ECONOMETRICA

VOLUME 52 | MARCH, 1984 | NUMBER 2

A METHOD FOR MINIMIZING THE IMPACT OF DISTRIBUTIONAL ASSUMPTIONS IN ECONOMETRIC MODELS FOR DURATION DATA

BY J. HECKMAN AND B. SINGER[1]

Conventional analyses of single spell duration models control for unobservables using a random effect estimator with the distribution of unobservables selected by *ad hoc* criteria. Both theoretical and empirical examples indicate that estimates of structural parameters obtained from conventional procedures are very sensitive to the choice of mixing distribution. Conventional procedures overparameterize duration models. We develop a consistent nonparametric maximum likelihood estimator for the distribution of unobservables and a computational strategy for implementing it. For a sample of unemployed workers our estimator produces estimates in concordance with standard search theory while conventional estimators do not.

ECONOMIC THEORIES of search unemployment (Lippman and McCall [**34**]; Flinn and Heckman [**14**]), job turnover (Jovanovic [**25**]), mortality (Harris [**17**]), labor supply (Heckman and Willis [**23**]) and marital instability (Becker [**3**]) produce structural distributions for durations of occupancy of states. These theories generate qualitative predictions about the effects of changes in parameters on these structural distributions, and occasionally predict their functional forms.[2]

In order to test economic theories about durations and recover structural parameters, it is necessary to account for population variation in observed and unobserved variables unless it is assumed *a priori* that individuals are homogeneous.[3] In every microeconomic study in which the hypothesis of heterogeneity is subject to test, it is not rejected. Temporally persistent unobserved components are an empirically important fact of life in microeconomic data (Heckman [**19**]).

Since the appearance of papers by Silcock [**39**] and Blumen, Kogan, and McCarthy [**5**], social scientists have been aware that failure to adequately control for population heterogeneity can produce severe bias in structural estimates of duration models. Serious empirical analysts attempt to control for these unob-

[1]This research was supported by NSF Grants SOC77-27136 and SES-8107963 and NIH Grant NIH-1-R01-HD16846-01. We are indebted to George Yates for numerous valuable comments and for excellent computational assistance. Takeshi Amemiya, Chuck Mansky, Kevin M. Murphy, Chris Sims, and especially Stephen Cosslett gave us valuable comments on this paper. Discussions with Aloisio Araujo substantially improved early drafts of the paper. He contributed generously to early versions of Section 2 of the paper. Portions of this work were presented as an invited paper at the World Meetings of the Econometric Society, Aix en Provence, September 1, 1980. This paper was presented (under a different title) at the summer meetings of the Econometric Society, San Diego, June, 1981, and at seminars at Princeton, Johns Hopkins, MIT, and Wisconsin.

[2]An example of the latter point is the theory of search in stationary environments with Poisson distributed wage offer arrival times. For this model, the distribution of unemployment durations is exponential. See Flinn and Heckman [**14**].

[3]Remarkably, Becker and Stigler [**4**] argue that homogeneity should be the maintained hypothesis in any economic study.

served components in securing estimates of econometric duration models. (See, e.g., Harris [**17**]), and the papers cited in Heckman and Singer [**22**].)

The standard procedure used to control for population heterogeneity in unobserved variables is the random effect estimator. (See, e.g., Heckman and Willis [**23**], Flinn and Heckman [**15**], Nickell [**37**], Harris [**17**], and Heckman [**20**]). For single spell data, it is the only available estimator (Heckman and Singer [**22**]).[4] In standard application, the random effect estimator is implemented by assuming a functional form for the structural duration distribution of interest given observed and unobserved variables and a functional form for the distribution of unobservables. Maximum likelihood is used to estimate the parameters of the structural duration distribution and the parameters of the distribution of unobservables.

Economic theory sometimes offers guidance on the functional form of the distribution of durations given values of observed and unobserved variables. It rarely offers guidance on the functional form of the distribution of unobservables. The choice of a particular distribution of unobservables is usually justified on the grounds of familiarity, ease of manipulation, and considerations of computational cost.

In this paper we demonstrate that currently used methods of controlling for unobserved heterogeneity over-parameterize econometric duration models. Given an empirical distribution of durations and an assumed functional form for the structural duration distribution, it is possible to consistently estimate the population distribution of unobservables along with the structural parameters of interest. We establish the existence of a consistent nonparametric maximum likelihood estimator (NPMLE) for the structural parameters and the distribution function of unobservables in a general class of proportional hazard duration models with censoring and time varying variables. In addition, we present a computational strategy for implementing the estimator. In a limited set of Monte Carlo experiments, we find that the NPMLE recovers the structural parameters of the underlying models very well but does not accurately estimate the distribution of unobservables even in very large samples. However, the estimated models produce accurate predictions of duration distributions on fresh samples of data.

Our analysis is of considerable practical interest. We present both theoretical and empirical examples of the sensitivity of estimates of the structural parameters of duration models to assumptions made about the distribution of model unobservables. Using the consistent NPMLE proposed in this paper, we analyze data on unemployment durations. Once the impact of arbitrary distributional assumptions is eliminated from the estimation procedure, the data are consistent with the declining reservation wage hypothesis of search theory. This conclusion is in sharp contrast to alternative analyses of the same data using *ad hoc* parameterizations for the unobservables.

This paper is in five parts. Section 1 demonstrates the sensitivity of parameter estimates achieved from duration models to assumptions made about the func-

[4] Andersen [**2**], Chamberlain [**7**, **8**], and Heckman [**20**] discuss conditioning schemes and fixed effect procedures that can be used in multiple spell models.

tional form of the distribution of unobservables. Section 2 establishes the existence of a consistent nonparametric maximum likelihood estimator for the structural parameters and for the distribution function of the unobservables for a general class of proportional hazard models. Section 3 characterizes this estimator and presents a computational algorithm for implementing it. Section 4 summarizes the results of a limited set of Monte Carlo experiments designed to evaluate the NPMLE. The last section of the paper reports estimates of an econometric model for durations of unemployment obtained using the NPMLE. The paper concludes with a brief summary of the main findings.

All of the analysis in this paper is for the case in which the analyst has only one spell of an event for each person in his sample. The analysis of multiple spell data is a task left for the future.

1. THE IDENTIFICATION PROBLEM

In this section we demonstrate that the standard strategy for controlling for population variation in unobserved variables produces estimates that are very sensitive to assumptions made about the distribution of model unobservables. Both theoretical and empirical examples are offered.

Define $G(t \mid \theta)$ as the structural distribution function of duration t conditional on scalar heterogeneity component θ with distribution function $\mu(\theta)$. For simplicity we ignore observed exogenous variables. Economic theory sometimes provides guidance on the functional form of this distribution. For example, continuous time Markovian decision models in stationary environments (e.g. Burdett and Mortensen [**6**]) predict that

$$G(t \mid \theta) = 1 - e^{-t\theta}.$$

As another example, the declining reservation wage hypothesis in search unemployment theory predicts that $G(t \mid \theta)$ exhibits positive duration dependence, i.e., that the hazard function

$$h(t \mid \theta) = -\frac{\partial \ln(1 - G(t \mid \theta))}{\partial t}$$

is an increasing function of t; in particular,

$$\frac{\partial h(t \mid \theta)}{\partial t} > 0.$$

The distribution function of observed durations is

$$F(t) = \int G(t \mid \theta)\, d\mu(\theta). \tag{1.1}$$

Assume the observation interval is sufficiently long that censoring can be ignored and further assume that the available data are sufficiently rich to estimate $F(t)$ to any desired degree of precision, so that sampling variation in the estimation of $F(t)$ can be safely ignored.

In terms of this notation, the key identification question discussed in this paper can now be stated. From knowledge of $F(t)$, is it possible to solve (1.1) for unique $G(t\mid\theta)$ and $\mu(\theta)$? Without further information, the answer to this question is "no." The following example illustrates this point.

We produce two dramatically different structural models which both generate an exponential distribution of observed durations. First, let

$$G_1(t\mid\theta)=1-e^{-t\theta},\qquad t\geq 0,\quad \theta\geq 0,$$

and let $\mu_1(\theta)$ put unit mass on $\theta=\eta$ so that there is no population heterogeneity. Then

$$F_1(t)=1-e^{-t\eta}.$$

Next, let

$$G_2(t\mid\theta)=1-\int_c^\infty \frac{2}{\sqrt{2\pi}}e^{-l^2/2}\,dl,\qquad t\geq 0,$$

where

$$c=t(2\theta)^{-1/2}$$

and

$$d\mu_2(\theta)=\eta^2e^{-\eta^2\theta}\,d\theta,\qquad \theta\geq 0.$$

The second model is characterized by positive duration dependence at the individual level, conditional on θ (e.g., a declining reservation wage model in search unemployment theory). Observe that

$$F_2(t)=1-e^{-t\eta}=F_1(t).$$

Two fundamentally different structural models explain the same data. The first explanation is one of no duration dependence at the individual level. The second explanation is one of positive duration dependence at the individual level which, when contaminated by population heterogeneity, generates an aggregate duration distribution function which exhibits no duration dependence. Without further identifying information one cannot choose between these observationally equivalent explanations.

The standard practice in much recent work assumes that $G(t\mid\theta)$ and $\mu(\theta)$ are members of simple parametric families of distributions. The goal of the econometrician then becomes estimation of the parameters of $G(t\mid\theta)$ and $\mu(\theta)$, usually by maximum likelihood procedures. The choice of the mixing distribution is often justified on the basis of computational cost or by appeal to familiarity with special functional forms.

Since the relevant economic theory can sometimes be used to suggest functional forms for $G(t \mid \theta)$, we consider solutions of equation (1.1) with $G(t \mid \theta)$ known except, possibly, for finitely many unknown parameters.[5]

If $G(t \mid \theta)$ were known, equation (1.1) could be solved for $\mu(\theta)$ subject to standard existence conditions in the theory of Fredholm integral equations of the first kind (Tricomi [**41**, p. 150]). It is unnecessary to assume a specific parametric functional form for $\mu(\theta)$ and it may be empirically dangerous to do so. Standard practice, which assumes arbitrary functional forms for $G(t \mid \theta)$ and $\mu(\theta)$ may produce estimates of key structural parameters that are wildly inaccurate. Current practice thus over-parameterizes duration models. Given assumed functional forms for $G(t \mid \theta)$, it is in principle possible to estimate $\mu(\theta)$ even when parameters of the structural duration distribution are estimated.

It might be thought that the identification issue raised in this paper is of purely theoretical concern. It is widely believed that once an analyst has selected a particular functional form for $G(t \mid \theta)$, specification of a parametric family for $\mu(\theta)$ will not affect the estimates of structural parameters "very much." The following empirical example demonstrates that this is not so.

Three models differing only in the functional form of the density of unobservables were fit to data on unemployment durations. The data used were the same as those utilized by Kiefer and Neumann [**28**] in their analysis of unemployment durations.

The empirical analysis reported here is not intended as a specific comment on the Kiefer–Neumann analysis.[6] In our experience the sensitivity of estimates of the parameters of structural duration models to the choice of distribution of unobservables that is demonstrated below is a general characteristic of econometric models for the analysis of duration data.

The hazard function is assumed to be of the Weibull functional form

$$\ln h(t \mid \theta, x) = \alpha x + \beta \ln t + c\theta \tag{1.2}$$

where x is a vector of control variables, $\ln t$ is log duration, and θ is a heterogeneity component with coefficient c. Thus

$$G(t \mid \theta, x) = 1 - \exp\left(- e^{\alpha x + c\theta} \frac{t^{\beta+1}}{\beta + 1} \right). \tag{1.3}$$

[5] In another paper (Heckman and Singer [**22**]), we discuss nonparametric procedures for testing hypotheses about $G(t \mid \theta)$ in the presence of heterogeneity.

[6] The structural model we fit is not directly comparable to their model because (a) we use a continuous time approach whereas they use a discrete time approach, (b) we do not utilize the wage in the previous job as a determinant of current search decisions as they do because the theoretical rationale for doing so is not clear, and its introduction into the model raises new econometric problems which we do not wish to discuss, (c) we do not follow previous econometric work on female labor supply, as they do and attempt to estimate a reservation wage function, and (d) they do not present the correct statistical model for the economic model they claim to estimate (see Flinn and Heckman [**16**]). In addition, we delete certain explanatory variables with coefficients that are statistically insignificant in the Kiefer–Neumann empirical analysis and in our own analysis.

If $\beta = 0$, there is no duration dependence. c is a scale parameter. Three distributions of θ were utilized to fit the structural model: standard normal, log normal, and gamma distributions.

Empirical estimates of the parameters of $G(t \mid \theta, x)$ are very sensitive to the specification of the distribution of unobservables. In the empirical results reported in the first three columns of Table I, there are as many different "structural models" as there are distributions of heterogeneity. (The results reported in the fourth column will be discussed in Section 5 below.) *Ad hoc* specifications of model unobservables critically affect the empirical estimates achieved from "structural" duration models.

In light of this demonstrated sensitivity of estimates and model inference to *ad hoc* identifying assumptions, it is natural to proceed cautiously before drawing firm conclusions from econometric duration models. In the remainder of this paper, we consider a procedure for jointly estimating the distribution of unobservables, and the structural parameters of the model. We utilize the fundamental analysis of Kiefer and Wolfowitz **[29]** who produce general conditions for the existence of a consistent estimator of the mixing distribution $\mu(\theta)$ and the structural parameters. We verify their conditions for a general class of proportional hazard models in the presence of time varying covariates and censoring. Kiefer and Wolfowitz do not characterize the nonparametric maximum likelihood estimator. Using the analysis of Lindsay **[31, 32, 33]** we present a computational algorithm for the nonparametric maximum likelihood estimator.

TABLE I

KIEFER–NEUMANN DATA[a]

Sample Size is 456; Reduced Form Estimates (Standard Errors in Parentheses)

	Normal Heterogeneity	Log Normal Heterogeneity	Gamma Heterogeneity	Non-parametric Maximum Likelihood Estimator
Intercept	− 3.92	− 13.2	5.90	—
	(2.8)	(4.7)	(3.4)	
ln Duration	−.066	−.708	−.576	.494
	(.15)	(.17)	(.17)	
Age	.0036	−.106	−.202	−.0396
	(.048)	(.03)	(.06)	
Education	.0679	−.322	−.981	−.156
	(.233)	(.145)	(.301)	
Tenure on Previous Job	−.0512	.00419	−.034	−.041
	(.0149)	(.023)	(.016)	
Unemployment Benefits	−.0172	.0061	−.003	−.0174
	(.0036)	(.0051)	(.004)	
Married (0, 1)	.833	.159	−.607	.124
	(.362)	(.30)	(.496)	
Unemployment Rate	− 26.12	25.8	− 17.9	− 24.61
	(9.5)	(10.3)	(11.2)	
Ed. × Age	−.00272	.00621	.0152	.0011
	(.0044)	(.034)	(.0053)	
Heterogeneity (c)	5.16	5.7	4.62	—
	(.567)	(.42)	(.790)	

[a] For definitions of variables, see Kiefer and Neumann **[27]**.

2. CONSISTENCY OF THE MAXIMUM LIKELIHOOD ESTIMATOR

In this section we first present a detailed verification of the Kiefer–Wolfowitz (KW) conditions for a Weibull duration model without censored observations but with a time invariant vector $\boldsymbol{x}$ of observed covariates and an unobserved scalar heterogeneity component Θ assumed to be distributed independently of $\boldsymbol{x}$. Our verification of the KW conditions for the simple Weibull model serves as a prototype for verification of the KW conditions in more general settings. We next discuss how the steps in the verification procedure for the simple Weibull model must be modified to account for right censoring, restricted classes of time varying covariates, and general proportional hazard models.

A. Notation

Let $\boldsymbol{x} = (x_1, \dots, x_l)$ denote a row vector in $\mathscr{X}$, a subset of l-dimensional Euclidean space; and let $\boldsymbol{\alpha}_2 = (\alpha_2^{(1)}, \dots, \alpha_2^{(l)}) \in \mathscr{A}_2$, where $\mathscr{A}_2$ is the Cartesian product of finite intervals. $\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}$ is the inner product $\alpha_2^{(1)}x_1 + \cdots + \alpha_2^{(l)}x_l$. Introduce α_1 and assume $0 < \alpha_1 < \alpha_1^* < \infty$ and define $\boldsymbol{\alpha} = (\alpha_1, \boldsymbol{\alpha}_2)$ to be a point in $\mathscr{A} = I_1 \times I_2 \times \cdots \times I_{l+1}$, where $I_1 = (0, \alpha_1^*)$ and I_j, $2 \le j \le l+1$ are each finite intervals in $\mathbb{R}_1$. Let θ denote a value of a random variable Θ taking values in an interval $\underline{\underline{\Theta}} \subset \mathbb{R}_1$, and define $\mathscr{M} = \{\mu\} \equiv$ set of probability distributions on $\underline{\underline{\Theta}}$. Let a point in the product space $\Gamma = \mathscr{A} \times \mathscr{M}$ be denoted by $\gamma = (\boldsymbol{\alpha}, \mu)$. Finally, define $h(\boldsymbol{x})$ to be the frequency function of $\boldsymbol{X}$. It is assumed to be a bounded density with respect to Lebesgue measure on the continuous coordinates of $\boldsymbol{x}$ and a probability mass function on the discrete coordinates.

With these ingredients at hand we introduce the random vector $(T, \boldsymbol{X})$ having frequency function $f_\gamma(t, \boldsymbol{x})$ represented as

$$f_\gamma(t, \boldsymbol{x}) = h(\boldsymbol{x}) \int_{\underline{\underline{\Theta}}} g_{\boldsymbol{\alpha}}(t \mid \boldsymbol{x}, \theta)\, d\mu(\theta) \tag{2.1a}$$

where

$$g_{\boldsymbol{\alpha}}(t \mid \boldsymbol{x}, \theta) = \alpha_1 t^{\alpha_1 - 1} \exp\left[\boldsymbol{\alpha}_2 \cdot \boldsymbol{x} + \theta - t^{\alpha_1} \exp(\boldsymbol{\alpha}_2 \cdot \boldsymbol{x} + \theta)\right]. \tag{2.1b}$$

We assume that $h(\boldsymbol{X})$ is functionally independent of the parameters in g_α.

Suppose that we observe $(t_i, \boldsymbol{x}_i)$, $1 \le i \le n$, viewed as possible values of the i.i.d. random vectors $(T_i, \boldsymbol{X}_i)$, $1 \le i \le n$, each of which is distributed according to (2.1a, b). Before stating and verifying the KW conditions we require some additional notation and definitions. Let $\boldsymbol{\alpha}^0$ and μ_0, respectively, be the "true" value of $\boldsymbol{\alpha}$ and the "true" distribution of Θ. Then bring in the metric

$$\begin{aligned}
\delta(\gamma_1, \gamma_2) &\equiv \delta\big((\boldsymbol{\alpha}^{(1)}, \mu_1), (\boldsymbol{\alpha}^{(2)}, \mu_2)\big) \\
&= \sum_{m=1}^{l+1} \left|\arctan \alpha_{(m)}^{(1)} - \arctan \alpha_{(m)}^{(2)}\right| \\
&\quad + \int_{\underline{\underline{\Theta}}} |\mu_1(\theta) - \mu_2(\theta)| e^{-|\theta|}\, d\theta
\end{aligned}$$

where

$$\boldsymbol{\alpha}^{(k)} = \left(\alpha_{(1)}^{(k)}, \ldots, \alpha_{l+1}^{(k)}\right) \equiv \left(\alpha_1, \alpha_2^{(1)}, \ldots, \alpha_2^{(l)}\right)_k, \qquad k = 1, 2,$$

and

$$\mu_i(\theta) = \mathrm{Prob}_i(\Theta \leq \theta).$$

Finally, let $\overline{\mathscr{A}} \times \overline{\mathscr{M}} = \overline{\Gamma}$ be the completion of $\mathscr{A} \times \mathscr{M} = \Gamma$ with respect to the metric $\delta(\cdot, \cdot)$—i.e., $\overline{\Gamma}$ contains Γ together with the limits of its Cauchy sequences in the sense of the metric $\delta(\cdot, \cdot)$.

