Semi-Nonparametric Identification of the Right Censored Mixed Proportional Hazard Model

Herman J. Bierens
Department of Economics
Pennsylvania State University

November 3, 2008

Abstract

Elbers and Ridder (1982) and Heckman and Singer (1984) have shown that under mild conditions the mixed proportional hazard (MPH) model is nonparametrically identified. However, Elbers and Ridder did not consider the case of right censored MPH models, whereas Heckman and Singer only considered right censoring in the Weibull baseline hazard case. In this lecture note I will explain the Elbers-Ridder aproach and extend it to the case of right censoring.

1 The mixed proportional hazard model

Let $\widetilde{T}$ be a duration, and let X be a vector of covariates. The conditional hazard function is defined as

$$\lim_{\delta \downarrow 0} \frac{\Pr\left[\widetilde{T} \in (t, t+\delta) \mid \widetilde{T} > t, X\right]}{\delta} = \frac{f(t|X)}{1 - F(t|X)} = \lambda(t, X),$$

where $F(t|X) = \Pr[\widetilde{T} \leq t|X]$, and $f(t|X)$ is the corresponding conditional density function. Since

$$\frac{\partial \ln(1 - F(t|X))}{\partial t} = \frac{-f(t|X)}{1 - F(t|X)} = -\lambda(t, X),$$

it follows that

$$1 - F(t|X) = \exp\left(-\int_0^t \lambda(\tau, X) d\tau\right)$$

The proportional hazard model assumes that

$$\lambda(t, X) = \varphi(X) \lambda_0(t),$$

were $\varphi(X) > 0$ is called the systematic hazard, and $\lambda_0(t)$ is called the baseline hazard. Usually, $\varphi(X)$ is parametrized as

$$\varphi(X) = \exp(\beta_0' X), \tag{1}$$

so that the conditional survival function takes the form

$$\begin{aligned} S(t|X) &= \Pr[\widetilde{T} > t|X] = \exp\left(-\exp(\beta_0' X) \int_0^t \lambda_0(\tau) d\tau\right) \\ &= \exp(-\exp(\beta_0' X) \Lambda_0(t)), \end{aligned}$$

where

$$\Lambda_0(t) = \int_0^t \lambda_0(\tau) d\tau$$

is the integrated baseline hazard. Note that for $\lim_{t\to\infty} \Pr[\widetilde{T} > t|X] = 0$ we need to require that $\lim_{t\to\infty} \Lambda_0(t) = \int_0^\infty \lambda_0(\tau) d\tau = \infty$, and for $\Pr[\widetilde{T} > t|X]$ to be monotonic decreasing on $(0, \infty)$ we need to require that $\lambda_0(t) > 0$ for $t \in (0, \infty)$.

Adopting the specification (1) for the systematic hazard, the mixed proportional hazard (MPH) model, proposed by Lancaster (1979), assumes that the conditional survival function takes the form

$$S(t|X) = E\left[\exp(-\exp(\beta_0' X + Y) \Lambda_0(t)) | X\right], \tag{2}$$

where Y represents unobserved heterogeneity, which is assumed to be independent of X. Denoting the distribution function of $V = \exp(Y)$ by $G_0(v)$, we have

$$\begin{aligned} S(t|X) &= \int_0^\infty \exp(-v . \exp(\beta_0' X) \Lambda_0(t)) \, dG_0(v) \\ &= \int_0^\infty \left(\exp(-\exp(\beta_0' X) \Lambda_0(t))\right)^v dG_0(v) \\ &= H_0(\exp(-\exp(\beta_0' X) \Lambda_0(t))), \end{aligned} \tag{3}$$

where

$$H_0(u) = \int_0^\infty u^v dG_0(v),\ \ u \in [0,1], \tag{4}$$

is a distribution function on $[0,1]$. Since for $u \in (0,1)$ and $n = 1,2,3,....,$

$$\sup_{v\geq 0}\left|\frac{\partial^n(u^v)}{(\partial u)^n}\right| < \infty$$

it follows from the dominated convergence theorem that

$$d^n H_0(u)/(du)^n = u^{-n}\int_0^\infty \prod_{j=0}^n (v-j)\, u^v dG_0(v)$$

Consequently, $H_0(u)$ is absolutely continuous, with density

$$h_0(u) = \int_0^\infty v u^{v-1} dG_0(v), \tag{5}$$

which is continuously differentiable of any order on $(0,1)$. Note that if $\lim_{v\uparrow 1} G_0(v) > 0$ then $\lim_{u\downarrow 0} h_0(u) = \infty$ but $\lim_{u\downarrow 0} u.h_0(u) = 0$, whereas $\lim_{u\uparrow 1} h_0(u) = \int_0^\infty v dG_0(v)$. Moreover, note that absence of unobserved heterogeneity, i.e., $\Pr[V=1]=1$, is equivalent to the case $h_0(u) \equiv 1$.

