Statistical Methods in Medical Research 1994; **3**: 279–299

Econometric mixture models and more general models for unobservables in duration analysis

James J Heckman and **Christopher R Taber** University of Chicago, USA

This paper considers models for unobservables in duration models. It demonstrates how cross-section and time-series variation in regressors facilitates identification of single-spell, competing risks and multiple spell duration models. We also demonstrate the limited value of traditional identification studies by considering a case in which a model is identified in the conventional sense but cannot be consistently estimated.

Econometricians have obtained new results on the identification and estimation of mixture models and more general statistical models with unobservables. These results have applications to models for the analysis of duration data when the possibility of omitted covariates is explicitly allowed for.

This paper summarizes these results. We place a special emphasis on identification of nonparametric or partially nonparametric models. A major insight from the econometrics literature is that introduction of observed covariates in a structured way solves major identification problems. For example, the standard proof of non-identifiability of the widely used competing risks model assumes no covariates. These proofs forfeit an important source of identifying information which is heavily exploited in the econometrics literature. Dependent competing risks models can be identified if covariates satisfy the conditions presented below in our general discussion of identification in nonparametric duration models. By taking a position on the way observables enter duration models, it is possible to account for unobservables as well and still recover scientifically interpretable duration models.

Econometricians have investigated the behaviour of a variety of estimators for partially nonparametric duration models. We present results from studies of consistency, rates of convergence and asymptotic normality of estimators for the parametric portions of nonparametric models. We also summarize results from Monte Carlo studies.

The plan of this paper is as follows. We first present results on identification of single spell, multiple spell and multiple destination (competing risks) models. We then consider results on estimation. In a concluding section, we briefly discuss the econometric research frontier.

1.1 Basic identification results for single spell and competing risks models

This section considers the identification of the competing risks model when covariates are part of the model specification. The single spell model is a special case of our multiple spell set up. Given widespread interest in the competing risks model in

Address for correspondence: Professor James J Heckman, the Henry Schultz Professor of Economics, Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637, USA.

medical statistics, we start with this model first. The classical competing risks model excludes covariates. In the classical model there are J competing causes of failure indexed by the integers 1 to J. Associated with each failure j there is a latent failure time, T_j, which is the time to failure from cause j. The observed quantities are the duration to the first failure and the associated cause of failure

$$(\mathbf{T},\mathbf{I}) = \left\{ \min_j (\mathbf{T}_j), \arg\min_j (\mathbf{T}_j) \right\},$$

the identified minimum for the problem. In biology, T is the waiting time to death and I is the cause of death. David and Moeschberger,[1] Kalbfleisch and Prentice[2] and Cox and Oakes[3] discuss such models. In economics, Flinn and Heckman[4] estimate a competing risks model for unemployed workers, where T is the waiting time to the end of unemployment and I indexes the reason for leaving unemployment, i.e. getting a job or dropping out of the workforce. The problem posed in the competing risks literature is to identify the joint distribution of latent failure times from the distribution of the identified minimum.

Cox[5] and Tsiatis[6] show that for any joint distribution of the latent failure times there exists a joint distribution with independent failure times which gives the same distribution of the identified minimum. This nonidentification theorem has led much empirical work on multistate duration models to be conducted within an independent risks paradigm.

In many applications of the competing risks model, there is considerable interest in identifying the underlying distribution of latent failure times. Yashin *et al.*[7] demonstrates the importance of accounting for dependence among causes of death in assessing the impact of eliminating one cause of death on overall mortality rates. In behavioural or biological models with covariates, there is additional interest in determining the impact of the regressors on specific marginal failure time distributions. Thus Yashin *et al.*[7] investigate how smoking, blood pressure and body weight differentially affect the marginal distributions of time to death attributable to cancer and other illnesses. Flinn and Heckman[8] discuss how unemployment benefits and other variables differentially affect exit rates from unemployment to out of the workforce and to employment.

As a consequence of the Cox–Tsiatis theorem, in competing risks models without regressors it is necessary to make functional form assumptions about the joint distribution of failure times in order to identify the distribution. Basu and Ghosh,[9] David and Moeschberger[1] and Arnold and Brockett[10] exemplify this approach.

The recent literature in econometrics establishes identifiability for models with covariates. (See Heckman and Honoré.[11]) It demonstrates conditions under which it is possible to identify the joint distribution of failure times without invoking distributional assumptions. The literature summarized below considers identification of competing risks models in which each marginal distribution is a nonparametric version of the Cox[12] proportional hazard model. It also presents identifiability results for an accelerated hazard competing risk model with covariates.

To simplify the exposition in this survey paper, we consider models with only two competing failure times. All of our results can easily be generalized to competing risks models with an arbitrary, but known, finite number of latent failure times. We follow the discussion Heckman and Honoré[11] rather closely.

The Cox[12] proportional hazard model specifies the survivor function conditional on the covariates to be

$$S(t \mid x) = \exp\{-Z(t)\phi(x)\}, \tag{1}$$

where $Z(\cdot)$ is the integrated hazard and $\phi(x)$ is usually specified as $e^{x\beta}$ where β is a vector of parameters. Assuming $Z(t)$ is differentiable, the associated hazard is $Z'(t)\phi(x)$. Differentiability is often only a convenience. Below we provide conditions under which it can be relaxed. Usually differentiability in a neighbourhood is all that is required.

One way to combine the Cox proportional hazard specification with the competing risks model is to assume that each of the potential failure time distributions has a proportional hazard specification, possibly with different integrated hazard functions and different functional forms for ϕ or different values of β when $\phi(x) = e^{x\beta}$. If independence is assumed, then it is straightforward to specify the resulting competing risks model (Kalbfleisch and Prentice,[2] Flinn and Heckman[13]).

Introduce dependence among latent failure times in the following way. In order to produce random variables from an independent competing risks model one could generate two independent random variables from a $U(0,1)$ distribution, U_1 and U_2, and then solve $S_1(T_1) = U_1$ and $S_2(T_2) = U_2$ for the potential failure times T_1 and T_2. This is equivalent to solving the equations

$$Z_1(T_1) = -\log U_1\{\phi_1(x)\}^{-1}, Z_2(T_2) = -\log U_2\{\phi_2(x)\}^{-1}, \tag{2}$$

for T_1 and T_2. Dependence between T_1 and T_2 can be introduced by assuming that U_1 and U_2 are not necessarily independent. This implies that the joint survivor function of T_1 and T_2 conditional on $X = x$ is

$$S(t_1, t_2 \mid x) = K[\exp\{-Z_1(t_1)\phi_1(x)\}, \exp\{-Z_2(t_2)\phi_2(x)\}], \tag{3}$$

where K is the distribution function for (U_1, U_2) and we assume that $Z_1(0) = 0$ and $Z_2(0) = 0$. If the marginal distributions are to be of the proportional hazard form, the marginal distributions associated with K must be of the form y^c for some $c > 0$.

Clayton and Cuzick[14] and Flinn and Heckman[4] consider generalizations of the proportional hazard model which are special cases of equation (3). The first generalization assumes that the true model is an independent competing risks model with $\phi(x) = e^{x\beta}$ but that one of the covariates is not observed. This implies the model

$$S(t_1, t_2 \mid x) = \int_{\underline{\underline{\theta}}} \exp\{-Z_1(t_1)e^{x\beta_1 + c_1\theta}\} \exp\{-Z_2(t_2)e^{x\beta_2 + c_2\theta}\} \, dG(\theta), \tag{4}$$

where G is the distribution of the unobserved covariate, assumed independent of X, and the integration is over the support of the unobserved covariate, θ. Defining

$$K(\eta_1, \eta_2) = \int_{\underline{\underline{\theta}}} \eta_1^{\exp(c_1\theta)} \eta_2^{\exp(c_2\theta)} \, dG(\theta)$$

shows that equation (4) is a special case of equation (3). Heckman and Honoré[11] produce a theorem for models more general than standard univariate mixture models.

