Large Duration Asymptotics in Multivariate Survival Models with Unobserved Heterogeneity

Yang Lu

Center of Research in Economics and Statistics (CREST, Paris) and SCOR

February 2015
INTRODUCTION
The importance of tail characteristics in survival analysis

- In Insurance/Finance, failure to capture the tail characteristics can lead to financial disasters.
- The longevity risk example:
  large duration (age) survivors = high costs for pension funds.
- This analysis has to also account for:
  1. Different individuals (male, female)
  2. Different causes of death (competing risks)
- Very often the estimation of asymptotic distribution is model-based.
- Moreover, this analysis should take into account the heterogeneity.
Effects of omitted individual characteristics (or unobserved heterogeneity, or frailty):

▸ For univariate survival models: “negative duration dependence”:
  Individuals with higher mortality risk leave faster, and the population becomes more and more homogenous.

▸ For multivariate survival models: the frailty also creates dependence between duration variables.

From a statistical point of view, the aim of introducing frailties is to correct these effects, and to increase the flexibility of the model.

From a risk management point of view: it induces more uncertainties, and avoids the underestimation of some risks.
I consider univariate and bivariate survival models, both with unobserved heterogeneity (frailty) and study large duration survivors.

I derive an appropriate “extreme competing risks" framework, characterize the limiting heterogeneity distribution among large duration survivors.

I discuss identification and large sample properties of the estimators under various observability assumptions.
UNIVARIATE SURVIVAL MODELS
The setting: the proportional hazard specification
Denote by $T$ the duration variable. Its conditional intensity is:

$$\theta(t|z, U, T > t) = \lambda(t, z)U,$$

where $z$ is the observable individual covariate, $U$ the unobserved frailty with distribution $F$, $\lambda$ the baseline hazard.

- For expository purpose we omit $z$.
- Denote $\Lambda$ the cumulative hazard.
- We are interested in the distribution of $T|T > t$, as well as in $\mathbb{P}[T > t] = \mathbb{E}[e^{-\Lambda(t)U}]$, for large $t$. 
A change of variable

\[
\mathbb{P}[T > t + \tau | T > t] = \frac{\int e^{-\Lambda(t+\tau)u}dF(u)}{\int e^{-\Lambda(t)x}dF(x)}
\]

\[
= \int e^{-[\Lambda(t+\tau) - \Lambda(t)]u} \frac{e^{-\Lambda(t)u}}{\int e^{-\Lambda(t)x}dF(x)} dF(u),
\]

where \(\int \frac{e^{-\Lambda(t)u}}{\int e^{-\Lambda(t)x}dF(x)} dF(u)\) is the distribution of frailty \(U\) for large duration \(T > t\), that is among the survivors at time \(t\).

This suggests the change of variable \(U^* = \Lambda(t)U\). We denote \(F_t\) the cumulative distribution of \(\Lambda(t)U | T > t\).
Let us introduce the time-change:

\[ X_t = \frac{\Lambda(T)}{\Lambda(t)} - 1, \]

when \( T > t \). This is both a change of time origin and an increasing nonlinear change of time unit. Then:

\[
\mathbb{P}[X_t > \frac{\Lambda(t + \tau)}{\Lambda(t)} - 1 | T > t] = \mathbb{P}[T > t + \tau | T > t] = \int e^{-\left[\frac{\Lambda(t+\tau)}{\Lambda(t)} - 1\right]} u^* \, dF_t(u^*),
\]

Let us first study the behavior of the standardized heterogeneity \( \Lambda(t)U | T > t \) for large \( t \).
Abbring and Van den Berg (2007) derive the relationship between the behavior at zero of frailty $U$ at time origin $t = 0$ and its distribution among large duration survivors.

**Definition 1**
The cdf of $U$ is regularly varying at zero ($RV_0$) if for all $a \in [0, 1]$,

$$
\lim_{a \to 0} \mathbb{P}[U < ax | U < a] = \lim_{a \to 0} \frac{F(ax)}{F(a)} = \nu(x).
$$

Then $\nu(x) = x^\rho$, where the regular variation index $\rho > 0$.

