

The unobserved heterogeneity distribution in duration analysis

BY JAAP H. ABBRING AND GERARD J. VAN DEN BERG

*Department of Economics, Free University Amsterdam, De Boelelaan 1105, 1081 HV
Amsterdam, The Netherlands*

jabbring@feweb.vu.nl gberg@feweb.vu.nl

SUMMARY

In a large class of hazard models with proportional unobserved heterogeneity, the distribution of the heterogeneity among survivors converges to a gamma distribution. This convergence is often rapid. We derive this result as a general result for exponential mixtures and explore its implications for the specification and empirical analysis of univariate and multivariate duration models.

Some key words: Duration analysis; Exponential mixture; Gamma distribution; Limit distribution; Mixed proportional hazard.

1. INTRODUCTION

It is well known that duration analysis produces incorrect results if unobserved heterogeneity is ignored (Lancaster, 1990). On average, subjects with relatively high hazard rates for unobserved reasons leave the state of interest first, so that samples of survivors are selected. Differences between such samples at different times reflect behavioural differences as well as this selection effect. Lancaster (1979) specified and estimated a proportional hazard model with multiplicative unobserved heterogeneity. This is called a mixed proportional hazard model and has subsequently become by far the most popular duration model in econometrics. Van den Berg (2001) presents a survey. The model is typically estimated using methods that require parametric functional form assumptions on the heterogeneity distribution. Lancaster (1979) assumes a gamma distribution, as do Vaupel et al. (1979), who introduced the model in demography. Nickell (1979) assumes a discrete distribution, and others have made other choices (Van den Berg, 2001).

Unfortunately, estimators of the mixed proportional hazard model are usually biased if the functional form of the heterogeneity distribution is misspecified. Extensive simulation evidence is provided by, for example, Baker & Melino (2000) and Bretagnolle & Huber-Carol (1988). Also, many empirical studies report that the estimates are sensitive to the functional form of the distribution (Heckman & Singer, 1984; Trussell & Richards, 1985; Hougaard et al., 1994; Keiding et al., 1997).

As a result, studies in which mixed proportional hazard models are estimated have wrestled with the choice of a functional form for the heterogeneity distribution; see for example Heckman & Singer (1984). In general, there is no argument in favour of one choice over the other. Also, formal results in the methodological studies by Heckman & Taber (1994), Kortram et al. (1995) and Horowitz (1999) indicate that duration data are rather uninformative about the shape of this distribution. In practice, researchers often choose a gamma mixing distribution for computational and expositional reasons;

all functions of interest have simple explicit expressions in this case (Lancaster, 1990). The mixed proportional hazard model with gamma heterogeneity is a preferred option in popular statistical packages like STATA, SAS, S-Plus and SPSS. Recently developed semiparametric estimators for the model also assume gamma heterogeneity; for examples see Clayton (1978), Meyer (1990), Nielsen et al. (1992), Murphy (1994, 1995), Petersen et al. (1996) and references in Andersen et al. (1993). The results in this paper rationalize this preference for the gamma distribution, and connect the many results that have been derived for the gamma case to a wider class of models.

2. A LIMIT RESULT FOR EXPONENTIAL MIXTURES

2.1. Exponential mixtures

Let Z and V be nonnegative random variables such that

$$\text{pr}(Z > z|V) = \exp(-Vz). \quad (1)$$

The marginal distribution of Z is therefore a mixture of exponential distributions with respect to the marginal distribution F of V :

$$\text{pr}(Z > z) = \int_0^\infty \exp(-vz)dF(v).$$

We examine the limiting behaviour of the distribution of V conditional on $Z \geq z$ as $z \rightarrow \infty$. In particular, we examine the limiting behaviour of

$$G_z(v) = \text{pr}(zV \leq v|Z \geq z).$$

2.2. Main result

We adopt the definitions of Feller (1971, §VIII.8) of slow variation and regular variation at 0.

DEFINITION 1. A positive function L defined on $(0, \infty)$ is slowly varying at 0 if $\lim_{y \downarrow 0} L(\alpha y)/L(y) = 1$ for every fixed $\alpha > 0$.

DEFINITION 2. A positive function k defined on $(0, \infty)$ is regularly varying with exponent $-\infty < \rho < \infty$ at 0 if

$$\lim_{y \downarrow 0} \frac{k(y)}{y^\rho L(y)} = 1$$

for a function L that is slowly varying at 0.

Also, let $\Gamma_{\alpha, \rho}$ denote the gamma distribution with density

$$\frac{\alpha^\rho}{\Gamma(\rho)} v^{\rho-1} \exp(-\alpha v), \quad \alpha, \rho > 0,$$

at v . We define the standard gamma distribution as $\Gamma_\rho := \Gamma_{1, \rho}$, with density denoted by γ_ρ . Finally, we define the limiting case Γ_0 such that $\Gamma_0(v) = 1$ for all $v \in [0, \infty)$. This is a degenerate distribution with all probability mass at zero.

We now state the main result.

PROPOSITION 1. *If $G_z \rightarrow G$ as $z \rightarrow \infty$, with G a proper distribution function, then $G = \Gamma_\rho$ for some $\rho \geq 0$. A necessary and sufficient condition for $G_z \rightarrow \Gamma_\rho$ ($\rho \geq 0$) is that F is regularly varying with exponent ρ at 0.*

Proof. The Laplace transform \mathcal{L}_{G_z} of G_z is given by

$$\mathcal{L}_{G_z}(s) = \int_0^\infty \exp(-sv) dG_z(v) = \frac{\mathcal{L}_F\{z(s+1)\}}{\mathcal{L}_F(z)}.$$

First, suppose that $G_z \rightarrow G$ as $z \rightarrow \infty$, with G a proper distribution function, and denote the Laplace transform of G by \mathcal{L}_G . Then $\mathcal{L}_{G_z} \rightarrow \mathcal{L}_G$ as $z \rightarrow \infty$ by the continuity of the Laplace transform. Thus,

$$\lim_{z \rightarrow \infty} \mathcal{L}_{G_z}(s) = \lim_{z \rightarrow \infty} \frac{\mathcal{L}_F\{z(s+1)\}}{\mathcal{L}_F(z)}$$

exists and is positive and nonincreasing on $(0, \infty)$. By Feller (1971, §VIII-8, Lemma 1), the latter limit then necessarily equals $(s+1)^{-\rho}$ for some $\rho \geq 0$. In turn, this implies that $G = \Gamma_\rho$ for some $\rho \geq 0$.