With these definitions, we now state and verify the five KW conditions for the simple Weibull model.

B. VERIFICATION OF CONSISTENCY CONDITIONS FOR THE UNCENSORED WEIBULL MODEL WITH REGRESSORS AND A SCALAR HETEROGENEITY COMPONENT

1. The conditional frequency function $f_\gamma(t, \boldsymbol{x} \mid \theta) = h(\boldsymbol{x}) g_{\boldsymbol{\alpha}}(t \mid \boldsymbol{x}, \theta)$ is a density with respect to a σ-finite measure ν defined on $\mathbb{R}_1^+ \times \mathbb{R}_l$, where l is the dimension of the vector of covariates.

Because of our assumptions about $h(\boldsymbol{x})$ and the special functional form of $g_{\boldsymbol{\alpha}}(t \mid \boldsymbol{x}, \theta)$, the necessary σ-finite measure is

$$d\nu = dt\, dx_{i_1}, \ldots, dx_{i_k}\, d\zeta_{i_{k+1}}, \ldots, d\zeta_{i_l}$$

where dt is Lebesgue measure on $\mathbb{R}_1^+$, $dx_{i_1}, \ldots, dx_{i_k}$ is Lebesgue measure on those coordinates of $\boldsymbol{x}$ which take on a continuum of values, and $d\zeta_{i_{k+1}}, \ldots, d\zeta_{i_l}$ is counting measure on the discrete coordinates of $\boldsymbol{x}$.

2 (Continuity). It must be possible to extend the definition of $f_\gamma(t, \boldsymbol{x})$ so that the range of γ will be $\overline{\Gamma} = \overline{\mathscr{A}} \times \overline{\mathscr{M}}$ and so that for any $\{\gamma_i\}$ and γ^* in $\overline{\mathscr{A}} \times \overline{\mathscr{M}}$, $\gamma_i \to \gamma^*$, $i \to \infty$ implies that $f_{\gamma_i}(t, \boldsymbol{x}) \to f_{\gamma^*}(t, \boldsymbol{x})$ except perhaps on a set of $(t, \boldsymbol{x})$ whose probability is 0 according to the frequency function of $f_{\gamma_0}(t, \boldsymbol{x})$ where $\gamma_0 = (\boldsymbol{\alpha}^0, \mu_0)$ is the true parameter vector.

First observe that points of the form $(0, \boldsymbol{x})$, with $\boldsymbol{x} \in \mathscr{X}$, comprise a set of $f_{\gamma_0}(t, \boldsymbol{x}) d\nu$ measure 0. This is a consequence of the fact that $f_{\gamma_0}(t, \cdot)$ is a continuous density with respect to Lebesgue measure dt on $\mathbb{R}_1^+$. These points will represent the exceptional set on which the continuity condition $\gamma_i \to \gamma^* \Rightarrow f_{\gamma_i}(t, \boldsymbol{x}) \to f_{\gamma^*}(t, \boldsymbol{x})$ is allowed to be violated. Then in order to characterize $\overline{\mathscr{A}} \times \overline{\mathscr{M}} \setminus \mathscr{A} \times \mathscr{M}$, first notice that $\overline{\mathscr{M}} \setminus \mathscr{M}$ consists of all measures μ which are limits of $\{\mu_i\}$ in the sense that $\lim_{i \to \infty} \int_{\underline{\Theta}} | \mu_i(\theta) - \mu(\theta)| e^{-|\theta|} d\theta = 0$ where limit measures μ need not necessarily be proper probability measures (this includes defective distributions).

Since all coordinates of $\boldsymbol{\alpha}_2$ are restricted to finite open intervals, $\overline{\mathscr{A}} \setminus \mathscr{A}$ consists of the finite valued end points of the intervals I_j, $2 \leq j \leq l + 1$, together with $\alpha_1 = 0$. Then for every $\gamma \in \overline{\mathscr{A}} \times \overline{\mathscr{M}} \setminus \mathscr{A} \times \mathscr{M}$ $f_\gamma(t, \boldsymbol{x})$ is well defined at all $(t, \boldsymbol{x})$ except $(0, \boldsymbol{x})$ which is a point of measure 0 with respect to $f_{\gamma_0}(t, \boldsymbol{x}) d\nu(t, \boldsymbol{x})$.

To verify continuity, first recall that $\gamma_i \to \gamma^*$ means that $\lim_{i \to \infty} \delta(\gamma_i, \gamma^*) = 0$

and, in particular, that

$$\lim_{i\to\infty}\int_{\underline{\underline{\Theta}}}|\mu_i(\theta)-\mu_*(\theta)|e^{-|\theta|}\,d\theta=0.$$

Now, using the triangle inequality we have

$$\begin{aligned}
(2.2)\qquad & |f_{\gamma_i}(t,x)-f_{\gamma^*}(t,x)| \\
&= h(x)\left|\int_{\underline{\underline{\Theta}}} g_{\alpha^{(i)}}(t\,|\,x,\theta)\,d\mu_i(\theta)-\int_{\underline{\underline{\Theta}}} g_{\alpha^*}(t\,|\,x,\theta)\,d\mu_*(\theta)\right| \\
&\le h(x)\int_{\underline{\underline{\Theta}}}|g_{\alpha^{(i)}}(t\,|\,x,\theta)-g_{\alpha^*}(t\,|\,x,\theta)|\,d\mu_i(\theta) \\
&\quad + h(x)\left|\int_{\underline{\underline{\Theta}}} g_{\alpha^*}(t\,|\,x,\theta)\,d\mu_i(\theta)-\int_{\underline{\underline{\Theta}}} g_{\alpha^*}(t\,|\,x,\theta)\,d\mu_*(\theta)\right|.
\end{aligned}$$

Since the function $g_\alpha(t\,|\,x,\theta)$ is jointly uniformly continuous in α and θ, we have that $\forall\epsilon>0$, $\exists I(t,x)$ such that if $i>I(t,x)$, then

$$(2.3)\qquad |g_{\alpha^{(i)}}(t\,|\,x,\theta)-g_{\alpha^*}(t\,|\,x,\theta)|<\epsilon.$$

As a consequence of (2.3), the first term on the right hand side of (2.2) is bounded above by

$$\epsilon h(x)\int_{\underline{\underline{\Theta}}} d\mu_i(\theta)=\epsilon h(x)$$

where $i>I(t,x)$.

To show that the second term vanishes as $i\to\infty$, we integrate by parts to obtain

$$\begin{aligned}
&\left|\int_{\underline{\underline{\Theta}}} g_{\alpha^*}(t\,|\,x,\theta)\,d\mu_i(\theta)-\int_{\underline{\underline{\Theta}}} g_{\alpha^*}(t\,|\,x,\theta)\,d\mu_*(\alpha)\right| \\
&\quad=\Bigg|\, g_{\alpha^*}(t\,|\,x,\theta)\big[\mu_i(\theta)-\mu_*(\theta)\big]\Big|_{\theta_0}^{\theta_1} \\
&\qquad-\int_{\underline{\underline{\Theta}}}\big[\mu_i(\theta)-\mu_*(\theta)\big]\frac{\partial}{\partial\theta}\,g_{\alpha^*}(t\,|\,x,\theta)\,d\theta\Bigg|
\end{aligned}$$

where $\underline{\underline{\Theta}}=(-\infty,\infty)$. Note that

$$g_{\alpha^*}(t\,|\,x,\theta)\big[\mu_i(\theta)-\mu_*(\theta)\big]\Big|_{-\infty}^{\infty}=0$$

since $\lim_{\theta\to\pm\infty}g_{\alpha^*}(t\,|\,x,\theta)=0$, except possibly on the exceptional set $(0,x)$.

Now set $v(t,x)=\alpha_1t^{\alpha_1-1}\exp(\alpha_2\cdot x)$ and observe that

$$\begin{aligned}
(2.4)\qquad \frac{\partial}{\partial\theta}\,g_{\alpha^*}(t\,|\,x,\theta) &= v(t,x)\frac{\partial}{\partial\theta}\exp\big[\theta-t^{\alpha_1}e^{(\alpha_2\cdot x+\theta)}\big] \\
&=\big[1-t^{\alpha_1}e^{\alpha_2\cdot x+\theta}\big]\exp\big[\theta-t^{\alpha_1}e^{(\alpha_2\cdot x+\theta)}\big]v(t,x).
\end{aligned}$$

Bring in the inequality $1 - \zeta \leq e^{-\zeta}$ for $\zeta > 0$. Then for an upper bound on (2.4) we have

$$\frac{\partial}{\partial\theta} g_{\alpha^*}(t \mid \boldsymbol{x}, \theta) \leq v(t, \boldsymbol{x}) \exp(-t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x} + \theta}) \exp\left[\theta - t^{\alpha_1} e^{(\boldsymbol{\alpha}_2 \cdot \boldsymbol{x} + \theta)}\right]$$

$$= v(t, \boldsymbol{x}) \exp\left[\theta - 2t^{\alpha_1} e^{(\boldsymbol{\alpha}_2 \cdot \boldsymbol{x} + \theta)}\right].$$

We also have the obvious lower bound

$$\frac{\partial g_{\alpha^*}(t \mid \boldsymbol{x}, \theta)}{\partial\theta} \geq -t^{\alpha_1} e^{(\boldsymbol{\alpha}_2 \cdot \boldsymbol{x} + \theta)} \exp\left[\theta - t^{\alpha_1} e^{(\boldsymbol{\alpha}_2 \cdot \boldsymbol{x} + \theta)}\right] v(t, \boldsymbol{x}).$$

Rewriting these inequalities we have

$$\left| \frac{\partial g_{\alpha^*}(t \mid \boldsymbol{x}, \theta)}{\partial\theta} \right| \leq g_{\alpha^*}(t \mid \boldsymbol{x}, \theta) \max\left(t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x} + \theta}, \exp\left[-t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x} + \theta}\right]\right)$$

$$= v(t, \boldsymbol{x}) \exp\left[\theta - t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x} + \theta}\right]$$

$$\times \max\left(t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x} + \theta}, \exp\left[-t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x} + \theta}\right]\right).$$

For $\theta < 0$ we have

$$\left| \frac{\partial g_{\alpha^*}(t \mid \boldsymbol{x}, \theta)}{\partial\theta} \right| < v(t, \boldsymbol{x}) \left[1 + t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}}\right] \exp\left[\theta - t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x} + \theta}\right]$$

$$< v(t, \boldsymbol{x}) \left[1 + t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}}\right] e^{\theta}$$

$$\equiv v(t, \boldsymbol{x}) \left[1 + t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}}\right] e^{-|\theta|}.$$

For $\theta > 0$, define

$$\theta^*(t, \boldsymbol{x}) = \sup(\theta > 0 : t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x} + \theta} \leq 3\theta).$$

Then for $\theta > \theta^*(t, \boldsymbol{x})$,

$$\exp\left[2\theta - t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x} + \theta}\right] \leq e^{-\theta}.$$

Define

$$C(t, \boldsymbol{x}) = \frac{\sup_{\theta \in [0, \theta^*(t, \boldsymbol{x})]} \exp\left[2\theta - t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x} + \theta}\right]}{e^{-\theta^*(t, \boldsymbol{x})}}.$$

Then for $\theta > 0$

$$\left|\frac{\partial g_{\alpha^*}(t\,|\,\boldsymbol{x},\theta)}{\partial\theta}\right| < v(t,\boldsymbol{x})\left[1 + t^{\alpha_1}e^{\alpha_2\cdot x}\right]\exp\left[2\theta - t^{\alpha_1}e^{\alpha_2\cdot x+\theta}\right]$$

$$< \begin{cases} v(t,\boldsymbol{x})\left[1 + t^{\alpha_1}e^{\alpha_2\cdot x}\right]e^{-\theta} & \text{for } \theta > \theta^*(t,\boldsymbol{x}), \\ v(t,\boldsymbol{x})\left[1 + t^{\alpha_1}e^{\alpha_2\cdot x}\right]C(t,\boldsymbol{x})e^{-\theta} & \text{for } 0 \leq \theta \leq \theta^*(t,\boldsymbol{x}). \end{cases}$$

Thus a general bound, valid for all θ, is

$$\left|\frac{\partial g_{\alpha^*}(t\,|\,\boldsymbol{x},\theta)}{\partial\theta}\right| < v(t,\boldsymbol{x})\left[1 + t^{\alpha_1}e^{\alpha_2\cdot x}\right]\max(1,C(t,\boldsymbol{x}))e^{-|\theta|}.$$

Therefore

$$\left|\int_{\underline{\underline{\Theta}}}\frac{\partial}{\partial\theta}\,g_{\alpha^*}(t\,|\,\boldsymbol{x},\theta)\left[\,\mu_i(\theta) - \mu_*(\theta)\,\right]d\theta\right|$$

$$\leq v(t,\boldsymbol{x})\left[1 + t^{\alpha_1}e^{\alpha_2\cdot x}\right]\max(1,C(t,\boldsymbol{x}))\int_{\underline{\underline{\Theta}}}|\,\mu_i(\theta) - \mu_*(\theta)|e^{-|\theta|}\,d\theta$$

$$\to 0 \qquad \text{as} \quad i \to \infty$$

and condition 2 is verified.

3. For any $\gamma \in \overline{\mathscr{A}} \times \overline{\mathscr{M}} = \overline{\Gamma}$ and any $\rho > 0$, $\omega_\gamma(t,\boldsymbol{x};\rho)$ must be a measurable function of $(t,\boldsymbol{x})$ where

$$\omega_\gamma(t,\boldsymbol{x};\rho) = \sup_{\gamma'\,:\,\delta(\gamma,\gamma')<\rho} f_{\gamma'}(t,\boldsymbol{x}).$$

To verify measurability observe that since $\overline{\mathscr{A}} \times \overline{\mathscr{M}}$ is separable,[7] we have for any $\gamma \in \overline{\Gamma}$ and $\rho > 0$

$$\sup_{\gamma'\,:\,\delta(\gamma,\gamma')<\rho} f_{\gamma'}(t,\boldsymbol{x}) = \sup_{\gamma_n'\,:\,\delta(\gamma\,,\,\gamma_n')<\rho} f_{\gamma_n'}(t,\boldsymbol{x})$$

where $\{\gamma_n'\}$ is a countable dense set in $\overline{\Gamma}$. Since $f_\gamma(t,\boldsymbol{x})$ is measurable for each $\gamma \in \overline{\Gamma}$, we have that $\omega_\gamma(t,\boldsymbol{x};\rho)$ is measurable due to the fact that it is the

[7]The required countable dense set in $\overline{\Gamma}$ to ensure separability is the collection of points $\boldsymbol{\alpha}_2$ in l-dimensional Euclidean space with rational coordinates, points α_1 which assume positive rational values, together with the probability distributions on $\underline{\underline{\Theta}}$ having only finitely many points of increase but with these points and the values of the distribution functions only allowed to be rational numbers. It is important to observe that separability of $\overline{\Gamma}$ is ensured via this set because of the special choice of metric $\delta(\cdot,\cdot)$. In particular, the term

$$\int_{\underline{\underline{\Theta}}}|\,\mu_1(\theta) - \mu_2(\theta)|e^{-|\theta|}\,d\theta$$

ensures that all distributions on $\underline{\underline{\Theta}}$ can be approximated arbitrarily closely by rational valued distributions with finitely many points of increase and only at rational points.

supremum of a countable sequence of measurable functions (see, e.g., Royden [**38**, p. 67]).

4 (Identifiability). If $\gamma_1 \in \bar{\Gamma}$ is different from γ_0 then for at least one $\boldsymbol{y}$ we must have

$$\int_{-\infty}^{y} f_{\gamma_1}(t,\boldsymbol{x})\,d\nu(t,\boldsymbol{x}) \neq \int_{-\infty}^{y} f_{\gamma_0}(t,\boldsymbol{x})\,d\nu(t,\boldsymbol{x}).$$

For this it is enough to show that $\gamma_1 \neq \gamma_0 \Rightarrow f_{\gamma_1}(t,\boldsymbol{x}) \neq f_{\gamma_0}(t,\boldsymbol{x})$ for points $(t,\boldsymbol{x})$ in some set of positive measure with respect to $d\nu(t,\boldsymbol{x}) \equiv dt\,dx_1, \ldots, dx_{i_k} d\zeta_{i_{k+1}}, \ldots, d\zeta_{i_l}$. We verify this by deriving a contradiction. To this end, suppose $\exists \gamma_1 = (\alpha_1^{(1)}, \boldsymbol{\alpha}_2^{(1)}, \mu_1) \neq (\alpha_1^{(0)}, \boldsymbol{\alpha}_2^{(0)}, \mu_0) = \gamma_0$ such that

$$\begin{aligned}(2.5)\qquad f_{\gamma_1}(t,\boldsymbol{x}) &= h(\boldsymbol{x}) \int_{\underline{\underline{\Theta}}} \alpha_1^{(1)} t^{\alpha_1^{(1)}-1} \exp\left[\boldsymbol{\alpha}_2^{(1)} \cdot \boldsymbol{x} + \theta - t^{\alpha_1^{(1)}} e^{(\boldsymbol{\alpha}_2^{(1)} \cdot \boldsymbol{x} + \theta)}\right] d\mu_1(\theta) \\ &= h(\boldsymbol{x}) \int_{\underline{\underline{\Theta}}} \alpha_1^{(0)} t^{\alpha_1^{(0)}-1} \exp\left[\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x} + \theta - t^{\alpha_1^{(0)}} e^{(\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x} + \theta)}\right] d\mu_0(\theta) \\ &= f_{\gamma_0}(t,\boldsymbol{x})\end{aligned}$$

for almost all $(t,\boldsymbol{x}) \in \mathbb{R}_1^+ \times \mathscr{X}$, where $\mathscr{X} \subset \mathbb{R}_l$ and $\mathscr{X}$ is the set of possible values of $\boldsymbol{x}$. Recall that $f_\gamma(t,\boldsymbol{x})$ is uniformly continuous so that we may replace "for almost all $(t,\boldsymbol{x})$" with $\forall(t,\boldsymbol{x})$.