**Theorem 1**
$F$ is $RV_0(\rho)$ if and only if the distribution of

$$
\Lambda(t)U \mid T > t,
$$

converges to a gamma distribution $\gamma(\rho, 1)$, when $t \to \infty$. 
Under this regular variation assumption,

\[ \mathbb{P}[X_t > x | T > t] \rightarrow \int e^{-xu^*} dF_{\rho,1}(u^*) = \frac{1}{(1 + x)^\rho}. \]

**Proposition 1**

\( X_t = \frac{\Lambda(T)}{\Lambda(t)} - 1 | T > t \) converges to a Pareto distribution with survivor function \( S(x) = \frac{1}{(1+x)^\rho} \).

Let us also look at the marginal tail of \( T \).
Marginal tail properties of $T$

1. $h(t) := -\frac{d}{dt} \log \mathbb{P}[T > t] = \lambda(t) \mathbb{E}[U | T > t] \sim \rho \frac{\lambda(t)}{\Lambda(t)} = \rho \frac{d}{dt} \log \Lambda(t)$

2. $\mathbb{P}[T > t] = \frac{L(\Lambda(t))}{\Lambda^\rho(t)}$, where $L$ is a slowly varying at infinity:

$$\lim_{t \to \infty} \frac{L(at)}{L(t)} = 1 \text{ for all } a > 0.$$ 

For large $t$, $\mathbb{P}[T > t]$ is characterized by both the asymptotic behavior of $\Lambda$, and more importantly, by the index $\rho$ (up to a slowly varying function).
Example 1 (gamma frailty)

If $U \sim \gamma(\rho, c)$, we have:

$$
P[T > t + \tau | T > t] = \frac{1}{1 + c \frac{\Lambda(t + \tau) - \Lambda(t)}{1 + c\Lambda(t)}}^\rho,
$$

that is, $X_t = c \frac{\Lambda(T) - \Lambda(t)}{1 + c\Lambda(t)} | T > t$ follows a Pareto distribution, for any $t$. For large $t$, we get the result in Proposition 1, since:

$$
X_t \sim \frac{\Lambda(T)}{\Lambda(t)} - 1.
$$
Figure: Evolution of the distribution of $\Lambda(t)U|T > t$. The cdf of $U$ is $F(x) = 0.4\gamma_{1.5,2}(x) + 0.6\mathbb{1}_{x \geq 5}\gamma_{1.5,2}(x - 5)$. 
BIVARIATE SURVIVAL MODELS
The setting: The duration variables $T_1$ and $T_2$ are assumed independent given $(U, V)$ and

$$
\theta_1(t|U, V, T_1 > t) = \lambda_1(t)U,
\theta_2(t|U, V, T_2 > t) = \lambda_2(t)V.
$$

$U$ and $V$ are not necessarily independent: they control for the association between $T_1$ and $T_2$.

Case 1 competing risks:

- $T_1$ the potential time of death due to cause 1;
- $T_2$ the potential time of death due to cause 2;

in this case $T_1 > t, T_2 > t$ means “still alive", and

$$
\min[T_1, T_2], \mathbb{1}_{T_1 < T_2}
$$

is observed, but not $\max[T_1, T_2]$. 
Case 2 semi-competing risks:

- $T_1$ the potential time of entering into long-term care (LTC)
- $T_2$ the time of death;

in this case $T_1 > t, T_2 > t$ means “still alive and autonomous". The observable is $(T_1 \perp_{T_1<T_2} T_2)$.

Case 3 complete observation:

- $T_1$ the time of death of the husband;
- $T_2$ the time of death of the wife;

in this case $T_1 > t, T_2 > t$ means “both spouses still alive". The observable is $(T_1, T_2)$.

Question:

- What is, for large $t$, the joint distribution of $(T_1, T_2)|T_1 > t, T_2 > t$?
- What about the asymptotic behavior of $\mathbb{P}[T_1 > t, T_2 > t]$?
Condition to avoid degenerate large duration asymptotics
We want to avoid

\[ \mathbb{P}[T_1 < T_2 | T_1 > t, T_2 > t, U, V] \rightarrow 0, \text{ or } 1, \]

that is, for each survivor, the two risks should be of comparable importance, even for large \( t \).

- If \( \frac{\lambda_1(t)}{\lambda_2(t)} \rightarrow 0 \), then \( \mathbb{P}[T_1 < T_2 | T_1 > t, T_2 > t, U, V] \rightarrow 0 \), for large duration survivors: \( T_2 \) arrives always first.
- If \( \frac{\lambda_1(t)}{\lambda_2(t)} \rightarrow \infty \), \( T_1 \) arrives always first.
- If \( \frac{\lambda_1(t)}{\lambda_2(t)} \rightarrow \ell \), then \( \mathbb{P}[T_1 < T_2 | T_1 > t, T_2 > t, U, V] \rightarrow \frac{\ell U}{\ell U + V} \).