2 Right-censoring

Usually the duration $\widetilde{T}$ is only observed up to an upper bound $\overline{T}$, which may vary per individual. This is called right-censoring, which is indicated by the observable dummy variable $C = I\left(\widetilde{T} > \overline{T}\right)$, where $I(.)$ is the indicator function.[1] As usual, it will be assumed that

Assumption 1. *Conditional on the covariates X, the actual duration $\widetilde{T}$ and the censoring time $\overline{T}$ are independent. The observed duration T is equal to $\widetilde{T}$-if $\widetilde{T} \leq \overline{T}$ and is equal to $\overline{T}$ if $\widetilde{T} > \overline{T}$. These events are observable in the form of a dummy variable $C = I\left(\widetilde{T} > \overline{T}\right)$.*

Also, for later reference it will be assumed that

[1] $I(true) = 1$, $I(false) = 0$.

Assumption 2. *The support of the distribution of* $\overline{T}$ *is* $(0, \overline{t})$,

where possibly but not necessarily $\overline{t} = \infty$. Then

$$\Pr[C = 1|X, \overline{T}] = S\left(\overline{T}|X\right) = H_0\left(\exp\left(-\exp(\beta_0'X)\Lambda_0\left(\overline{T}\right)\right)\right) \tag{6}$$

and

$$\begin{aligned} \Pr[T \leq t|X, \overline{T}, C = 0] &= \frac{\Pr[T \leq t, T \leq \overline{T}|X]}{\Pr[C = 0|X, \overline{T}]} = \frac{\Pr[T \leq \min\left(t, \overline{T}\right)|X]}{\Pr[C = 0|X, \overline{T}]} \\ &= \frac{1 - H_0\left(\exp\left(-\exp(\beta_0'X)\Lambda_0\left(\min\left(t, \overline{T}\right)\right)\right)\right)}{1 - H_0\left(\exp\left(-\exp(\beta_0'X)\Lambda_0\left(\overline{T}\right)\right)\right)}. \end{aligned}$$

Thus the survival function of T given X, $\overline{T}$ and $C = 0$ takes the form:

$$\begin{aligned} &S\left(t|X, \overline{T}, C = 0\right) \qquad (7) \\ &= \frac{H_0\left(\exp\left(-\exp(\beta_0'X)\Lambda_0\left(t\right)\right)\right) - H_0\left(\exp\left(-\exp(\beta_0'X)\Lambda_0\left(\overline{T}\right)\right)\right)}{1 - H_0\left(\exp\left(-\exp(\beta_0'X)\Lambda_0\left(\overline{T}\right)\right)\right)} \\ &\times I\left(\overline{T} > t\right) \end{aligned}$$

3 Nonparametric identification

Elbers and Ridder (1982) have shown that under some conditions the MPH model is nonparametrically identified. Heckman and Singer (1984) provide an alternative identification proof based on the more general conditions in Kiefer and Wolfowitz (1956), and propose to parametrize G_0 as a discrete distribution: $G_0(v) = \sum_{i=1}^{q} I\left(v \leq \theta_i\right) p_i$, with $I(.)$ the indicator function, where $\theta_i > 0$, $p_i > 0$, and $\sum_{i=1}^{q} p_i = 1$. Thus, Heckman and Singer (1984) implicitly specify $h_0(u) = \sum_{i=1}^{q} \theta_i u^{\theta_i - 1} p_i$. However, they assume that the support S of the systematic hazard $\exp\left(\beta_0'X\right)$ contains an open interval, which excludes the case that all the components of X are discrete. Elbers and Ridder (1982) derive the nonparametric identification of the MPH model under this condition as well as for the case that X is a dummy variable, with $\beta_0 \neq 0$.

In this section I will explain the Elbers and Ridder (1982) approach, in a slightly different way than in that paper.

3.1 Identification of the systematic hazard

Suppose that there exist a parameter vector β, an integrated baseline hazard $\Lambda(t)$, and a distribution function $H(u)$ on $[0,1]$ with density $h(u)$ such that

$$\sup_{t\in\left[0,\overline{T}\right]} \left|H\left(\exp\left(-\exp(\beta' X).\Lambda\left(t\right)\right)\right) - H_0\left(\exp\left(-\exp(\beta_0' X).\Lambda_0\left(t\right)\right)\right)\right|$$
$$= 0 \text{ a.s.} \tag{8}$$

Then

Lemma 1. *Under Assumption 2,* (8) *implies that for all* $t \in (0,\overline{t})$

$$H\left(\exp\left(-\exp(\beta' X).\Lambda\left(t\right)\right)\right) = H_0\left(\exp\left(-\exp(\beta_0' X).\Lambda_0\left(t\right)\right)\right) \;\; a.s. \tag{9}$$