Secondly, again by continuity of the Laplace transform,

$$G_z \rightarrow \Gamma_\rho \iff \lim_{z \rightarrow \infty} \frac{\mathcal{L}_F\{z(s+1)\}}{\mathcal{L}_F(z)} = (s+1)^{-\rho},$$

so that $G_z \rightarrow \Gamma_\rho$ if and only if \mathcal{L}_F is regularly varying with exponent $-\rho$ at infinity. In turn, it follows from an Abelian/Tauberian theorem, like Theorem 3 of Feller (1971, §XIII-5), that this is true if and only if F varies regularly with exponent ρ at 0. \square

Examples of continuous distributions that are regularly varying at 0 with exponent $\rho > 0$ are all distributions with densities that have finite positive limits at 0, such as the exponential, uniform and truncated normal distributions, and all gamma and beta distributions. Examples with $\rho > 0$ also include some discrete distributions with dense support near 0. The case $\rho = 0$ includes all distributions, including finitely discrete distributions, with a point mass at 0.

An obvious example of a distribution that is not regularly varying at 0 is a distribution without support near 0. Let $v_0 := \inf\{v : F(v) > 0\}$ be the largest lower bound on the support of F . Let F^0 be the distribution of $V - v_0$ and G_z^0 the distribution of $z(V - v_0)$ conditional on $Z \geq z$. Then Proposition 1 applies without change with F replaced by F^0 and G_z replaced by G_z^0 .

2.3. Speed of convergence

In statistical applications results about the rate of convergence of G_z to G would be useful. The following example shows that no general result about this rate can be derived under the conditions of Proposition 1, notably under regular variation of F with exponent $0 \leq \rho < \infty$ at 0, alone. First, suppose that $F(v) = v^k$ on $(0, 1)$ for some $k > 0$. Then $\rho = k$ and $G_z \rightarrow \Gamma_k$ by Proposition 1. Note that this convergence is uniform. It is easy to show that

$$\lim_{z \rightarrow \infty} \frac{\sup_v \{G_z(v) - G(v)\}}{z^{k-1} \exp(-z)/\Gamma(k)} = 1.$$

This result does not generalize to all distributions that are regularly varying with exponent ρ . For example, let $F(v) = v\{1 - \log(v)\}$ on $(0, 1)$. Then F is regularly varying

with exponent 1 at 0, but convergence is much slower than for the linear case $k = 1$ above. In particular, it can be shown that

$$\lim_{z \rightarrow \infty} \frac{\sup_v \{G_z(v) - G(v)\}}{c/\log(z)} = 1,$$

for some constant $0 < c < \infty$. The working paper version of this article, which is available upon request, provides details.

Rather than pursuing general results on the speed of convergence under additional assumptions, we examine a range of specific cases. We focus on the distribution G_z^* of $(z+1)V|Z \geq z$ rather than on G_z . The distribution G_z^* has the same limit as G_z , but has the additional expositional advantage that $G_z^* = \Gamma_\rho$ for all z if $F = \Gamma_\rho$ (Lancaster, 1990). In all cases, F is taken to be a beta distribution, $\text{Be}(\mu, \nu; \bar{v})$. This family covers a wide range of density shapes, in particular around 0. Its densities are defined by

$$\beta_{\mu, \nu; \bar{v}}(v) \propto v^{\mu-1}(\bar{v} - v)^{\nu-1}, \quad \text{for } 0 < v < \bar{v},$$

and $\beta_{\mu, \nu; \bar{v}}(v) = 0$ otherwise, for $\mu, \nu > 0$. The density is increasing if $\mu > 1$ and $\nu < 1$, decreasing if $\mu < 1$ and $\nu > 1$, U-shaped if $\mu < 1$ and $\nu < 1$, and bell-shaped if $\mu > 1$ and $\nu > 1$. It includes the uniform density on $(0, \bar{v})$ for $\mu = \nu = 1$. The corresponding cumulative distribution function is regularly varying with exponent $\mu > 0$, which implies that $G_z^* \rightarrow \Gamma_\mu$ according to Proposition 1. The parameter \bar{v} is a scale parameter: we can write $V = \bar{v}V^1$, with V^1 distributed with density $\beta_{\mu, \nu; 1}$. We ensure that the examples are mutually comparable by fixing the value of \bar{v} for each given μ and ν such that $E(\log V) = 0$.

Figure 1 displays the densities g_z^* of $(z+1)V|Z \geq z$ corresponding to $\beta_{\mu, \nu; \bar{v}}$ for values of μ and ν that generate the various density shapes mentioned above, and for $z = 0, 0.5, 1, 2, 5$. Obviously, in each case, $g_0^* = \beta_{\mu, \nu; \bar{v}}$. The figures also display the limiting density $g_\infty^* = \gamma_\mu$. In all cases, we observe convergence to the gamma density. To assess whether or not convergence is rapid we need to obtain some insight into what constitutes a large or a small value of z . By equation (1), the normalization $E(\log V) = 0$ implies that $E(\log Z) = -0.577$. In addition, note that $x \mapsto \exp(-x)$ is convex, so that $E(1/V) = E[\exp\{-\log(V)\}] \geq \exp[-E\{\log(V)\}] = 1$ by Jensen's inequality, and as a result $E(Z) = E(1/V) \geq 1$. To be more precise, if $\mu \leq 1$ then $E(Z) = \infty$, whereas otherwise $E(Z) = (1/\bar{v})(\mu + \nu - 1)/(\mu - 1)$. Given all this, it is fair to state that the convergence is rapid: in most cases depicted g_z^* is close to the density of its limiting distribution for z as small as 0.5 or 1.