There is a converging probability of observing \( T_1 \) first. It depends on \((U, V)\).
This motivates the following

**Assumption 1**

\[
\lim_{t \to \infty} \frac{\lambda_1(t)}{\lambda_2(t)} = \ell > 0, \text{ which implies } \lim_{t \to \infty} \frac{\Lambda_1(t)}{\Lambda_2(t)} = \ell.
\]

Equivalently, \( \exists a_1, a_2 > 0, a_1/a_2 = \ell, \text{ and } \Lambda(t) > 0: \)

\[
\Lambda_1(t) \sim a_1 \Lambda(t), \quad \Lambda_2(t) \sim a_2 \Lambda(t),
\]

or

\[
a_2 \Lambda_1(t) - a_1 \Lambda_2(t) = o(\Lambda(t)).
\]

This is an asymptotic (deterministic) co-integration relationship.

Consider the time change, for \( T_1 > t, T_2 > t: \)

\[
(X_t, Y_t) = \left( \frac{\Lambda_1(T_1)}{\Lambda_1(t)} - 1, \frac{\Lambda_2(T_2)}{\Lambda_2(t)} - 1 \right) \sim \left( \frac{\Lambda(T_1)}{\Lambda(t)} - 1, \frac{\Lambda(T_2)}{\Lambda(t)} - 1 \right).
\]

Thus, asymptotically, \( \Lambda(\cdot) \) is the common nonlinear time change for \( T_1, T_2. \)
Can there still be truly competing risks at the macro level, for large $t$? In other words, we have to avoid

$$P[T_1 < T_2 | T_1 > t, T_2 > t] \to 0, \text{ or } 1.$$ 

Similarly, it suffices that the ratio of macro level intensities $h_1(t)/h_2(t)$ converges, where $h_j(t) := E[\theta_j(t, U) | T_1 > t, T_2 > t]$, $j = 1, 2$. We get:

$$\frac{h_1(t)}{h_2(t)} = \frac{\lambda_1(t)E[U | T_1 > t, T_2 > t]}{\lambda_2(t)E[V | T_1 > t, T_2 > t]} \sim \frac{E[\Lambda(t)U | T_1 > t, T_2 > t]}{E[\Lambda(t)V | T_1 > t, T_2 > t]}.$$ 

Thus the convergence of the distribution $(\Lambda(t)U, \Lambda(t)V) | T_1 > t, T_2 > t$ implies the deterministic co-integration of $h_1, h_2$, thus a fortiori ensures truly competing risks at the macro level.
As in the univariate case, we get an integral formula:

\[
\mathbb{P}[T_1 > t + \tau_1, T_2 > t + \tau_2 | T_1 > t, T_2 > t] = \int \int e^{-\left[\frac{\Lambda_1(t+\tau_1)}{\Lambda_1(t)} - 1\right]u^*} - \left[\frac{\Lambda_2(t+\tau_2)}{\Lambda_2(t)} - 1\right]v^* \; dF_t(u^*, v^*),
\]

where \( F_t \) is the cdf of \( (\Lambda_1(t)U, \Lambda_2(t)V) | T_1 > t, T_2 > t \).

If this cdf converges to \( F_\infty \), then the distribution of \( (X_t, Y_t) | T_1 > t, T_2 > t \) converges to \( (X, Y) \) such that:

\[
\mathbb{P}[X > x, Y > y] := \Sigma(x, y) = \int \int e^{-xu-yv} \; dF_\infty(u, v).
\]
Bivariate regular variation

Theorem 2
Under the co-integration assumption between \( \Lambda_1, \Lambda_2 \), the distribution of \((U, V)\) is regularly varying at zero (BRV), that is there exists \( \nu > 0 \) such that for all \( x, y \geq 0 \):

\[
\lim_{a \to 0} \frac{F(ax, ay)}{F(a, a)} = \lim_{a \to 0} \frac{\mathbb{P}[U < ax, V < ay]}{\mathbb{P}[U < a, V < a]} = \nu(x, y),
\]

if and only if the distribution of

\[
\Lambda(t)(U, V) \mid T_1 > t, T_2 > t
\]

converges to a non degenerate distribution:

\[
dF_\infty(u, v) = \frac{1}{c} e^{-a_1 u - a_2 v} \nu(du, dv),
\]

where \( c = \int \int e^{-a_1 u - a_2 v} \nu(du, dv) \) is a normalizing constant.
Properties of $\nu$.