Proof: Appendix

Taking derivatives of both sides of (9) to $t \in (0,\overline{t})$ yield

$$\begin{aligned} &h\left(\exp\left(-\exp(\beta' X).\Lambda\left(t\right)\right)\right)\exp\left(-\exp(\beta' X).\Lambda\left(t\right)\right) \\ &\quad \times \exp(\beta' X)\lambda\left(t\right) \\ &\quad = h_0\left(\exp\left(-\exp(\beta_0' X).\Lambda_0\left(t\right)\right)\right)\exp\left(-\exp(\beta_0' X)\Lambda_0\left(t\right)\right) \\ &\qquad \times \exp(\beta_0' X)\lambda_0\left(t\right) \end{aligned}$$

hence

$$\begin{aligned} &\frac{h\left(\exp\left(-\exp(\beta' X).\Lambda\left(t\right)\right)\right)}{h_0\left(\exp\left(-\exp(\beta_0' X).\Lambda_0\left(t\right)\right)\right)} \\ &\quad = \frac{\exp\left(-\exp(\beta_0' X).\Lambda_0\left(t\right)\right)}{\exp\left(-\exp(\beta' X).\Lambda\left(t\right)\right)}\exp\left(\left(\beta_0-\beta\right)' X\right)\frac{\lambda_0\left(t\right)}{\lambda\left(t\right)}. \end{aligned} \tag{10}$$

Following Elbers and Ridder (1982) and Bierens (2008) I will now assume that

Assumption 3: *The distribution function* $G_0(v)$ *of the unobserved heterogeneity variable* V *is confined to the class* $\mathcal{G}$ *of distribution functions* G *on* $(0,\infty)$ *satisfying* $\int_0^\infty v dG(v) = 1$.

The actual assumption involved is that for all $G \in \mathcal{G}$, $\int_0^\infty v dG(v) = \mu$ for some common $\mu \in (0, \infty)$. Then without loss of generality we may normalize $\mu = 1$.

With $h_0(u) = \int_0^\infty v u^{v-1} dG_0(v)$ and $h(u) = \int_0^\infty v u^{v-1} dG(v)$, where $G_0, G \in \mathcal{G}$, it follows from Assumption 3 that

$$h_0(1) = h(1) = 1. \tag{11}$$

Regardless wether or not $h(u)$ is of the form $\int_0^\infty v u^{v-1} dG(v)$ with $G \in \mathcal{G}$, it will be assumed that condition (11) holds. How to implement (11) in practice has been shown in Bierens (2008).

Taking the limit of (10) for $t \downarrow 0$, it follows now from (11) that

$$1 = \frac{h(1)}{h_0(1)} = \exp\left((\beta_0 - \beta)' X\right) \lim_{t \downarrow 0} \frac{\lambda_0(t)}{\lambda(t)}. \tag{12}$$

hence

$$(\beta - \beta_0)' X = \ln\left(\lim_{t \downarrow 0} \frac{\lambda_0(t)}{\lambda(t)}\right) \text{ a.s.} \tag{13}$$

Next, following Bierens (2008), suppose that

Assumption 4: *$E[X'X] < \infty$ and $\Sigma = E\left[(X - E[X])(X - E[X])'\right]$ is non-singular.*[2]

The condition $E[X'X] < \infty$ implies that $E[X]$ exists and is finite, hence $\ln\left(\lim_{t \downarrow 0} \lambda_0(t)/\lambda(t)\right) = (\beta - \beta_0)' E[X]$ exists and is finite and thus by (13)

$$(\beta - \beta_0)'(X - E[X]) = 0 \text{ a.s.} \tag{14}$$

It follows now from (14) and Assumption 4 that $(\beta - \beta_0)' \Sigma (\beta - \beta_0) = 0$, hence $\beta = \beta_0$ and

$$\lim_{t \downarrow 0} \frac{\lambda_0(t)}{\lambda(t)} = 1. \tag{15}$$

3.2 Identification of the baseline hazard in the Weibull case

Note that if $\lambda_0(t)$ and $\lambda(t)$ are of the Weibull type,

$$\lambda_0(t) = \exp(\alpha_0)\, \omega_0 t^{\omega_0 - 1}, \ \ \lambda(t) = \exp(\alpha)\, \omega t^{\omega - 1},$$

[2]Note that the nonsingularity of Σ excludes the presence of a constant covariate.

