Discussion Paper No. 960

THE CORRELATION OF DURATIONS IN MULTIVARIATE HAZARD RATE MODELS

by

Gerard J. van den Berg*

Ton Steerneman**

October 15, 1991

We would like to thank Maarten Lindenboom for many valuable discussions on the topic of this paper and related issues. Also thanks to Guido Imbens and Jan de Leeuw for some helpful comments. Financial support from the Royal Dutch Academy of Sciences is gratefully acknowledged.

---

*Dept. of Economics, Groningen University, P.O. Box 800, 9700 AV Groningen, The Netherlands, and Dept. of Economics, Northwestern University.

**Dept. of Economics, Groningen University.

Abstract

The empirical analysis of multiple durations using multivariate mixed proportional hazard rate models is widespread. In such models, the duration variables are dependent if their unobserved determinants are dependent on each other. In this paper it is shown that these models restrict the magnitude of the correlation of the duration variables. For example, if the baseline hazards are constant, then this correlation necessarily lies between -1/3 and 1/2. Similar results hold for more general models. The usefulness for empirical analysis is twofold. First, the results can be used to assess the ability of the model to describe certain phenomena, relative to the models that impose less restrictions on the values the correlation can attain. Secondly, they suggest that, in parametric analyses, it is important to take a family of heterogeneity distributions that is flexible in the sense that it does not restrict the values the correlation can attain either further. We show that some frequently used parametric families are much more restrictive than others.

Key words: Multivariate hazard rate models, competing risks, proportional hazards, correlation of nonnegative random variables.

JEL classification: C41, C50.

1. Introduction

This paper examines the correlation of durations in multivariate hazard rate models. In particular, sharp bounds are derived in the general case as well as in cases in which the distribution of unobserved heterogeneity in the model is assumed to belong to a specific parametric family.

By now, the empirical analysis of multiple durations (or, equivalently, failure times) is widespread. Most of the empirical studies at least partly focus on whether the durations are independent or not, or, more specifically, on the degree in which they are dependent on each other. Generally, the analysis is based on the use of multivariate mixed proportional hazard models (for a list references, see the next section). In such models, the hazard rates associated with the different durations have mixed proportional hazard specifications. The durations are allowed to be dependent by way of stochastically related unobserved covariates (i.e. unobserved heterogeneity).

The correlation of the duration variables is generally considered to be a parameter of interest, since it is informative on the strength of the linear relationship between these variables. It is a commonly used measure that is readily understood. In this paper we examine the range of values that the correlation of the duration variables can attain in the model context outlined above. We derive sharp upper and lower bounds in the general case. Also, we derive bounds in cases in which the joint distribution of the unobserved heterogeneity terms is specified to belong to a parametric family of distributions. We pay particular attention to the families of distributions that are generally chosen in the empirical studies on bivariate hazard rate models, such as normal distributions and discrete distributions.

The results are of practical interest. First of all, they may indicate a limitation of ability of the general model to describe the distribution of dependent durations, relative to other possible models. Secondly, they can be used to compare the flexibility of different parametric families of distributions as representations of the distribution of the unobserved heterogeneity terms. It turns out that some of the frequently used families are more restrictive than other popular families.

The outline of the paper is as follows. In Section 2 we introduce the multivariate hazard rate model and derive sharp bounds for the correlation of the durations. We also derive such bounds in cases in which the heterogeneity distribution is assumed to belong to popular parametric families. This will be used to compare these popular families. Section 3 generalizes the results of

Section 2 to more general models. In particular, we consider (generalizations of) models that can be expressed as multivariate log-linear regression models. Section 4 concludes. In the appendix to this paper we derive some general inequalities in terms of the first few joint moments of nonnegative random variables.

2. The correlation of two duration variables with mixture distributions

2.1. *Dependent mixtures of exponentials*

The use of hazard rate models for the empirical analysis of durations or failure times is widespread. Hazard rate models specify the rate at which failure occurs at a time t conditional on survival up to t, as a function of t and, possibly, of explanatory variables. Depending on the particular application, the hazard rate is sometimes called the failure rate (in reliability analysis), or the exit rate (e.g. in the analysis of the duration spent in a labour market state).