- $\nu$ is a cdf for $x, y \in [0, 1]$.
- (Homogeneity): there exists $\rho > 0$ such that:

$$\nu(cx, cy) = c^\rho \nu(x, y), \quad \forall x, y \in [0, 1], \quad \forall c > 0.$$ 

Thus $\nu$ is the distribution function of a measure,
$\nu(x, y) = \nu([0, x] \times [0, y])$, with the integrability condition $\nu(1, 1) = 1$.

- Assume $\nu$ has a density $\mu$, then $\mu$ is also homogeneous:

$$\mu(cx, cy) = c^{\rho-2} \mu(x, y).$$

Let us denote by $\mathcal{F}_\rho$ the set of functions satisfying these three properties.

- Can all functions $\nu \in \mathcal{F}_\rho$ be written as the limit in Theorem 2 for some joint distribution of heterogeneity $(U, V)$?
- Is it a large family?
Do all $\nu \in \mathcal{F}_\rho$ have interpretations as limit?
Consider the distribution: $f_0(u, \nu) = \frac{1}{c_0} e^{-b(u, \nu)} \mu(u, \nu)$, where $\nu(du, dv) = \mu(u, \nu) dudv \in \mathcal{F}_\rho$ and $b$ is a function such that $\lim_{t \to \infty} b(u/t, v/t) = 0$.

- The pdf of $(U, V) | T_1 > t, T_2 > t$ is:

  \[
  f_t(u, \nu) = \frac{1}{c_t} e^{-\Lambda_1(t)u - \Lambda_2(t)v - b(u, \nu)} \mu(u, \nu).
  \]

- The pdf of $\Lambda(t)(U, V) | T_1 > t, T_2 > t$ converges to:

  \[
  f_\infty(u, \nu) = \frac{1}{c''} e^{-a_1u - a_2v} \mu(u, \nu),
  \]

  where $\Lambda_1(t) \sim a_1 \Lambda(t)$, $\Lambda_2(t) \sim a_2 \Lambda(t)$.

- Thus all continuous measures $\nu \in \mathcal{F}_\rho$ have interpretations as limit.

- Indeed all $BRV_0(\nu)$ functions admit this representation.
Is $\mathcal{F}_\rho$ a large family?

- Spherical representation: $\mu(x, y) = r^{\rho-2} s(\phi)$. Thus the continuous distributions of $\mathcal{F}_\rho$ are characterized by a scalar $\rho$ and a positive function $s$, with a joint integrability constraint.

- Conversely, any nonnegative homogeneous function of order $\rho - 2$ that is integrable on $[0, 1] \times [0, 1]$ defines a cdf via its primitive, up to a multiplier:

$$
\nu := \frac{1}{c'} \int_0^1 \int_0^1 \mu(u, v) dudv \in \mathcal{F}_\rho.
$$

- $\mathcal{F}_\rho$ is convex:

$$
\nu_1, \nu_2 \in \mathcal{F}_\rho \implies \omega \nu_1 + (1 - \omega) \nu_2 \in \mathcal{F}_\rho, \quad \forall \omega \in [0, 1].
$$
Properties of the limit heterogeneity distribution $f_\infty(u, v) = \frac{1}{c} e^{-a_1 u - a_2 v} \mu(u, v)$.

- It is $BRV_0$ with $b(u, v) = a_1 u + a_2 v$.
- $\mathbb{E}[e^{-x(a_1 U + a_2 V)}] = \frac{1}{(1+x)^\rho}$, thus $a_1 U + a_2 V$ follows a gamma distribution $\gamma(\rho, 1)$.
- If $\rho > 1$, then for $\delta > 0$, the following conditional density:

$$f_\rho(u \mid U = \delta V) \propto u^{\rho-2} e^{-(a_1 + a_2 \delta)u} \mu(1, \delta)$$