where α_0 and α are scale parameters, and $\omega_0 > 0$, $\omega > 0$, then (15) implies $\alpha = \alpha_0$, $\omega = \omega_0$, hence $\lambda(t) = \lambda_0(t)$ for all $t > 0$. Thus in this case (9) reads

$$\begin{aligned} &H\left(\exp\left(-\exp(\alpha_0+\beta_0'X).\overline{\Lambda}_0(t)\right)\right) \\ &\quad = H_0\left(\exp\left(-\exp(\alpha_0+\beta_0'X).\overline{\Lambda}_0(t)\right)\right) \text{ a.s.} \end{aligned} \tag{16}$$

for all $t \in [0, \bar{t})$, where

$$\overline{\Lambda}_0(t) = t^{\omega_0}.$$

However, this result by itself does not pin down α_0, because (16) also holds for $H^*(u) = H\left(u^{1/c}\right)$ and $H_0^*(u) = H_0\left(u^{1/c}\right)$, where $c > 0$ is arbitrary:

$$\begin{aligned} &H^*\left(\exp\left(-\exp(\alpha_0+\ln(c)+\beta_0'X).\overline{\Lambda}_0(t)\right)\right) \\ &\quad = H_0^*\left(\exp\left(-\exp(\alpha_0+\ln(c)+\beta_0'X).\overline{\Lambda}_0(t)\right)\right) \text{ a.s.} \end{aligned}$$

On the other hand, if $H_0^*(u) = \int_0^\infty u^v dG_0^*(v)$ where $G_0^* \in \mathcal{G}$, then by Assumption 3 and (11),

$$1 = \lim_{u\uparrow 1} \frac{dH_0^*(u)}{du} = \lim_{u\uparrow 1} h_0\left(u^{1/c}\right)\frac{1}{c}u^{1/c-1} = \frac{1}{c}.$$

and similarly for H^*. Thus, Assumption 3 pins down the scale of the integrated hazard in the Weibull case.

3.3 Identification of the integrated hazard

3.3.1 Continuous covariates

Under Assumptions 2-3, (9) now reads

$$H\left(\exp\left(-Z.\Lambda(t)\right)\right) = H_0\left(\exp\left(-Z.\Lambda_0(t)\right)\right) \tag{17}$$

a.s. for $t \in (0, \bar{t})$, where $Z = \exp(\beta_0'X)$. Elbers and Ridder (1982) assume in first instance that the support of Z contains an open interval, for example the interval $(z_0 - \varepsilon, z_0 + \varepsilon)$. Then

$$H\left(\exp\left(-z.\Lambda(t)\right)\right) = H_0\left(\exp\left(-z.\Lambda_0(t)\right)\right)$$

on $(z_0 - \varepsilon, z_0 + \varepsilon) \times (0, \bar{t})$. Hence, taking the derivative to $z \in (z_0 - \varepsilon, z_0 + \varepsilon)$ and letting $z \to z_0$, it follows that

$$\begin{aligned} &h\left(\exp\left(-z_0.\Lambda(t)\right)\right)\exp\left(-z_0.\Lambda(t)\right)\Lambda(t) \\ &\quad = h_0\left(\exp\left(-z_0.\Lambda_0(t)\right)\right)\exp\left(-z_0.\Lambda_0(t)\right)\Lambda_0(t), \end{aligned} \tag{18}$$

whereas if we take the derivative to $t \in (0, \bar{t})$ and set $z = z_0$ then

$$\begin{aligned} & h\left(\exp\left(-z_0.\Lambda\left(t\right)\right)\right)\exp\left(-z_0.\Lambda\left(t\right)\right)\lambda\left(t\right) \\ & \quad = h_0\left(\exp\left(-z_0.\Lambda_0\left(t\right)\right)\right)\exp\left(-z_0.\Lambda_0\left(t\right)\right)\lambda_0\left(t\right). \end{aligned} \tag{19}$$

Dividing (19) by (18) yields

$$\frac{\lambda\left(t\right)}{\Lambda\left(t\right)} = \frac{\lambda_0\left(t\right)}{\Lambda_0\left(t\right)} \text{ for } t \in \left(0, \bar{t}\right),$$

which is equivalent to

$$\frac{d\ln\left(\Lambda\left(t\right)\right)}{dt} = \frac{d\Lambda_0\left(t\right)}{dt} \text{ for } t \in \left(0, \bar{t}\right).$$

Integrating both derivatives and using the boundary condition $\Lambda\left(0\right) = \Lambda_0\left(0\right) = 0$ it follows that

$$\Lambda\left(t\right) = \Lambda_0\left(t\right) \text{ for } t \in \left[0, \bar{t}\right). \tag{20}$$

Thus, the integrated baseline hazard is now identified on $[0, \bar{t})$ up to a multiplicative constant. However, this constant is identified. So see this, denote

$$\begin{aligned} \alpha_0 &= \ln\left(\Lambda_0\left(1\right)\right) = \ln\left(\Lambda\left(1\right)\right), \\ \overline{\Lambda}_0\left(t\right) &= \Lambda_0\left(t\right)/\Lambda_0\left(1\right),\ \overline{\Lambda}_0\left(t\right) = \Lambda_0\left(t\right)/\Lambda_0\left(1\right) \end{aligned}$$

so that $\overline{\Lambda}\left(1\right) = \overline{\Lambda}_0\left(1\right) = 1$. Then it follows similar to the Weibull case that α_0 is identified.