3. SINGLE-SPELL DURATION ANALYSIS

3.1. *The mixed proportional hazard model*

We first discuss the implications of Proposition 1 for the mixed proportional hazard model as popularized by Lancaster (1979) and Vaupel et al. (1979). The mixed proportional hazard model is a model for the distribution of a continuous random duration T conditional on a vector X of observed covariates. Under some regularity conditions, it is straightforward to extend the analysis to the case of time-varying explanatory variables, but for ease of exposition we do not take this up here. The model specifies the distribution of $T|X$ as a mixture of the distribution of $T|(X, V)$ over the marginal distribution F of V . Here, V is a nonnegative random unobserved heterogeneity factor that is independent of X . The

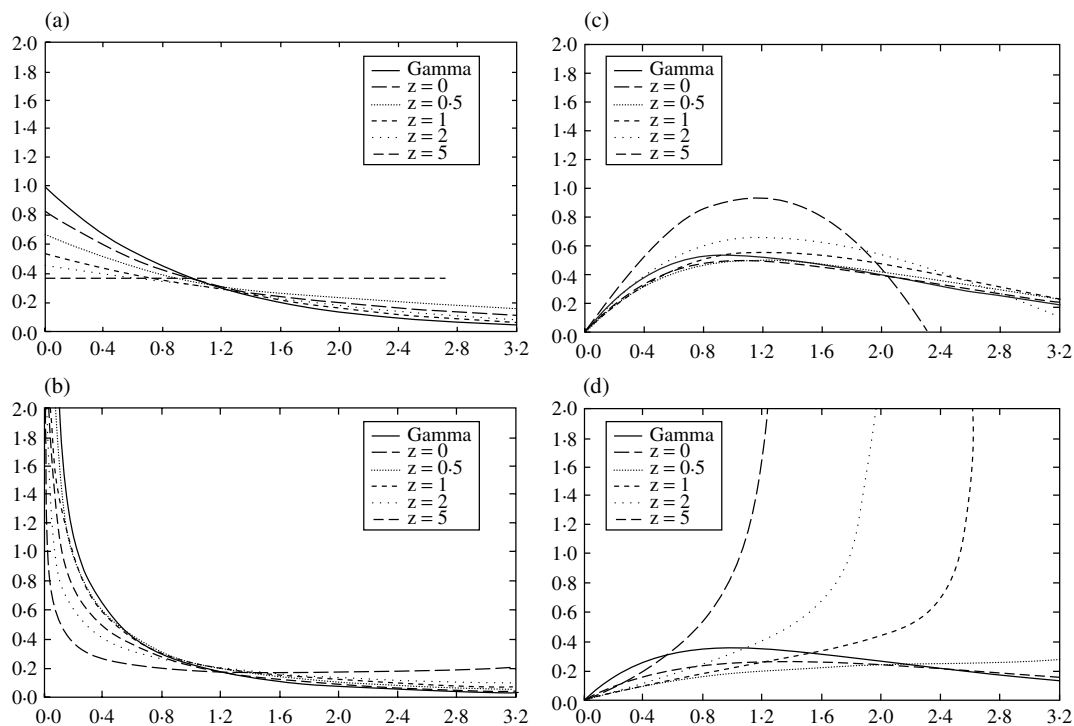


Fig. 1. Densities g_z^* of $(z+1)V|Z \geq z$ for (a) $V \sim \text{Be}(1, 1, e^1)$, i.e. $\text{Un}(0, e)$, with limiting density γ_1 , (b) $V \sim \text{Be}(1/2, 1/2, 4)$, with limiting density $\gamma_{1/2}$, (c) $V \sim \text{Be}(2, 2, e^{5/6})$, with limiting density γ_2 , (d) $V \sim \text{Be}(2, 1/2, e^{5/3}/4)$, with limiting density γ_2 .

distribution of $T|(X, V)$ is specified in terms of its hazard rate, which is defined by

$$\theta(t|X, V) := \lim_{dt \downarrow 0} \frac{\text{pr}(t \leq T < t + dt | T \geq t, X, V)}{dt}$$

for almost all t . In particular, the mixed proportional hazard model specifies that

$$\theta(t|X, V) = \lambda(t)\phi(X)V.$$

The ‘baseline hazard’ $\lambda : [0, \infty) \rightarrow [0, \infty)$ is integrable on bounded intervals, with integral $\Lambda(t) := \int_0^t \lambda(y)dy$, and $\lim_{t \rightarrow \infty} \Lambda(t) = \infty$. The function $\phi : \mathcal{X} \rightarrow (0, \infty)$ is a measurable function, with \mathcal{X} the support of X .