corresponds to a gamma distribution $\gamma(\rho - 1, a_1 + a_2 \delta)$.
- If $\mu(x, y) = x^\alpha y^\beta$, then $U$ and $V$ are independent gamma’s, and $(X_t, Y_t)$ are independent for large $t$. 
Figure: Iso-contour of the density $f(u, v) = \frac{1}{c} e^{-0.5u - 0.31v \frac{u^2 v^2}{u+v}}$. The marginals are not gamma but are $RV_0$. $U$ and $V$ are positively correlated: $\text{Corr}(U, V) = 0.06$. 
Figure: $f(x, y) = \frac{1}{c} e^{-0.4u-0.31v}(u^3 + v^3)$. Marginals are not gamma but $RV_0$. $U$ and $V$ are negatively correlated: $\text{Corr}(U, V) = -0.49$. 
Figure: Evolution of the density of $\Lambda(t)(U, V) | T_1 > t, T_2 > t$, where $\Lambda_1(t) = \Lambda_2(t) = \Lambda(t)$ and the initial distribution is $f_0(u, v) = \frac{1}{c} \left[ g_1(u, v) + g_2(u, v) \right]$, with $g_1(u, v) = e^{-u-v} \frac{u^2 v^2}{u+v}$ and $g_2(u, v) = g_1(u - 1, v - 1) 1_{u>1,v>1}$. 
Examples of bivariate regular varying distributions

- If $U, V$ follow marginal $\gamma(\alpha, 1)$, with an Archimedean copula $C(x, y) = \psi^{-1}(\psi(x) + \psi(y))$, such that
  \[
  \lim_{x \to 0} \frac{\psi(ax)}{\psi(x)} = a^{-\kappa},
  \]
  then $\nu(x, y) = \left(\frac{x^{-\kappa \alpha}}{2} + \frac{y^{-\kappa \alpha}}{2}\right)^{-\frac{1}{\kappa}}$, $\nu$ is differentiable, thus $f(u, \nu) = e^{-b_1 u - b_2 \nu} \mu(u, \nu)$ is $BRV_0$.

- If $U \in RV_0(\alpha)$, $V \in RV_0(\beta)$ and $(U, V)$ has a normal copula with correlation $r$, then $\nu(x, y) = x^{\alpha r + 1} y^{\beta r + 1}$. Thus we should avoid prior specification of $(U, V)$ since it may provide restrictive large duration dependence properties.

- If $\nu(x, y) = \int_0^\rho \Pi(\alpha) x^\alpha y^{\rho - \alpha} d\alpha$ with $\int_0^\rho \Pi(\alpha) d\alpha = 1$, where $\alpha \in (0, \rho)$ for integrability reason. Then $f(u, v) = e^{-b_1 u - b_2 v} \mu(u, v)$ is a gamma mixture.
IDENTIFICATION OF ASYMPTOTIC PARAMETERS
Is it possible to estimate the asymptotic parameters $\mu, a_1, a_2$ as well as those linked to the asymptotic behavior of $\Lambda$, from i.i.d. samples of $T_1, T_2$?

Recall that the distribution of \( \left( \frac{\Lambda(T_1)}{\Lambda(t)} - 1, \frac{\Lambda(T_2)}{\Lambda(t)} - 1 \right) \mid T_1 > t, T_2 > t \) converges in distribution to \((X, Y)\) such that:

\[
P[X > x, Y > y] = \int \int \frac{1}{c} e^{-(x+1)u-(y+1)v} \mu(u/a_1, v/a_2) dudv.
\]

**Scale normalization**

Since we can replace \((U, V, \Lambda_1, \Lambda_2)\) by \((a_1 U, a_2 V, \frac{1}{a_1} \Lambda_1, \frac{1}{a_2} \Lambda_2)\), without loss of generality, we assume $\ell = \lim_{t \to \infty} \frac{\Lambda_2(t)}{\Lambda_1(t)} = 1$ and $a_1 = a_2 = 1$. 
Identification of $\Lambda$.
We can prove: $\mathbb{P}[\min(T_1, T_2) > t] = \frac{L(\Lambda(t))}{\Lambda^\rho(t)}$, and as a consequence:

$$\log \Lambda(t) \sim -\frac{1}{\rho} \log S(t, t),$$  \hspace{1cm} (1)

thus $\Lambda$ is identified up to an unknown power parameter (as well as a slowly varying part).

**Proposition 2**

It is not possible to identify more: for any $\alpha \geq 1$, we can replace $\Lambda$ by $\Lambda^\alpha$ (and correspondingly change the distribution of $(U, V)$) without modifying the asymptotic distribution of $T_1, T_2$.

- This result is based on Bernstein’s theorem.
- For identification, we can fix $\rho = 1$.
- Since the asymptotic equivalence (1) depends only on $\min(T_1, T_2)$, $\Lambda$ is identified for all the three cases of observability.
Identification of $\mu$.