3.3.2 Discrete covariates

In the Appendix of their paper, Elbers and Ridder (1982) also consider the case where the systematic hazard takes two values. Thus, let again $Z = \exp\left(\beta_0' X\right)$ and suppose that there exists a pair z_1, z_2 such that

$$\Pr[Z = z_1] > 0,\ \Pr[Z = z_2] > 0,\ 0 < z_1 < z_2 < \infty.$$

Then for $0 \leq t \leq \bar{t}$ and $j = 1, 2$,

$$H\left(\exp\left(-z_j.\Lambda\left(t\right)\right)\right) = H_0\left(\exp\left(-z_j.\Lambda_0\left(t\right)\right)\right),$$

hence

$$\Lambda(t) = \frac{\ln\left(H^{-1}\left(H_0\left(\exp\left(-z_j.\Lambda_0(t)\right)\right)\right)\right)}{-z_j},\ j = 1, 2. \tag{21}$$

Next, assume that

Assumption 5. $\lambda_0(t) > 0$ *on* $(0, \infty)$.

Then $\Lambda_0(t)$ is strictly monotonic increasing on $(0, \infty)$, and since $H_0(u)$ and $H(u)$ are strictly monotonic increasing on $(0, 1)$ it follows from (21) that $\Lambda(t)$ is strictly monotonic increasing on $(0, \infty)$. Therefore, both $\Lambda_0(t)$ and $\Lambda(t)$ are invertible on $(0, \infty)$, with inverses denoted by $\Lambda_0^{-1}(.)$ and $\Lambda^{-1}(.)$, respectively.

Given an arbitrary $t \in (0, \overline{t})$, let

$$t_2 = \Lambda_0^{-1}\left(\frac{z_1}{z_2}\Lambda_0(t)\right) = \Lambda_0^{-1}\left(\rho\Lambda_0(t)\right), \tag{22}$$

where

$$\rho = z_1/z_2 < 1,$$

and note that by the monotonicity of $\Lambda_0(t)$, $t_2 < t$. Then it follows from (21) that

$$\begin{aligned}
\Lambda(t_2) &= \Lambda\left(\Lambda_0^{-1}\left(\rho\Lambda_0(t)\right)\right) \\
&= \frac{\ln\left(H^{-1}\left(H_0\left(\exp\left(-z_2.\Lambda_0(t_2)\right)\right)\right)\right)}{-z_2} \\
&= \frac{\ln\left(H^{-1}\left(H_0\left(\exp\left(-z_2.\Lambda_0\left(\Lambda_0^{-1}\left(\frac{z_1}{z_2}\Lambda_0(t)\right)\right)\right)\right)\right)\right)}{-z_2} \\
&= \frac{\ln\left(H^{-1}\left(H_0\left(\exp\left(-z_1.\Lambda_0(t)\right)\right)\right)\right)}{-z_2} = \frac{z_1}{z_2}\Lambda(t) \\
&= \rho\Lambda(t),
\end{aligned} \tag{23}$$

hence

$$\rho\Lambda(t) = \Lambda\left(\Lambda_0^{-1}\left(\rho\Lambda_0(t)\right)\right) \tag{24}$$

and thus also

$$\rho\Lambda_0(t) = \Lambda_0\left(\Lambda^{-1}\left(\rho\Lambda(t)\right)\right). \tag{25}$$

More generally, it follows by induction that

Lemma 2. *For $n = 1, 2, 3, ...$ and $t \in (0, \bar{t})$, $\rho^n \Lambda_0(t) = \Lambda_0(\Lambda^{-1}(\rho^n \Lambda(t)))$.*

Proof: Appendix.

Lemma 2 implies that

$$\frac{\Lambda_0(t)}{\Lambda(t)} = \frac{\Lambda_0(\Lambda^{-1}(\rho^n \Lambda(t)))}{\rho^n \Lambda(t)}.$$