Define $Z := \Lambda(T)\phi(X)$. Then ZV is a unit-exponentially distributed random variable that is independent of (X, V) . Conditional on X , Z is distributed as a mixture of exponential distributions with mixing distribution F . Thus, Proposition 1 applies to the distribution G_z of $zV|(Z \geq z, X)$. In particular, assume that F is regularly varying with exponent $\rho > 0$ at 0. Then Proposition 1 implies that the distribution of

$$\{c_0 + c_1\Lambda(t)\phi(x)\} V | (T \geq t, X = x)$$

converges to a $\Gamma_{1/c_1, \rho}$ distribution as $\Lambda(t)\phi(x) \rightarrow \infty$, for any $c_0 \in \mathbb{R}$ and $c_1 > 0$. This in turn implies that the distribution of $V|(T \geq t, X = x)$ can be approximated by a gamma distribution with parameters $(c_0/c_1) + \Lambda(t)\phi(x)$ and ρ . Here, the value of c_0/c_1 is arbitrary, apart from the requirement that $c_0/c_1 > -\Lambda(t)\phi(x)$: it is not determined by the limit result nor by properties of F . For t small we require $c_0 > 0$, however, so that $c_0/c_1 > 0$. Exactly

the same distribution for $V|(T \geq t, X = x)$ is also obtained if a gamma mixing distribution $F = \Gamma_{c_0/c_1, \rho}$ is adopted.

Note that we can achieve $\Lambda(t)\phi(x) \rightarrow \infty$ by letting $t \rightarrow \infty$ for given $x \in \mathcal{X}$. However, our result is not only a ‘large t ’ result. If $\{\phi(x); x \in \mathcal{X}\}$ includes a sequence that diverges to ∞ , we can also achieve $\Lambda(t)\phi(x) \rightarrow \infty$ along the corresponding sequence of covariate values for fixed t such that $\Lambda(t) > 0$.

3.2. Estimation of the baseline with left-truncated data

These results can be applied to the empirical analysis of mixed proportional hazard models with left-truncated data. Duration data are left-truncated if a spell only enters the sample if its duration exceeds some $t_0 > 0$. Left-truncation frequently arises in economic applications and poses some hard and mostly unresolved problems.

In general, mixed proportional hazard models that are identified from complete data will not be identified from left-truncated data. However, under the assumption that V has a gamma distribution, some interesting features of the model can still be identified. Consider the two-sample case in which X is binary. Let $S_x(t) := \text{pr}(T > t_0 + t | X = x, T > t_0)$. Note that S_0 and S_1 can be estimated from data that are left-truncated at t_0 . If $F = \Gamma_{\alpha, \rho}$ then

$$S_x(t) = \left\{ 1 + \tilde{\Lambda}(t)\tilde{\phi}(x) \right\}^{-\rho}, \quad (2)$$

with $\tilde{\Lambda}(t) := \Lambda(t_0 + t) - \Lambda(t_0)$ and $\tilde{\phi}(x) := \phi(x)/\{\alpha + \Lambda(t_0)\phi(x)\}$. Thus, the model for (S_0, S_1) reduces to a mixed proportional hazard model with integrated baseline $\tilde{\Lambda}$, regressor effects $\tilde{\phi}$ and $\Gamma_{1, \rho}$ -distributed heterogeneity. Elbers & Ridder’s (1982) identification result implies that ρ is identified from (S_0, S_1) , and that $\tilde{\Lambda}$ and $\tilde{\phi}$ are identified up to a scale normalization, provided that $\tilde{\phi}(0) \neq \tilde{\phi}(1)$. This, in turn, identifies λ up to scale almost everywhere on (t_0, ∞) .

The regressor effects $\tilde{\phi}$ confound dynamic selection effects and the structural covariate effects embodied in ϕ . Therefore, we cannot separately identify ϕ . However, we can identify the sign of $\phi(1) - \phi(0)$, because it equals the sign of $\tilde{\phi}(1) - \tilde{\phi}(0)$. We return to this in §3.3.

Our limit result implies that (2) holds approximately in a much wider class of models. This suggests that we adopt the gamma specification (2) and use estimates of $\tilde{\Lambda}$ to estimate Λ with truncated data. We expect this estimator often to outperform alternative estimators such as those based on a flexible discrete approximation of the heterogeneity distribution in the truncated sample.

We illustrate this point with some Monte Carlo analysis. We generate data from two-sample mixed proportional hazard models with linear Λ , and compare baseline estimates of the models with respectively gamma and two-point heterogeneity. For expositional convenience, we exploit our knowledge that the baseline is in the Weibull class and specify $\tilde{\Lambda}(t) = (t_0 + t)^{\exp(\delta)} - t_0^{\exp(\delta)}$. Table 1 reports simulated root mean squared errors of the maximum likelihood estimator of δ for three data-generating processes differing only in the heterogeneity distribution used. Each row in the table corresponds to a different data-generating process. They are all mixed proportional hazard models with linear Λ , $\text{pr}(X = 0) = \text{pr}(X = 1) = 0.5$, $\phi(1) = 2\phi(0)$, and they vary only by the distribution of the heterogeneity. ‘Uniform and discrete’ corresponds to $V \sim \text{Un}(0, 0.2)$ with probability 0.2, and $\text{pr}(V = 5) = 0.8$. The scale of each data-generating process’s hazard is calibrated so that $\text{pr}(T > 1) = 0.5$. The first two columns report root mean squared errors among 100 simulated samples of 5000 observations each. The last two columns report root mean

Table 1. *Simulated root mean squared errors of the maximum likelihood estimator of the Weibull parameter δ , for different types of heterogeneity in the data-generating process and for gamma and discrete model specifications*

Heterogeneity	Complete data		Left-truncated data	
	Gamma model	Discrete model	Gamma model	Discrete model
Exponential ($\Gamma_{\alpha,1}$)	0.021	0.358	0.159	1.182
Uniform ($\beta_{1,1;\bar{v}}$)	0.100	0.240	0.626	1.111
Uniform and discrete	0.269	0.251	0.404	0.922

squared errors among 100 simulated samples of 10 000 observations left-truncated at 1, which leaves 5000 observations on average.