First restrict attention to $\mathscr{M} \cap \{\mu : \int e^\theta d\mu(\theta) < \infty\}$, a class of uniformly integrable distribution functions, and assume that $\mathbf{0} = (0, \ldots, 0) \in \mathscr{X}$. If $\mathbf{0} \notin \mathscr{X}$ we redefine the old coordinates by a translation so that in the new coordinates $\mathbf{0} \in \mathscr{X}$. At $\boldsymbol{x} = (0, \ldots, 0)$ suppose there exist two pairs $\gamma_0 = (\alpha_1^{(0)}, \mu_0)$ and $\gamma_1 = (\alpha_1^{(1)}, \mu_1)$ such that $f_{\gamma_1}(t) = f_{\gamma_0}(t)$. This implies that

$$(2.6)\qquad 1 \equiv \frac{f_{\gamma_1}(t)}{f_{\gamma_0}(t)} = \frac{\alpha_1^{(1)}}{\alpha_1^{(0)}} t^{\alpha_1^{(1)} - \alpha_1^{(0)}} \frac{\int_{\underline{\underline{\Theta}}} \exp\{\theta - t^{\alpha_1^{(1)}} e^\theta\}\,d\mu_1(\theta)}{\int_{\underline{\underline{\Theta}}} \exp\{\theta - t^{\alpha_1^{(0)}} e^\theta\}\,d\mu_0(\theta)}.$$

Since $E(e^\theta) < \infty$, by the monotone convergence theorem when $t \to 0$ the right hand side of (2.6) tends to 0 or ∞ according as $\operatorname{sgn}(\alpha_1^{(1)} - \alpha_1^{(0)}) = -1$ or $+1$ so that

$$1 \neq \frac{f_{\gamma_1}(t)}{f_{\gamma_0}(t)} \qquad \text{as} \quad t \to 0$$

unless $\alpha_1^{(1)} = \alpha_1^{(0)}$. Given the truth of this equality, the uniqueness theorem for LaPlace transforms (see, e.g., Widder [**41**]) implies that $\mu_1(\theta) = \mu_0(\theta)$. Thus we have established that in a model without regressors (α_1, μ) is identified.

If there are regressors, we need to demonstrate that $\boldsymbol{\alpha}_2^{(1)} = \boldsymbol{\alpha}_2^{(0)}$. To show this, assume the contrary and suppose $\exists \boldsymbol{x}_0$ such that $\boldsymbol{\alpha}_2^{(1)} \cdot \boldsymbol{x}_0 = \boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}_0$. Assume that the distribution of $\boldsymbol{x}$ is nondegenerate in the neighborhood of $\boldsymbol{x}_0$ and of full rank.[8] At such a value of x, suppose that $\alpha_2^{(1)} \cdot x < \alpha_2^{(0)} \cdot x$. Then if $f_{\gamma_1}(t, x) = f_{\gamma_0}(t, x)\ \forall (t, x) \in \mathbb{R}_1^+ \times \mathscr{X}$, we have

$$0 = \int_1^\infty \left[f_{\gamma_1}(s, x) - f_{\gamma_0}(s, x) \right] ds$$

$$= h(x) \int_{\underline{\underline{\Theta}}} \left\{ \exp\left[-e^{\alpha_2^{(1)} \cdot x + \theta} \right] - \exp\left[-e^{\alpha_2^{(0)} \cdot x + \theta} \right] \right\} d\mu(\theta)$$

with the integrand nonnegative. But this implies that the integrand $= 0$ a.e. $d\mu(\theta)$ which, in turn, implies that $\alpha_2^{(1)} \cdot x = \alpha_2^{(0)} \cdot x$. This is a contradiction; and we have thus verified that $\alpha_2^{(1)} = \alpha_2^{(0)}$, thereby completing the proof of identifiability when $E(e^\theta) < \infty$. Assuming that $\mathscr{M}$ is a uniformly integrable class rules out limiting random variables with $\int_{\underline{\underline{\Theta}}} e^\theta d\mu(\theta) = +\infty$ that arise when forming the closure of $\mathscr{M}$ and that destroy identifiability. (For a discussion of uniform integrability see Meyer [**36**].)

Suppose next that $E(e^\theta) = \infty$. Provided that the tail behavior of the true $\mu_0 \in \mathscr{M}$ is restricted, the model can still be identified. To show this, we initially restrict our discussion to the class of absolutely continuous probability distributions where $d\mu(\theta) = m(\theta)\, d\theta$. We consider more general classes of distributions below. We require that the density $m(\theta)$ satisfy the following tail condition:

$$m(\theta) = \frac{c_i}{|\theta|^{\beta_i} \exp(\epsilon_i |\theta|) L_i(\theta)} \qquad \text{for} \quad \theta^* < \infty \quad \text{and} \quad |\theta| > \theta^* \tag{2.7}$$

$$\theta \to (-1)^i \infty, \quad c_i > 0, \quad 0 \leq \epsilon_i < 1, \quad \beta_i > 0, \quad i = 1, 2,$$

where θ^* is a uniform bound for all admissible densities in $\mathscr{M}$ and where the $L_i(\theta)$ are slowly varying in the sense of Karamata.[9]

We assume that at least one regressor is defined on $(-\infty, +\infty)$ and that $\mathbf{0} = (0, \ldots, 0) \in \mathscr{X}$. If $\mathbf{0} \notin \mathscr{X}$, we redefine the old coordinates by a translation so that in the new coordinates $\mathbf{0} \in \mathscr{X}$.

Define

$$W_\gamma(g) = \int_{-\infty}^\infty e^\theta e^{-g e^\theta} m(\theta)\, d\theta \qquad \text{for} \quad g \geq 0. \tag{2.8}$$

Let $e^\theta = \phi$ and define

$$\bar{v}(\phi) = m(\ln \phi).$$

Then

$$W_\gamma(g) = \int_0^\infty e^{-g\phi} \bar{v}(\phi)\, d\phi$$

[8] More precisely we assume that $E(|\alpha_2^{(1)} \cdot x|) = 0$ implies that $\alpha_2^{(1)} = 0$. An equivalent statement of this condition is that the population covariance of x is nonsingular.

[9] For a discussion of slowly varying functions, see Feller [**13**, p. 275].

and note that

$$\bar{v}(\phi) \sim \frac{c_2}{(\ln\phi)^{\beta_2}\phi^{\epsilon_2}L_2(\ln\phi)}, \qquad \phi \to \infty.$$

Now

$$\bar{V}(\phi) \sim \int_0^{\phi} \bar{v}(\delta)\,d\delta \sim \int_{c'}^{\phi} \bar{v}(\delta)\,d\delta \qquad \text{for some } c'$$

$$\sim c_2\phi \int_{c'/\phi}^{1} \frac{du}{(\ln\phi u)^{\beta_2}(\phi u)^{\epsilon_2}L_2(\ln\varphi u)}$$

$$\sim \frac{c_2^*}{(\ln\phi)^{\beta_2}\phi^{\epsilon_2-1}L_2(\ln\varphi)}, \qquad \phi \to \infty$$

where $c_2^* = c_2/(1-\epsilon_2)$ since

$$\frac{\ln(\phi u)}{\ln(\varphi)} \to 1, \qquad \phi \to \infty, \quad \forall u > 0,$$

and

$$\frac{L_2(\ln\phi u)}{L_2(\ln\phi)} \to 1, \qquad \phi \to \infty, \quad \forall u > 0.$$

The convergence is in fact uniform on $(\eta, 1)$, $0 < \eta < 1$.

By a standard Abelian theorem for Laplace transforms (see, e.g., Feller [**13**, p. 445])

$$\bar{V}(\phi) \sim \frac{c_2^*}{(\ln\phi)^{\beta_2}\phi^{\epsilon_2-1}L_2(\phi)}, \qquad \phi \to \infty,$$

implies that

$$W_\gamma(g) \sim \frac{c_2^*}{(g)^{1-\epsilon_2}\left[\ln\left(\frac{1}{g}\right)\right]^{\beta_2}L_2\left(\frac{1}{g}\right)}, \qquad g \to 0. \tag{2.9}$$

Now consider the ratio

$$\frac{f_{\gamma_0}(t,\mathbf{x})}{f_{\gamma_0}(t,\mathbf{0})} = e^{\boldsymbol{\alpha}_2^{(0)}\cdot\mathbf{x}}\,\frac{W_{\gamma_0}\left(t^{\alpha_1^{(0)}}e^{\boldsymbol{\alpha}_2^{(0)}\cdot\mathbf{x}}\right)}{W_{\gamma_0}\left(t^{\alpha_1^{(0)}}\right)}.$$

Then by relation (2.9)

$$\lim_{t\to 0}\frac{f_{\gamma_0}(t,\mathbf{x})}{f_{\gamma_0}(t,\mathbf{0})} = e^{\boldsymbol{\alpha}_2^{(0)}\cdot\mathbf{x}(\epsilon_2-1)}.$$

So

$$\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x} = \lim_{t \to 0} \left(\frac{1}{\epsilon_2 - 1} \right) \ln \left(\frac{f_{\gamma_0}(t, \boldsymbol{x})}{f_{\gamma_0}(t, \mathbf{0})} \right).$$

Thus for a fixed ϵ_2, assuming that the distribution of $\boldsymbol{x}$ is nondegenerate and of full rank, $\boldsymbol{\alpha}_2$ is uniquely identified.[10] (Note that the regressors may be categorical or continuous.)

We introduce the survivor function

$$S_{\gamma_i}(t, \boldsymbol{x}) = \int_{\underline{\underline{\Theta}}} \exp\left[-t^{\alpha_1^{(i)}} e^{\boldsymbol{\alpha}_2^{(i)} \cdot \boldsymbol{x} + \theta} \right] d\mu_i(\theta).$$

Set $\boldsymbol{x} = (0, \ldots, x_i, 0, \ldots, 0)$ where x_i is assumed to traverse an interval.[11] At $t = 1$,

$$S_{\gamma_i}(1, \boldsymbol{x}) = \int_{\underline{\underline{\Theta}}} \exp\left[-e^{\boldsymbol{\alpha}_2^{(i)} \cdot \boldsymbol{x}} e^{\theta} \right] d\mu_i(\theta).$$

Since $\boldsymbol{\alpha}_2^{(0)}$ is determined from the preceding argument, and x_i is defined on the continuum, μ is uniquely determined as a consequence of the uniqueness theorem for Laplace transforms.

Finally set $\boldsymbol{x} = \mathbf{0}$ and observe that $S_{\gamma_i}(t, \mathbf{0})$ has the representation

$$S_{\gamma_i}(t, \mathbf{0}) = \int_{\underline{\underline{\Theta}}} \exp\left[-t^{\alpha_1^{(i)}} e^{\theta} \right] d\mu_0(\theta).$$

Observe that $S_{\gamma_i}(t, \mathbf{0})$ may be viewed as a composite of monotone functions

$$H\left(t^{\alpha_1^{(i)}}\right) = \int_{\underline{\underline{\Theta}}} \exp\left[-t^{\alpha_1^{(i)}} \right] e^{\theta} \, d\mu_0(\theta).$$

To solve for $t^{\alpha_1^{(i)}}$ write $w = H(t^{\alpha_1^{(i)}})$ and observe that

$$t^{\alpha_1^{(i)}} = H^{-1}(w)$$

is uniquely determined by strict monotonicity and continuity of H. Then set $w = S_{\gamma_0}(t, \mathbf{0})$ and take logs to reach

$$\alpha_1^{(i)} = \ln H^{-1}\left[S_{\gamma_0}(t, \mathbf{0}) \right].$$

Thus

$$\gamma_0 = \left(\alpha_1^{(0)}, \boldsymbol{\alpha}_2^{(0)}, \mu_0 \right)$$

is uniquely determined given density (2.7) with a preassigned value of ϵ_2.

[10]Note that it is uniquely identified given ϵ_2. More precisely it is $(\epsilon_2 - 1)\boldsymbol{\alpha}_2$ that is uniquely identified.

[11]So x_i is a "continuous regressor."

Now note that if $d\mu(\theta)$ is a discrete mass distribution with

$$(2.10)\qquad d\mu(\theta)=\begin{cases}P_k & \text{if } \quad \theta=\theta_k,\ldots,\theta_{-1}<\theta_0<\theta_1<\ldots,\\ 0 & \text{otherwise}, \quad -\infty<k<\infty,\end{cases}$$

and $\theta_k \sim c|k|$, $k\to\infty$, then assuming that

$$P_k=\frac{c_i}{|k|^{\beta_i}e^{\epsilon_i k}L(k)}\qquad \text{for} \quad |k|>k^*,\quad k^*<\infty$$

where k^* is a uniform bound for all admissible mass distributions $\mathscr{M}$ and for $\beta_i>0$, $1>\epsilon_i\geq 0$, identifiability can be established by an argument identical in all essential details to the one just made. This completes our discussion of identifiability.

5 (Integrability). For any $\gamma\in\overline{\mathscr{A}}\times\overline{\mathscr{M}}$ we require

$$(2.11)\qquad \lim_{\rho\downarrow 0}E_{\gamma_0}\left(\ln\left[\frac{\sup_{\gamma':\,\delta(\gamma,\gamma')<\rho}f_{\gamma'}(T,\boldsymbol{X})}{f_{\gamma_0}(T,\boldsymbol{X})}\right]\right)^+<+\infty.$$

To verify this condition, we utilize the bound

$$f_\gamma(t,\boldsymbol{x})<\alpha_1 e^{-1}t^{-1}h(\boldsymbol{x})\qquad \forall\gamma\in\overline{\mathscr{A}}\times\overline{\mathscr{M}}.$$

(The bound is obtained by selecting the value of θ that maximizes $g_\alpha(t\,|\,\boldsymbol{x},\theta)$ in equation (2.1b). There is a unique maximizing value of θ.)

With this bound we then have a bound on the numerator of the argument of the logarithm in (2.11) given by

$$\sup_{\gamma':\,\delta(\gamma,\gamma')<\rho}f_{\gamma'}(t,\boldsymbol{x})<(\alpha_1+\rho)e^{-1}t^{-1}h(\boldsymbol{x})$$

and thus

$$(2.12)\qquad \left(\ln\left[\frac{\sup_{\gamma':\,\delta(\gamma,\gamma')<\rho}f_{\gamma'}(t,\boldsymbol{x})}{f_{\gamma_0}(t,\boldsymbol{x})}\right]\right)^+$$

$$<\left[\ln(\alpha_1+\rho)-1-\ln\alpha_1^{(0)}-\alpha_1^{(0)}\ln t-\boldsymbol{\alpha}_2^{(0)}\cdot\boldsymbol{x}-\ln W_{\gamma_0}(t,\boldsymbol{x})\right]^+$$

where

$$W_{\gamma_0}(t,\boldsymbol{x})=\int_{-\infty}^{\infty}\exp\left[\theta-t^{\alpha_1^{(0)}}e^{\alpha_2^{(0)}\cdot\boldsymbol{x}+\theta}\right]d\mu_0(\theta).$$

Utilizing the inequality

$$[f(X)+g(X)]^+\leq[f(X)]^+ + [g(X)]^+$$

where f and g are arbitrary real valued measurable functions,

$$(2.13)\qquad E\big[f(X)+g(X)\big]^{+}\leq E\big[f(X)\big]^{+}+E\big[g(X)\big]^{+}.$$

Combining (2.12) and (2.13) we conclude that

$$(2.14)\qquad E_{\gamma_0}\left(\ln\left[\frac{\sup_{\gamma':\delta(\gamma,\gamma')<\rho} f_{\gamma'}(t,\boldsymbol{x})}{f_{\gamma_0}(t,\boldsymbol{x})}\right]\right)^{+}$$

$$< E_{\gamma_0}\big[\ln(\alpha_1+\rho)-1-\ln\alpha_1^{(0)}\big]^{+}$$

$$+E_{\gamma_0}\Big[-\ln T^{\alpha_1^{(0)}}e^{\boldsymbol{\alpha}_2^{(0)}\cdot \boldsymbol{x}}W_{\gamma_0}(T,\boldsymbol{X})\Big]^{+}.$$

The first term on the right hand side is obviously bounded since the integrand is a constant. Moreover for all distributions of θ for which $\int_{-\infty}^{\infty}e^{\theta}\,d\mu(\theta)<\infty$, the second term is bounded provided that $\max_{1\leq i\leq l}E|X_i|<+\infty$. Thus the only class of distributions of θ that require special consideration are those for which

$$\int_{-\infty}^{\infty}e^{\theta}\,d\mu(\theta)=+\infty.$$

Initially we restrict attention to the class of absolutely continuous distributions, $d\mu(\theta)=m(\theta)d\theta$, satisfying (2.7). We assume in addition that

$$(2.15a)\qquad \overline{\lim_{\theta\to\infty}}\ \frac{(\ln\theta)^{2+\delta}}{L_2(\theta)}<\infty,$$

$$(2.15b)\qquad \overline{\lim_{\theta\to-\infty}}\ \frac{\big[\ln(-\theta)\big]^{2+\delta}}{L_1(\theta)}<\infty,$$

for some $\delta>0$.

Utilizing a standard Abelian theorem (see, e.g., Feller [**13**, p. 445]), densities satisfying (2.7) for $\epsilon_i=0$ and $\beta_i=1$ dominate those satisfying (2.7) for more general values of β_i and ϵ_i in the sense that

$$(2.16)\qquad \lim_{t^{\alpha_1}e^{\boldsymbol{\alpha}_2\cdot \boldsymbol{x}}\to 0\text{ or }\infty}\left(\frac{W_{\gamma}(t,x)}{W_{\gamma^*}(t,x)}\right)=0$$

where $\gamma^*=(\alpha_1^{(0)},\boldsymbol{\alpha}_2^{(0)},\mu)$ with μ having a density that satisfies (2.7) for $\beta=1$ and $\epsilon_i=0$ and $\gamma=(\alpha_1^{(0)},\boldsymbol{\alpha}_2^{(0)},\mu)$ with μ satisfying (2.7) with $1>\epsilon_i>0$ and $\beta_i>1$. Limit (2.16) plays a key role in the verification of (2.11).

Our strategy for verifying condition (2.11) in the case in which $\int_{-\infty}^{\infty}e^{\theta}\,d\mu(\theta)=+\infty$ is as follows. (a) We verify by explicit calculations that densities satisfying (2.7) with $\epsilon_i=0$ and $\beta_i=1$ yield finite upper bounds for (2.14). (b) Since $t^{\alpha_1}e^{\alpha_2\cdot x}W_{\gamma}(t,x)\to 0$ as $t^{\alpha_1}e^{\alpha_2\cdot x}\to 0$ or $+\infty$ for all densities satisfying (2.7), even for $\epsilon_i = 0$, we have as a consequence of (2.16) that

$$(2.17) \qquad 0 < \int_{\mathscr{B}} - \frac{t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} W_\gamma(t, \boldsymbol{x}) \ln\left[t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} W_\gamma(t, \boldsymbol{x})\right] h(\boldsymbol{x})\, dt\, d\boldsymbol{x}}{t}$$

$$< \int_{\mathscr{B}} - \frac{t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} W_{\gamma^*}(t, \boldsymbol{x}) \ln\left[t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} W_{\gamma^*}(t, \boldsymbol{x})\right] h(\boldsymbol{x})\, dt\, d\boldsymbol{x}}{t}$$

where $\mathscr{B}$ is either $\{(t, \boldsymbol{x}) : t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} < z_1\}$ or $\{(t, \boldsymbol{x}) : t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} > z_2\}$ for appropriately small values of z_1 and large values of z_2, respectively, where γ and γ^* are as defined above.