Taking the limit for $n \to \infty$ yields

$$\begin{aligned}
\frac{\Lambda_0(t)}{\Lambda(t)} &= \lim_{n\to\infty} \frac{\Lambda_0(\Lambda^{-1}(\rho^n \Lambda(t)))}{\rho^n \Lambda(t)} \\
&= \lim_{n\to\infty} \frac{\Lambda_0(\Lambda^{-1}(\rho^n \Lambda(t))) - \Lambda_0(\Lambda^{-1}(0))}{\rho^n \Lambda(t)} \\
&= \lim_{\tau\downarrow 0} \left.\frac{d\Lambda_0(\Lambda^{-1}(\tau))}{d\tau}\right| \\
&= \lim_{\tau\downarrow 0} \left.\frac{d\Lambda_0(\Lambda^{-1}(\tau))}{d\Lambda^{-1}(\tau)} \times \frac{d\Lambda^{-1}(\tau)}{d\Lambda(\Lambda^{-1}(\tau))}\right| \\
&= \lim_{\tau\downarrow 0} \left.\frac{\lambda_0(\Lambda^{-1}(\tau))}{\lambda(\Lambda^{-1}(\tau))}\right| = \lim_{\tau\downarrow 0} \left.\frac{\lambda_0(\tau)}{\lambda(\tau)}\right| \\
&= 1
\end{aligned}$$

where the latter follows from (15). Since $t \in (0, \bar{t})$ was chosen arbitrarily, it follows that $\Lambda(t) = \Lambda_0(t)$ on $[0, \bar{t})$. Consequently, the result (27) carries over.

3.3.3 Mixed continuous-discrete covariates

As is well-known [see for example Chung (1974, Sect. 1.3)], the distribution function $\Psi(z)$ of $Z = \exp(\beta_0' X)$ can always be written as a unique convex combination

$$\Psi(z) = \Pr[Z \leq z] = \delta_1 \Psi_c(z) + \delta_2 \Psi_d(z) + \delta_3 \Psi_s(z)$$

where $\delta_1 \geq 0$, $\delta_2 \geq 0$, $\delta_3 \geq 0$, $\delta_1 + \delta_2 + \delta_3 = 1$, $\Psi_c(z)$ is an absolutely continuous distribution function, $\Psi_d(z)$ is a discrete distribution function, and $\Psi_s(z)$ is a singular continuous distribution function.

Elbers and Ridder (1982) assume that either the support of Z contains an open interval, which corresponds to $\delta_1 > 0$ and $\Psi_c'(z) > 0$ on an open interval, or that $\delta_1 = 0$ and $\delta_2 > 0$ where $\Psi_d(z)$ is non-degenerated (i.e., $\Psi_d(z)$ is not the distribution function of a constant, so that $\Psi_d(z)$ has at least two jumps). These two assumptions can be combined as follows:

Assumption 6. *The distribution function $\Psi(z)$ of $Z = \exp(\beta_0' X)$ satisfies $\Psi(z) = \delta_1 \Psi_c(z) + \delta_2 \Psi_d(z)$, where $\delta_1 \geq 0$, $\delta_2 \geq 0$, $\delta_1 + \delta_2 = 1$, $\Psi_c(z)$ is an absolutely continuous distribution function on $(0, \infty)$ with support containing an open interval, and $\Psi_d(z)$ is a non-degenerated discrete distribution function.*

Summarizing, it has been shown:

Theorem 1. *Let Assumptions 1-6 hold. Suppose that the conditional survival function has two equivalent representations,*

$$\begin{aligned} S(t|X) &= H_0\left(\exp\left(-\exp\left(\alpha_0 + \beta_0' X\right) \Lambda_0(t)\right)\right) \\ &= H\left(\exp\left(-\exp\left(\alpha + \beta' X\right) \Lambda(t)\right)\right) \end{aligned} \tag{26}$$

for all $t \in (0, \bar{t})$, where $\bar{t}$ is the upper bound of the support of the censoring time $\overline{T}$, the integrated hazards $\Lambda_0(t)$ and $\Lambda(t)$ are normalized such that $\Lambda(1) = \Lambda_0(1) = 1$, and the density h of the distribution function H is normalized by $h(1) = 1$. Then $\alpha = \alpha$, $\beta = \beta_0$ and $\Lambda(t) = \Lambda_0(t)$ for all $t \in (0, \bar{t})$.

3.4 Identification of the unobserved heterogeneity distribution

It remains to show that under the conditions of Theorem 1,

$$\begin{aligned} \Pr\Bigg[\sup_{t \in [0, \bar{t})} &\left| H\left(\exp\left(-\exp\left(\alpha_0 + \beta_0' X\right) \Lambda_0(t)\right)\right) \right. \\ &\left. - H_0\left(\exp\left(-\exp\left(\alpha_0 + \beta_0' X\right) \Lambda_0(t)\right)\right) \right| = 0 \Big| X \Bigg] \\ &= 1 \end{aligned} \tag{27}$$

implies $H(u) = H_0(u)$ on $[0,1]$. If $\bar{t} = \infty$ this is trivial. If $\bar{t} < \infty$ but $H(u)$ is of the type $H(u) = \int_0^\infty u^v dG(v)$ then $H = H_0$ follows from the following lemma.