With exponential, i.e. $\Gamma_{\alpha,1}$, heterogeneity in the data, the gamma model is correctly specified and performs well. The misspecified discrete-mixing model performs much worse. With uniform, i.e. $\beta_{1,1;\bar{v}}$, heterogeneity in the data, both models are misspecified. However, the gamma model approximates the data-generating process of the truncated sample and outperforms the discrete model, even though the latter has one extra parameter to be fitted in the heterogeneity distribution. The third data-generating process uses a heterogeneity distribution that resembles a two-point distribution but also satisfies the conditions for our limit result. Indeed, in this case the gamma model performs slightly worse with complete data, but much better with truncated data. We conclude that the gamma approximation can be fruitfully applied to reduce the error in baseline estimation with truncated data.

In empirical practice, the Weibull-gamma model may be too restrictive. First, one typically does not know that the baseline is in the Weibull class. Then, estimators that do not require parametric specification of $\tilde{\Lambda}$ can be applied. Section 1 provides some references. Secondly, recall from §2.2 that Proposition 1 applies more generally to the distribution G_z^0 of $z(V - v_0)|Z \geq z$. The corresponding approximation for the distribution of V is a transposed gamma distribution. This family of distributions has densities $v \in [v_0, \infty) \mapsto \gamma_{\alpha,\rho}(v - v_0)$, with three parameters, namely $\alpha > 0$, $\rho > 0$ and $v_0 \geq 0$. In practice, one could adopt this family if the gamma family were considered to be too restrictive.

3.3. Testing for unobserved heterogeneity

In the absence of unobserved heterogeneity, we could avoid the difficult problem of estimating a mixed proportional hazard model with left-truncated data. Therefore, we now focus on testing for unobserved heterogeneity with truncated data.

Consider again the two-sample case of §3.2, without parametric restrictions on the baseline. Denote the observable hazards

$$\theta(t|X = x) = \lambda(t)\phi(x)E(V|T \geq t, X = x) \quad (3)$$

by $\theta_x(t)$. First note that, under Elbers & Ridder's (1982) restriction to mixed proportional hazard models such that $E(V) < \infty$, testing for unobserved heterogeneity is equivalent to testing for proportionality of θ_0 and θ_1 . Thus, we can build on tests for proportional hazards developed by, for example, Gill & Schumacher (1987), Dabrowska et al. (1989), Dabrowska et al. (1992) and Deshpande & Sengupta (1995). None of these concentrates on mixed proportional hazard alternatives. Here, we outline and illustrate a way to adjust,

in particular, the Gill-Schumacher test to have power against mixed proportional hazard alternatives.

The Gill-Schumacher test is based on comparing different estimators of the relative risk θ_1/θ_0 . It is a two-sided test that is consistent against alternatives in which this hazard ratio is monotone. For now, suppose that we know the sign of $\phi(1) - \phi(0)$ and, without loss of generality, let $\phi(1) > \phi(0)$. Then, with gamma heterogeneity, θ_1/θ_0 is decreasing. This implies that a one-sided version of the Gill-Schumacher test can be constructed that is consistent against the gamma mixed proportional hazard alternative. This test will be more powerful against this alternative than the two-sided test. Since the implied model for a truncated sample is again a gamma mixed proportional hazard model, these results carry over to truncated data.

For general heterogeneity distributions, θ_1/θ_0 is decreasing in a neighbourhood of 0, but may be nonmonotone at higher durations (Van den Berg, 2001). Therefore, it is less obvious that a one-sided version of the Gill-Schumacher test will outperform the two-sided test against general mixed proportional hazard alternatives. Moreover, with left-truncated data, we cannot exploit the fact that θ_1/θ_0 is decreasing in a neighbourhood of 0. However, our approximation result shows that θ_1/θ_0 will quickly be decreasing in a wide class of mixed proportional hazard models. This in particular suggests that our one-sided version of the Gill-Schumacher test will outperform the two-sided test with truncated data.

We illustrate this point with Monte Carlo simulations of the rejection rates of the one-sided and two-sided Gill-Schumacher tests under various data-generating processes. Table 2 tabulates these rates for tests with a nominal size of 5%. The tests are based on the weight functions recommended by Gill & Schumacher (1987, p. 294). The one-sided tests are chosen to be consistent against the mixed proportional hazard model with gamma heterogeneity. As in Table 1, each row in the table corresponds to a different data-generating process. The data-generating processes are all mixed proportional hazard models with linear Λ , $\text{pr}(X = 0) = \text{pr}(X = 1) = 0.5$, $\phi(1) = 2\phi(0)$, and they vary only by the distribution of the heterogeneity. ‘Discrete’ corresponds to $\text{pr}(V = 1) = \text{pr}(V = 2) = 0.5$. The scale of each data-generating process’s hazard is calibrated so that $\text{pr}(T > 1) = 0.5$. The first two columns report rejection rates among 1000 simulated samples of 1000 observations each. The last two columns report rejection rates among 1000 simulated samples of 2000 observations left-truncated at 1, which leaves 1000 observations on average. The first row shows that the actual size of the one-sided tests is slightly below that of the two-sided tests and the nominal size. Under all three alternatives considered, the one-sided Gill-Schumacher test outperforms the two-sided test. As expected, the tests have relatively low power against the alternative with discrete heterogeneity, and perform very poorly with

Table 2. *Simulated rejection rates of one-sided and two-sided versions of the Gill-Schumacher test for proportional hazards, for different types of heterogeneity in the data-generating process*

Heterogeneity	Complete data		Left-truncated data	
	One-sided test	Two-sided test	One-sided test	Two-sided test
None	4.5%	4.8%	4.6%	6.2%
Exponential ($\Gamma_{\alpha,1}$)	92.7%	87.5%	43.2%	32.9%
Uniform ($\beta_{1,1;\overline{\tau}}$)	97.5%	96.0%	78.5%	66.9%
Discrete	14.7%	9.3%	7.3%	5.8%

truncated data under this alternative. The tests perform very well against the exponential and uniform alternatives. Note that, in particular with truncated data, substantial power against the uniform model is gained from moving from a two-sided to a one-sided test.