A particularly popular hazard rate model is the Mixed Proportional Hazard (MPH) Model. In this model, the hazard rate is written as a multiplicative function of observed explanatory variables x, the elapsed duration t, and a random term v representing unobserved explanatory variables. Specifically,

$$(2.1) \qquad \theta(t|v,x) = \lambda(t).\theta_0(x).v \qquad \text{with } \theta_0(x) = \exp(x'\beta)$$

is the hazard rate of $t|x,v$. It is related to the distribution function F of $t|x,v$ by $\theta(t|x,v) = -\,\mathrm{dlog}(1-F(t|x,v))/\mathrm{d}x$. The distribution of $t|x$ follows by integration of $F(t|x,v)$ w.r.t. the density of v. (See e.g. Kalbfleisch & Prentice (1980) and Lancaster (1990) for extensive surveys).

Most survival studies are concerned with univariate failure times only. However, the literature on the simultaneous analysis of multiple dependent failure times is vastly growing. In general, these studies focus on modelling and estimating the dependence of the failure times. Multiple failure times may be consecutive (e.g. the duration of unemployment and the subsequent job duration) but they may also occur jointly (e.g. the duration of unemployment and the duration of participation in a panel survey). If in the latter case all failure times start at the same moment, and only the realization of the smallest failure time is observed, then the model is called a competing risks model.

Again, a particularly popular model in empirical analysis is the model in which the failure times follow an MPH model (see e.g. Flinn & Heckman (1982), Flinn & Heckman (1983), Newman & McCulloch (1984), Heckman, Hotz & Walker (1985), Heckman & Walker (1987), Butler, Anderson & Burkhauser (1989), Ham & Lalonde (1990), Visser (1990), Heckman & Walker (1990), and Van den Berg, Lindeboom & Ridder (1991)). The failure times are allowed to be dependent by way of (stochastically) related unobserved covariates. Consider two failure times, t_1 and t_2. We assume that all individual differences in the joint distribution of t_1 and t_2 can be characterized by variables x, v_1, and v_2. The fundamental assumption is that, conditionally given x, v_1, and v_2, the variables t_1 and t_2 are independent. Further, the distribution of $t_i|x,v_1,v_2$ equals the distribution of $t_i|x,v_i$ $(i=1,2)$, so to explain individual differences in t_i, the variable v_j $(j\neq i)$ does not give information that is not available in v_i. The hazard rates of t_1 and t_2 conditionally given the explanatory variables x, v_1 and v_2 can be expressed as

$$\theta_1(t_1|v_1,x) = \lambda_1(t_1).\theta_{01}(x).v_1 \qquad \text{with } \theta_{01}(x) = \exp(x'\beta_1) \tag{2.2}$$
$$\theta_2(t_2|v_2,x) = \lambda_2(t_2).\theta_{02}(x).v_2 \qquad \text{with } \theta_{02}(x) = \exp(x'\beta_2)$$

In practice, the empirical analysis is conditional on the observed explanatory variables, which constitute x. For ease of exposition, we take x to be independent of v_1 and v_2. Also, we assume that x is not time-varying. (It should be noted that in a few of the papers listed above x is allowed to depend on time. Also, if x is a linear function of t_i then the part of x depending on t_i can be thought of as being part of $\lambda_i(t_i)$.) The variables $t_1|x$ and $t_2|x$ can only be dependent if the unobserved explanatory variables v_1 and v_2 associated with t_1 and t_2 are dependent. Heckman & Honoré (1989) and Honoré (1991) prove the nonparametric identifiability of such multivariate duration models.

A relatively simple and frequently used version of the model is obtained by imposing $\lambda_1(t_1)=\lambda_2(t_2)=1$. In this version, $t_i|x,v_i$ has an exponential distribution, and $t_i|x$ is distributed as a mixture of exponentials. In empirical applications, a parametric family of distributions for v_1,v_2 is chosen to close the model. Let f be a generic symbol for a density. The likelihood function is based on $f(t_1,t_2|x)$. This density depends on the unknown β_1, β_2 and $f(v_1,v_2)$ in the following way.

(2.3) $$f(t_1,t_2|x) = \int_{v_1}\int_{v_2} f(t_1|x,v_1)\,.f(t_2|x,v_2)\,.f(v_1,v_2)\;dv_1\;dv_2$$

in which $$f(t_i|x,v_i) = v_i.\exp(x'\beta_i)\;.e^{-v_i.\exp(x'\beta_i).t_i} \qquad i=1,2$$

Of course, v_1 and v_2 can be discrete as well.