By virtue of this inequality, verification of integrability for densities satisfying (2.7) for $\beta_i = 1$ and $\epsilon_i = 0$ implies integrability for all densities for which (2.16) holds. In particular, integrability in (2.14) is satisfied for all densities characterized by (2.7) when $\beta_i > 1$ and $1 \geq \epsilon_i > 0$.

The proof of integrability is as follows. For $m(\theta) \sim c_2/\theta L_2(\theta)$, $\theta \to \infty$, with

$$\overline{\lim_{\theta \to \infty}} \frac{(\ln \theta)^{2+\delta}}{L_2(\theta)} < \infty,$$

let $e^\theta = \phi$ and define $\bar{v}(\phi) = m(\ln \phi)$ as in the verification of identifiability.

Then

$$W_\gamma(t, \boldsymbol{x}) = \int_{-\infty}^{\infty} \exp(-t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}} e^\theta) e^\theta m(\theta)\, d\theta$$

$$= \int_0^\infty \exp(-t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}} \phi) \bar{v}(\phi)\, d\phi.$$

By the Abelian theorem used in (2.9) for $\epsilon_2 = 0$,

$$W_\gamma(t, \boldsymbol{x}) \sim \frac{c_2}{t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}} \ln\left(\frac{1}{t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}}}\right) L_2\left(\ln\left(\frac{1}{t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}}}\right)\right)}$$

for $t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}} \to 0$. Given $\epsilon' > 0$ arbitrary, $\exists z_1(\epsilon')$ such that $t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}} < z_1$ implies

$$\left| \frac{t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}} W_\gamma(t, \boldsymbol{x})}{\dfrac{c_1}{\ln\left(\frac{1}{t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}}}\right) L_2\left(\ln\left(\frac{1}{t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}}}\right)\right)}} - 1 \right| < \epsilon'.$$

Define $\mathscr{B}_0 = \{(t, \boldsymbol{x}) : t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}} < z_1\}$. Since $t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}} W_\gamma(t, \boldsymbol{x}) \to 0$ as $t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}} \to 0$

$$\left[-\ln t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}} W_\gamma(t, \boldsymbol{x})\right]^+ = -\ln t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}} W_\gamma(t, \boldsymbol{x})$$

for $t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}}$ sufficiently small.

Define $\mathscr{B} = \mathscr{B}_0 \cap \{(t,\boldsymbol{x}): -\ln t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}} W_\gamma(t,\boldsymbol{x}) > 0\}$. The first step in verifying that

$$E_{\gamma_0}\Big[-\ln T^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} W_{\gamma_0}(T,\boldsymbol{X})\Big]^+ < \infty$$

is to show that

$$\int_{\mathscr{B}}\Big[-\ln t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} W_{\gamma_0}(t,\boldsymbol{x})\Big] f_{\gamma_0}(t,\boldsymbol{x})\,dt\,d\boldsymbol{x} < \infty. \tag{2.18}$$

Then an analogous estimate for $t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} \to \infty$ combined with the fact that, for this product bounded away from 0 and ∞, the quantity

$$\Big[-\ln t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} W_{\gamma_0}(t,\boldsymbol{x})\Big]^+$$

is also bounded is enough to establish that (2.14) is bounded. Details are given below but we first establish (2.18).

For this observe that

$$\begin{aligned}
&\int_{\mathscr{B}}\Big[-\ln t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} W_{\gamma_0}(t,\boldsymbol{x})\Big] f_{\gamma_0}(t,\boldsymbol{x})\,dt\,d\boldsymbol{x} \\
&\quad= \int_{\mathscr{B}}\Big[-\ln t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} W_{\gamma_0}(t,\boldsymbol{x})\Big] \frac{\alpha_1^{(0)} t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} W_{\gamma_0}(t,\boldsymbol{x})}{t} h(\boldsymbol{x})\,dt\,d\boldsymbol{x} \\
&\quad= \int_{\mathscr{B}}\left\{-\ln\left[\frac{c_2}{\ln\left(\frac{1}{t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}}}\right) L_2\left(\ln\left(\frac{1}{t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}}}\right)\right)}\right]\right\} \\
&\qquad \frac{\alpha_1 c_2 h(\boldsymbol{x})\,dt\,d\boldsymbol{x}}{t \ln\left(\frac{1}{t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}}}\right) L_2\left(\ln\left(\frac{1}{t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}}}\right)\right)}.
\end{aligned} \tag{2.19}$$

The term in braces can be rewritten as

$$\begin{aligned}
&-\left\{\ln c_2 - \ln\Big[-\ln\big(t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}}\big)\Big] - \ln L_2\left(\ln\left(\frac{1}{t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}}}\right)\right)\right\} \\
&\qquad \sim \left\{\ln\Big[-\ln\big(t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}}\big)\Big]\right\}, \qquad t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} \to 0,
\end{aligned}$$

since

$$\frac{L_2\left(\ln\left(\frac{1}{t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}}}\right)\right)}{\ln\left(\frac{1}{t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}}}\right)} \to 0, \qquad t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} \to 0.$$

Thus the integral (2.19) can be written asymptotically as

$$(2.20)\quad \int_{\mathscr{B}} \frac{\alpha_1^{(0)} c_2 h(\boldsymbol{x}) \ln\left[-\alpha_1^{(0)} \ln t - \boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}\right] dt\, d\boldsymbol{x}}{t\left[-\alpha_1^{(0)} \ln t - \boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}\right] L_2\left(-\alpha_1^{(0)} \ln t - \boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}\right)}$$

$$= \alpha_1^{(0)} c_2$$

$$\cdot \int_{\mathscr{X}} h(\boldsymbol{x}) \int_0^{\bar{z} e - \left(\frac{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}}{\alpha_1^{(0)}}\right)} \frac{\ln\left[-\alpha_1^{(0)} \ln t - \boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}\right] dt\, d\boldsymbol{x}}{t\left[-\alpha_1^{(0)} \ln t - \boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}\right] L_2\left(-\alpha_1^{(0)} \ln t - \boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}\right)}$$

where $\bar{z} = (z_1)^{1/\alpha_1^{(0)}}$. The last double integral is an explicit formulation of the limits of integration on the set

$$\mathscr{B} = \left\{(t, \boldsymbol{x}) : t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} < z_1\right\}$$

with z_1 small enough so that

$$\ln t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} W_{\gamma_0}(t, \boldsymbol{x}) < 0.$$

Bringing in the change of variable,

$$t = \bar{z} \exp\left(-\frac{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}}{\alpha_1^{(0)}}\right) u,$$

and letting $\ln u = y$, equation (2.20) becomes

$$\int_{-\infty}^{0} \frac{\ln\left[-\alpha_1^{(0)} \ln \bar{z} - \alpha_1^{(0)} y\right] dy}{\left[-\alpha_1^{(0)} \ln \bar{z} - \alpha_1^{(0)} y\right] L_2\left(-\alpha_1^{(0)} \bar{z} - \alpha_1^{(0)} y\right)}$$

$$< \int_{-\infty}^{0} \frac{\ln\left[-\alpha_1^{(0)} \ln \bar{z} - \alpha_1^{(0)} y\right] dy}{\left[-\alpha_1^{(0)} \ln \bar{z} - \alpha_1^{(0)} y\right]\left[\ln\left(-\alpha_1^{(0)} \bar{z} - \alpha_1^{(0)} y\right)\right]^{2+\delta} c''} < \infty$$

where the last inequality is a consequence of

$$\overline{\lim_{\theta \to \infty}} \frac{(\ln \theta)^{2+\delta}}{L_2(\theta)} < \infty$$

so that

$$(\ln \theta)^{2+\delta} < L_2(\theta)(c' + \epsilon') \qquad \text{for} \quad \theta > \theta^*(\epsilon')$$

where $c'' = 1/(c' + \epsilon')$, and $c' > 0$.

Since the integral with respect to t in (2.20) is finite and independent of $\boldsymbol{x}$, the full double integral (2.20) is finite because

$$\int_{\mathscr{X}} h(\boldsymbol{x})\,d\boldsymbol{x} = 1.$$

Now consider $t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2 \cdot \boldsymbol{x}} \to \infty$ and observe that

$$m(\theta) \sim \frac{c_1}{-\theta L_1(\theta)}, \qquad \theta \to -\infty,$$

implies

$$\bar{V}(\phi) \sim \frac{c_1 \phi}{-\ln \phi L_1(\ln \phi)}, \qquad \phi \to 0.$$

Exactly the same Abelian theorem used before implies that

$$W_\gamma(t, \boldsymbol{x}) \sim \frac{c_1}{-t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} \ln\left(\frac{1}{t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}}}\right) L_1\left(\ln\left(\frac{1}{t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}}}\right)\right)}$$

for $t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} \to \infty$.

Repeating the previous calculations with $\mathscr{B}^*$ defined as

$$\mathscr{B}^* = \left\{(t, \boldsymbol{x}) : t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} > z_2\right\}$$

yields

$$(2.21) \qquad \int_{\mathscr{B}^*} \left[-\ln t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} W_{\gamma_0}(t, \boldsymbol{x})\right] f_{\gamma_0}(t, \boldsymbol{x})\,dt\,d\boldsymbol{x} < \infty.$$

For the final step it is useful to define

$$\mathscr{B}^{**} = \left\{(t, \boldsymbol{x}) : z_1 \le t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} \le z_2\right\}.$$

Note that

$$(2.22) \qquad \sup_{\mathscr{B}^{**}} \left[-\ln t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} W_{\gamma_0}(t, \boldsymbol{x})\right]^+ < \infty.$$

Combining (2.18), (2.21), and (2.22) we conclude that

$$E_{\gamma_0}\left[-\ln T^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} W_{\gamma_0}(T, \boldsymbol{X})\right]^+ < \infty$$

for densities satisfying condition (2.7) with $\epsilon_i = 0$. Condition (2.16) follows from application of the previously utilized Abelian theorems using the asymptotic formula (2.9) in the general case. Hence we have the inequality in (2.17), and integrability holds for all densities satisfying (2.7).

A parallel proof can be constructed for frequencies satisfying (2.10). For the sake of brevity we delete this verification. We also note that it appears necessary to make such asymptotic assumptions in the case $E(e^{\theta}) = \infty$ since bounds on the rate of growth of $W_{\gamma_0}(t, \boldsymbol{x})$ for

$$t^{\alpha_1^{(0)}} e^{\boldsymbol{\alpha}_2^{(0)} \cdot \boldsymbol{x}} \to 0 \text{ and } \infty$$

are required in order to establish integrability.

This completes the proof of consistency for the uncensored Weibull model. In summary, the following assumptions are critical. (1) $\boldsymbol{\alpha}_2$ and α_1 are restricted to lie in finite open intervals. (2) $\max_{1 \leq i \leq l} E|X_i| < \infty$, and the distribution of $\boldsymbol{X}$ is of full rank. (3) μ_0 is nondefective. (4) To verify identifiability either $E(e^{\theta}) < \infty$ and the class of admissible distributions is uniformly integrable but no regressor need appear in the model or else $E(e^{\theta}) = \infty$, at least one regressor defined on the continuum is present in the model, and the tail behavior of the mixing distribution is restricted. (5) To verify integrability either $E(e^{\theta}) < \infty$ or else $E(e^{\theta}) = \infty$ and the true distribution of unobservables satisfies all the restrictions on tail behavior needed for identifiability for $E(e^{\theta}) = \infty$ plus a few additional conditions. Note that if $E(e^{\theta}) < \infty$, the proof of consistency is drastically simplified.

C. The Censored Weibull Model

Let $(T, \boldsymbol{X})$ be distributed as in (2.1a) and (2.1b), and define C to be a positive random variable independent of $(T, \boldsymbol{X})$ whose values represent censoring times for T. Define $v_C(t)$ to be the probability density associated with C. We assume $v_C(t)$ is known.[12] Furthermore, let T^* be a fixed positive number interpreted as the duration of the observation period, or equivalently, the length of a study on individuals whose dynamics are governed by (2.1a) and (2.1b). With this apparatus at hand, the observed random vector is defined to be $(Y, \boldsymbol{X})$ where $Y = \min(T, C)$ when $\min(T, C) < T^*$ and $Y = T^*$ when $\min(T, C) \geq T^*$. The corresponding frequency function is

$$\bar{f}_{\gamma}(t, \boldsymbol{x}) = \begin{cases} f_{\gamma}(t, \boldsymbol{x}) P(C > t) + S_{\gamma}(t, \boldsymbol{x}) v_C(t) & \text{for} \quad 0 \leq t < T^*, \\ S_{\gamma}(T^*, \boldsymbol{x}) P(C > T^*) & \text{for} \quad t = T^*, \end{cases}$$

where

$$S_{\gamma}(t, \boldsymbol{x}) = \int_t^{\infty} f_{\gamma}(s, \boldsymbol{x})\, ds.$$

Verification of the KW conditions: 1.

$$\bar{f}_{\gamma}(t, \boldsymbol{x} \mid \theta) = \begin{cases} f_{\gamma}(t, \boldsymbol{x} \mid \theta) P(C > t) + S_{\gamma}(t, \boldsymbol{x} \mid \theta) v_C(t) & \text{for } 0 \leq t < T^*, \\ S_{\gamma}(T^*, \boldsymbol{x} \mid \theta) P(C > T^*) & \text{for } t = T^*, \end{cases}$$

[12] Analysis of the case for which $v_C(t)$ is unknown is not attempted in this paper.

is a density with respect to $\overline{dt}dx_{i_1}, \ldots, dx_{i_k}d\zeta_{i_{k+1}}, \ldots, d\zeta_{i_l}$ where $dx_{i_1}, \ldots, d\zeta_{i_l}$ is defined as in the uncensored case while $\overline{dt}$ is Lebesgue measure on $[0, T^*)$ augmented by unit point mass δ_{T^*} at T^*.

2 (Continuity).

$$(2.23)\qquad |\bar{f}_{\gamma_i}(t,x) - \bar{f}_\gamma(t,x)| \le h(x)\left|\int_{\underline{\underline{\Theta}}} g_{\alpha^{(i)}}(t \mid x,\theta)\, d\mu_i(\theta) - \int_{\underline{\underline{\Theta}}} g_\alpha(t \mid x,\theta)\, d\mu(\theta)\right|$$
$$\times P(C > t) + v_C(t)|S_{\gamma_i}(t,x) - S_\gamma(t,x)|$$
$$\text{for}\quad 0 \le t < T^*,$$

and

$$|\bar{f}_{\gamma_i}(T^*,x) - \bar{f}_\gamma(T^*,x)| = P(C > T^*)|S_{\gamma_i}(T^*,x) - S_\gamma(T^*,x)|.$$

The first term on the right hand side of (2.23) vanishes as $i \to \infty$ by the argument used in the uncensored case. Thus it is enough to show that $\delta(\gamma_i, \gamma) \to 0$, $i \to \infty$, implies that

$$\lim_{i\to\infty} |S_{\gamma_i}(t,x) - S_\gamma(t,x)| = 0, \qquad \forall (t,x) \in (0, T^*] \times \mathscr{X}.$$

To this end recall that

$$S_\gamma(t,x) = h(x)\int_{\underline{\underline{\Theta}}} S_\alpha(t \mid x,\theta)\, d\mu(\theta)$$

where

$$S_\alpha(t \mid x,\theta) = \exp\left[-t^{\alpha_1} e^{(\alpha_2 \cdot x + \theta)}\right].$$

Proceeding as in the uncensored case we have that

$$(2.24)\qquad |S_{\gamma_i}(t,x) - S_\gamma(t,x)|$$
$$\le h(x)\int_{\underline{\underline{\Theta}}} |S_{\alpha^{(i)}}(t \mid x,\theta) - S_\alpha(t \mid x,\theta)|\, d\mu_i(\theta)$$
$$+ h(x)\left|\left\{ S_\alpha(t \mid x,\theta)\left[\mu_i(\theta) - \mu(\theta)\right]\Big|_{\theta_0}^{\theta_1}\right.\right.$$
$$\left.\left. - \int_{\underline{\underline{\Theta}}} \left[\mu_i(\theta) - \mu(\theta)\right] \frac{\partial}{\partial\theta} S_\alpha(t \mid x,\theta)\, d(\theta)\right\}\right|$$

for $\theta_1 = \infty$ and $\theta_0 = -\infty$. Then, since $S_\alpha(t \mid x,\theta)$ is uniformly continuous in $\boldsymbol{\alpha}$

and θ, we have

$$\int_{\underline{\underline{\Theta}}} |S_{\alpha^{(i)}}(t \mid x, \theta) - S_{\alpha}(t \mid x, \theta)| \, d\mu_i(\theta) \to 0, \qquad i \to \infty,$$

as in our earlier argument. The only potential source of difficulty is the last integral where we require a bound on

$$\left| \frac{\partial S_{\boldsymbol{\alpha}}(t \mid x, \theta)}{\partial \theta} \right|$$

which decays exponentially in $|\theta|$ as $|\theta| \to \infty$. But

$$\left| \frac{\partial S_{\boldsymbol{\alpha}}(t \mid x, \theta)}{\partial \theta} \right| = t^{\alpha_1} e^{(\boldsymbol{\alpha}_2 \cdot x + \theta)} \exp\left[- t^{\alpha_1} e^{(\boldsymbol{\alpha}_2 \cdot x + \theta)} \right].$$

Thus if we define $\theta^*(t, x) = \sup(\theta > 0 : t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot x + \theta} \leq 2\theta)$, then for $\theta > \theta^*(t, x)$, $\exp[\theta - t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot x + \theta}] \leq e^{-\theta}$. Then we have

$$\left| \frac{\partial S_{\boldsymbol{\alpha}}(t \mid x, \theta)}{\partial \theta} \right| < \begin{cases} t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot x} e^{-|\theta|} & \text{for} \quad \theta < 0 \quad \text{and} \quad \theta > \theta^*(t, x), \\ t^{\alpha_1} d^{\boldsymbol{\alpha}_2 \cdot x} C(t, x) e^{-|\theta|} & \text{for} \quad 0 \leq \theta \leq \theta^*(t, x), \end{cases}$$

where

$$C(t, x) = \frac{\sup_{\theta \in [0, \theta^*(t, x)]} \exp\left[\theta - t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot x + \theta} \right]}{e^{-\theta^*(t, x)}}.$$

Therefore,

$$\left| \int_{\underline{\underline{\Theta}}} \left[\mu_i(\theta) - \mu(\theta) \right] \frac{\partial S_{\alpha}(t \mid x, \theta)}{\partial \theta} \, d\theta \right|$$

$$\leq \left[1 + C(t, x) \right] t^{\alpha_1} e^{\boldsymbol{\alpha}_2 \cdot x} \int_{\underline{\underline{\Theta}}} | \mu_i(\theta) - \mu(\theta) | e^{-|\theta|} \, d\theta \to 0 \quad \text{as } i \to \infty.$$

3. Measurability of $\omega_{\gamma}(t, x; \rho) = \sup_{\gamma' : \delta(\gamma, \gamma') < \rho} \bar{f}_{\gamma'}(t, x)$ follows from the identical argument used in the uncensored case.