Lemma 3. *Let $H(u) = \int_0^\infty u^v dG(v)$ and $H_0(u) = \int_0^\infty u^v dG_0(v)$ for all $u \in [0,1]$, where $G(v)$ and $G_0(v)$ are distribution functions with non-negative support. If $H(u) = H_0(u)$ on an arbitrary interval $(\underline{u}, \overline{u}) \subset [0,1]$ then $G(v) = G_0(v)$ on $[0,\infty)$, hence $H(u) = H_0(u)$ on $[0,1]$.*

Proof: Appendix

Remark. Note that the condition that H takes the form $H(u) = \int_0^\infty u^v dG(v)$ for all $u \in [0,1]$ is necessary. If $H_0(u) = \int_0^\infty u^v dG_0(v)$ a.e. on $[0,1]$ but only $H(u) = \int_0^\infty u^v dG(v) = H_0(u)$ a.e. on $(\underline{u}, \overline{u})$ then Lemma 3 implies that $H(u) = \int_0^\infty u^v dG_0(v)$ a.e. on $(\underline{u}, \overline{u})$ but not a.e. on $[0,1]$ if $\underline{u} > 0$ or $\overline{u} < 1$. The latter follows easily from the fact that we can always extend $H(u)$ beyond $(\underline{u}, \overline{u})$ such that $H(u) \neq \int_0^\infty u^v dG_0(v)$ with positive Lebesgue measure.

Theorem 2. *In addition to the conditions of Theorem 1, suppose that either the support of the censoring time $\overline{T}$ is $(0,\infty)$, or $H(u) = \int_0^\infty u^v dG(v)$, where G has nonnegative support. Then $H(u) = H_0(u)$ for all $u \in [0,1]$.*

4 Proofs

4.1 Lemma 1

It is trivial that (8) implies that for any $\tau \in (0, \bar{t})$,

$$\sup_{t\in[0,\tau)} \left| H\left(\exp\left(-\exp(\beta' X).\Lambda\left(t\right)\right)\right) - H_0\left(\exp\left(-\exp(\beta_0' X).\Lambda_0\left(t\right)\right)\right)\right| \times I\left(\overline{T} \leq \tau\right) = 0 \text{ a.s.}$$

Letting $\tau \to \bar{t}$, the lemma now follows from the fact that by Assumption 2, $I\left(\overline{T} \leq \bar{t}\right) = 1$ a.s.

4.2 Lemma 2

Suppose that for an $n \geq 1$, $\rho^n \Lambda_0(t) = \Lambda_0\left(\Lambda^{-1}\left(\rho^n \Lambda(t)\right)\right)$ on $(0, \overline{t})$. Let $t_2 = \Lambda^{-1}\left(\rho^n \Lambda(t)\right)$ and note that $t_2 \leq t$ because $\rho \in (0,1)$. Then

$$\begin{aligned}
\rho^{n+1}\Lambda_0(t) &= \rho^n.\Lambda_0\left(\Lambda_0^{-1}\left(\rho\Lambda_0(t)\right)\right) \\
&= \rho^n.\Lambda_0(t_2) = \Lambda_0\left(\Lambda^{-1}\left(\rho^n \Lambda(t_2)\right)\right) \\
&= \Lambda_0\left(\Lambda^{-1}\left(\rho^{n+1}\Lambda(t)\right)\right).
\end{aligned}$$

4.3 Lemma 3

The proof of Lemma 3 is an adaptation of the results of Abbring and van den Berg (2003). It is given in the separate appendix (Bierens and Carvalho 2007b, Lemma A.1) to Bierens and Carvalho (2007a), but will be reproduced here again.

First, observe that for $u \in (0,1)$ and non-negative integers m,

$$\sup_{v \geq 0} v^m u^{v-1} < \infty. \tag{28}$$

Take the derivative of $H(u)$ and $H_0(u)$ to $u \in (\underline{u}, \overline{u})$. Then by (28) and dominated convergence we may take the derivatives inside the integrals involved:

$$\int_0^\infty v u^{v-1} dG(v) = \int_0^\infty v u^{v-1} dG_0(v). \tag{29}$$

Multiply (29) by u, and then take the derivatives to $u \in (\underline{u}, \overline{u})$ again, which by (28) implies that

$$\int_0^\infty v^2 u^{v-1} dG(v) = \int_0^\infty v^2 u^{v-1} dG_0(v).$$

Repeating this procedure it follows by induction that

$$\int_0^\infty v^m u^v dG(v) = \int_0^\infty v^m u^v dG_0(v) \text{ for } m = 0, 1, 2, ... \tag{30}$$

hence

$$\int_0^\infty \sum_{m=0}^k \frac{(t.v)^m}{m!} u^v dG(v) = \int_0^\infty \sum_{m=0}^k \frac{(t.v)^m}{m!} u^v dG_0(v) \text{ for } k = 0, 1, 2, ... \tag{31}$$