A second approach taken by Clayton and Cuzick[14] specifies

$$S(t_1, t_2 \mid x) = \begin{cases} [\exp\{\gamma Z_1(t_1)\phi_1(x)\} + \exp\{\gamma Z_2(t_2)\phi_2(x)\} - 1]^{-1/\gamma} & (\gamma > 0) \\ \exp\{-Z_1(t_1)\phi_1(x) - Z_2(t_2)\phi_2(x)\} & (\gamma = 0) \end{cases} \tag{5}$$

This specification is also a special case of equation (3). In this case

$$K(\eta_1, \eta_2 \mid x) = \begin{cases} (\eta_1^{-\gamma} + \eta_2^{-\gamma} - 1)^{-1/\gamma} & (\gamma > 0) \\ (\eta_1 \eta_2) & (\gamma = 0). \end{cases} \tag{6}$$

This specification of K has uniform marginal distributions for all γ and therefore equation (5) has marginal distributions that are consistent with a proportional hazard specification. The independent competing risks model with proportional marginal hazards is a special case where $\gamma = 0$.

The following theorem proved by Heckman and Honoré[11] gives sufficient conditions for the identifiability of Z_1, Z_2, ϕ_1 and ϕ_2 as well as K for the model given by equation (3).

THEOREM 1

Assume that (T_1, T_2) has the joint survivor function as given in equation (3). Then Z_1, Z_2, ϕ_1, ϕ_2 and K are identified from the identified minimum of (T_1, T_2) under the following assumptions.

- (i) *K is continuously differentiable with partial derivatives K_1 and K_2 and for $i = 1,2$ the limit as $n \to \infty$ of $K_i(\eta_{1n}, \eta_{2n})$ is finite for all sequences of η_{1n}, η_{2n} for which $\eta_{1n} \to 1$ and $\eta_{2n} \to 1$ for $n \to \infty$. We also assume that K is strictly increasing in each of its arguments in all of $[0,1] \times [0,1]$.*
- (ii) *$Z_1(1) = 1$, $Z_2(1) = 1$, $\phi_1(x_0) = 1$, $\phi_2(x_0) = 1$ for some fixed point x_0 in the support X.*
- (iii) *The support of $\{\phi_1(x), \phi_2(x)\}$ is $(0, \infty) \times (0, \infty)$.*
- (iv) *Z_1 and Z_2 are nonnegative, differentiable, strictly increasing functions, except that we allow them to be ∞ for finite t.*

Proof The proof is instructive and we present the main outlines. (See Heckman and Honoré,[11] for more details). By assumption we know

$$Q_1(t) = p(T_1 > t, T_2 > T_1), \; Q_2(t) = p(T_2 > t, T_1 > T_2),$$

for all t and x. For notational convenience we suppress the dependence of Q_1 and Q_2 on x. It follows from Theorem 1 of Tsiatis[6] that

$$Q_1'(t) = \left[\frac{\partial S}{\partial t_1}\right]_{t_1=t_2=t} \qquad Q_2'(t) = \left[\frac{\partial S}{\partial t_2}\right]_{t_1=t_2=t}.$$

From the expression for S it follows that

$$Q_1'(t) = -K_1[\exp\{-Z_1(t)\phi_1(x)\}, \exp\{-Z_2(t)\phi_2(x)\}]\exp\{-Z_1(t)\phi_1(x)\}Z_1'(t)\phi_1(x).$$

Calculation of the ratio between Q_1' at an arbitrary $x \neq x_0$ in the support of X and Q_1' at x_0 gives

$$\frac{K_1[\exp\{-Z_1(t)\phi_1(x)\}, \exp\{-Z_2(t)\phi_2(x)\}] \exp\{-Z_1(t)\phi_1(x)\} Z_1'(t)\phi_1(x)}{K_1[\exp\{-Z_1(t)\phi_1(x_0)\}, \exp\{-Z_2(t)\phi_2(x_0)\}] \exp\{-Z_1(t)\phi_1(x_0)\} Z_1'(t)\phi_1(x_0)}.$$

Cancelling $Z_1'(t)$ and taking the limit as $t \to 0$ we get $\phi_1(x)$. We can thus identify $\phi_1(x)$ for all x in the support of X. Using a parallel result for Q_2' we can identify ϕ_2. (Notice that only (i) and part of (ii) are used to identify $\phi_1(x)$ and $\phi_2(x)$. Differentiability of Z_1 and Z_2 is used but the property need only be local. Z_1 and Z_2 need not be strictly increasing. The support of $\phi_1(x)$ and $\phi_2(x)$ can be any intervals including points).

Next observe that by setting $t = 1$ and letting $\phi_1(x)$ and $\phi_2(x)$ range over the set $(0,\infty) \times (0,\infty)$, which can be done as a consequence of assumption (iii), we trace out K. (This is the first place in the proof where (ii) and (iii) are both used).

To identify $Z_2(t)$, let $\phi_2(x)$ go to 0 holding $\phi_1(x)$ fixed. We can do this as a consequence of assumption (iii). Then $S(t,t)$ goes to a function $H[\exp\{-Z_1(t)\phi_1(x)\}]$, where H is a known increasing function since K is known and is increasing in its argument. Since ϕ_1 is already identified, and $Z_1(t) = 1$ by assumption, Z_1 can be identified; Z_2 is identified in the same way.

The assumptions made in Theorem 1 deserve a few additional comments. Observe that fewer assumptions are required to identify ϕ than are required to identify K, and identification of K requires fewer assumptions than does the identification of the $Z_i(.)$. Assumption (ii) is an innocuous normalization. Multiplying by a positive number and dividing ϕ_1 by the same number has no effect on the survivor function. Thus without loss of generality we can assume that $Z_1(1) = 1$. With this normalization we can divide ϕ_1 by a positive number α and define a new K, say $\tilde{K}$, by $\tilde{K}(\eta_1, \eta_2) = K(\eta_1^{\alpha}, \eta_2)$. This redefinition has no effect on the survivor function, so we can assume $\phi_1(x_0) = 1$ for some x_0 in the support of x. The normalizations on Z_2 and ϕ_2 are justified in the same way. The assumption that Z_1 and Z_2 are strictly increasing and differentiable is necessary only in a neighbourhood of zero. Continuity of Z_1 and Z_2 implies that the potential failure times T_1 and T_2 have continuous distributions, and if Z_1 and Z_2 are strictly increasing then T_1 and T_2 both have convex support. Observe that Z_1 or Z_2 can be ∞. Thus the failure times are permitted to have bounded support. We also do not need to assume that either Z_1 or Z_2 goes to ∞ as t goes to ∞, which implies that we allow the potential failure times to be infinite with positive probability, so we do not exclude defective duration distributions.

Assumption (iii) is satisfied in the case where $\phi_i(x) = \exp(x\beta_i)$ and there is one covariate which enters both equations but with different coefficients and for which the support is all of the real line. Yashin *et al.*[7] and Manton *et al.*[15] use normal covariates in a competing risks model and argue the plausibility of assuming that different causes affect the marginal distributions in different ways, so the β_i are distinct across specific causes.