These results indicate that we can safely use Gill-Schumacher tests to test for heterogeneity with truncated data. They strongly suggest that we use the one-sided version of this test. A practical problem is that the one-sided test used so far assumed the sign of $\phi(1) - \phi(0)$ to be known. Since $\phi(0)$ and $\phi(1)$ are unknown parameters, this test is not feasible. A feasible test can be constructed by empirically determining the sign of $\phi(1) - \phi(0)$. In §3.2 we have seen that, in the gamma case and with left-truncated data, the sign of $\phi(1) - \phi(0)$ equals the identified sign of $\tilde{\phi}(1) - \tilde{\phi}(0)$. This suggests that we empirically determine this sign by estimating a gamma mixed proportional hazard model. This approach to the construction of a feasible test should still work well under our gamma approximation with truncated data. In theory, it can fail in more general cases, but our Monte Carlo evidence suggests that this is not a problem. Estimates of the sign of $\phi(1) - \phi(0)$ using a gamma mixed proportional hazard model, not reported in detail here, are correct in all samples simulated for Table 2. This suggests that, in general, the feasible statistic is very close to the infeasible one. In particular, it would have produced exactly the same Monte Carlo results.

The problem of estimating the sign of $\phi(1) - \phi(0)$ is of more general interest and is closely related to the results on the monotonicity of the relative risk θ_1/θ_0 . With complete data, the sign of $\phi(1) - \phi(0)$ can be inferred directly from a comparison of the survival curves in the two samples. With left-truncated data, we can exploit the fact that the sign of $\phi(1) - \phi(0)$ equals the signs of both $\theta_1 - \theta_0$ and $-d \log\{\theta_1(t)/\theta_0(t)\}/dt$ in the gamma case. Our limit result suggests that this result can be used more generally with truncated data. In theory, with some distributions of V this approach to establishing the sign of $\phi(1) - \phi(0)$ can be misleading (Van den Berg, 2001). However, our simulation results suggest that this is not very likely to be a problem with data that are left-truncated but not right-censored. Practical problems may arise if there is both truncation and heavy right-censoring, and we can only observe survival on a bounded positive interval. Our limit result may then provide some guidance in the interpretation of relative risks.

3.4. More general single-spell duration models

Throughout this section, we have focused on the popular mixed proportional hazard model. However, our results apply without change to more general models. In particular, suppose that, conditional on (X, V) , T is continuously distributed with hazard rate $\xi(t, X)V$ at time t . Thus, we maintain the separability of V , but relax the assumption that $\xi(t, X)$ is proportional in t and X . The function $\xi : [0, \infty) \times \mathcal{X} \rightarrow [0, \infty)$ is a measurable function such that $\int_0^t \xi(y, x)dy$ exists for all $t \in [0, \infty)$ and that $\lim_{t \rightarrow \infty} \int_0^t \xi(y, x)dy = \infty$, for all $x \in \mathcal{X}$. Then our analysis of the mixed proportional hazard model applies with $Z := \int_0^T \xi(y, X)dy$.

In addition, we can allow for general dependence of X and V if we focus on the limit as $t \rightarrow \infty$. Then, we would not impose any structure on the way the observed covariates X enter the model. We could think of the entire analysis as being conditional on X . However, if we want to achieve the limit along a sequence of covariate values for given t then we need independence of X and V .

4. MULTIVARIATE DURATION ANALYSIS

4.1. *The multivariate mixed proportional hazard model*

Nowadays, the study of multiple dependent durations is widespread (Van den Berg, 2001). The vast majority of applications use multivariate generalizations of the mixed proportional hazard model in which the durations associated with a certain unit or subject are dependent if the corresponding unobserved determinants are dependent. Often it is natural and/or convenient to assume that such durations have identical unobserved heterogeneity terms V . For example, V may be thought to capture unobserved genetic determinants, so that in studies on lifetime durations of identical twins V is identical within twin pairs. Hougaard et al. (1992b) present an example of this. In mixed proportional hazard models for multiple-spell data, the multiple durations that a single subject spends in the same state are dependent because they are affected by the same realisation of V ; some examples are Newman & McCulloch (1984) on birth intervals, Coleman (1989) on unemployment durations and Lillard (1993) and Lillard & Panis (1996) on marriage durations. In general, multivariate duration models may concern successive spells in a given state, or successive spells in different states, as well as competing risks in a single state.

For expositional reasons we restrict ourselves to two possibly dependent duration variables T_1 and T_2 , and we suppress the covariates X throughout. As in §3.4, we can think of the entire analysis as being conditional on X . Thus, we do not impose any structure on the way X enters the model. We assume that $T_1 \perp\!\!\!\perp T_2 | V$, for some nonnegative random variable V . We will refer to V as a ‘subject-specific effect’, although a subject may consist of two individuals, as in the case of twins. Conditional on V , T_1 and T_2 are continuously distributed with hazard rates $\lambda_1(t)V$ and $\lambda_2(t)V$ at time t , respectively. We adopt regularity assumptions similar to those in §3, and define $\Lambda_i(t) := \int_0^t \lambda_i(y)dy$ and $Z_i := \Lambda_i(T_i)$.

Conditional on V , Z_1V and Z_2V are independently and unit-exponentially distributed. Thus, by analogy with the single-spell case in §3, Proposition 1 implies that $zV | Z_i \geq z$ converges in distribution to a gamma distribution as $z \rightarrow \infty$, that $(z_1 + z_2)V | (Z_1 \geq z_1, Z_2 \geq z_2)$ converges to a gamma distribution as z_1 and/or z_2 go to ∞ , and that $z_1V | (Z_1 \geq z_1, Z_2 = z_2)$ converges to a gamma distribution as $z_1 \rightarrow \infty$. As a result, we may approximate the distributions of these random variables by gamma distributions.

