991(.003)
	β_1	1.829(.006)	1.878(.030)	1.928(.003)
	β_2	2.774(.005)	2.665(.037)	2.976(.004)
n=1000	γ	.982(.002)	.835(.025)	.992(.001)
	β_1	1.838(.006)	1.646(.039)	1.935(.002)
	β_2	2.778(.004)	2.466(.051)	2.987(.001)

[1] Values in ( ) represent standard deviations.