Case 1: $(T_1, T_2)$ observable

Informally,

$$S_t(t + \tau_1, t + \tau_2) := \mathbb{P}[T_1 > t + \tau_1, T_2 > t + \tau_2 | T_1 > t, T_2 > t]$$

$$\approx \frac{1}{c} \int \int \exp \left( - \frac{\Lambda(t + \tau_1)}{\Lambda(t)} u - \frac{\Lambda(t + \tau_2)}{\Lambda(t)} v \right) \mu(u, v) dudv$$

where $\frac{\Lambda(t+\tau_1)}{\Lambda(t)} \sim \frac{S(t,t)}{S(t+\tau_1, t+\tau_1)} := \frac{1}{S_t(\tau_1, \tau_1)}, \frac{\Lambda(t+\tau_2)}{\Lambda(t)} \sim \frac{1}{S_t(\tau_2, \tau_2)}$.

Thus $\mu$ is identified.

Formally, The conditional distribution

$$\left( 1/S_t(T_1, T_1) - 1, 1/S_t(T_2, T_2) - 1 \right) | T_1 > t, T_2 > t$$

converges also to that of $(X, Y)$ when $t$ goes to infinity.
Case 2: Semi-competing risks: only \((1_{T_1 < T_2} T_1, T_2)\) is observed

- We identify the corresponding truncated distribution of \((\tilde{x}_t, \tilde{y}_t) = \left(1/S_t(T_1, T_1) - 1, 1/S_t(T_2, T_2) - 1\right)\), thus we identify the limit distribution of \((X \mathbb{1}_{X < Y}, Y)\).

- It has a density with respect to \(m_{D_1} + m_{D_2}\), where \(m\) is the Lebesgue measure,

\[
D_1 = \{(x, y), 0 < x < y\}, \quad D_2 = \{(0, y), 0 < y\},
\]

and the density is:

\[
\begin{align*}
\sigma_{\mu}(x, y) &= \frac{1}{c} \int \int e^{-(1+x)u-(1+y)v} uv \mu(u, v) \, dudv \quad \text{on} \quad D_1, \\
\sigma_{\mu}(y) &= \frac{1}{c} \int \int e^{-(1+x)u-(1+y)v} v \mu(u, v) \, dv \quad \text{on} \quad D_2.
\end{align*}
\]

- Thus \(\mu\) is identified.
Case 3 (competing risks)

- Only \((\min(T_1, T_2), \mathbb{1}_{T_1 < T_2})\) is observed, thus only the distribution of \((\min(X, Y), \mathbb{1}_{X < Y})\) is identified.

- It is easily checked that \(\min(X, Y)\) follows \(\gamma(\rho, 1)\), and

\[
p = \mathbb{P}[X < Y | \min(X, Y) = m] = \frac{\int \int e^{-u-v} u \mu(u, v) dudv}{\int \int e^{-u-v}(u + v) \mu(u, v) dudv}.
\]

does not depend on \(m\). Indeed we have: \(\mathbb{1}_{\tilde{x}_t < \tilde{y}_t} = \mathbb{1}_{T_1 < T_2}\), and because of the “macro level” co-integration between \(h_1(t), h_2(t)\), the share of the two risks converges to a constant.

- \(p\) can be identified.

- The identification of \(\mu\) is only achieved by introducing covariates.
NON-PARAMETRIC ESTIMATION AND LARGE SAMPLE PROPERTY
Estimation of the structural parameters.
Let us define the Population-at-Risk:

$$\mathcal{P}_t := \{i \mid T_{1,i} > t, T_{2,i} > t\}.$$ 

There are two alternative estimation approaches:

- **Two-step approach**: First estimate $S_t(t + \tau, t + \tau)$ (for instance by the empirical survivor function).

- This estimator can then be used to compute "pseudo" samples from the distribution

  $$\left( \frac{1}{\hat{S}_t(T_1, T_1)} - 1, \frac{1}{\hat{S}_t(T_2, T_2)} - 1 \right)|T_1 > t, T_2 > t,$$

  which can be used to estimate $\mu$.

- This approach is simple, consistent, but likely inefficient.

- **Joint approach**: Estimate jointly $S_t(t + \tau, t + \tau)$ and $\mu$, but the implementation will be more complicated. We focus on the first approach.
(Two-stage) estimation of $\mu$: Case 1 & 2

- Compute $(\tilde{x}_t = 1/\hat{S}_t(T_1, T_1) - 1, \tilde{y}_t = 1/\hat{S}_t(T_2, T_2)) - 1$ for individuals in $\mathcal{P}_t$. We get a sample whose pdf is approximately $\sigma_\mu(x, y)$.