2.2. *The values that the correlation of the duration variables can attain*

The correlation of $t_1|x$ and $t_2|x$ is generally considered to be a parameter of interest, since it is informative on the strength of the linear relationship between these variables. Because $E(t_i|x,v_i)$ and $V(t_i|x,v_i)$ are proportional to $1/v_i$ and $1/v_i^2$, respectively, there holds that $COV(t_1,t_2|x)$ and $V(t_i|x)$ (and therefore $CORR(t_1,t_2|x)$) can be expressed in terms of moments of $1/v_1$ and $1/v_2$. For example,

$$\begin{aligned} COV(t_1,t_2|x) &= E(E(t_1.t_2|x,v_1,v_2)) - E(E(t_1|x,v_1)).E(E(t_2|x,v_2)) \\ &= E(\frac{1}{\theta_{01}(x).\theta_{02}(x).v_1.v_2}) - E(\frac{1}{\theta_{01}(x).v_1}).E(\frac{1}{\theta_{02}(x).v_2}) \end{aligned}$$

As a result, it is easily obtained that

(2.4) $$CORR(t_1,t_2|x) = \frac{COV(1/v_1,1/v_2)}{\left[\left[V(1/v_1)+E(1/v_1^2)\right]\;.\;\left[V(1/v_2)+E(1/v_2^2)\right]\right]^{1/2}}$$

Note that this expression does not depend on x. We assume that $P(0<v_1<\infty,0<v_2<\infty)=1$ and that $1/v_1$ and $1/v_2$ have finite variances. For convenience, we will restrict attention to cases in which the variances of $1/v_1$ and $1/v_2$ are positive (if they are not, then $CORR(t_1,t_2|x)=0$).

Because $V(1/v_i)+E(1/v_i^2) = 2.V(1/v_i)+E^2(1/v_i)$, it is clear that $CORR(t_1,t_2|x) < 1/2$. This result is also derived in Cantor & Knapp (1985) for the specific case in which $v_1 \equiv v_2$ with probability one, and in Lindeboom & Van den Berg (1991).

Denote $CORR(1/v_1,1/v_2)$ as ρ_{12}, $V(1/v_i)$ as σ_i^2, and $E(1/v_i)$ as μ_i. We can then rewrite equation (2.4) as

(2.5) $$CORR(t_1,t_2|x) = \frac{1}{3}\cdot\frac{\rho_{12}}{\left[\frac{2}{3}+\frac{1}{3}\cdot\frac{\mu_1^2}{\sigma_1^2}\right]^{1/2}\cdot\left[\frac{2}{3}+\frac{1}{3}\cdot\frac{\mu_2^2}{\sigma_2^2}\right]^{1/2}}$$

From Lemma 2 in Appendix 1 it follows directly that $\mathrm{CORR}(t_1,t_2|x) > -1/3$ (take $p=2/3$). Note that the lower bound does not depend on the distribution of $1/v_1,1/v_2$. Also note that we apply Lemma 2 to the distribution of $1/v_1,1/v_2$ rather than the distribution of $t_1,t_2|x$. (In the notation of the appendix, we have taken $X=1/v_1$ and $Y=1/v_2$.) So, in a way, we have derived an inequality for $\mathrm{CORR}(t_1,t_2|x)$ by using an inequality for $\mathrm{CORR}(1/v_1,1/v_2)$. It can easily be shown that by applying Lemma 2 directly to $\mathrm{CORR}(t_1,t_2|x)$ one gets an inequality that is inferior to the one above.

From the results in Appendix 2 it follows that the lower bound $-1/3$ for $\mathrm{CORR}(t_1,t_2|x)$ is strict in the sense that it can be approximated arbitrarily well by taking certain distributions of v_1,v_2. It can also be shown that the upper bound $1/2$ is strict (see Lindeboom & Van den Berg (1991)). Consequently, the bounds in (2.6) are the best possible nonparametric bounds; bounds closer to zero would exclude certain distributions of v_1,v_2.

Summarizing, we have the following result.

Proposition 1

In the bivariate duration model set up above, with the assumptions mentioned above, there holds that

$$(2.6) \qquad -\frac{1}{3} < \mathrm{CORR}(t_1,t_2|x) < \frac{1}{2}$$

regardless of the values of β_1 *and* β_2, *and regardless of the shape of* $f(v_1,v_2)$. *The inequalities in* (2.6) *are sharp in the sense that for every* $\varepsilon_1>0$ ($\varepsilon_2>0$) *there are distributions* $f(v_1,v_2)$ *such that* $\mathrm{CORR}(t_1,t_2|x)<(-1/3+\varepsilon_1)$ ($\mathrm{CORR}(t_1,t_2|x)>(1/2-\varepsilon_2)$).