Assumption (i) is a technical assumption which has to be either assumed or verified in specific cases. In the model given by equation (4), assumption (i) is satisfied if

$$\int_{\Theta} e^{c_i\theta} \, dG(\theta) < \infty.$$

The finiteness of this expectation is exactly the condition on unobservables required for nonparametric identification of the proportional hazard model that appears in Elbers and Ridder.[16]

1.2 Competing risks in an accelerated hazard model

We next consider the identifiability of a competing risks version of the accelerated hazard model. The survivor function for the accelerated hazard model is given by

$$S(t \mid x) = \exp[-Z\{t\phi_1(x)\}]. \tag{7}$$

Using the same procedure as was used for the proportional hazard model, Heckman and Honoré[11] introduce dependence between two potential failure times by assuming that they are are generated by solving $U_1 = S_1(T_1)$ and $U_2 = S_2(T_2)$, and where U_1 and U_2 are not necessarily independent uniform $U(0,1)$ random variables. If the joint distribution of U_1 and U_2 is K then the joint survivor function for T_1 and T_2 is

$$S(t_1, t_2 \mid x) = K(\exp[-Z_1\{t_1\phi_1(x)\}], \exp[-Z_2\{t_2\phi_2(x)\}]). \tag{8}$$

Notice that for all K the bivariate survival model has marginal distributions with accelerated hazards.

Defining

$$\tilde{K}(\eta_1, \eta_2) = K(\exp[-Z_1\{-\log(\eta_1)\}], \exp[-Z_2\{-\log(\eta_2)\}] \tag{9}$$

we can write (7) in the same form as equation (3):

$$S(t_1, t_2 \mid x) = \tilde{K}[\exp\{-\tilde{Z}_1(t_1)\phi_1\}, \exp\{-\tilde{Z}_2(t_2)\phi_2\}],$$

where $\tilde{Z}_1(t) = t$ and $\tilde{Z}_2(t) = t$. This means that the specification (3) is general enough to cover dependent accelerated hazard models as a special case. Under the conditions of Theorem 1 we can identify $\tilde{K}$, ϕ_1 and ϕ_2. If it is further assumed that the marginal distribution of K in equation (7) are uniform then we can also identify Z_1 and Z_2. The uniformity of the marginal distribution of K implies that the marginal distribution of $\tilde{K}$ is given by

$$\tilde{K}(\eta_1, 1) = K(\exp[-Z_1\{-\log(\eta_1)\} - 1] = \exp[-Z_1\{-\log(\eta_1)\}],$$

and hence $Z_1(t) = -\log\{\tilde{K}(e^{-t}, 1)\}$ and by a similar argument $Z_2(t) = -\log\{\tilde{K}(e^{-t}, 1)\}$. Thus the model given by equation (7) is identified if it is assumed that K has uniform marginal distributions. Moreover it is clear that identification of K and of the Z_1 can be established if the marginals of K are specified to be any other known distribution. Note that equation (3) can be interpreted as arising from an accelerated hazard model if and only if $Z_1(t)$ and $Z_2(t)$ are power functions.

1.3 Proportional hazard models for single spells

The logic underlying the proof of Theorem 1 can be utilized to establish the identifiability of competing risks models with an arbitrary but known number of risks. With only one risk, this implies that we can identify single spell duration models of the type

$$S(t \mid x) = K[\exp\{-Z(t)\phi(x)\}].$$

This model includes the proportional hazard model with unobserved heterogeneity and the accelerated hazard model as special cases. A more familiar representation for the proportional hazards model which has been widely used since Elbers and Ridder[16] writes this as a Laplace transform with θ as an unobservable. Thus we write

$$S(t \mid x) = \int_0^\infty e^{-Z(t)\phi(x)\theta}\, dG(\theta) = L(Z(t)\phi(x)) \tag{10}$$

where L is the Laplace transform. A number of results are available for this widely used special case which we now present. Many are applications of standard results in the theory of Laplace transforms. In this section we restate results already implicit in section one, in terms of more familiar-looking Laplace transform theory.

The model given in equation (10) is exactly the model studied by Elbers and Ridder[16] and Heckman and Singer.[17] It is clear that (10) cannot be identified without some normalization. Since only the product of $Z(t), \phi(x)$ and θ appear in (10), we can change the scale of $Z(t), \phi(x)$ and θ and still be consistent with equation (10). We continue to make the conventional normalizations of the form $Z(t_0) = 1$ and $\phi(x_0) = 1$ for some t_0 and x_0. To simplify notation we assume in this subsection that X is one-dimensional. This restriction is of no consequence, for if X is of higher dimension, then we can always split X in (X_1, X_2) where X_1 is one-dimensional. We can then perform all of the analysis conditional on X_2, treating only X_1 as the covariate.

In the analysis of equation (10) we will make use of some of the properties of the Laplace transform first used in this context by Honoré.[18] For completeness we state the most important of these in the following lemma.

LEMMA

Assume that $L(t)$ is a Laplace transform. Then $\tilde{L}(t) = L(at^b)$ is also a Laplace transform if $a > 0$ and $0 < b \leq 1$.

Proof Follows from Feller[19] Theorem 1 (page 439) and Criterion 2 (page 441).

The identifiability of equation (10) is investigated in the following theorem. Theorem 2 presented next shows that (10) is in general not identified, because if Z and ϕ are consistent with (10), then so is Z^b and ϕ^b for $b \geq 1$. Theorem 1 shows that these models are identified given that we exclude power transformations.

THEOREM 2

If (Z, ϕ, G) is consistent with (10) then for any $\alpha \geq 1$ there exists a $G^\star$ such that $(Z^\alpha, \phi^\alpha, G^\star)$ is also consistent with (1).

Proof Follows directly from Lemma 1. See also Heckman and Singer,[17] and Heckman and Singer,[20] page 64.

THEOREM 3

(Ridder,[21] Honoré[18]). Assume that the support of X is connected, and that ϕ and Z are differentiable and non-constant in the support of X and T. Then any two specifications (Z_1, ϕ_1, G_1) and (Z_2, ϕ_2, G_2) consistent with (10) must satisfy

$$Z_1(t) = mZ_2(t)^k \tag{11}$$

$$\phi_1(x) = n\phi_2(x)^k \tag{12}$$

and

$$L_{G_2}(t) = L_{G_1}(ct^k) \tag{13}$$

for some positive, real numbers k, m and n, and with $c = mn$.

Equation (13) in Theorem 3 is very useful because it gives the relationship between the mixing distributions that are consistent with (11). One way to get identification of (10) is therefore to make assumptions that guarantee that (10) cannot be satisfied for different specifications. We now follow Honoré[18] and show how the identification results of Elbers and Ridder[16] and Heckman and Singer[17] can be derived in this way.

Elbers and Ridder[16] show that the model is identified if it is assumed that G has finite mean. This result can be easily derived as a corollary to Theorem 3.

Corollary 1. (Honoré[18]) The model is identified if G has a finite mean.

It is interesting to note that if a specification (Z_1, ϕ_1, G_1) consistent with equation (10) has finite mean, then any other specification (Z_2, ϕ_2, G_2) must be related to (Z_1, ϕ_1, G_1) by equations (11), (12), and (13) with $k < 1$.

In a generalization of Elbers and Ridder,[16] Heckman and Singer[17] prove that (10) is identified if it is also assumed that

$$W(\ell) \equiv L_G'(\ell) \sim \frac{c}{(\ln l/\ell)^{\delta}(1/\ell)^{\varepsilon-1}L(1/\ell)} \quad \text{as } \ell \to 0 \tag{14}$$

where ε is known, $0 < \varepsilon \leqslant 1$, $\delta \geqslant 0$, $c > 0$, and L satisfies the condition that for any fixed $\kappa > 0$, $L(\kappa t)/L(t) \to 1$ as $t \to \infty$. (Following the notation of Feller,[19] we write $u \sim v$ if $u/v \to 1$.) Equation (14) is equation (5b) in Heckman and Singer.[17]

Theorems 2 and 3 tells us that even though equation (10) is identified if either moment or tail conditions are imposed on G, there are infinitely many specifications that are consistent with (10) if such conditions are not imposed. Most of the identification theorems produced thus far in the literature are for multiplicative, separable hazards. ($\lambda(t,x,\theta) = Z'(t)\phi(x)\theta$). Heckman[22] considers identification of non-separable hazard $\lambda(t,x,\theta) = m(t,x)\theta$ and shows that if $E(\theta) < \infty$ and is normalized to some value, $m(t,x)$ is identified in the neighbourhood of $t = 0$.