Since

$$\begin{aligned}\sup_{k\geq 1}\left|\sum_{m=0}^{k}\frac{(t.v)^m}{m!}u^v\right| &\leq \sum_{m=0}^{\infty}\frac{(|t|.v)^m}{m!}\exp\left(-v.\ln(1/u)\right)\\ &= \exp\left(\left(|t|-\ln(1/u)\right).v\right)\\ &\leq 1 \text{ if } |t|<\ln(1/u)\end{aligned}$$

it follows from (31) and bounded convergence that

$$\int_0^\infty \exp\left(t.v\right)u^v dG(v) = \int_0^\infty \exp\left(t.v\right)u^v dG_0(v) \text{ for } |t|<\ln(1/u). \tag{32}$$

Now denote

$$F(x|u) = \frac{\int_0^x u^v dG(v)}{\int_0^\infty u^v dG(v)},\ F_0(x|u) = \frac{\int_0^x u^v dG_0(v)}{\int_0^\infty u^v dG_0(v)} \tag{33}$$

for $u \in (\underline{u}, \overline{u})$ and $x > 0$. Then it follows from (32) and (33) that

$$\int_0^\infty \exp\left(t.v\right)dF(v|u) = \int_0^\infty \exp\left(t.v\right)dF_0(v|u) \text{ for } |t|<\ln(1/u).$$

Hence it follows from the uniqueness of moment-generating functions that $F(x|u) = F_0(x|u)$ for $u \in (\underline{u}, \overline{u})$ and $x > 0$, and thus

$$\int_0^x u^v dG(v) = \int_0^x u^v dG_0(v). \tag{34}$$

Moreover, similar to (30) it follows from (34) that for $x > 0$, $m, k = 0, 1, 2, ...$ and $u \in (\underline{u}, \overline{u})$,

$$\int_0^x v^{m+k}u^v dG(v) = \int_0^x v^{m+k}u^v dG_0(v),$$

hence

$$\begin{aligned}\int_0^x v^m dG(v) &= \int_0^x v^m\sum_{k=0}^{\infty}\frac{\left(v.\ln(1/u)\right)^k}{k!}u^v dG(v)\\ &= \sum_{k=0}^{\infty}\frac{\left(\ln(1/u)\right)^k}{k!}\int_0^x v^{m+k}u^v dG(v)\\ &= \sum_{k=0}^{\infty}\frac{\left(\ln(1/u)\right)^k}{k!}\int_0^x v^{m+k}u^v dG_0(v)\end{aligned}$$

$$
\begin{aligned}
&= \int_0^x \sum_{k=0}^{\infty} \frac{(v.\ln(1/u))^k}{k!} v^m u^v dG_0(v) \\
&= \int_0^x \exp\left(-v.\ln(u)\right) u^v v^m dG_0(v \\
&= \int_0^x v^m dG_0(v).
\end{aligned}
$$

Thus, for $x > 0$ and $m = 0, 1, 2, ...,$

$$
\int_0^\infty (v.I(v \leq x))^m \, dG(v) = \int_0^\infty (v.I(v \leq x))^m \, dG_0(v), \tag{35}
$$

where $I(.)$ is the indicator function.

Now use the well-known fact that distributions of bounded random variables are equal if and only if all their moments are equal. Then, with V a random drawing from $G(v)$ and V_0 a random drawing from $G_0(v)$, it follows from (35) that for $x > v > 0$,

$$
\Pr\left[V.I(V \leq x) \leq v\right] = \Pr\left[V_0.I(V_0 \leq x) \leq v\right].
$$

This implies that $G(x) - G(v) = G_0(x) - G_0(v)$. Hence, letting $x \to \infty$, it follows that

$$
G(v) = G_0(v) \ \text{ for } v \geq 0.
$$

References

Abbring, J. H., and G. J. van den Berg (2003), "The Identifiability of the Mixed Proportional Hazards Competing Risks Model", *Journal of the Royal Statistical Society B* 65, 701-710.

Bierens, H. J. (2008), "Semi-Nonparametric Interval Censored Mixed Proportional Hazard Models: Identification and Consistency Results", *Econometric Theory* 24, 749-794.

Bierens, H.J. and J. R. Carvalho (2007a), "Semi-Nonparametric Competing Risks Analysis of Recidivism", *Journal of Applied Econometrics* 22, 971-993.

Bierens, H.J. and J. R. Carvalho (2007b), "Separate appendix to: Semi-Nonparametric Competing Risks Analysis of Recidivism" (http://econ.la.psu.edu/~hbierens/RECIDIVISM_APP.PDF).

Chung, K. L. (1974), *A Course in Probability Theory*, New York, Academic Press.

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

Heckman, J. J., and B. Singer (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica*, 52, 271-320.

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

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