4 (Identifiability). First observe that for $\bar{f}_{\gamma}(t, x)$, $P(C > t)$ and $v_C(t)$ are known functions when $0 \leq t \leq T^*$, and we may interpret the defining relation for $\bar{f}_{\gamma}(t, x)$ as a partial differential equation for $S_{\gamma}(t, x)$. In particular we have

$$\text{(2.25)} \qquad \frac{-\partial S_{\gamma}(t, x)}{\partial t} + \frac{v_C(t)}{P(C > t)} S_{\gamma}(t, x) = \frac{\bar{f}_{\gamma}(t, x)}{P(C > t)}$$

for $0 \leq t < T^*$, $x \in \mathscr{X}$, and

$$S_\gamma(T^*, x) = \frac{\bar{f}_\gamma(T^*, x)}{P(C > T^*)}.$$

For each $x \in \mathscr{X}$, the unique solution of (2.25) is an analytic function on $[0, T^*]$ having the representation

$$(2.26) \qquad S_\gamma(t, x) = \int_{\underline{\Theta}} \exp\left[-t^{\alpha_1} e^{\alpha_2' \cdot x + \theta}\right] d\mu(\theta).$$

However, we are assuming that a representation of the form (2.26) holds for all t in $(0, +\infty)$, despite the fact that $S_\gamma(t, x)$ which solves (2.25) is only determined on $[0, T^*]$. The values of $S_\gamma(t, x)$ are uniquely determined for $t \in (T^*, \infty)$ via analytic continuation of its power series on $[0, T^*]$. With this full survivor function at hand, we establish uniqueness of α_1, $\boldsymbol{\alpha}_2$, and μ via the same argument as in the uncensored case.

From an empirical point of view, recovery of $S_\gamma(t, x)$ on $(0, +\infty)$ from a knowledge of $S_\gamma(t, x)$ on $[0, T^*]$ is only possible if one assumes that the unobserved portion of the survivor function has the representation (2.26) and is thus the analytic continuation of the observed portion of the function.

5 (Integrability). The verification of integrability in the censored case proceeds exactly as in the uncensored case except that now t is bounded above by truncation point T^*.

D. General Proportional Hazard Models With and Without Time Varying Covariates

In this subsection we first consider the conditions required for identification of a general proportional hazard model. We then briefly indicate how to verify the remaining KW conditions for this model.

Our identification analysis is of interest in its own right because it supplements and complements the recent important analysis of identification in proportional hazard models presented by Elbers and Ridder [**12**]. Their analysis requires that $E(e^\theta) < \infty$ and so rules out a variety of candidate mixing distributions (e.g., Cauchy, certain members of the Gamma family, and other distributions without moment generating functions).[13] Imposing this restriction in a computational algorithm is difficult. Our analysis does not require this condition and so in this regard is more general. On the other hand we impose a condition that they do not and so our criteria and theirs are complementary.

[13] Thus for

$$f(\theta) = \frac{\lambda e^{-\lambda\theta}(\lambda\theta)^{r-1}}{\Gamma(r)},$$

if $\lambda < 1$, $E(e^\theta) = \infty$.

Identifiability for Classes of Proportional Hazard Functions

Consider the class of frequency functions

$$(2.27)\qquad f_\gamma(t,\boldsymbol{x}) = h(\boldsymbol{x})z'(t)e^{\zeta(\boldsymbol{x})}\int_{\underline{\underline{\Theta}}}\exp\left[\theta - z(t)e^{(\zeta(\boldsymbol{x})+\theta)}\right]d\mu(\theta)$$

$$= h(\boldsymbol{x})\int_{\underline{\underline{\Theta}}} g_\alpha(t\mid\zeta(\boldsymbol{x}),\theta)\,d\mu(\theta)$$

so

$$z'(t)e^{\zeta(\boldsymbol{x})}\exp\left[\theta - z(t)e^{(\zeta(\boldsymbol{x})+\theta)}\right] = g_\alpha(t\mid\zeta(\boldsymbol{x}),\theta)$$

where $\gamma = (z(t), \zeta(\boldsymbol{x}), \mu)$.

Elbers and Ridder **[12]** establish the identifiability of (2.27) under the following conditions. Let $\mathscr{X}$ be an open subset of $\mathbb{R}_l$ and assume that (i) $\int_{\underline{\underline{\Theta}}} e^{\theta}\,d\mu(\theta) = 1$, (ii) $\zeta(\boldsymbol{x})$ is differentiable and nonconstant on $\mathscr{X}$ with at least one argument of $\boldsymbol{x}$ defined on the continuum,[14] (iii) $z(t)$ is the integral of a nonnegative integral function and $\lim_{t\to\infty} z(t) = \infty$. Their conditions supplemented by a restriction of the admissible class of distributions to be uniformly integrable, satisfy the Kiefer–Wolfowitz identifiability condition.

In this paper we present alternative criteria derived in a companion paper (Heckman and Singer **[21]**) that permit (i) to be violated so $E(e^{\theta}) = \infty$. The analysis presented in this section generalizes our identification analysis presented in the Weibull case. We make the following assumptions:

(a) $z \in \underline{\underline{Z}} = \{z(t), t \geq 0 : z(t)$ is a continuous nonnegative increasing function with $z(0) = 0$, and $\exists c > 0$ and $t_+ > 0$ *not* depending on the function $z(t)$ such that $z(t_+) = c$, where c is a known constant$\}$.

We label the condition that $z(t_+) = c$ for some fixed $c > 0$ and $t_+ > 0$ for all admissible z the *cross over condition*. In particular this condition is satisfied for the Weibull integrated hazard $z(t) = \exp t^{\alpha}$ with $t_+ = 1$ and $c = e$.

(b) $\zeta = \underline{\underline{\mathscr{Z}}} = \{\zeta(\boldsymbol{x}), \boldsymbol{x} \in \mathscr{X} : \zeta$ is nonconstant on $\mathscr{X}$, $\exists$ at least one coordinate x_i defined on an interval such that $\zeta((0, \ldots, 0, x_i, 0, \ldots, 0))$ traverses $(-\infty, +\infty)$ as x_i traverses its domain of definition, $\mathbf{0}$ is in the support of $\boldsymbol{X}$ and $\zeta(\mathbf{0}) = 0$. The distribution of $\boldsymbol{X}$ is of full rank$\}$.

(c) $\mu \in \mathscr{M}_0 =$ all probability distributions on $\underline{\underline{\Theta}}$ which are restricted to be a member of the class of distributions specified by (2.7) or (2.10).

[14] In their Appendix they relax the assumption that at least one component of $\boldsymbol{X}$ is defined on the continuum. However ζ must be nonconstant on $\mathscr{X}$.

The assumptions about $\underline{\underline{\Theta}}$, $\mathscr{X}$, and $h(\boldsymbol{x})$ are retained from the analysis of the Weibull case presented in subsection A.

For the sake of brevity we restrict μ_0 to be a member of the class of absolutely continuous probability densities (2.7). The analysis for μ_0 characterized by (2.10) proceeds along similar lines.

Consider the ratio

$$\frac{f_{\Upsilon_0}(t,\boldsymbol{x})}{f_{\Upsilon_0}(t,\boldsymbol{0})} = e^{\zeta(\boldsymbol{x})} \frac{W_{\Upsilon_0}\left(z(t)e^{\zeta(\boldsymbol{x})}\right)}{W_{\Upsilon_0}(z(t))}.$$

From the relation (2.9),

$$\lim_{t\to 0} \frac{f_{\Upsilon_0}(t,\boldsymbol{x})}{f_{\Upsilon_0}(t,\boldsymbol{0})} = e^{\zeta(\boldsymbol{x})(\epsilon_2 - 1)}$$

so

$$\zeta(\boldsymbol{x}) = \frac{1}{(\epsilon_2 - 1)} \lim_{t\to 0} \ln\left(\frac{f_{\Upsilon_0}(t,\boldsymbol{x})}{f_{\Upsilon_0}(t,\boldsymbol{0})}\right).$$

Thus from assumptions (b) and (c), $\zeta(x)$ is uniquely determined.

Next set $\boldsymbol{x} = (0, \dots, x_i, \dots, 0)$ where x_i is defined on the continuum and $t = t_+$. Without loss of generality set $c = 1$ in condition (a). The survivor function may be written as

$$S(t_+, x) = \int_{\underline{\underline{\Theta}}} \exp\left[-z(t_+)e^{\zeta(x)+\theta}\right] d\mu(\theta).$$

Since $\zeta(x)$ has already been uniquely determined, $\mu(\theta)$ is uniquely determined from the uniqueness theorem for LaPlace transforms.

Finally, set $\boldsymbol{x} = 0$ and observe that $S(t, \boldsymbol{0})$ has the representation

$$S(t,\boldsymbol{0}) = \int_{\underline{\underline{\Theta}}} \exp\left(-z(t)e^{\theta}\right) d\mu(\theta).$$

Now $S(t, \boldsymbol{0})$ may be viewed as a composite of monotone functions, $H(z(t))$, where

$$H(z) = \int_{\underline{\underline{\Theta}}} \exp\left(-z(t)e^{\theta}\right) d\mu(\theta).$$

To solve for $z(t)$ write $w = H(z)$ and observe that $z = H^{-1}(w)$ is uniquely determined by strict monotonicity and continuity of $H(\cdot)$. Then set $w = S(t, \boldsymbol{0})$ and we have that

$$z(t) = H^{-1}(S(t, \boldsymbol{0}))$$

is uniquely determined. This completes our proof.

The preceding analysis assumed that the regressors are time invariant. We briefly consider the consequences of relaxing this assumption. The conditional survivor function for a proportional hazard model with time varying covariates may be defined as

$$(2.28)\qquad S_{\gamma}(t \mid \boldsymbol{x}) = \int_{\underline{\underline{\Theta}}} \exp - \left[z(t) \Phi(\boldsymbol{x}(t), t, \boldsymbol{x}_0) e^{\theta} \right] d\mu(\theta)$$

where $z(t)$ is a nonnegative increasing function and

$$(2.29)\qquad \Phi(\boldsymbol{x}(t), t, \boldsymbol{x}_0) = \int_0^t \phi(\boldsymbol{x}(u))\, du + \eta(\boldsymbol{x}(0))$$

for $\boldsymbol{x}(0) = \boldsymbol{x}_0$ and where

$$\Phi \in \underline{\underline{\Phi}} = \{ \Phi(\boldsymbol{x}(t), t, \boldsymbol{x}_0), \text{ a positive function, } \boldsymbol{x}(t) \in \mathscr{X} : \Phi(\cdot) \text{ is a}$$

$$\text{mapping from } R_l \to R_1 \text{ which is nonconstant on } \mathscr{X},$$

$$\exists \text{ at least one coordinate } x_i(t) \text{ defined on an interval so that}$$

$$\Phi((0, \ldots, x_i(t), \ldots, 0)) \text{ traverses } (-\infty, \infty) \text{ as } x_i(t)$$

$$\text{traverses its domain of definition, } \mathbf{0} \text{ is in the support of } \boldsymbol{X},$$

$$\text{and } \Phi(\mathbf{0}) = 1 \text{ and the distribution of } \boldsymbol{x} \text{ is of full rank} \}.$$

This definition is sufficiently general to encompass constants for some or all components of $\boldsymbol{x}(t)$. The components of $\boldsymbol{x}(t)$ may be deterministic time trends or realizations of stochastic processes with initial conditions $\boldsymbol{x}_0$. In all cases we proceed conditionally on realized values.

Assume that sample values of $\boldsymbol{x}_0$ are drawn from density $h(\boldsymbol{x}_0)$. This excludes the case of common time trended variables for all observations. With these assumptions and definitions in hand, we could mimic the proof presented above. For the sake of brevity we delete the proof.

For empirically relevant cases in which $\boldsymbol{x}(t)$ consists of time trended variables common to all observations, Φ cannot be varied independently of z. For assumed functional forms for z and Φ it is sometimes possible to identify both functions using the argument presented above provided that there exists at least one coordinate of x, defined on a continuum that varies across individuals. For example, if $z = t^{\alpha_1}$ and $\Phi = \beta_1 \ln t e^{\beta_2 \boldsymbol{x}}$ where $\boldsymbol{x}$ is a constant, we may redefine z in the previous proof as $z^* = t^{\alpha_1} \cdot \beta_1 \ln t$ and hence can separately identify α_1, β_1, and β_2 provided that $\boldsymbol{x}$ is defined on a continuum and varies over individuals. In the case $z = t^{\alpha_1}$, $\Phi = t^{\alpha_1} e^{\beta_2 \boldsymbol{x}}$, α_1 and β_1 cannot be separately identified while $\alpha_1 + \beta_1$, and β_2 can be identified in these combinations. There is little in a general way that can be said about this case, except to say that with no restrictions placed on z and Φ, the model is not identified. This concludes our discussion of identifiability in the general model.

For the sake of brevity, we work with equation (2.27) in considering verification of the remaining KW conditions. We briefly indicate how to modify the analysis to incorporate time varying exogenous variables.

Continuity. In order to verify the Kiefer–Wolfowitz conditions in this more general setting, we require that z and ζ be parameterized in terms of a finite set of parameters α.

We impose the following restrictions: (i) $\mathscr{A} = I_1 \times \cdots \times I_l$ where each I_j is a finite interval $1 \leq j \leq l$; (ii) for $\boldsymbol{\alpha} \in \overline{\mathscr{A}} \setminus \mathscr{A}$ and $\forall \mu \in \underline{\underline{\mathscr{M}}}$, $\int_{\underline{\underline{\Theta}}} g_{\boldsymbol{\alpha}}(t \mid \boldsymbol{x}, \theta)\, d\mu(\theta) = 0$ for all $(t, \boldsymbol{x})$ except possibly a set of points of $d\nu = dt\, dx_{i_1} \ldots dx_{i_k}\, d\zeta_{i_{k+1}} \ldots d\zeta_{i_l}$ measure zero; (iii) $g_{\boldsymbol{\alpha}}(t \mid \boldsymbol{x}, \theta)$ is uniformly continuous in $\boldsymbol{\alpha}$ and θ; (iv) $\lim_{\theta \to \infty} g_{\boldsymbol{\alpha}}(t \mid \boldsymbol{x}, \theta) = 0$ except possibly on a set of points $(t, \boldsymbol{x})$ of $d\nu$-measure 0.

These assumptions enable us to duplicate the continuity argument presented for the Weibull model. At the time of this writing, it is not obvious to us how to weaken these assumptions and still establish continuity.

The only exception is for a special family of kernels for which (iii) is not satisfied but where one can show by direct calculation that

$$\lim_{i \to \infty} \int_{\underline{\underline{\Theta}}} \left| g_{\boldsymbol{\alpha}^{(i)}}(t \mid \boldsymbol{x}, \theta) - g_{\boldsymbol{\alpha}}(t \mid \boldsymbol{x}, \theta) \right| d\mu_i(\theta) = 0.$$

The only nonobvious generalization of the Weibull proof for the general model is a bound on

$$\left| \frac{\partial g_{\boldsymbol{\alpha}}(t \mid \boldsymbol{x}, \theta)}{\partial \theta} \right|$$

which is necessary to complete the continuity proof. For this, observe that

$$\begin{aligned} \frac{\partial g_{\boldsymbol{\alpha}}(t \mid \boldsymbol{x}, \theta)}{\partial \theta} &= g_{\boldsymbol{\alpha}}(t \mid \boldsymbol{x}, \theta)\left[1 - z(t) e^{\zeta(\boldsymbol{x})}\right] \\ &= z'(t) e^{\zeta(\boldsymbol{x})}\left[1 - z(t) e^{\zeta(\boldsymbol{x})}\right] \exp\left[\theta - z(t) e^{(\zeta(\boldsymbol{x}) + \theta)}\right]. \end{aligned}$$

Then arguing as in the Weibull case we have

$$\left| \frac{\partial}{\partial \theta} g_{\boldsymbol{\alpha}}(t \mid \boldsymbol{x}, \theta) \right| < z'(t) e^{\zeta(\boldsymbol{x})}\left[1 + z(t) e^{\zeta(\boldsymbol{x})}\right] \max(1, C(t, \boldsymbol{x})) e^{-|\theta|}$$

where

$$C(t, x) = \frac{\sup_{\theta \in [0, \theta^*(t, \boldsymbol{x})]} \exp\left[2\theta - z(t) e^{(\zeta(\boldsymbol{x}) + \theta)}\right]}{e^{-\theta^*(t, \boldsymbol{x})}}$$

where $\theta^*(t, \boldsymbol{x}) = \sup(\theta > 0 : z(t) e^{\zeta(\boldsymbol{x}) + \theta} \leq 3\theta)$.

Note that this bound is correct even if $\boldsymbol{x}$ depends on time. This concludes our discussion of continuity in the general model.

Integrability. Verification of integrability in the general case for which $\int_{-\infty}^{\infty} e^{\theta}\, d\mu(\theta) = \infty$ hinges on the asymptotic behavior of

$$\int_{-\infty}^{\infty} \exp\left[-z(t) e^{\zeta(\boldsymbol{x})}\right] e^{\theta}\, d\mu(\theta).$$

An Abelian argument similar to the one used in the Weibull case can be used except now we investigate the asymptotic behavior of functions as $z(t)e^{\zeta(x)} \to 0$ and ∞. To produce explicit formulae requires adopting a parametric specification for $z(t)$ and $\zeta(\boldsymbol{x})$. However the general strategy—bounding integrals over sets $(t, \boldsymbol{x})$ for which $z(t)e^{\zeta(x)} < z_1$ or $z(t)e^{\zeta(x)} > z_2$—used in the Weibull case is a general method for establishing integrability.

As in the Weibull case if $\int_{-\infty}^{\infty} e^{\theta}\, d\mu(\theta) < \infty$ and $E|\zeta(\boldsymbol{x})| < \infty$, it is trivial to establish integrability.

E. Standard Errors of the Estimators

The preceding analyses have established the existence of a consistent maximum likelihood estimator for a general class of duration models. They are silent on the derivation of the sampling distribution of the estimators. The derivation of these sampling distributions is a nontrivial task left for future work.

We refer to the analysis of Efron [**11**] and recommend versions of his jackknife and bootstrap estimators of the variance. In our problem, these procedures amount to using subsamples of the data to maximize the likelihood function and to estimate the sampling distribution of the estimators from the histogram of the estimators obtained from the subsamples.

3. CHARACTERIZATION AND COMPUTATION OF THE NONPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATOR

Kiefer and Wolfowitz [**29**] offer no guidance on the computation of the NPMLE.[15] For special models, the NPMLE is characterized in papers by Simar [**40**] and Laird [**30**]. These authors also present explicit algorithms for computing the NPMLE. Lindsay [**31, 32, 33**] presents a very general characterization of the NPMLE which motivates the analysis presented in this paper.

In the notation of the preceding section, the maximum likelihood estimator for $(\boldsymbol{\alpha}, \mu)$ is the solution to the problem

$$(3.1)\qquad \operatorname*{Sup}_{\gamma \in \Gamma} \mathscr{L} = \operatorname*{Sup}_{\gamma \in \Gamma} \sum_{i=1}^{I} \ln f_\gamma(t_i \mid \boldsymbol{x})$$

$$= \operatorname*{Sup}_{\mu \in \mathscr{M}, \alpha \in A} \sum_{i=1}^{I} \ln\Big(\int g_\alpha(t_i \mid \boldsymbol{x}, \theta)\, d\mu(\theta) \Big)$$

where $\mathscr{M} = \{ \mu : \mu(\theta) \geq 0$ is nondecreasing, right continuous and $\int_\Theta d\mu(\theta) = 1\}$ and I denotes the number of observations.