1.4 Results for models with single spells without covariates

With additional functional form assumptions on the base hazard, it is not necessary to have access to covariates to identify the model. For specificity we first consider identifiability for the class of Box–Cox hazards introduced in Flinn–Heckman[8]:

$$Z'(t) = \exp\left(\gamma\left(\frac{t^{\lambda-1}}{\lambda}\right)\right).$$

For $\lambda = 0$, a Weibull hazard model is produced. For $\lambda = 1$, a Gompertz hazard model is obtained. $\gamma = 0$ produces the exponential model. This class of hazard models subsumes a wide variety of models used in applied duration analysis. (For further discussion see Flinn and Heckman[8].)

For this class of hazard models there is an interesting trade-off between the interval of admissible λ and the number of bounded moments that is assumed to characterize $G(\theta)$. More precisely, the following propositions can be proved.

Proposition 1 (Heckman and Singer[17]). *For the true value of λ, λ_0, defined so that $\lambda_0 \leq 0$, if $E(\theta) < \infty$ for all admissible G, and for all bounded γ, the triple $(\gamma_0, \lambda_0, G_0)$ is uniquely identified.*

Proposition 2 (Heckman and Singer[17]). *For the true value of λ, λ_0, such that $0 < \lambda_0 < 1$, if all admissible G are restricted to have a common finite mean that is assumed to be known a priori ($E(\theta) = \mu_1$) and a bounded (but not necessarily common) second moment $E(\theta^2) < \infty$, and all admissible γ are bounded, the triple $(\gamma_0, \lambda_0, G_0)$ is uniquely identified.*

Proposition 3 (Heckman and Singer[17]). *For the true value of λ, λ_0, restricted so that $0 < \lambda_0 < j$, j a positive integer, if all admissible G are restricted to have a common finite mean that is assumed to be known a priori ($E(\theta) = \mu_1$) and a bounded (but not necessarily common) $j+1$-st moment ($E(\theta^{j+1}) < \infty$), and all admissible γ are bounded, the triple $(\gamma_0, \lambda_0, G_0)$ is uniquely identified.*

Thus for $|\gamma| < \infty$ if $\lambda_0 \leq 0$, finiteness of the mean of θ ($E(\theta) < \infty$) is all that is required in order to secure identification $(\gamma_0, \lambda_0, G_0)$. For $j > \lambda_0 > 0$, the admissible G are restricted to have a common finite mean ($E(\theta) = \mu_1$) and a bounded but not necessarily common $j+1$-st moment ($E(\theta^{j+1}) < \infty$).

The general strategy of specifying a flexible functional form for the hazard and placing moment restrictions on the admissible G works in other models besides the Box–Cox class of hazards. As an example, we consider a nonmonotonic log logistic model:

$$Z'(t) = \frac{(\lambda\alpha)(\lambda t)^{\alpha-1}}{1 + (\lambda t)^{\alpha}}, \quad \infty > \lambda, \alpha > 0.$$

Proposition 4 *For the log logistic model with multiplicative non-negative heterogeneity θ, the triple $(\lambda_0, \alpha_0, G_0)$ is uniquely identified provided that the admissible G are restricted to have a common finite mean $E(\theta) = \mu_1 < \infty$.*

Proof: (*See Heckman and Singer*[17]).

1.5 Models with time-varying covariates

The next theorem demonstrates that identification of single spell models (and, by extension, multiple state competing risks models) is facilitated by access to time-varying variables. Before stating any formal results on this topic, it is necessary to be precise about what we mean by duration model with time-varying variables.

Kalbfleisch and Prentice[2] present a vague and confusing taxonomy of duration models with time-varying variables. (See their discussion of 'external' and 'internal' covariates.) Fortunately, Yashin and Arjas[23] have clarified the issue. Let $\{X(u)\}_0^t$ be the sample path of a continuous-time stochastic process up to time t. Realizations of this process are independent of θ. For the hazard written in terms of $X(t)$, θ and t, where θ is an invariant random variable and $x(t)$ is the sample realization (at time t) of the stochastic covariate process:

$$\lambda(t, x(t), \theta),$$

Yashin and Arjas demonstrate that the conventional exponential representation of the survivor function in the time-varying case:

$$S(t \mid \{X(u)\}_0^t, \theta) = \exp - \int_0^t \lambda(u, X(u), \theta)\, du \qquad (15)$$

is valid under one of two *sufficient* conditions:

(a) T given $(\{X(u)\}_0^t, \theta)$ is a random variable with an absolutely continuous distribution function

or

(b) $P(T \leq t \mid \{X(u)\}_0^t, \theta)$ is *predictable* with respect to $\{(X(u)\}_0^t, \theta)$ (i.e. measurable with respect to the sub σ-algebra of F_{t-}, events up to t- but not up to t) and

$$P(T \leq t \mid \{X(u)\}_0^t, \theta) = P(T \leq t \mid \{X(u)\}_0^\infty, \theta).$$

Condition (b) is called Granger noncausality of the X process. Unless one of these conditions is satisfied, one cannot guarantee that the representation of the survivor function as minus an exponentiated integrated hazard (i.e. in the form (15)) is valid. Either set of conditions ensures that $P(T \leq t \mid \{(X(u)\}_0^t, \theta)$ is a martingale. These conditions rule out contemporaneous feedback between $X(t)$ and t. They also rule out the case that future values of X, not known at time t, predict the probability of exit from the state at time t.

These conditions are to be distinguished from those that arise when $X(t)$ is a deterministic function of t for all sample paths. Then any distinction between t and $X(t)$ is arbitrary in any sample with a common starting point ($t = 0$) for all observations. Heckman and Singer[20] discuss how access to successive samples with different real time starting points with the process observed at different points in times (in real time) may afford identification of the model in this situation.

THEOREM 4

(Honoré[18]). *Suppose that there are two types of covariates each satisfying condition (a) or (b) from Yashin and Arjas[23]: either the covariate is time-invariant and satisfies the conditions* *of Theorem 1, or for some fixed $t^\star$ the covariate is x_1 for $t<t^\star$ and x_2 for $t \geq t^\star$. If $\phi(x_1) \neq \phi(x_2)$, Z satisfies the conditions of Theorem 1, and $Z' > 0$ in a neighbourhood of $t^\star$, then (10) is identified. If $t^\star > 0$, no moments of θ need exist.*

Proof: For $t > t^\star$ compare the survivor function conditional on the covariate being x_1 for all t, $L(Z(t)\phi(x_1))$, to the survivor function conditional on the covariate being x_1 for $t \geq t^\star$ and x_2 for $t \geq t^\star$, $L(Z(t^\star)\phi(x_1) + (Z(t) - Z(t^\star))\phi(x_2))$. The ratio of the derivative of the former with respect to t to the derivative of the latter with respect to t is

$$\frac{L'(Z(t)\phi(x_1))Z'(t)\phi(x_1)}{L'(Z(t^\star)\phi(x_1) + (Z(t) - Z(t^\star))\phi(x_2))Z'(t)\phi(x_2)}.$$

The limit of this as $t \to t^\star$ from the right is $\phi(x_1)/\phi(x_2)$. The noteworthy feature of this result is that finiteness of the mean of the unobservable θ need not be assumed.

As in the proof of Theorem 3, we can use the data with time-invariant covariates to identify a $\bar{Z}(t)$ and a $\bar{\phi}(x)$ such that $Z(t) = \bar{Z}(t)^\alpha$ and $\phi(x) = \bar{\phi}(x)^\alpha$ for some unknown $\alpha > 0$.