Of course, (2.6) does not depend on the way $\theta_{01}(x)$ and $\theta_{02}(x)$ are parameterized.

The result stated in (2.6) is of practical interest. First of all, it indicates a limitation of the model set up here as a general model for dependent failure times. For example, the two failure times cannot be almost equal in a probabilistic sense. Whether this makes the model inappropriate depends of course on the particular application at hand. It should be noted that most of the alternative models proposed for bivariate failure times restrict $\mathrm{CORR}(t_1,t_2|x)$ to be nonnegative (see e.g. Hougaard (1987)), which of course may also be too restrictive.

A second practical aspect of the result concerns its use as a guide for the parameterization of the model in empirical analysis. Not every family of distributions for $f(v_1,v_2)$ contains elements for which the resulting $\mathrm{CORR}(t_1,t_2|x)$ is close to the bounds $-1/3$ or $1/2$. If there is no compelling reason for $f(v_1,v_2)$ to be in a particular parametric family of distributions, then for reasons of flexibility one should choose a family that allows $\mathrm{CORR}(t_1,t_2|x)$ to attain every value in $(-1/3,1/2)$. In the next subsection, we will examine the values $\mathrm{CORR}(t_1,t_2|x)$ can attain for specific popular families of distributions of $f(v_1,v_2)$.

2.3. *Specific families of distributions*

Empirical analyses of multivariate failure times often assume that one can write

$$(2.7) \qquad v_1 = \exp(\alpha.z) \qquad\qquad v_2 = \exp(\beta.z)$$

for some random variable z and nonzero parameters α and β. Flinn & Heckman (1982) (among others) take a normal distribution for z, while Ham & Lalonde (1990) (among others) take a discrete distribution with two points of support for z. Note that if z has a normal distribution, then the t_i are lognormal mixtures of exponentials; also note that in both cases $V(1/v_i)$ exists. In Lindeboom & Van den Berg (1991) it is shown that if the distribution of z belongs to the family of discrete distributions with two points of support, then $\mathrm{CORR}(t_1,t_2|x)$ can attain every value in $(-1/3,1/2)$. Of course, this therefore also holds for the families of discrete distributions with a larger number of points of support.

On the other hand, if the distribution of z belongs to the class of normal distributions, then, by elaborating on equation (3.4), one can show that $\mathrm{CORR}(t_1,t_2|x)$ can only attain values in $[-1/(3+2\sqrt{2}),1/2)$ (the lower bound equals about -0.17). This is a remarkable result. In social sciences, the family of normal distributions is by far the most popular choice as a model for distributions of unobservables. This is partly due to the high level of generality and flexibility assigned to this family. Our result shows that, in the present model setting, choosing the family of normal distributions for z results in a model that is more restrictive than models in which certain other families of distributions are chosen for z.

Equation (2.7) implies that there is an exact linear relationship between $\log v_1$ and $\log v_2$. We now turn to cases in which v_1,v_2 has a genuine bivariate distribution. Van den Berg, Lindeboom & Ridder (1991) estimate a model in which v_1,v_2 has a genuine bivariate discrete distribution with two positive points of support for each v_i. From the previous paragraph it immediately follows that in this model $\text{CORR}(t_1,t_2|x)$ can attain every value in $(-1/3,1/2)$.

On the other hand, it can be shown that if $v_1=\exp(\alpha.z_1)$ and $v_2=\exp(\beta.z_2)$, with $z\equiv(z_1,z_2)$ having a bivariate normal distribution and $\alpha\neq0$ and $\beta\neq0$ (see e.g. Butler, Anderson & Burkhauser (1986)), then $\text{CORR}(t_1,t_2|x)$ can only attain values in $(-1/(3+2\sqrt{2}),1/2)$. If we allow the correlation of z_1 and z_2 to be equal to -1 (in which case the bivariate distribution is degenerate) then the lower bound $-1/(3+2\sqrt{2})$ can be attained. This again illustrates that, in the present context, models in which the family of heterogeneity distributions is based on normal distributions are relatively restrictive.