Lindsay [**33**] provides the following characterization of the NPMLE.

[15]For this reason, Malinvaud [**35**, p. 402] states that the analysis of Kiefer and Wolfowitz " . . . remain[s] purely theoretical for the present."

THEOREM 3.1: *For a fixed* $\boldsymbol{\alpha}$ *and a finite sample, the* NPMLE *of a mixing distribution in an identified model is a finite mixture with at most* I^* *points of increase where* I^* *is the number of distinct values of* $(t_i, \boldsymbol{x}_i)$ *in the sample. For this property to hold it is required that* $g_\alpha(t \mid \boldsymbol{x}, \theta)$ *be a bounded function of* θ *for fixed* $\boldsymbol{\alpha} \in A$ *and* $\boldsymbol{x} \in X$. *The computational form of the conditional density is thus*

$$f_Y(t \mid \boldsymbol{x}) = \sum_{j=1}^{I^*} g_\alpha(t \mid \boldsymbol{x}, \theta_j) P_j . \tag{3.2}$$

$\boldsymbol{P} \in \mathscr{M}_{I^*}$ *where*

$$\mathscr{M}_{I^*} = \left\{ P = (P_1, \dots, P_{I^*});\ P_j \geq 0,\ 1 \leq j \leq I^*,\ \sum_{j=1}^{I^*} P_j = 1 \right\}.$$

For the proof see Lindsay [**32**, Theorem 3.2].

Lindsay analyzes mixing models with no regressors and with no structural parameters. However, his analysis can be applied to the most general class of survivor models with separable hazards considered in this paper

$$P_{\boldsymbol{\alpha}}(T > t \mid \boldsymbol{x}, \theta) = \exp\left[-z(t)\Phi(\boldsymbol{x}(t), t, \boldsymbol{x}_0) e^{\theta} \right]. \tag{3.3}$$

For fixed $\boldsymbol{\alpha}$, z, and Φ, define a generalized duration t^* as

$$t^* = z(t)\Phi(\boldsymbol{x}(t), t, \boldsymbol{x}_0).$$

Replacing t with t^* in all expressions for the frequency function $g_{\boldsymbol{\alpha}}(t \mid \boldsymbol{x}, \theta)$ produces a frequency function for t^*, $g_{\boldsymbol{\alpha}}(t^* \mid \theta)$, which satisfies the assumptions about frequency functions utilized in Lindsay's analysis. Note further that the survivor function for T^* is the survivor function of an exponential distribution

$$P_{\boldsymbol{\alpha}}(T^* > t^* \mid \boldsymbol{x}, \theta) = \exp\left[-t^* e^{\theta} \right]. \tag{3.4}$$

For fixed $\boldsymbol{\alpha}$, the frequency function of t^* may be written as

$$f_Y(t^*) = \sum_{j=1}^{I^*} g_{\boldsymbol{\alpha}}(t^* \mid \theta_j) P_j . \tag{3.2$'$}$$

As a consequence of Theorem 3.1, optimization problem (3.1) simplifies to

$$\sup_{\boldsymbol{\alpha} \in A, \theta \in \underline{\underline{\Theta}}, P \in \mathscr{M}} \mathscr{L} = \sup_{\boldsymbol{\alpha} \in A, \theta \in \underline{\underline{\Theta}}, \boldsymbol{P} \in \mathscr{M}_{I^*}} \sum_{i=1}^{I} \ln\left(\sum_{j=1}^{I^*} g_{\boldsymbol{\alpha}}(t_i \mid \boldsymbol{x}_i, \theta_j) P_j \right) \tag{3.5}$$

for some $I^* \leq I$. For fixed $\boldsymbol{\alpha}$, the optimization problem simplifies to

$$(3.5)' \qquad \underset{\theta \in \underline{\underline{\Theta}}, \boldsymbol{P} \in \mathscr{M}}{\operatorname{Sup}} \mathscr{L} = \underset{\theta \in \underline{\underline{\Theta}}, \boldsymbol{P} \in \mathscr{M}_{I^*}}{\operatorname{Sup}} \sum_{i=1}^{I} \ln\left(\sum_{j=1}^{I^*} g_\alpha(t_i^* \mid \theta_j) P_j \right).$$

The Weibull kernels (2.1b) as well as the more general kernels (2.21) and (2.24) are bounded functions of θ so that representation (3.2) applies for all the kernels considered in this paper.

For fixed $\boldsymbol{\alpha}$, the NPMLE of $\boldsymbol{P}, \boldsymbol{\theta}$, the vector of support points, is unique for kernel densities from the exponential family. More precisely, we have the following theorem.

THEOREM 3.2: *For each $\boldsymbol{\alpha} \in A$, $\boldsymbol{x} \in X$ if $g_\alpha(t^* \mid \theta)$ is in the exponential family and provided one condition is met, then*

$$\underset{\theta \in \underline{\underline{\Theta}}, P \in \mathscr{M}_{I^*}, I^* \leq I}{\operatorname{Sup}} \sum_{i=1}^{I^*} \ln\left(\sum_{j=1}^{I^*} g_\alpha(t_i^* \mid \theta_j) P_j \right)$$

is attained for a unique mixing distribution $(\boldsymbol{\theta}, \boldsymbol{P})$. The required condition is that no points of support come from the boundary of $\underline{\underline{\Theta}}$.

For proof see Lindsay [**33**]. Note that by virtue of representation (3.4), Theorem 3.2 applies to all of the survivor models considered in this paper including the censored models introduced in Section 2.

The computational strategy suggested by Theorems 3.1 and 3.2 utilizes characterization (3.2) to form the finite mixture likelihood shown in (3.5). For fixed $\boldsymbol{\alpha}$ and $\boldsymbol{x}$, the EM algorithm (Dempster, Laird, and Rubin [**9**]; Laird [**30**] as corrected by Wu [**43**]) may be used to estimate $\boldsymbol{\theta}$ and $\boldsymbol{P}$.[16] The corrected EM algorithm is guaranteed to converge to a stationary point. One possible algorithm for computation estimates $\boldsymbol{\theta}$ and $\boldsymbol{P}$ for each $\boldsymbol{\alpha}$ and then iterates on $\boldsymbol{\alpha}$ using the estimates of $\boldsymbol{\theta}$ and $\boldsymbol{P}$ secured from the preceding step. See Heckman and Singer [**22**] for details on the application of this algorithm to the problem at hand.

A difficulty with any computing algorithm that attempts to maximize (3.5′) with respect to $\boldsymbol{\alpha}$, $\boldsymbol{\theta}$, and $\boldsymbol{P}$ is the fact that likelihood surfaces in mixture models typically contain multiple local maxima. It is the global optimum that is the consistent estimator.

In particular, estimates obtained from the EM algorithm applied to the mixtures problem are sensitive to starting values (Laird [**30**], Heckman and Singer [**22**]). The danger of convergence to a local optimum is especially acute in mixture models no matter what computational algorithm is used.

The following three theorems aid in securing a global optimum for the model and are the basis for the computing strategy presented in this paper. Theorem (3.3) provides a bound on the range of estimated θ.

[16] Laird's analysis assumes no regressors. As a consequence of the separability of θ from t and $\boldsymbol{x}$, and for fixed $\boldsymbol{\alpha}$, her analysis applies to the more general separable hazard models considered in this paper. (Simply redefine the model in terms of t^*.)

THEOREM 3.3: *Suppose that for each i, and for* $\boldsymbol{\alpha} \in A$, $g_\alpha(t_i^* \mid \theta)$ *has a unique maximum in* $\theta, \theta_i' \in \underline{\underline{\Theta}}$. *Let* $\theta^- = \min(\theta_1', \ldots, \theta_I')$ *and* $\theta^+ = \max(\theta_1', \ldots, \theta_I')$. *Then the support of* $(\overline{\boldsymbol{\theta}}, \boldsymbol{P})$ *must lie in* $[\theta^-, \theta^+]$.

For proof, see Lindsay [**31**, p. 102].

For the general survival model (3.4) the unique mode of each observation is

$$\theta_i' = -\ln t_i^*$$

so

$$\left[\theta^-, \theta^+\right] = \left[-\ln t^*_{\max}, -\ln t^*_{\min}\right].$$

The next theorem further reduces the region of $\underline{\underline{\Theta}}$ in which estimated points of support may lie. Before the theorem is introduced it is useful to order the I^* distinct values of t_i^* and denote the ordered values by $t_{[i]}$. Thus

$$t^*_{\min} = t^*_{[1]} \leq \cdots \leq t^*_{[I^*]} = t^*_{\max}.$$

The mixture quadratic $M_{ij}(\theta)$ is defined as

$$M_{ij}(\theta) = E((T^* - t^*_{[i]})(T^* - t^*_{[j]}) \mid \theta).$$

For the general separable model with survivor function given in (3.4),

$$M_{ij}(\theta) = 2e^{-2\theta} - (t^*_{[i]} + t^*_{[j]})e^{-\theta} + t^*_{[i]}t^*_{[j]}.$$

For each adjacent pair of values $(t^*_{[i]}, t^*_{[i+1]})$, $i = 1, \ldots, I^* - 1$, denote the roots of $M_{i,i+1}(\theta)$ as $r^{(i)}_{[i]}, r^{(i)}_{[i+1]}$ where for real valued roots $r^{(i)}_{[i+1]} > r^{(i)}_{[i]}$. If the roots are real, define the interval $\{z_i\} = (r^{(i)}_{[i]}, r^{(i)}_{[i+1]})$. Otherwise, $\{z_i\} = \{\phi\}$.

With this notation in hand, we can now state the following theorem.

THEOREM 3.4: *If for each* $\boldsymbol{\alpha} \in A$ *and* $\boldsymbol{x} \in \boldsymbol{X}$, $g_\alpha(t^* \mid \theta)$ *is in the exponential family, then the points of increase* $\theta_1, \ldots, \theta_{I^*}$ *must lie outside the intervals* $\{z_i\}$, $1 \leq i \leq I^* - 1$. *Moreover, for all i and j there can be no more than one point of support in any interval where* $M_{1I^*}(\theta)$ *is positive.*

For proof, see Lindsay [**33**].

The final theorem presented in this section provides a check on any alleged optimum to likelihood (3.1). Before presenting this theorem, it is necessary to define the concept of a Gateaux differential.

A functional T acting on probability distributions is Gateaux differentiable at F if there exists a linear functional $L = L_f$ such that for all $G \in \mathscr{M}$

$$\lim_{\rho \to 0} \frac{T(F_\rho) - T(F)}{\rho} = L_F(G - F)$$

with $F_\rho = (1 - \rho)F + \rho G$. Of particular interest in our case is the situation in which $dG = \delta_x$, unit mass at point x.

If it is further assumed that L_F can be represented by a measurable function $\mathscr{X}_F$ standardized such that

$$\int \mathscr{X}_F(\theta)\, dF(\theta) = 0.$$

Then

$$L_F(G - F) = \int \mathscr{X}_F(\theta)\, dG(\theta)$$

and $\mathscr{X}_F$ is the Gateaux derivative.

The natural notion of derivative for maximization problem (3.1) for fixed $\boldsymbol{\alpha}$ and $\boldsymbol{x}$ is the Gateaux derivative, specializing $\mathscr{X}_F$ to be $D(\theta, \mu)$. We obtain

$$\begin{aligned} D(\theta;\, \mu) &= \sum_{i=1}^{I} \frac{g_{\alpha}(t^* \mid \theta)}{f_{\gamma}(t^*)} - 1 \\ &= \lim_{\rho \to 0} \frac{1}{\rho} \left[\ln \mathscr{L}_{(1-\rho)\mu + \rho\delta_\theta} - \ln \mathscr{L}_{\mu} \right] \end{aligned}$$

where

$$\mathscr{L}_{\mu} = \prod_{i=1}^{I} f_{\gamma}(t_i^*) = \prod_{i=1}^{I} \int_{\underline{\underline{\Theta}}} g_{\alpha}(t_i^* \mid \theta)\, d\mu(\theta).$$

With these definitions in hand, we state the final theorem of this section.

THEOREM 3.5: *The estimator* $\hat{\mu} = (\boldsymbol{\theta}, \boldsymbol{P})$ *maximizes likelihood* (3.1) *for fixed* $\boldsymbol{\alpha}$ *iff* $D(\theta;\, \hat{\mu}) \le 0$ *for all* $\theta \in \Theta$. *Furthermore, a maximizing measure* $\hat{\mu}$, *has support in the set* $\{\theta : D(\theta,\, \hat{\mu}) = 0\}$.

For proof, see Lindsay [**31**, p. 100].

Our experience to date suggests that the following computational strategy, based on Theorems 3.1–3.5, works well in practice. We offer no proof that the procedure converges to a global optimum but the limited Monte Carlo evidence on this approach reported in Section 4 indicates that it does.

Step 1. From initial values for $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$, maximize likelihood (3.5) for a model with one point of support, $\hat{\boldsymbol{\theta}}^{(1)}$.[17] Let $\hat{\boldsymbol{\alpha}}^{(1)}$ denote the estimated parameter vector for $\boldsymbol{\alpha}$.

Step 2. For fixed $\hat{\alpha}^{(1)}$ scan $[\theta^-, \theta^+]^{\hat{\alpha}^{(1)}} \backslash \bigcup_{i=1}^{I^*-1} \{z_i\}^{\hat{\alpha}^{(1)}}$ where the superscript $\hat{\boldsymbol{\alpha}}^{(j)}$ denotes intervals generated conditional on the value of $\hat{\boldsymbol{\alpha}}^{(j)}$. Locate the intervals in which $D(\theta;\, \hat{\mu}) > 0$. Add one point of support to the model in the interval with the highest value of $D(\theta;\, \hat{\mu})$ at the θ value with the highest value of the Gateaux derivative. If $D(\theta;\, \hat{\mu}) \le 0$ over the entire admissible interval, stop. Otherwise estimate a model with two points of support. Although a variety of procedures can be used, the most effective locates $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{P}})$ for fixed $\hat{\boldsymbol{\alpha}}^{(1)}$, then

[17] One may use a gradient procedure, a simplex computing algorithm, or the EM algorithm.

iterates to locate a new value of $\hat{\boldsymbol{\alpha}}$, then for fixed $\hat{\boldsymbol{\alpha}}$ locates a new $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{P}})$. When this iteration sequence converges, denote the new estimates by $(\hat{\boldsymbol{\alpha}}^{(2)}, \hat{\boldsymbol{\theta}}^{(2)}, \hat{\boldsymbol{P}}^{(2)})$.

Step 3. For fixed $\hat{\boldsymbol{\alpha}}^{(2)}$ scan $[\theta^-, \theta^+]^{\hat{\alpha}^{(2)}} \backslash \bigcup_{i=1}^{I^*-1} \{z_i\}^{\hat{\alpha}^{(2)}}$. Locate intervals in which $D(\theta; \hat{\mu}) > 0$. If there are none stop. If there are, continue along the lines indicated in step 2 until the iteration sequence converges to $(\hat{\boldsymbol{\alpha}}^{(3)}, \hat{\theta}^{(3)}, \hat{P}^{(3)})$.

Go on to subsequent steps until there is termination by the criterion that $D(\theta; \hat{\mu}) \leq 0$.

We reiterate that we have not proved that this procedure converges to a global optimum. Nonetheless the available Monte Carlo evidence suggests that it does and that variants of the above iteration cycle are considerably less computationally efficient. For example, if step 2 is modified to introduce multiple points of support in situations in which $D(\theta; \hat{\mu}) > 0$ over multiple disjoint intervals, the computational efficiency of the algorithm is considerably reduced.

4. EMPIRICAL EVIDENCE ON THE NONPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATOR

This section presents some evidence on the empirical performance of the NPMLE of μ and $\boldsymbol{\alpha}$ for the Weibull model (2.2) without censoring. The computing algorithm presented at the end of Section 3 is applied to samples of artificially generated data in an attempt to recover the structural parameters and mixing distribution. Due to the cost involved, we have not been able to perform an exhaustive Monte Carlo study. Because of the limited number of sampling experiments performed, the numerical results reported here can only be considered as suggestive of the performance of the estimator.

Three principal conclusions emerge from this analysis. First, for both finite and continuous mixing distributions and for samples of five hundred or more observations, the NPMLE estimates the structural parameters $(\boldsymbol{\alpha})$ very well. The proposed procedure enables the analyst to avoid the sensitivity of structural parameter estimates to the choice of mixing measure that is demonstrated in the examples in Section 1. Second, for both finite and continuous mixing measures, the NPMLE never estimates μ very well. This is true even when samples of size 5000 are used to generate NPMLE estimates. Third, even though the underlying mixing measure is never well estimated, estimated values of $\boldsymbol{\alpha}$ and μ generate excellent forecasts of sample distributions of durations. This is not only true for the samples used to produce the estimates, but also for fresh samples with the same distribution of $\boldsymbol{x}$ as is used to generate the sample from which the estimates are extracted. It is also true for fresh samples in which different distributions are assumed for the $\boldsymbol{x}$ variables.

Normal, gamma, and multinomial densities were used to generate θ values:

$$d\mu(\theta) = (2\pi\sigma^2)^{-1/2} \exp - (\theta^2/2\sigma^2) d\theta, \qquad \sigma^2 > 0. \tag{4.1}$$

$$d\mu(\theta) = (\exp(\Delta\theta)) \exp - (e^\theta/\beta) d\theta, \qquad \Delta, \beta > 0. \tag{4.2}$$

$$d\mu(\theta_i) = P_i \qquad \text{for} \quad i = 1, \ldots, I, \quad P_i \geq 0, \quad \sum P_i = 1. \tag{4.3}$$

TABLE II

RESULTS FROM A TYPICAL ESTIMATION

(θ Generated from Eqn. (4.2) with $\Delta = 1/2$, $\beta = 1$)

	α_1	α_2
True Model	$\alpha_1 = 1$	$\alpha_2 = 1$
Estimated Model	0.9852	0.9846
	(0.0738)[a]	(0.1022)[a]

Sample Size	$L = 500$
Log Likelihood	-1886.47

Estimated Mixing Distribution

Estimated θ_i	Estimated P_i	Estimated CDF	True θ CDF	Observed CDF
-12.9031	0.008109	0.008109	0.001780	0.0020
-7.0938	0.06524	0.07335	0.03250	0.0400
-4.0107	0.1887	0.2621	0.1510	0.1620
-1.7898	0.3681	0.6302	0.4366	0.4280
-0.0338	0.3698	1.0000	0.8356	0.8320

Estimated Cumulative Distribution of Duration vs. Actual ($\hat{F}(t)$ vs. $F(t)$)

Value of T	Estimated t CDF	Observed CDF
0.25	0.1237	0.102
0.50	0.2005	0.186
1.00	0.3005	0.296
3.00	0.4830	0.484
5.00	0.5661	0.556
10.00	0.6675	0.660
20.00	0.7512	0.754
40.00	0.8169	0.818
99.00	0.8800	0.880

[a] The numbers reported below the estimates are standard errors from the estimated information matrix for ($\boldsymbol{\alpha}$, $\boldsymbol{P}$, $\boldsymbol{\theta}$) given I^*. As noted in Section 2, these have no rigorous justification.