For any $t > t^\star$ we can find a $t^0 > t^\star (t^0 \neq t)$, such that

$$L(Z(t)\phi(x_1)) = L(Z(t^\star)\phi(x_1) + (Z(t^0) - Z(t^\star))\phi(x_2))$$

or equivalently

$$Z(t)\phi(x_1) = Z(t^\star)\phi(x_1) + (Z(t^0) - Z(t^\star))\phi(x_2)$$

or

$$Z(t) - Z(t^\star) = (Z(t^0) - Z(t^\star))(\phi(x_2)/\phi(x_1))$$

or

$$\bar{Z}(t)^\alpha - \bar{Z}(t^\star)^\alpha = (\bar{Z}(t^0)^\alpha - \bar{Z}(t^\star)^\alpha)(\phi(x_2)/\phi(x_1)).$$

As discussed above, an arbitrary normalization is necessary. Let $Z(t^\star) = 1$. Then

$$\bar{Z}(t)^\alpha - 1 = (\bar{Z}(t^0)^\alpha - 1)(\phi(x_2)/\phi(x_1)).$$

Since $\phi(x_1)/\phi(x_2)$ is identified by the argument above, this identifies α, and hence $Z(t)$ and $\phi(x)$. $G(\theta)$ can be obtained by inverting the Laplace transform.

The preceding theorem can clearly be extended to consider cases where the $X(t)$ have countable discrete jumps. The main point is that by allowing the covariates to vary in a simple way drastically changes the nature of the identifiability of proportional hazard models with a proportional unobserved component. The known discrete jump points take the place of the origin in Theorem 1 and so allow us to identify models without assuming that the mean is finite. Honoré[18] conjectures that the sensitivity of the parameter estimates of ϕ to different specifications of Z and G will depend on whether or not there are time-varying covariates.

It is straightforward to extend Honoré's result to models with covariates that are realizations from stochastic processes with continuous sample paths.

THEOREM 5

For a model with regressors that are realizations from continuous time stochastic processes with continuous sample paths (e.g. diffusions), it is possible to identify $Z(t)$ *and* $\phi(x)$ *in* $\lambda(t, x(t), \theta) = Z'(t)\phi(x(t))\theta$ *– a separable form of the model and to identify* $G(\theta)$.

Proof: Invoke Yashin–Arjas condition (a) or (b) and write:

$$S(t \mid \{x(u)\}_0^t) = \int_\theta \exp\left[-\theta \int_0^t \lambda(u,x(u))\,du\right] dG(\theta).$$

Differentiating with respect to t

$$\frac{\partial S(t \mid \{x(u)\}_0^t)}{\partial t} = \left[\int_\theta \theta \exp - \left[\theta \int_0^t \lambda(u,x(u))\,du\right] dG(\theta)\right] \lambda(t,x(t)).$$

Evaluating this derivative at the same t for two different values of $x(t)$ (i.e. $x'(t)$ and $x''(t)$) with the same sample paths up to t- (i.e. $\{x'(u)\}_0^{t-} = \{x''(u)\}_0^{t-}$) we form the ratio of the derivatives of the survivor functions

$$\frac{\dfrac{\partial S(t \mid \{x'(u)\}_0^t)}{\partial t}}{\dfrac{\partial S(t \mid \{x''(u)\}_0^t)}{\partial t}} = \frac{\lambda(t,x'(t))}{\lambda(t,x''(t))} = \frac{\phi(x'(t))}{\phi(x''(t))}.$$

With one normalization e.g. $\phi(x(0)) = 1$, we can recover $\phi(x(t))$ over the support of $x(t)$, $t \in (0,\infty)$ without invoking finiteness of the mean of θ. Only at $t = 0$ do we need to invoke the finiteness of the mean. The identification of the remainder of the model follows using an argument like that given in Theorems 4 and 1.

It is clear how to combine Theorems 4 and 5 to generate identification for models with both jump covariate processes and processes with continuous sample paths. Invoking a finiteness of mean assumption, ($E(\theta) < \infty$). McCall[24] presents a set of conditions for the identification of models with time-varying variables that possess a special structure. McCall[25] also presents identifiability conditions for models with time-varying coefficients.

1.6 Multi-spell duration models

We next consider the identification of models that have two observations for each individual. Extension of the results to models with more than two spells is in most cases straightforward. We draw heavily on Honoré[26] who has pioneered in this area.

First, we will show that a multi-spell model can be identified even if there are no covariates and even if we do not invoke specific functional form assumptions of the sort invoked in Section 4. Multi-spell models without covariates have been used for example by Heckman *et al.*[27]

Assume that observations are independently distributed conditional on the individual, but that there is an individual-specific θ component common across spells. Specifically,

$$S(t_1, t_2) = \int_0^\infty e^{-\theta Z_1(t_1) - \theta Z_2(t_2)} \, dG(\theta) = L(Z_1(t_1) + Z_2(t_2)). \tag{16}$$

THEOREM 6

(Honoré[26]). *Suppose that for $i = 1,2, Z_i$ is differentiable and non-constant, then Z_1, Z_2 and G are identified except for a normalizing constant.*

It is clear from equation (16) that some kind of normalization is necessary. Alternative normalizations are $Z_2'(t_0) = k$, $Z_1(t_0) = k$ or $Z_2(t_0) = k$. We could also impose $E(\theta) = 1$. The latter is more restrictive, as it imposes the restriction that $E[\theta] < \infty$.

The result in Theorem 6 is interesting in that it highlights the benefit of having an additional observation on the same person, even if it is not assumed that the two observations have the same baseline hazard. It is also worth noting that no assumptions are needed about the moments of θ. From the proof of Theorem 6 it follows that if we observe $S(t_1, t_2)$ only for $t_1 + t_2 \leqslant \bar{T}$, then we can identify $Z_i(t)$ for $t_i \leqslant \bar{T}$, $i = 1,2$, as well as G.

Covariates are essential for identification in a single-spell model. We will now show how covariates in a multi-spell duration model can help relax some of the assumptions needed for identification of more general versions of (16).

First consider the case where θ is allowed to be different for different spells. Then we have

$$S(t_1, t_2 \mid x) = \int_0^\infty \int_0^\infty e^{-\theta_1 Z_1(t_1)\phi_1(x)} e^{-\theta_2 Z_2(t_2)\phi_2(x)} \, dG(\theta_1, \theta_2) \tag{17}$$

where G is the joint distribution (θ_1, θ_2). It follows from our discussion of single spell models that if the conditions of Theorem 3 and the conditions below the theorem are satisfied, then we can identify Z_1, Z_2, ϕ_1, ϕ_2, as well as the marginal distributions of G, by considering the marginal distributions of (T_1, T_2). It then allows from the uniqueness of the multi-dimensional Laplace transform that G is identified as well. Honoré[26] establishes that:

THEOREM 7

Let the conditional distribution of (T_1, T_2) given X be given by equation (17). If for $i = 1,2$, (θ_i, ϕ_i, Z_i) satisfies the assumptions of Theorem 3 and the first corollary below the Theorem (e.g. the Elbers–Ridder conditions) then ϕ_1, ϕ_2, Z_1, Z_2 and G are uniquely identified (except for the scale normalizations discussed earlier).

We next turn to extensions of equation (17) that allow the specification for the hazard in the second spell to depend on the outcome of the first spell. Thus we write the density of (T_1, T_2) as

$$f(t_1, t_2 \mid x) = \int_0^\infty \int_0^\infty \theta_1 Z'(t_1)\phi_1(x) \, e^{-\theta_1 Z_1(t_1)\phi_1(x)} \, \theta_2 Z'(t_2 \mid t_1)\phi_2(x \mid t_1) \cdot e^{-\theta_2 Z_2(t_2 \mid t_1)\phi_2(x \mid t)} \, dG(\theta_1, \theta_2). \tag{18}$$

Models like (18) have been used, for example, in Heckman and Borjas[28] and Heckman *et al.*[27] The dependence of ϕ_2 and Z_2 on T_1 is usually called 'lagged duration dependence'.