- The (non-parametric) MLE is (informally):

$$\hat{s}_n := \arg \max_s \sum_{i \in \mathcal{P}_t} \log \sigma_\mu(\tilde{x}_{it}, \tilde{y}_{it})$$

with constraint $\int_0^1 \int_0^1 \mu(u, v) dudv = 1$, and $\mu(u, v) = r^{\rho - 2} s(\omega)$, where $r = \sqrt{u^2 + v^2}$, $\tan \omega = v/u$.

- The estimation in Case 2 is similar.

$$\hat{s}_n := \arg \max_s \left[ \sum_{i \in \mathcal{P}_t, \tilde{x}_i < \tilde{y}_i} \log \sigma_\mu(\tilde{x}_{it}, \tilde{y}_{it}) + \sum_{i \in \mathcal{P}_t, \tilde{x}_i > \tilde{y}_i} \log \sigma_\mu(\tilde{y}_{it}) \right].$$
Joint Estimation
Principle: Estimate \((S, \mu)\) jointly by minimizing the criterion

\[
-\omega \sum_{i \in P_t} \log \sigma_\mu \left( \frac{1}{S(T_{1i})} - 1, \frac{1}{S(T_{2i})} - 1 \right) + (1 - \omega) d(S, S_t)
\]

where \(\omega\) is a weight and \(d(S, S_t)\) denotes an appropriate distance between the functional \(S\) and the theoretical (or empirical) distribution of \(T_1, T_2|T_1 > t, T_2 > t\).

- For instance, if we replace \(d(S, S_t)\) by \(d(S, \hat{S}_t)\), then \(S = \hat{S}_t\), which corresponds to the 2-stage estimator, minimizes the second term, but not necessarily the sum.

- For instance, denote by \(\hat{f}(h)\) a kernel-based pdf estimator of \(T_1, T_2|T_1 > t, T_2 > t\) and \(\hat{S}(h)\) its survivor function, and define \(d(\hat{S}(h), S_t)\) as the mean-square error of the kernel pdf estimator.

- We should define the optimal choice of the weight coefficient (further research).
CONCLUSION
Contributions

▶ A framework for “extreme multivariate survival variables" with observed/unobserved heterogeneity.
▶ A by-product: a new non-parametric family of bivariate distribution able to capture the dependence structure between different duration variables.

Perspectives

▶ The competing risks: modelling cohort mortality data disaggregated by causes.
▶ Covariate is essential both for identification and for capturing the longevity phenomenon.
Thanks for attention. Questions/comments welcome.
Abbring and Van den Berg (2003). The Identifiability of the Mixed Proportional Hazards Competing Risks Model, JRSS B.

Abbring and Van Den Berg (2007). The Unobserved Heterogeneity Distribution in Duration Analysis, Biometrika.


Ledford and Tawn (1997) Modelling Dependence Within Tail Regions, JRSS B.

Appendix
Non identifiability of $\kappa$.

**Theorem 3**

*Given a proportional hazard model $(\Lambda_1, \Lambda_2, U, V)$, and a constant $\alpha \geq 1$, we can construct an observationally equivalent model $(\Lambda_1^\alpha, \Lambda_2^\alpha, U_\alpha, V_\alpha)$ such that:*

$$\mathcal{L}(U_\alpha, V_\alpha)(x, y) = \mathcal{L}(U, V)(x^{1/\alpha}, y^{1/\alpha})$$

Moreover, if $(U, V)$ is $BRV_0(\nu)$, then $(U_\alpha, V_\alpha)$ is also $BRV_0(\nu^*)$ such that:

$$\int\int e^{-xu-uy} \nu^*(du, dv) = \int\int e^{-x^{1/\alpha}u-y^{1/\alpha}v} \nu(du, dv).$$

In particular, $\rho^* = \frac{\rho}{\alpha}$. 
Identification of MPH competing risks models

Theorem 4 (See Abbring and Van den Berg (2003))

In a MPH competing risks model, assume that there exists a continuum of covariate values z, and that \((\phi_1(z), \phi_2(z))\) covers a non empty open set of \((\mathbb{R}^+)^2\). If \(E[U] = E[V] = 1\), then \((\phi_1(z), \phi_2(z), \lambda_1, \lambda_2)\) as well as the distribution of \((U, V)\) are nonparametrically identified.