The artificial samples utilized in our analysis are generated by the following procedure. Draw a uniform variable in the interval [0, 1]. From the appropriate cumulative density function of θ, we solve for the implied θ. Values of scalar x are taken from a standard normal random number generator. Another uniform random number in the interval [0, 1] is drawn. Given θ, x, and specified values of $\boldsymbol{\alpha}$, solve the appropriate distribution for the implied t. Because the gamma (for e^{θ}) is a very flexible family of distributions, more attention was devoted to models with unobservables derived from this density.

Table II reports results from a typical experiment. The θ are generated from density (4.2) with $\Delta = 1/2$ and $\beta = 1$. The agreement between the fitted and true values of $\boldsymbol{\alpha}$ is very high. The estimates of P_i and θ_i are reported below the subheading "Estimated Mixing Distribution." For a variety of starting values, the algorithm converged to the estimates displayed in the first two columns.[18] A

[18] A remarkable finding in all of our runs based on the EM algorithm is a clustering phenomenon. In those runs the optimization is started with ten distinct values of θ_i. The algorithm converges to fewer than ten distinct values. There are several values of θ_i that become indistinguishable given the numerical accuracy of the computational procedure. A typical complete iteration sequence documenting this phenomenon for the EM algorithm is reported in Heckman and Singer [**22**].

comparison of the estimated cdf, the true cdf, and the sample cdf (based on realized values of θ) is given in the last three columns. The agreement between the estimated and true distributions is rather poor.

Next consider the agreement between the calculated cdf for duration times and the sample cdf. The calculated cdf is computed conditional on realized x values. For each observation we compute the probability that durations are less than the values of t reported in the table. Estimated values of θ_i, P_i, and $\boldsymbol{\alpha}$ are used to generate the distribution, using equation (2.2). Summing over all observations produces the numbers reported in the first column under the subheading "Estimated Cumulative Distribution of Durations vs. Actual ($\hat{F}(t)$ vs. $F(t)$)." The sample distribution function is recorded in the second column. The agreement between the estimated and empirical distribution functions is quite good.

We have investigated the predictive power of the estimated model.[19] For the true values of $\boldsymbol{\alpha}$ and μ recorded in Table II, but for different x distributions, fresh duration samples were generated. Using the estimated $\boldsymbol{\alpha}$ and μ reported in Table II, we predict durations in the new samples using equation (2.2). The model produces very accurate forecasts of sample duration distributions even when the distributions of x are quite different from the distributions used to produce the sample of durations from which estimated values of $\boldsymbol{\alpha}$ and μ are obtained.[20]

Table III summarizes our evidence on the performance of the NPMLE as an estimator of the structural parameters $\boldsymbol{\alpha}$ in the uncensored Weibull model. The agreement between estimated and true structural parameters is generally very high. In each of the runs, the mixing distribution is poorly estimated and the fit of the estimated cdf to the sample cdf is quite good.[21]

In order to investigate the rate of convergence of the estimated mixing measure to the true (continuous) mixing measure as a function of sample size, we estimate a mixture of exponentials model ($\alpha_1 = 1$, $\alpha_2 = 0$) for samples of size 500, 1000, and 5000. Results for this investigation are reported in Heckman and Singer **[22]**, and in Appendix C. As sample size increases, more points of increase are estimated. But even in a sample as large as 5000 the true cdf is poorly estimated.

The EM algorithm is not guaranteed to produce a global optimum of the likelihood (Wu **[43]**), nor is the algorithm proposed in Section 3.[22] Our experience indicates that the optima achieved from the EM procedure are sometimes sensitive to starting values. Table IV provides an example of this phenomenon. The estimates reported there are for the same model and for the same data used to produce the estimates reported in Table II. The format of Table IV is identical to that of Table II. The starting values used to obtain these estimates are reported below the subheading "Estimated Mixing Distribution."

[19] A detailed description of the runs described in this paragraph is presented in Appendix B.

[20] The original sample for x is $N(0,1)$. The fresh samples are $N(10,1)$, $N(-10,1)$, $N(0,25)$, $N(10,25)$, $N(-10,25)$.

[21] A detailed description of these runs in the format of Table II is presented in Appendix A.

[22] Laird **[30]** offers some empirical evidence on this point. It is important to find the global optimum of the likelihood because estimators achieved from nonglobal optima are not consistent.

TABLE III

COMPARISON OF ESTIMATED STRUCTURAL PARAMETERS WITH TRUE VALUES[a]
(Based on Numbers from Table II and Appendix A)

Gamma Density for e^{θ} eqn. (4.2)

True α_1	Estimated α_1	True α_2	Estimated α_2	Δ	β	Sample Size
2	1.947(.12)	.5	.484(.05)	1	1	1000
1	.985(.07)	1	.984(.10)	1/2	1	500
2	2.004(.096)	0	−.008(.04)	4	1	500
4	4.165(.183)	4	4.196(.193)	2	3	500
2	2[b]	.5	.496(.044)	1	1	1000
.5	.624(.031)	−1	−1.149(.081)	1/2	2	500
1	.859(.032)	0	[d]	2	1	100

Normal Density for θ eqn. (4.1)

True α_1	Estimated α_1	True α_2	Estimated α_2	σ^2	α_0[c]	Sample Size
1	.965(.06)	−1	−1.041(.12)	.25	3	500
1	.922(.10)	−1	−.870(.10)	.25	0	500

Multinomial Density for θ eqn. (4.3)

True α_1	Estimated α_1	True α_2	Estimated α_2	Sample Size
1	1.176(.152)	−1	−1.17(.16)	500
1	1.01 (.045)	−1	−.97(.06)	500

[a] The numbers in parentheses following each estimated coefficient are standard errors from the estimated information matrix. As noted in Section 2, there is no rigorous justification for these standard errors.
[b] In this run, α_1 is fixed at 2.
[c] In the estimates of the normal models, α_0 in equation (2.1b) is fixed at different values. This shifts the mean of the mixing distribution.
[d] Fixed at zero.

TABLE IV

AN EXAMPLE OF A LOCAL OPTIMUM THAT IS NOT A GLOBAL OPTIMUM
(The Model and Sample Are Identical to the One Reported in Table II.)

True Model	$\alpha_1 = 1$	$\alpha_2 = 1$
Estimated Model	0.4995	0.5907
	(0.0219)	(0.0454)

Sample Size	$L = 500$
Log Likelihood	−1942.53

Estimated Mixing Distribution
Starting values for θ_i are $\theta = (1, 2, \dots, 5)$; $P_i = 0.2$; $i = 1, \dots, 5$

Estimated θ_i	Estimated P_i	Estimated CDF	True θ CDF	Observed CDF
−4.3260	0.08031	0.08031	0.1292	0.1340
−0.9475	0.9197	1.0000	0.6215	0.5820

Estimated Cumulative Distribution of Duration vs. Actual ($\hat{F}(t)$ vs. $F(t)$)

Value of t	Estimated t CDF	Observed CDF
0.25	0.1794	0.1020
0.50	0.2396	0.1860
1.00	0.3138	0.2960
3.00	0.4577	0.4840
5.00	0.5317	0.5560
10.00	0.6326	0.6600
20.00	0.7259	0.7540
40.00	0.8038	0.8180
99.00	0.8758	0.8800

The estimated structural parameters reported in Table IV are badly biased, and the agreement between $\hat{F}(t)$ and $F(t)$ is decidedly inferior to that reported in Table II. The value of the log likelihood is markedly lower in Table IV than in Table II.[23] When the EM algorithm was used in conjunction with the algorithm suggested in Section 3, convergence to the optimum reported in Table II occurred.

Much further Monte Carlo work remains to be done before any firm conclusions on the empirical performance of the NPMLE can be reached. Our experience to date suggests the following conclusions:

(i) The NPMLE estimates the structural parameters rather well. (ii) The NPMLE does not produce reliable estimates of the underlying mixing distribution. (iii) The estimated cdf for duration times produced by using estimated values of α and μ and actual values of x in equation (2.1b) predicts the sample cdf of durations quite well even in fresh samples with markedly different distributions of the x variables. (iv) The computing strategy outlined in Section 3 appears to protect against local optima.

The NPMLE can be used to check the plausibility of any particular parametric specification of the distribution of unobserved variables. If the estimated parameters of a structural model achieved from a parametric specification of the distribution of unobservables are not "too far" from the estimates of the same parameters achieved from the NPMLE proposed in this paper, the econometrician would have much more confidence in adopting the particular specification of the mixing measure. Development of a formal test statistic to determine how far is "too far" is a topic for the future. However, because of the consistency of the nonparametric maximum likelihood estimator a test based on the difference between α estimated via the NPMLE and α estimated under a particular assumption about the functional form of the mixing distribution would be consistent.

The fact that we produce a good estimator of α while producing a poor estimator for μ suggests that it might be possible to protect against the consequences of misspecification of the mixing distribution by fitting duration models with mixing distributions from parametric families, such as members of the Pearson system, with more than the usual two parameters. The failure of the NPMLE to estimate more than four or five points of increase for μ can be cast in a somewhat more positive light. A finite mixture model with five points of increase is, after all, a nine (independent) parameter model for the mixing measure. Imposing a false, but very flexible, mixing distribution may not cause much bias in estimates of the structural coefficients.

5. REANALYSIS OF THE KIEFER-NEUMANN DATA WITH THE NPMLE

The first three columns of Table I in Section 1 demonstrate the extreme sensitivity of parameter estimates of a structural unemployment duration model

[23] Parenthetically, we observe that a smaller number of points of increase are estimated for the mixing measure at a local optimum than at a global optimum. In our limited experience, this phenomenon always occurs when the EM algorithm converges to a nonglobal optimum.

to the functional form of the distribution of unobservables. If log normal or gamma mixing distributions are used to secure estimates, the random effect estimator indicates negative duration dependence (i.e., the coefficient on ln duration is negative). In the language of search theory, the estimates reported in columns two and three of Table I imply that unemployed workers have a reservation wage that *increases* with the length of an unemployment spell. Estimates obtained from imposing a normal heterogeneity assumption indicate no duration dependence. The estimated coefficients of the exogenous variables in the model are very sensitive to the choice of functional form for the mixing distribution.

Column four of Table I presents empirical results obtained from the NPMLE discussed in this paper. The most striking feature of these estimates is the positive coefficient on ln duration. In the language of search theory, the Kiefer–Neumann data now indicate that a declining reservation wage describes an unemployed worker's strategy. This finding accords with the predictions of several models of search unemployment (see Lippman and McCall [**34**]). The coefficient estimates of the other variables display a distinct pattern of signs and magnitudes that are not captured by any single parametric model presented in Table I.

6. SUMMARY AND CONCLUSIONS

This paper demonstrated the sensitivity of parameter estimates obtained from econometric models for duration data to assumed functional forms for the distribution of unobserved variables. Provided that there is information about the functional form of a duration model conditional on values of unobserved variables, it is possible to utilize observed duration data to estimate the distribution function of the unobservables and the structural parameters of the model using a nonparametric maximum likelihood procedure. For this reason we claim that current practice, which assumes that the functional form of the distribution of unobserved variables is known, overparameterizes econometric duration models. This overparameterization dramatically affects the estimates and inferences achieved from structural duration models.

We discussed a prototypical structural duration model and produced a consistent nonparametric maximum likelihood estimator (NPMLE) for the distribution of unobservables and the structural parameters of this model. We also presented a general discussion of conditions that must be satisfied to ensure the existence of a consistent nonparametric maximum likelihood estimator in a general proportional hazards model. An algorithm for estimating the NPMLE is proposed.

A limited set of Monte Carlo experiments was conducted to evaluate the performance of the estimator. The NPMLE estimated the parameters of the structural duration model very well for samples as small as 500. Estimation of the distribution of the unobservables was less successful. Nonetheless, the estimated model produced reliable forecasts of durations in fresh samples even when the distribution of the covariates was quite different from the distribution in the sample from which the estimates were derived.

Using the NPMLE we reanalyzed some data on unemployment durations.

When the impact of distributional assumptions was minimized, the data were consistent with the declining reservation wage model of search theory. Estimates obtained from conventional methods were very sensitive to the assumptions made about the distributions of unobserved variables.

University of Chicago
and
Columbia University

Manuscript received July, 1981; final revision received March, 1983.

APPENDIX

This Appendix is in three parts. Appendix A reports in detail the runs summarized in Table III of the text. Appendix B reports the results of forecasting with the model of Table II on fresh samples of data. Appendix C reports some evidence on the rate of convergence of estimated μ to actual μ as a function of sample size.

APPENDIX A

THE EMPIRICAL PERFORMANCE OF THE NPMLE

The format of the ten tables in this Appendix is identical to the format of Table II in the text of our paper. The particular choice of $d\mu(\theta)$ used to generate each sample is given below the table titles. $x \sim N(0,1)$ is the distribution generating the data used to form the samples in all tables. The data generation procedure is discussed in Section 4 of the paper.

TABLE A-I

GAMMA MIXING DISTRIBUTION $\Delta = 1$, $\beta = 1$
(See Equation (4.2) in the Text.)

True Model	$\alpha_1 = 2$	$\alpha_2 = .5$
Estimated Model	$\hat{\alpha}_1 = 1.947$	$\hat{\alpha}_2 = .484$
	(.12)	(.053)

Sample Size	$L = 1000.$
Log Likelihood	-3493.57

Estimated Mixing Distribution

Estimated θ_i	Estimated P_i	Estimated CDF	True CDF	Observed CDF
-3.030	.0063	.0063	.0011	.002
-1.2968	.0918	.0981	.0312	.042
0.2132	.6057	.7038	.3509	.370
1.3992	.2961	1.000	.9152	.917

Estimated Cumulative Distribution of Durations vs. Actual ($\hat{F}(t)$ vs. $F(t)$)

Value of t	Estimated t cdf	Observed cdf
.25	.1321	.1380
.50	.3775	.3720
.75	.5878	.5890
1.00	.7339	.7290
1.25	.8285	.8270
1.50	.8876	.8970
1.75	.9242	.9280
2.00	.9471	.9430
3.00	.9833	.9850
4.00	.9933	.9920

TABLE A-II

GAMMA MIXING DISTRIBUTION, $\Delta = 4$, $\beta = 1$
(See Equation (4.2) in the Text.)

	α_1	α_2	α_0
True Model	$\alpha_1 = 2$	$\alpha_2 = 0$	$\alpha_0 = 0$
Estimated Model	2.0448	−.0084	—
	(.0964)	(.0488)	

Sample Size	$L = 500$
Log Likelihood	13.0639

Estimated Mixing Distribution

Estimated θ_i	Estimated P_i	Estimated CDF	True θ CDF	Observed CDF
0.2921	.0448	.0448	.0471	.0460
1.3507	.9552	1.0000	.5387	.5380

Estimated Cumulative Distribution of Durations vs. Actual ($\hat{F}(t)$ vs. $F(t)$)

Value of t	Estimated t CDF	Observed CDF
.1000	.0332	.0420
.2000	.1300	.1420
.3000	.2727	.2580
.4000	.4355	.4440
.5000	.5928	.5860
.6000	.7264	.7300
.7000	.8280	.8240
.9000	.9422	.9320
1.1000	.9824	.9840

TABLE A-III

GAMMA MIXING DISTRIBUTION $\Delta = 2$, $\beta = 3$
(See Equation (4.2) in the Text.)

	α_1	α_2	α_0
True Model	$\alpha_1 = 4$	$\alpha_2 = 4$	$\alpha_0 = 0$
Estimated Model	4.1653	4.1961	—
	(.1833)	(.1931)	

Sample Size	$L = 500$
Log Likelihood	−138.3063

Estimated Mixing Distribution

Estimated θ_i	Estimated P_i	Estimated CDF	True θ CDF	Observed CDF
−2.1839	.0698	.0698	.0457	.0440
−0.6206	.9302	1.0000	.4792	.4680

Estimated Cumulative Distribution of Durations vs. Actual ($\hat{F}(t)$ vs. $F(t)$)

Value of t	Estimated t CDF	Observed CDF
.2500	.0906	.1020
.5000	.2357	.2200
.7500	.3587	.3640
1.0000	.4599	.4580
1.5000	.6142	.6100
2.0000	.7243	.7340
3.0000	.8484	.8440
5.0000	.9459	.9540
10.000	.9919	.9900

TABLE A-IV

GAMMA MIXING DISTRIBUTION $\Delta = 1, \beta = 1$
(See Equation (4.2) in the Text.)

True Model	$\alpha_1 = 2$	$\alpha_2 = .5$	$\alpha_0 = 0$
Estimated Model	—	.496	—
		(.044)	

$L = 1000$

Estimated Mixing Distribution

Estimated θ_i	Estimated P_i	Estimated CDF	True θ CDF	Observed CDF
− 3.1442	.0062	.0062	.0009	.0002
− 1.3591	.0901	.0963	.0279	.0390
0.1007	.4098	.5062	.3031	.3240
0.5953	.2781	.7842	.5412	.5630
1.688	.2157	1.0000	.9713	.9740

Estimated Cumulative Distribution of Durations vs. Actual ($\hat{F}(t)$ vs. $F(t)$)

Value of t	Estimated t CDF	Observed CDF
.2500	.1318	.1380
.5000	.3776	.3720
.7500	.5867	.5890
1.0000	.7333	.7290
1.2500	.8286	.8270
1.5000	.8879	.8970
1.7500	.9244	.9280
2.0000	.9471	.9430
3.0000	.9834	.9850
4.0000	.9934	.9920

TABLE A-V

GAMMA MIXING DISTRIBUTION ($\Delta = 1/2, \beta = 2$)
(See Equation (4.2) in the Text.)

True Model	$\alpha_1 = .5$	$\alpha_2 = -1$	$\alpha_0 = 0$
Estimated Model	$\alpha_1 = .6244$	$\alpha_2 = -1.149$	—
	(.0312)	(.0809)	

Sample Size	$L = 500$
Log Likelihood	− 3648.74

Estimated Mixing Distribution

Estimated θ_i	Estimated P_i	Estimated CDF	True θ CDF	Observed CDF
− 16.5659	.00807	.00807	.000404	.0040
−7.0597	.2482	.2562	.0468	.0540
−3.1966	.3784	.6347	.3141	.3260
−0.6014	.3653	1.0000	.8613	.8360

Estimated Cumulative Distribution of Duration vs. Actual ($\hat{F}(t)$ vs. $F(t)$)

Value of t	Estimated t CDF	Observed CDF
.2500	.1100	.1100
.5000	.1483	.1500
1.0000	.1941	.2020
3.0000	.2781	.2800
5.0000	.3198	.3160
10.0000	.3768	.3560
20.0000	.4332	.4220
40.0000	.4885	.4760
80.0000	.5426	.5340

TABLE A-VI

NORMAL MIXING DISTRIBUTION WITH $\sigma^2 = .25$
(See Equation (4.1) in the Text.)