4.2. *The cross-ratio and current-versus-alive functions*

The results of the previous subsection have implications for how certain observable dependency measures change with the elapsed or realized durations t_1 and t_2 . To demonstrate this, we consider the observable hazard rates $\theta_i(t_i)$, $\theta_i(t_i | T_j > t_j)$ and $\theta_i(t_i | T_j = t_j)$ of the distributions of T_i , $T_i | T_j > t_j$ and $T_i | T_j = t_j$, respectively. It is straightforward to show that these can be expressed in terms of the model determinants by way of equations similar to (3), as is done for example in Lancaster (1990). Now consider the relative effect of the realization of one duration variable on the hazard rate of the other, and the way this changes over the durations. In the literature, this is captured by the cross-ratio function

$$\Theta_{\text{cr}}(t_1, t_2) := \frac{\theta_1(t_1 | T_2 = t_2)}{\theta_1(t_1 | T_2 > t_2)}$$

and the current-versus-alive function

$$\Theta_{\text{cva}}(t_1, t_2) := \frac{\theta_1(t_1|T_2 = t_2)}{\theta_1(t_1|T_2 > t_1)}.$$

The cross-ratio function, which is also called the odds-ratio function, captures to what extent the hazard rate of T_1 at t_1 depends on knowledge that T_2 is realized at a certain point of time t_2 , relative to when T_2 is realized after t_2 . The current-versus-alive function captures to what extent the hazard rate of T_1 at t_1 depends on knowledge that T_2 is realized at a certain point of time t_2 , relative to when T_2 is not yet realized. Relevant references are Clayton (1978), Oakes (1989), Anderson et al. (1992) and Yashin & Iachine (1999) on Θ_{cr} , Hougaard et al. (1992a) on Θ_{cva} , and Hougaard et al. (1992b) and Klein et al. (1992) on both. These studies provide characterizations and properties for the general case, and they also discuss how the functions are affected by F . Both functions can be estimated nonparametrically from data on T_1 and T_2 .

The functions $\Theta_{\text{cr}}(t_1, t_2)$ and $\Theta_{\text{cva}}(t_1, t_2)$ are informative about the way in which the dependence of two duration variables changes over time. For example, if Θ_{cr} decreases in t_1 for a given t_2 , with $t_1 > t_2$, then the knowledge that T_2 has been realized at t_2 becomes less important as time proceeds, so the local dependence between T_1 and T_2 decreases as $t_1 - t_2$ increases. Anderson et al. (1992) show that

$$\Theta_{\text{cr}}(t_1, t_2) = 1 + \{\text{cv}(t_1, t_2)\}^2,$$

where $\text{cv}(t_1, t_2)$ is the coefficient of variation of $V|(T_1 > t_1, T_2 > t_2)$. The larger this coefficient, the larger is the part of the variation in the truncated duration variables that is explained by V , so the stronger their dependence.

If $F = \Gamma_{\alpha, \rho}$ then $\Theta_{\text{cr}}(t_1, t_2) = (\rho + 1)/\rho$ and $\{\text{cv}(t_1, t_2)\}^2 = 1/\rho$, so then $\Theta_{\text{cr}}(t_1, t_2)$ and $\text{cv}(t_1, t_2)$ are constants. In fact, Oakes (1989) shows that constancy of $\Theta_{\text{cr}}(t_1, t_2)$ for all t_1 and t_2 characterizes the gamma distribution for V . This can be further strengthened by showing that $\Theta_{\text{cr}}(t, t)$ is constant for all t if and only if V has a gamma distribution. Note that the constancy of $\text{cv}(t_1, t_2)$ in the gamma case reflects the fact that the relative amount of heterogeneity among the survivors remains constant when V has a gamma distribution. This is of course to be expected given the results in §2.

Similarly, if $F = \Gamma_{\alpha, \rho}$ then

$$\Theta_{\text{cva}}(t_1, t_2) = \left(\frac{\rho + 1}{\rho} \right) \frac{\alpha + \Lambda_1(t_1) + \Lambda_2(t_1)}{\alpha + \Lambda_1(t_1) + \Lambda_2(t_2)},$$

implying that $\Theta_{\text{cva}}(t_1, t_2)$ increases in t_1 for t_1 close to t_2 . With other distributions of V , the derivatives of $\Theta_{\text{cva}}(t_1, t_2)$ do not necessarily have the signs that they have in the gamma case, as shown in the above literature. For example, in many cases, $\Theta_{\text{cva}}(t_1, t_2)$ decreases in t_1 .

We can now apply the convergence results from earlier in the paper. The behaviour of the dependency functions can be expected to hold for long-term survivors for any bivariate duration data with a common heterogeneity term V whether or not it is gamma distributed. Consequently, such behaviour can be examined empirically from estimates of these functions using long-term survivors to test for heterogeneity.

ACKNOWLEDGEMENT

The authors wish to thank, for helpful comments, the editor, an anonymous referee, Richard Blundell, Karsten Hansen, Jim Heckman, Andries Lenstra, Geert Ridder and participants in workshops at the University of Chicago and meetings of the Midwest Econometrics Group, the Econometric Society and the U.K. Econometric Study Group. The research of Jaap Abbring was supported by a fellowship of the Royal Netherlands Academy of Arts and Sciences. Both authors are also affiliated to the Tinbergen Institute, the Institute for the Study of Labour in Bonn and the Institute for Fiscal Studies in London. Part of this paper was written when Abbring was at the Department of Economics of the University of Chicago and Van den Berg was at the Swedish Institute for Social Research at Stockholm University.