Conditional on T_1, θ_2 is not independent of X, so we cannot identify ϕ_2 and Z_2 from the conditional distribution of T_2 given T_1 using the previous results. A separate analysis is necessary. The next theorem gives conditions sufficient to guarantee identification of equation (18).

THEOREM 8

The functions Z_1, Z_2, ϕ_1, ϕ_2 and G in (18) are uniquely identified (except for scale-normalizations) if

(1) *(Z_1, ϕ_1, θ_1) satisfies the conditions of Theorem 3 and Corollary 1.*
(2) *For given t_1, (Z_2, ϕ_2, θ_2) satisfies the conditions of Theorem 3 and Corollary 1, and $h(t_1) = Z_1'(t)\phi_1(x)\theta > 0$ for all t_1.*
(3) *θ_1 and θ_2 are positive random variables with $E(\theta_1) = 1$, $E(\theta_1\theta_2) = 1$, and $h(t_1^*) = 1$ for some known t_1^*.*

Proof: See (Honoré,[26] Theorem 3).

1.7 A qualification of the preceding results on identification

Before concluding our discussion of identification, it is important to note that the concept of identifiability employed in this and other papers is the requirement that the mapping from a space of (conditional hazards) X (a restricted class of probability distributions) to (a class of joint frequency functions for durations and covariates) be one to one and onto. This formulation of identifiability is standard. In the literature on identification there is no requirement of a metric on the spaces or of completeness. Such requirements are essential if consistency of a partially parametric estimator is desired. In this connection, Kiefer and Wolfowitz[29] propose a definition of identifiability in a metric space whereby the above-mentioned mapping is 1:1 on the completion (with respect to a given metric) of the original spaces. Without some additional restriction in defining the original space, undesirable distributions can appear in the completions.

As an example, consider a Weibull hazard model with conditional survivor function given an observed k-dimensional covariate x defined as

$$S(t \mid x) = \int_0^\infty \exp(-t^{\alpha_0}(\exp x'\beta_0)\theta)\, dG_0(\theta), \tag{19}$$

where

$$0 < \alpha_0 \leqslant A < +\infty,$$

$\beta \in$ compact subset of k-dimensional Euclidean space, and G_0 is restricted to be a probability distribution on $[0, +\infty)$ with $\int_0^\infty \theta dG_0(\theta) = 1$. As a specialization of Elbers and Ridder's[16] general proof, α_0, β_0 and G_0 are identified. Now consider the completion with respect to the Kiefer–Wolfowitz[29] metric of the Cartesian product of the parameter space of allowed α and β values, and the probability distributions on $[0, +\infty)$ satisfying $\int \theta\, dG_0(\theta) = 1$. The completion contains distributions G_1 on $[0, +\infty)$ satisfying $\int_0^\infty \theta\, dG_1(\theta) = \infty$. Now observe that if $S(t \mid x)$ has a representation as defined above for some $\alpha \in (0,1)$ and G_0 with mean 1, then it is also a completely monotone function of t. Thus we also have the representation

$$S(t \mid x) = \int_0^\infty [\exp(-t(\exp(x'\beta_1))\theta]\, dG_1(\theta),$$

but now G_1 must have an infinite mean. This implies that (α_0, β_0, G_0) and $(1, \beta_1, G_1)$ generate the same survivor function. Hence the model is not identifiable on the completion of a space where probability distributions are restricted to have a finite mean. For further discussion see Heckman and Singer.[30]

This difficulty can be eliminated by further restricting G_0 to belong to a uniformly integrable family of distribution functions. Then all elements in the completion with respect to the Kiefer–Wolfowitz and a variety of other metrics will also have a finite mean, and identifiability is again ensured. The comparable requirement for the case when $E_0(\theta) = \infty$ is that the density with specified tail condition given in equation (14) converges uniformly to its limit.

The *a priori* restriction of identifiability considerations to complete metric spaces is central to establishing consistency of estimation methods in semiparametric models.

2.1 Nonparametric estimation

Securing identifiability of a nonparametric model is only the first step toward estimating the model. At the time of this writing, no nonparametric estimator has been devised that consistently estimates the general proportional hazard model (10). There are results for semiparametric versions of such models.

Heckman and Singer[30] consider consistent estimation of the proportional hazard model when $Z'(t)$ and $\phi(x)$ are specified up to a finite number of parameters but $G(\theta)$ is unrestricted, except that it must have either a finite mean and belong to a uniformly integrable family or satisfy a tail condition with uniform convergence (e.g. condition (15)). They verify sufficiency conditions due to Kiefer and Wolfowitz[29] which, when satisfied, guarantee the existence of a consistent nonparametric maximum likelihood estimator. They analyse a Weibull model for censored and uncensored data and demonstrate how to verify the sufficiency conditions for more general models. Meyer[31] has verified the Kiefer–Wolfowitz conditions for a model with grouped data in finite intervals. His analysis applies to models with a finite and known number of spline knots with known location.

These analyses only ensure the existence of a consistent estimator. The asymptotic distribution of the estimator is unknown but recent bounds on rates of convergence have been obtained by Honoré[32] and Ishwaran.[33] These are discussed below.

Drawing on results by Lindsey,[34,35] we characterize the computational form of the nonparametric maximum likelihood estimator.* To state these results most succinctly, we define

$$t^\star = \phi(x)Z(t).$$

*In computing the estimator it is necessary to impose all of the identifiability conditions in order to secure consistent estimators. For example, in a Weibull model with $E(\theta) < \infty$, it is important to impose this requirement in securing estimates. As our example in the preceding subsection indicated, there are other models with $E(\theta) = \infty$ that will explain the data equally well. In large samples, this condition is imposed, for example, by picking estimates of $G(\theta)$ such that $\mid \int (1 - \hat{G}(\theta))d\theta \mid < \infty$ or equivalently $\mid \int (1 - \hat{G}(\theta))d\theta \mid^{-1} > \infty$. Similarly, if identification is secured by tail condition this must be imposed in selecting a unique estimator.

For any fixed value of the parameters determining $\phi(x)$ and $Z(t)$, $t^\star$ conditional on θ is an exponential random variable, i.e.

$$f(t^\star \mid \theta) = \theta \exp(-t^\star\theta) \quad \theta \geq 0.$$

For this model, the following propositions can be established for the nonparametric maximum likelihood estimator (NPMLE).

Proposition 5 Let $I^\star$ be the number of distinct $t^\star$ values in the sample of $I(\geq I^\star)$ observations. Then the NPMLE of $\mu(\theta)$ is a finite mixture with at most $I^\star$ points of increase, i.e. for censored and uncensored data (with $d=1$ for uncensored observations)

$$f(t^\star) = \sum_{i=1}^{I^\star} \theta_i^d \exp(-t^\star\theta_i)P_i,$$

where

$$P_i \geq 0, \sum_{i=1}^{I^\star} P_i = 1.$$

Thus the NPMLE is a finite mixture but in contrast to the usual finite mixture model, $I^\star$ is estimated along with the P_i and θ_i. Other properties of the NPMLE are as follows.