Comment: Key to the identification is the finite mean assumption of \(U, V\), which is violated by the counter example in the previous page.
Further implications
The family $f(u, v) = e^{-b_1 u - b_2 v} \mu(u, v)$ can be used to specify heterogeneity distribution. This is especially relevant in two situations:

1. “stationary" population. For instance, in finance, the risks of firms/banks does not depend much on the age of the firm, but rather on the current economic factors (e.g. GDP growth rates, interest rates).

2. Small sample size.
Specification of bivariate frailty

Under the bivariate regular varying assumption, we can approximate the initial distribution $(U, V)$ by a mixture of:

- A continuous component with cdf:

  $$f(u, v) = \mathbb{1}_{u < x_1, v < y_1} \mu(u, v),$$

  where the homogeneity parameter $\rho - 2$ can be fixed.

- A finite discrete mixture with masses $(x_i, y_i), i = 2, \ldots, n$, such that $x_i \geq x_1, y_i \geq y_1$. 

Unconditional dependence structure.

Case 1: Extremal dependence

\[
\lim_{t \to \infty} \frac{P[T_1 > t]}{P[T_1 > t, T_2 > t]} > 0
\]

and similarly for \( T_2 \).

Case 2: Hidden regular variation

1. \( P[T_1 > t] > P[T_1 > t, T_2 > t] > P[T_1 > t]P[T_2 > t] \): positively associated.

2. \( P[T_1 > t]P[T_2 > t] > P[T_1 > t, T_2 > t] \): negatively associated.
Marginal of $\min(T_1, T_2)$.

$$\theta(t|T > t, U, V) = \lambda_1(t)U + \lambda_2(t)V \sim \lambda(t)(a_1U + a_2V).$$

Thus $T$ (approximately) satisfies a univariate proportional hazard model. We get similar results as in the univariate case:

**Proposition 3**

Under the bivariate regular variation assumption,

- $a_1 U + a_2 V \in RV_0(\rho)$, for all $a_1, a_2 > 0$.
- (We know already:) $a_1 U + a_2 V \mid T_1 > t, T_2 > t$ converges to $\gamma(\rho, 1)$.
- $\mathbb{P}[T_1 > t, T_2 > t] = \frac{L(t)}{\Lambda^\rho(t)}$, where $L$ is slowly varying.
- $h(t) := -\frac{d}{dt} \log \mathbb{P}[T_1 > t, T_2 > t] \sim \rho \frac{\lambda(t)}{\Lambda(t)}$.

Again the probability $\mathbb{P}[T_1 > t, T_2 > t]$ is characterized by the function $\rho$ as well as the asymptotic behavior of $\Lambda$. 
Interpretations

- For fixed $\Lambda_1, \Lambda_2$ and $\nu$, distributions of $(U, V)$ that are $BRV(\nu)$ are "asymptotically equivalent", among large duration survivors, $T_1 > t, T_2 > t$.

- But $U$ and $V$ can admit different marginal behavior at 0, leading to different marginal tail properties for $T_1, T_2$.

- Thus the set of $BRV(\nu)$ distribution is extremely large. However if we are solely interested in the joint tail, we can approximate the cdf around zero of such a $(U, V)$ by a "natural" representative, namely by $f(u, v) = e^{-a_1 u - a_2 v} \mu(u, v)$.

- There is little loss of generality since we need not (and perhaps more importantly, should not) specify the marginal distributions.
Consider the MPH model:

\[
\theta_1(t|U, V, z) = \lambda_1(t)\phi_1(z)U \\
\theta_2(t|U, V, z) = \lambda_2(t)\phi_2(z)V,
\]

then \( \min(X, Y) | z \) follows \( \gamma(\rho, 1) \) and:

\[
P[X < Y|z] = \frac{\iint \phi_1(z)ue^{-\phi_1(z)u-\phi_2(z)v}\mu(u, v)dudv}{\iint (\phi_1(z)u + \phi_2(z)v)e^{-\phi_1(z)u-\phi_2(z)v}\mu(u, v)dudv}.
\]

\( (\phi_1, \phi_2, \mu) \) still unidentified from asymptotic samples.

\( (\phi_1(z), \phi_2(z)) \) is identified from non asymptotic samples (see Appendix). It can then be used to identify \( \mu \) (inversion of analytic functions).

Estimators/large sample theory trickier since the number of covariate values should also go to infinity.

Suggests that further constraints will help.