REFERENCES

- ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. & KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer.
- ANDERSON, J. E., LOUIS, T. A., HOLM, N. V. & HARVALD, B. (1992). Time-dependent association measures for bivariate survival distributions. *J. Am. Statist. Assoc.* **87**, 641–50.
- BAKER, M. & MELINO, A. (2000). Duration dependence and nonparametric heterogeneity: a Monte Carlo study. *J. Econ.* **96**, 357–93.
- BRETAGNOLLE, J. & HUBER-CAROL, C. (1988). Effect of omitting covariates in Cox's model for survival data. *Scand. J. Statist.* **15**, 125–38.
- CLAYTON, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of family tendency in chronic disease incidence. *Biometrika* **65**, 141–51.
- COLEMAN, T. S. (1989). Unemployment behaviour: evidence from the CPS work experience survey. *J. Hum. Resour.* **24**, 1–38.
- DABROWSKA, D. M., DOKSUM, K. A., FEDUSKA, N. J., HUSING, R. & NEVILLE, P. (1992). Methods for comparing cumulative hazard functions in a semi-proportional hazard model. *Statist. Med.* **11**, 1465–76.
- DABROWSKA, D. M., DOKSUM, K. A. & SONG, J. K. (1989). Graphical comparison of cumulative hazards for two populations. *Biometrika* **76**, 763–73.
- DESHPANDE, J. V. & SENGUPTA, D. (1995). Testing for the hypothesis of proportional hazards in two populations. *Biometrika* **82**, 251–61.
- ELBERS, C. & RIDDER, G. (1982). True and spurious duration dependence: the identifiability of the proportional hazard model. *Rev. Econ. Stud.* **49**, 403–9.
- FELLER, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. II, 2nd ed. New York: Wiley.
- GILL, R. & SCHUMACHER, M. (1987). A simple test for the proportional hazards assumption. *Biometrika* **74**, 289–300.
- HECKMAN, J. J. & SINGER, B. (1984). Econometric duration analysis. *J. Economet.* **24**, 63–132.
- HECKMAN, J. J. & TABER, C. R. (1994). Econometric mixture models and more general models for unobservables in duration analysis. *Statist. Meth. Med. Res.* **3**, 279–302.
- HOROWITZ, J. L. (1999). Semiparametric estimation of a proportional hazard model with unobserved heterogeneity. *Econometrica* **67**, 1001–28.
- HOUGAARD, P., HARVALD, B. & HOLM, N. V. (1992a). Assessment of dependence in the life times of twins. In *Survival Analysis: State of the Art*, Ed. J. P. Klein & P. K. Goel, pp. 77–97, Dordrecht: Kluwer Academic Publishers.
- HOUGAARD, P., HARVALD, B. & HOLM, N. V. (1992b). Measuring the similarities between the lifetimes of adult Danish twins born between 1881–1930. *J. Am. Statist. Assoc.* **87**, 17–24.
- HOUGAARD, P., MYGLEGAARD, P. & BORCH-JOHNSEN, K. (1994). Heterogeneity models of disease susceptibility, with an application to diabetic nephropathy. *Biometrics* **50**, 1178–88.
- KEIDING, N., ANDERSEN, P. K. & KLEIN, J. P. (1997). The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statist. Med.* **16**, 215–24.
- KLEIN, J. P., MOESCHBERGER, M. L., LI, Y. H. & WANG, S. T. (1992). Estimating random effects in the Framingham heart study. In *Survival Analysis: State of the Art*, Ed. J. P. Klein & P. K. Goel, pp. 99–120, Dordrecht: Kluwer Academic Publishers.
- KORTRAM, R. A., LENSTRA, A. J., RIDDER, G. & VAN ROOIJ, A. C. M. (1995). Constructive identification of the mixed proportional hazards model. *Statist. Neer.* **49**, 269–81.

- LANCASTER, T. (1979). Econometric methods for the duration of unemployment. *Econometrica* **47**, 939–56.
- LANCASTER, T. (1990). *The Econometric Analysis of Transition Data*. Cambridge: Cambridge University Press.
- LILLARD, L. A. (1993). Simultaneous equations for hazards. *J. Econ.* **56**, 189–217.
- LILLARD, L. A. & PANIS, C. W. A. (1996). Marital status and mortality: the role of health. *Demography* **33**, 313–27.
- MEYER, B. D. (1990). Unemployment insurance and unemployment spells. *Econometrica* **58**, 757–82.
- MURPHY, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22**, 712–31.
- MURPHY, S. A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23**, 182–98.
- NEWMAN, J. L. & MCCULLOUGH, C. E. (1984). A hazard rate approach to the timing of births. *Econometrica* **52**, 939–61.
- NICKELL, S. J. (1979). Estimating the probability of leaving unemployment. *Econometrica* **47**, 1249–66.
- NIELSEN, G. G., GILL, R. D., ANDERSEN, P. K. & SØRENSEN, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist.* **19**, 25–43.
- OAKES, D. (1989). Bivariate survival models induced by frailties. *J. Am. Statist. Assoc.* **84**, 487–93.
- PETERSEN, J. H., ANDERSEN, P. K. & GILL, R. D. (1996). Variance components models for survival data. *Statist. Neer.* **50**, 193–211.
- TRUSSELL, J. & RICHARDS, T. (1985). Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer procedure. In *Sociological Methodology 1985*, Ed. N. Tuma, pp. 242–76, San Francisco: Jossey-Bass.
- VAN DEN BERG, G. J. (2001). Duration models: specification, identification, and multiple durations. In *Handbook of Econometrics*, Vol. V, Ed. J. J. Heckman & E. Leamer, pp. 3381–460, Amsterdam: North Holland.
- VAUPEL, J. W., MANTON, K. G. & STALLARD, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**, 439–54.
- YASHIN, A. I. & IACHINE, I. A. (1999). What difference does the dependence between durations make? Insights for population studies of aging. *Lifetime Data Anal.* **5**, 5–22.

[Received January 2004. Revised July 2006]