Proposition 6 Assuming that no points of support $\{\theta_i\}$ come from the boundary of θ the NPMLE is unique. (See Lindsay.[34,35])

Proposition 7 For uncensored data, $\hat{\theta}_{\min} = 1/t^\star_{\max}$ and $\hat{\theta}_{\max} = 1/t^\star_{\min}$ where $\hat{}$ denotes the NPMLE estimate, and $t^\star_{\max}$ and $t^\star_{\min}$ are, respectively, the sample maximum and minimum values for $t^\star$. For censored data $\hat{\theta}_{\min} = 1$ and $\hat{\theta}_{\max} = 1/t^\star_{\min}$. (See Lindsay.[34,35])

These propositions show that the NPMLE for $G(\theta)$ in the proportional hazard model is in general unique and the estimated points of support lie in a region with known bounds (given $t^\star$). In computing estimates one can confine attention to this region. Further characterization of the NPMLE is given in Lindsay.[34,35]

It is important to note that all of these results are for a given $t^\star = Z(t)\phi(x)$. The computational strategy fixes the parameters determining $Z(t)$ and $\phi(x)$ and estimates $G(\theta)$. For each estimate of $G(\theta)$ so achieved $Z(t)$ and $\phi(x)$ are estimated by traditional parametric maximum likelihood methods. Then fresh $t^\star$ are generated and a new $G(\theta)$ is estimated until convergence occurs. There is no assurance that this procedure converges to a global optimum.

In a series of Monte Carlo runs reported in Heckman and Singer[30] the following results emerge.

(i) The NPMLE recovers the parameters governing $Z(t)$ and $\phi(x)$ rather well.
(ii) The NPMLE does not produce reliable estimates of the underlying mixing distribution.
(iii) The estimated c.d.f. for duration times $F(t \mid x)$ produced via the NPMLE predicts the sample c.d.f. of durations quite well, even in fresh samples of data with different distributions for the x variables.

Table 1 Results from a typical estimation

$g(\theta) = [\exp(\Delta\theta)\exp - (e^{\theta}/\beta)d\theta]\Gamma(1/2)$ with $\Delta = 1/2$ $\beta = 1$		
True model	$\alpha = 1$	$\beta = 1$
Estimated model	0.9852	0.9846
	(0.0738)*	(0.1022)*

where $Z(t) = t^{\alpha_1}$ and $\phi(x) = \exp(\alpha_2 x)$

Sample size $L = 500$

Log likelihood −1886.47

Estimated mixing distribution

Estimated θ	*Estimated* P_i	*Estimated c.d.f.*	*True c.d.f.*	*Observed c.d.f.*
−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	1.000	0.8356	0.8320

Estimated cumulative distribution of duration versus actual ($\hat{G}(t)$ versus $G(t)$)

Value of t	*Estimated t c.d.f.*	*Observed c.d.f.*
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

* The numbers reported below the estimates are standard errors from the estimated information matrix for (α, P, θ) given I^*. As noted in the text these have no rigorous justification unless the number of points is fixed in advance.

A typical run is reported in Table 1. The structural parameters (α, β) are estimated rather well. The mixing distribution is poorly estimated but the within sample agreement between the estimated c.d.f. of T and the observed c.d.f. is good. Table 2 records the results of perturbing a model by changing the mean of the regressors from 0 to 10. There is still close agreement between the estimated model (with parameters estimated on a sample where $X \sim N(10,1)$).

The NPMLE can be used to check the plausibility of any particular parameter 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

Table 2 Predictions on a fresh sample, $X \sim N(10,1)$. (The model used to fit the parameters is $X \sim N(0,1)$.)

Estimated cumulative distribution of duration versus actual $\hat{F}(t)$ versus $F(t)$

Value of t ($\times 10^5$)	*Estimated t c.d.f.*	*Observed c.d.f.*
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
100.0	0.7607	0.7640
300.0	0.8543	0.8620
5000.0	0.9615	0.9660

from the NPMLE, the econometrician would have much more confidence in adopting a particular specification of the mixing distribution. Development of a formal test statistic to determine how far is 'too far' awaits development of a distribution for the nonparametric maximum likelihood estimator. However, because of the consistency of the nonparametric maximum likelihood estimator a test based on the difference between the parameters of $Z(t)$ and $\phi(x)$ estimated via the NPMLE, and the same parameters 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 the structural parameters while producing a poor estimator for $G(\theta)$ suggests that it is possible to protect against the consequences of misspecification of the mixing distribution by fitting duration models with mixing distributions from parametric families, such as finite mixtures models, with more than the usual two parameters. Thus the failure of the NPMLE to estimate more than four or five points of increase for $G(\theta)$ can be cast in a somewhat more positive light. A finite mixture model with five points of increase is a nine (independent) parameter model for the mixing distribution. Imposing a false, but very flexible, mixing distribution does not seem to cause much bias in estimates of the structural coefficients. Moreover, for small $I^\star$, computational costs are *lower* for the NPMLE than they are for traditional parametric maximum likelihood estimators of $G(\theta)$. The computational costs of precise evaluation of $G(\theta)$ over 'small enough' intervals of θ are avoided by estimating a finite mixtures model.

Heckman *et al.*[36] present Monte Carlo evidence on a nonparametric method of moments estimator for $G(\theta)$. Their study again shows very slow rates of convergence. The method of moments estimator appears to be much less reliable than the maximum likelihood estimator.

In important unpublished research, Ishwaran[33] has demonstrated that the lower bound on the rate of convergence of *any* estimator of (α_0) in Weibull model (19) with $\beta = 0$ depends on the finiteness of

$$E(e^{j\theta}) = \int_{\underline{\underline{\theta}}} e^{j\theta} \, dG(\theta). \tag{20}$$

As j increases, and $E(e^{j\theta})$ remains finite, the convergence rate increases. Thus if

$$E(e^{\theta}) < \infty$$

but

$$E(e^{(1+\varepsilon)\theta}) = \infty, \quad \varepsilon > 0,$$

convergence is at rate $\log N$ where N is sample size. If equation (20) is finite for $j = 2$, convergence is bounded below at rate $N^{1/3}$. As j increases and (20) remains finite, the lower bound on the rate of convergence approaches $N^{1/2}$. These results indicate that the tail behaviour of $G(\theta)$ vitally affects the rate of convergence of the parametric portion of the maximum likelihood estimator. Honoré[32] has some related results.

An important alternative approach to the estimation of the Weibull model when the mixing distribution is unknown and is not specified parametrically has been developed by Honoré.[37] A great advantage of his approach over that taken by Heckman and Singer[30] is that with his estimator both the rates of convergence and the asymptotic distribution theory are known.

Building on an insight by Arnold and Brockett,[10] Honoré notes that in a Weibull model with no regressors ($\beta = 0$ in equation (19) except for a constant) that

$$\alpha = \lim_{t \to 0} \frac{\ln(-\ln(S(t)))}{\ln t}$$

provided $E(\theta) < \infty$. Without this assumption, α is not identified (recall the discussion in the Theorem 3 and the example in the preceding subsection). Honoré constructs an order statistic sample analogue to this condition to produce an estimator of α.

Specifically, Honoré lets t be the mth order statistic of the sample where $m \to \infty$, but $m/N \to 0$ where N is sample size. $S(t) = 1 - m/N$. By choosing $m = N^{1-d}$, $0 < d < 1$, it is possible to produce an order statistics estimator of $\alpha, \hat{\alpha}(d)$ with a known asymptotic distribution. Honoré assumes that $E(\theta^2) < \infty$, $P(\theta > 0) > 0$ and $P(\theta \geqslant 0) = 1$. These final conditions ensure that durations are finite.

The estimator is generated from

$$\hat{\alpha}(d) = \frac{f}{\ln(T_m; N)}$$

where $m = N^{1-d}$, $f = \ln(-\ln(1 - m/N))$ and where $T_{m;N}$ denotes the mth order statistic $T_1, \ldots, T_N$ and T_i is the survival time of the ith observation. Picking two different values of d, (d_1, d_2) with $d_1 > d_2$ we may write

$$\hat{\alpha}(d_1, d_2) = -\frac{\rho \ln(N)(d_1 - d_2)}{\ln(T_{m_1;N}) - \ln(T_{m_2;N})}$$

where m_i corresponds to the choice of d_i, and where

$$\rho = 1 - \frac{1}{2} \frac{N^{-d_1} - N^{-d_2}}{(d_1 - d_2)(\ln N)}.$$

This estimator converges to a normal random variable at rate

$$(\ln N) N^{(1-d_1)/2}.$$

Honoré shows that restrictions required to produce a finite variance lead to $d_1 + 2d_2 > 1$, which, coupled with the restriction $0 < d_2 < d_1 < 1$ leads to a rate of convergence that can be made arbitrarily close to $\sqrt[3]{N}$. He extends the model to allow for regressors ($\beta \neq 0$ in (19)). An extensive Monte Carlo analysis shows that his estimator performs well especially if the coefficient of variation of θ is not too large.