True Model	$\alpha_1 = 1$	$\alpha_2 = -1$	$\alpha_0{}^{a} = 3$
Estimated Model	0.9642	$-$ 1.041	—
	(0.0911)	(0.117)	

Sample Size	$L = 500$
Log Likelihood	946.48

Estimated Mixing Distribution

Estimated $\eta_i = e^{\theta}$	Estimated P_i	Estimated CDF	True η CDF	Observed CDF
2.3184	0.2670	0.2670	0.0864	.0881
2.7146	0.3169	0.5839	0.2840	.2829
3.3731	0.4161	1.0000	0.7723	.7721

Estimated CDF of Duration vs. Actual ($\hat{F}(t)$ vs. $F(t)$)

Value of t	Estimated t CDF	Observed CDF
0.001	0.0422	0.0480
0.005	0.1596	0.1560
0.010	0.2620	0.2640
0.020	0.4001	0.4020
0.040	0.5607	0.5540
0.060	0.6547	0.6560
0.100	0.7626	0.7600
0.200	0.8761	0.8740
0.500	0.9605	0.9680

[a] In the estimates of the normal models, α_0 in equation (2.1b) is fixed at different values. This shifts the mean of the mixing measure.

TABLE A-VII

NORMAL MIXING DISTRIBUTION WITH $\sigma^2 = .25$
(See Equation (4.1) in the Text.)

True Model	$\alpha_1 = 1$	$\alpha_2 = -1$	$\alpha_0 = 0$
Estimated Model	0.9224	$-$ 0.8647	—
	(0.0984)	(0.1025)	

Sample Size	$L = 500$
Log Likelihood	$-$ 556.88

Estimated Mixing Distribution

Estimated $\eta_i = e^{\theta}$	Estimated P_i	Estimated CDF	True η CDF	Observed CDF
$-$ 0.1108	0.8904	0.8904	0.4123	0.4340
0.1843	0.1096	1.0000	0.9207	0.9120

Estimated CDF of Duration vs. Actual ($\hat{F}(t)$ vs. $F(t)$)

Value of t	Estimated t CDF	Observed CDF
0.10	0.1474	0.1280
0.20	0.2452	0.2280
0.30	0.3208	0.3040
0.50	0.4338	0.4340
0.75	0.5331	0.5320
1.00	0.6051	0.6200
1.50	0.7033	0.7260
2.00	0.7670	0.7900
3.00	0.8441	0.8660

TABLE A-VIII

MULTINOMIAL DENSITY

$P_1 = .1, \theta_1 = .25; P_2 = .9, \theta_2 = .75.$

True Model	$\alpha_1 = 1$	$\alpha_2 = -1$	$\alpha_0 = 0$
Estimated Model	1.0123	−.9736	—
	(.0456)	(.0631)	

Sample Size	$L = 500$
Log Likelihood	−713.6520

Estimated Mixing Values

Estimated θ_i	Estimated P_i	Estimated CDF	True θ CDF	Observed CDF
−1.8376	.0307	.0307	.1000	0.094
−0.3757	.9693	1.0000	1.0000	1.0000

Estimated CDF of Duration vs. Actual ($\hat{F}(t)$ vs. $F(t)$)

Value of t	Estimated t CDF	Observed CDF
.1000	.0955	.0980
.5000	.3442	.3500
1.0000	.5168	.5260
2.0000	.6962	.6820
3.0000	.7879	.7780
4.0000	.8422	.8400
5.0000	.8774	.8760
10.0000	.9506	.9460
20.0000	.9836	.9880

TABLE A-IX

MULTINOMIAL DENSITY

$(P_1 = .1, \theta_1 = .25; P_2 = .2, \theta_2 = .75; P_3 = .3, \theta_3 = 1.00;$
$P_4 = .3, \theta_4 = 1.25; P_5 = .1, \theta_5 = 1.75.)$

True Model	$\alpha_1 = 1$	$\alpha_2 = -1$	$\alpha_0 = 0$
Estimated Model	1.176	−1.166	—
	(.152)	(.159)	

Sample Size	$L = 500$
Log Likelihood	−567.54

Estimated Mixing Distribution

Estimated θ_i	Estimated P_i	Estimated CDF	True θ CDF	Observed CDF
−2.5889	.01711	.01711	.1000	0.0940
−0.9068	.3566	.3737	.3000	0.2800
0.5522	.2388	.6125	.6000	0.5600
0.5861	.2063	.8189	.9000	0.8800
0.5966	.1811	1.0000	1.0000	1.0000

Estimated CDF of Duration vs. Actual ($\hat{F}(t)$ vs. $F(t)$)

Value of t	Estimated t CDF	Observed CDF
.1000	.1301	.1480
.5000	.4399	.4340
1.0000	.6156	.6140
2.0000	.7707	.7740
3.0000	.8421	.8340
4.0000	.8829	.8740
5.0000	.9089	.9080
10.0000	.9624	.9600
20.0000	.9869	.9880

TABLE A-X

GAMMA MIXING DISTRIBUTION ($\Delta = 2$, $\beta = 1$)

True Model	$\alpha_1 = 1$
Estimated Model	$\beta_1 = 0.8589$ (0.0941)
Sample Size	100
Log Likelihood	-67.35

Estimated Mixing Distribution

Estimated θ_i	Estimated P_i	Estimated CDF	True CDF
-0.7606	0.0935	0.0935	0.0805
0.4408	0.8313	0.9248	0.4600
1.1521	0.0752	1.	0.8241

Estimated Cumulative Distribution of Duration vs. Actual ($\hat{F}(t)$ vs. $F(t)$)

Value of t	Estimated t CDF	Observed CDF
0.025	0.05754	0.04
0.050	0.1071	0.05
0.075	0.1510	0.09
0.10	0.1903	0.13
0.25	0.3672	0.41
0.50	0.5451	0.62
0.75	0.6536	0.69
1.0	0.7259	0.74
2.5	0.8951	0.86
5.0	0.9594	0.94
7.5	0.9793	0.97
10.	0.9878	0.99
25.	0.9984	1.

APPENDIX B

EVALUATION OF FORECASTS ON FRESH SAMPLES

This appendix reports results from the following experiments. The model that generates the sample used to produce the estimates reported in Table II is used to generate fresh samples. Different normal distributions for the X variables (indicated at the heading of each table) are used for different tables. (Recall that the X variables used to generate the sample from which estimates reported in Table II are generated is a standard normal ($N(0, 1)$).) Using the estimated α and μ from Table II in the paper

TABLE B-I

PREDICTIONS ON A FRESH SAMPLE. $X \sim N(10, 1)$

Estimated Cumulative Distribution of Duration vs. Actual ($\hat{F}(t)$ vs. $F(t)$)

Value of t ($\times 10^5$)	Estimated t CDF	Observed CDF
1.0	0.1118	0.1000
4.0	0.2799	0.2800
8.0	0.3924	0.3920
10.0	0.4300	0.4360
25.0	0.5802	0.5740
50.0	0.6795	0.6740
100.0	0.7607	0.7640
300.0	0.8543	0.8620
5000.0	0.9615	0.9660

with the fresh X variables indicated in each table, the sample distribution function is estimated ($\hat{F}(t)$) and compared to the observed cdf ($F(t)$). A remarkable feature of these experiments is the ability of the estimated model to forecast well outside the range of the data found in the sample used to generate the estimates of α and μ. This is all the more remarkable in light of the fact that μ is so poorly estimated.

TABLE B-II

PREDICTIONS ON A FRESH SAMPLE $X \sim N(-10, 1)$

Estimated CDF of Duration vs. Actual ($\hat{F}(t)$ vs. $F(t)$)		
Value of t ($\times 10^{-3}$)	Estimated t CDF	Observed CDF
1.0	0.0295	0.0320
3.0	0.0770	0.0660
10.0	0.1890	0.1800
50.0	0.4372	0.4400
100.0	0.5517	0.5400
200.0	0.6552	0.6440
400.0	0.7413	0.7380
700.0	0.7975	0.7920
999.999	0.8276	0.8300

TABLE B-III

PREDICTIONS ON A FRESH SAMPLE $X \sim N(0, 25)$

Estimated CDF of Duration vs. Actual ($\hat{F}(t)$ vs. $F(t)$)		
Value of t	Estimated t CDF	Observed CDF
0.01	0.1377	0.1440
0.10	0.2452	0.2340
0.50	0.3444	0.3460
1.00	0.3916	0.3820
5.00	0.5089	0.5260
10.00	0.5599	0.5820
100.00	0.7158	0.7240
1000.00	0.8370	0.8340
10000.00	0.9161	0.9140

TABLE B-IV

PREDICTIONS ON A FRESH SAMPLE $X \sim N(10, 25)$

Estimated CDF of Duration vs. Actual ($\hat{F}(t)$ vs. $F(t)$)		
Value of t ($\times 10^{3}$)	Estimated t CDF	Observed CDF
0.001	0.1836	0.1740
0.010	0.3119	0.2840
0.040	0.4013	0.3980
0.100	0.4627	0.4840
0.500	0.5736	0.5820
3.000	0.6952	0.6940
20.000	0.8024	0.8060
100.000	0.8703	0.8660
5000.000	0.9595	0.9540

APPENDIX C

SOME EVIDENCE ON THE RATE OF COVERGENCE OF ESTIMATED μ TO ACTUAL μ AS A FUNCTION OF SAMPLE SIZE

Estimators of distribution functions are known to converge slowly. We examined the behavior of the EM estimator for three sample sizes: $L = 500$, $L = 1000$, $L = 5000$. We fix $\alpha_0 = 0$, $\alpha_1 = 1$, $\alpha_2 = 0$ so that the model that is estimated is a mixture of exponentials. The mixing density is (4.2) with $\Delta = \beta = 1$. The same set of starting values for θ_i and P_i was employed in all runs. Results are presented in Table C-I. More points of increase are estimated in the larger samples. But even in a sample as large as 5000, the estimated cdf of θ is a poor indicator of the true underlying cdf. In all samples, the estimated distribution of t agrees rather closely with the empirical distribution. As expected, the fit improves with sample size. The performance of the more general estimator discussed in the paper was similar.

TABLE C-I

ESTIMATES OF THE MIXING MEASURE FOR A MIXTURE OF EXPONENTIALS MODEL
Starting Values for θ_i are $\theta = (1, \ldots, 10)$, $P_i = .1$ $(i = 1, \ldots, 10)$,
$\alpha_0 = 0$, $\alpha_1 = 1$, $\alpha_2 = 0$ Fixed.

Sample Size 500

Estimated θ_i	Estimated P_i	Estimated CDF	True CDF	Observed CDF
.1256	.0499	.0498	.0073	.0080
1.0416	.5817	.6316	.2796	.2820
4.0510	.3683	1.000	.9120	.9140

Estimated Cumulative Distribution of Durations vs. Actual ($\hat{F}(t)$ vs. $F(t)$)

Value of t	Estimated t CDF	Observed CDF
.2500	.3694	.3720
.5000	.5590	.5480
.7500	.6706	.6820
1.0000	.7443	.7560
1.2500	.7968	.7860
1.5000	.8359	.8240
1.7500	.8656	.8580
2.0000	.8886	.8840
3.0000	.9402	.9520
4.0000	.9608	.9580

Sample Size 1000

Estimated θ_i	Estimated P_i	Estimated CDF	True CDF	Observed CDF
.0429	.00613	.00613	.0009	.0020
.2525	.0877	.0938	.0270	.0380
1.1609	.5487	.6426	.3232	.3410
3.3978	.3127	.9553	.8529	.8490
12.4452	.0446	1.0000	.9999	1.0000

Estimated Cumulative Distribution of Durations vs. Actual ($\hat{F}(t)$ vs. $F(t)$)

Value of t	Estimated t CDF	Observed CDF
.2500	.3653	.3580
.5000	.5522	.5450
.7500	.6673	.6790
1.0000	.7436	.7510
1.2500	.7971	.7920
1.5000	.8361	.8240
1.7500	.8651	.8580
2.0000	.8872	.8850
3.0000	.9366	.9420
4.0000	.9577	.9580

TABLE C-I—*Continued*

		Sample Size 5000		
Estimated θ_i	Estimated P_i	Estimated CDF	True CDF	Observed CDF
.0839	.0141	.0141	.0033	.0032
.4323	.1324	.1465	.0704	.0686
1.0823	.3639	.5104	.2944	.2966
2.9211	.4399	.9503	.7887	.7856
4.5012	.0497	1.0000	.9389	.9380

Estimated Cumulative Distribution of Durations vs. Actual ($\hat{F}(t)$ vs. $F(t)$)

Value of t	Estimated t CDF	Observed CDF
.2500	.3616	.3582
.5000	.5606	.5612
.7500	.6785	.6814
1.0000	.7535	.7572
1.2500	.8044	.8036
1.5000	.8410	.8392
1.7500	.8682	.8648
2.0000	.8892	.8872
3.0000	.9386	.9412
4.0000	.9616	.9606

REFERENCES

[1] AHLFORS, L.: *Complex Analysis*, Third Edition. New York: McGraw Hill, 1979.

[2] ANDERSEN, E. B.: *Conditional Inference and Models for Measuring*. Copenhagen: Mentalhygieinisk Forlag, 1973.

[3] BECKER, G. S.: "A Theory of Marriage," *Economics of the Family*, ed. by T. W. Schultz. Chicago: University of Chicago Press, 1974, pp. 299–345.

[4] BECKER, G., AND G. STIGLER: "De Gustibus Non Est Disputandum," *American Economic Review*, 67(1977), 76–90.

[5] BLUMEN, I., M. KOGAN, AND P. MCCARTHY: *The Industrial Mobility of Labor as a Probability Process*. Ithaca, N.Y.: Cornell Studies in Industrial and Labor Relations, Vol. 6, 1955.

[6] BURDETT, K., AND D. MORTENSEN: "Labor Supply Under Uncertainty" in *Research in Labor Economics*, Vol. 2, ed. by R. Ehrenberg. London, CT: JAI Press, 1978, pp. 109–157.

[7] CHAMBERLAIN, G.: "Heterogeneity, Omitted Variable Bias, and Duration Dependence," Department of Economics, Harvard, 1979, forthcoming in *The Longitudinal Analysis of Labor Market Data*, ed. by J. Heckman and B. Singer. New York: Academic Press, 1984.

[8] ———: "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, 47(1980), 225–238.

[9] DEMPSTER, A. P., N. LAIRD, AND D. B. RUBIN: "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, Series B, 39(1977), 1–38.

[10] DOETSCH, G.: *Introduction to the Theory and Application of the Laplace Transformation*. Berlin: Springer Verlag, 1974.

[11] EFRON, B.: *The Jacknife, the Bootstrap and Other Resampling Plans*. Philadelphia: SIAM, 1982.

[12] ELBERS, C., AND G. RIDDER: "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model," *Review of Economic Studies*, 49(1982), 403–410.

[13] FELLER, W.: *An Introduction to Probability Theory and Its Applications, Vol.* 2. New York: Wiley, 1971.

[14] FLINN, C., AND J. HECKMAN: "New Methods for Analyzing Structural Models of Labor Force Dynamics," *Journal of Econometrics*, 18(1982), 115–168.

[15] ———: "Models for the Analysis of Labor Force Dynamics," in *Advances in Econometrics*, Vol. 1, ed. by R. Basmann and G. Rhodes. Greenwich, Conn.: JAI Press, 1982, pp. 35–95.

[16] ———: "Comment on 'Individual Effects in a Nonlinear Model: Explicit Treatment of Heterogeneity in the Empirical Job Search Literature'," unpublished manuscript, University of Chicago, August, 1982.

[17] HARRIS, J.: "Prenatal Medical Care and Infant Morality," in *Economic Aspects of Health*, ed. by V. Fuchs. Chicago: University of Chicago Press, 1982.

[18] HECKMAN, J.: "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1979), 73–92.

[19] ———: "Heterogeneity and State Dependence," in *Studies in Labor Markets*, ed. by S. Rosen. Chicago: University Chicago Press, 1981.

[20] ———: "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. Manski and D. McFadden. Cambridge: MIT Press, 1981.

[21] HECKMAN, J., AND B. SINGER: "The Identifiability of the Proportional Hazard Model," forthcoming in *Review of Economic Studies*, 1984.

[22] ———: "The Identification Problem in Econometric Models for Duration Data," invited paper in *Advances in Econometrics*, Proceedings of the Fourth World Congress of the Econometric Society, ed. by W. Hildenbrand. Cambridge: Cambridge University Press, 1982, pp. 39–77.

[23] HECKMAN, J., AND R. WILLIS: "A Beta Logistic Model for the Analysis of Sequential Labor Force Participation of Married Women," *Journal of Political Economy*, 85(1977), 27–58.

[24] HUBER, P.: *Robust Statistics*. New York: John Wiley and Sons, 1981.

[25] JOVANOVIC, B.: "Job Matching and the Theory of Turnover," *Journal of Political Economy*, 87(1979), 972–990.

[26] KALBFLEISCH, J., AND R. PRENTICE: *Statistical Analysis of Failure Time Data*. New York: Wiley, 1980.

[27] KIEFER, N., AND G. NEUMANN: "An Empirical Job Search Model with a Test of the Constant Reservation Wage Hypothesis," *Journal of Political Economy*, 87(1979), 69–82.

[28] ———: "Individual Effects in a Nonlinear Model: Explicit Treatment of Heterogeneity in the Empirical Job Search Model," *Econometrica*, 49(1981), 965–980.

[29] KIEFER, J., AND J. WOLFOWITZ: "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters," *Annals of Mathematical Statistics*, 27(1956), 887–906.

[30] LAIRD, N.: "Nonparametric Maximum Likelihood Estimation of a Mixing Distribution," *Journal of the American Statistical Association*, 73(1978), 805–811.

[31] LINDSAY, B.: "Properties of the Maximum Likelihood Estimator of a Mixing Distribution," *Statistical Distributions in Scientific Work*, ed. by C. Tallie. Hingham, Mass.: Kluwer Academics Publishers, 1981, pp. 95–109.

[32] ———: "The Geometry of Mixture Likelihoods, Part I," *Annals of Statistics*, 11(1983), 86–94.

[33] ———: "The Geometry of Mixture Likelihoods, Part II," *Annals of Statistics*, 11(1983), 783–792.

[34] LIPPMAN, S., AND J. MCCALL: "The Economics of Job Search: A Survey," *Economic Inquiry*, 14(1976), 113–126.

[35] MALINVAUD, E.: *Statistical Methods in Econometrics*. Amsterdam-London: North Holland, 1970.

[36] MEYER, P.: *Probability and Potentials*. Waltham, Mass.: Blaisdell Publishing, 1966.

[37] NICKELL, S.: "Estimating the Probability of Leaving Unemployment," *Econometrica*, 47(1979), 1249–1264.

[38] ROYDEN, H.: *Real Analysis*, Second Edition. New York: MacMillan Company, 1979.

[39] SILCOCK, H. R.: "The Phenomenon of Labor Turnover," *Journal of the Royal Statistical Society*, Series (A), 117(1954), 112–143.

[40] SIMAR, L.: "Maximum Likelihood Estimation of a Compound Poisson Process," *Annals of Statistics*, 4(1976), 1200–1209.

[41] TRICOMI, F.: *Integral Equations*. New York: Interscience Press, 1957.

[42] WIDDER, DAVID V.: *The Laplace Transform*. Princeton: Princeton University Press, 1941.

[43] WU, C. F.: "On the Convergence Properties of the EM Algorithm," *Annals of Statistics*, 11(1983), 95–103.