2.2 Econometric computer algorithms

A computer program estimating maximum likelihood-based general multi-state duration models with time-varying variables, general hazard rates (including the Box–

Cox class of models and spline models for hazards) and nonparametric mixing distributions was initially developed by Heckman, Yates and Honoré at Chicago. An application of these programs to birth process data is made by Heckman and Walker.[38] These programs have been greatly refined by Steinberg and Colla[39] and are now part of the Systat Library. They are distributed through Salford Systems, 5952 Bernadette Lane, San Diego, California, 92120, USA. Versions are available for personal computers, work stations and main frame computers.

3 Summary

This survey presents results from the recent econometrics literature on duration models based on mixtures and more general models for unobservables. We have focused on continuous time duration models. There is a related discrete time literature in econometrics which we have not surveyed here. (See Heckman and Taber[40].) A major theme of the econometrics literature has been to establish how the introduction of regressors in a structured way aids in securing identification of models. There is an extensive literature on semiparametric estimation of duration models.

Theorem 1 and the other theorems suggest that it should be possible to estimate more general models with unobservables invoking fewer parametric assumptions than are conventional. Standard approaches to the competing risks problem are in fact quite restrictive and unnecessary. The development of more robust semiparametric and nonparametric estimation methods is a very active topic of research in econometrics, and medical statisticians would be well advised to keep abreast of developments in this field.

Acknowledgements

This research was supported by NGF grant SBR-9309525 to the Economic Research Center, NORC. We gratefully acknowledge comments by Tim Conley and Bo Honoré.

References

1 David HA, Moeschberger ML. *The theory of competing risks*. High Wycombe: Griffin, 1978.
2 Kalbfleisch JD, Prentice RL. *The statistical analysis of failure time data*. New York: Wiley, 1980.
3 Cox DR, Oakes D. *Analysis of survival data*. London: Chapman and Hall, 1984.
4 Flinn CJ, Heckman JJ. Are unemployment and out of the labor force behaviorally distinct labor force states? *Journal of Labor Economics* 1983; **1**: 28–42.
5 Cox DR. *Renewal theory*. London: Methuen, 1962.
6 Tsiatis A. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences* 1975; **72**: 20–2.
7 Yashin AI, Manton KG, Stallard E. Dependent competing risks: a stochastic process model. *Journal of Mathematical Biology* 1986; **24**: 119–64.
8 Flinn CJ, Heckman JJ. Models for the analysis of labor force dynamics. In *Advances in econometrics*, **1**, Ed. Bassman R, Rhodes G, 35–95, Greenwich, CT: JAI Press, 1982.
9 Basu AP, Ghosh JK. Identifiability of the multinormal and other distributions under competing risks model. *Journal of Multivariate Analysis* 1978; **8**: 413–29.
10 Arnold B, Brockett P. Identifiability for dependent multiple decrement/competing risk models. *Scandinavian Actuarial Journal* 1983; **10**: 117–27.
11 Heckman JJ, Honoré B. The identifiability of the competing risks model. *Biometrika* 1989; **76**: 325–30.
12 Cox DR. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society Series B* 1972; **34**: 187–202.
13 Flinn CJ, Heckman JJ. The likelihood

function for the multistate-multiepisode model in 'Models for the analysis of labor force dynamics'. In *Advances in econometrics*, **3**, Ed. Bassman R, Rhodes G. pp. 225–31, Greenwich, CT: JAI Press, 1983.
14 Clayton D, Cuzick J. Multivariate generalizations of the proportional hazard model. *Journal of the Royal Statistical Society Series A* 1985; **148**: 82–117.
15 Manton KG, Stallard E, Woodbury M. Chronic disease evolution and human aging: a general model for assessing the impact of chronic disease in human populations. *International Journal of Mathematical Modelling* 1986; **17**: 406–52.
16 Elbers C, Ridder G. True and spurious duration dependence: the identifiability of the proportional hazard model. *Review of Economic Studies* 1982; **49**: 403–10.
17 Heckman JJ, Singer B. The identifiability of the proportional hazard model. *Review of Economic Studies* 1984; **51**(2): 231–43.
18 Honoré B. Identification of duration models with unobserved heterogeneity. Unpublished manuscript, Northwestern University, 1990.
19 Feller W. *An introduction to probability theory and its applications.* New York: Wiley, Vol. II, 1971.
20 Heckman JJ, Singer B. Social science duration analysis. In *Longitudinal analysis of labor market data.* Ed. Heckman JJ, Singer B. New York: Cambridge University Press, 1985.
21 Ridder G. The non-parametric identification of generalized hazard models. *Review of Economic Studies* 1990; **57**: 167–82.
22 Heckman JJ. Identifying the hand of past: distinguishing state dependence from heterogeneity. *American Economic Review.* Papers and Proceedings 1991; **106**(3): 75–79.
23 Yashin AI, Arjas A. A note on random intensities and conditional survivor functions. *Journal of Applied Probability* 1988; **25**: 630–35.
24 McCall B. Identifying state dependence in duration models with time-varying regressors, Industrial Relations Section. University of Minnesota, 1993.
25 McCall B. The identifiability of the mixed proportional hazards model with time-coefficients, unpublished manuscript, Industrial Relations Section, University of Minnesota, 1993.
26 Honoré B. Identification results for duration models with multiple spells. *Review of Economic Studies* 1993; **60**(1): 241–46.
27 Heckman JJ, Hotz VJ, Walker JR. New evidence on the timing and spacing of births. *American Economic Review* 1985; **72**: 179–84.
28 Heckman JJ, Borjas GJ. Does unemployment cause future unemployment? Definitions, questions and answers for a continuous time model of heterogeneity and state dependence. *Economica* 1980; **47**: 247–83.
29 Kiefer J, Wolfowitz J. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Annals of Mathematical Statistics* 1956; **27**: 363–66.
30 Heckman JJ, Singer B. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 1984; **52**(2): 271–320.
31 Meyer B. Semiparametric estimation of hazard models. *Econometrica* 1992; **60**(3): 64–67.
32 Honoré 1993. Rates of convergence in Weibull mixture models, unpublished manuscript, Northwestern, 1993.
33 Ishwaran H. Rates of convergence in semiparametric mixture models, published Ph.D. Thesis, Yale University, Department of Statistics, 1993.
34 Lindsey B. The geometry of mixture likelihoods, Part I. *Annals of Statistics* 1983; **11**: 86–94.
35 Lindsey B. The geometery of mixture likelihoods, Part II. *Annals of Statistics* 1983; **11**(3): 783–92.
36 Heckman JJ, Robb R, Walker J. Testing the mixture of exponentials hypothesis and estimating the mixing distribution by the method of moments. *Journal of The American Statistical Association* 1990; **85**: 410, 582–89.
37 Honoré B. Simple estimation of a duration model with unobserved heterogeneity. *Econometrica* 1990; **58**: 453–74.
38 Heckman JJ, Walker J. The relationship between wages and income and the timing and spacing of births: evidence from Swedish longitudinal data. *Econometrica* 1990; **58**(6): 1411–41.
39 Steinberg D, Colla P. *"CTM" A supplementary module for SYSTAT.* Evanston, Illinois: SYSTAT, 1994.
40 Heckman JJ, Taber C. Identification in binary choice models and their extensions, unpublished manuscript, University of Chicago, 1994.