Butler, Anderson & Burkhauser (1989) estimate a model in which v_1,v_2 has a bivariate discrete distribution with points of support that are fixed in advanced. This means that the only parameters of $f(v_1,v_2)$ to be estimated are the probabilities associated with these points of support. It may be interesting to examine to what extent fixing the points of support narrows the range of values $\text{CORR}(t_1,t_2|x)$ can attain. Suppose that both v_1 and v_2 have two points of support, denoted by v_1^u and v_1^l, and by v_2^u and v_2^l, respectively. All points of support are positive and finite. As a normalization, we take $v_1^l\leq v_1^u$ and $v_2^l\leq v_2^u$. In Lindeboom & Van den Berg (1991), it is shown that to obtain a $\text{CORR}(t_1,t_2|x)$ close to its limiting values $-1/3$ or $1/2$, it is necessary that $v_1^l/v_1^u\downarrow0$ and $v_2^l/v_2^u\downarrow0$. This implies that for fixed values of v_1^u, v_1^l, v_2^u, and v_2^l, $\text{CORR}(t_1,t_2|x)$ can not attain all values in $(-1/3,1/2)$. In Appendix 3 we prove the following result:

Suppose that v_1^l/v_1^u equals v_2^l/v_2^u. Denote this ratio by c. There holds that

$$(2.8) \qquad \frac{-(1-c)^2}{3c^2-2c+3} \leq \text{CORR}(t_1,t_2|x) \leq \frac{(1-c)^2}{2c^2+2}$$

for all values of the other parameters in the model. The inequalities in (2.8) are sharp in the sense that they can be attained, for every $c\in(0,1]$.

Figure 1 below shows the upper and the lower bound of $\text{CORR}(t_1,t_2|x)$ as functions of c on $(0,1]$. If $c\downarrow0$ then these bounds go to $1/2$ and $-1/3$, respectively, which are the bounds of Proposition 1. If $c\uparrow1$ then the dispersion of v_1 and v_2 vanishes and, as a result, $\text{CORR}(t_1,t_2|x)$ goes to zero. The upper bound is decreasing in c while the lower bound is increasing in c.

In fact, the bounds can be expressed in terms of each other in a simple way. Let U and L denote the upper and lower bound; then $-L = U/(U+1)$. So, if $U = 1/k$, with $k>2$, then $L = -1/(k+1)$. As a consequence, L is always closer to zero than U is.

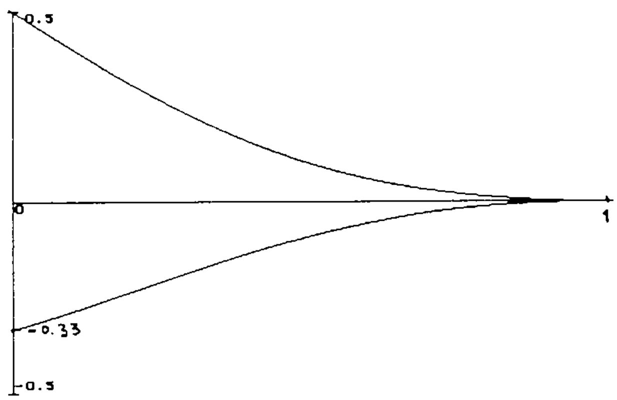

Figure 1. Sharp bounds for the correlation of the duration variables when unobserved heterogeneity has a bivariate discrete distribution, as a function of the ratios of the points of support.

These results show that it is restrictive to fix the points of support of the discrete distribution of v_1,v_2. In particular, when the points for v_1 or v_2 are relatively close to each other then the range of values $\text{CORR}(t_1,t_2|x)$ can attain is very small.

From Appendix 3 it follows that in the discrete distributions for v_1,v_2 discussed here, the bounds for $\text{CORR}(t_1,t_2|x)$ are attained in cases in which $\log v_1$ and $\log v_2$ are linear functions of each other. This means that the result above is also true in case equation (2.7) holds with z having a discrete distribution with two points of support.

The results in the present subsection so far may suggest that taking a particular family of nondegenerate bivariate distributions for v_1,v_2 results in the same amount of flexibility as taking the corresponding family of marginal distributions for v_1 and assuming that there is some non-stochastic relationship between v_1 and v_2. In other words, it seems that one-dimensional random variation in the unobserved heterogeneity terms is suffucient to get the maximum amount of flexibility. This is true if flexibility is defined in terms of the range of values $\text{CORR}(t_1,t_2|x)$ can attain. However, if flexibility is also related to certain properties of the marginal distributions of $t_1|x$ and $t_2|x$, then non-degenerate bivariate distributions for v_1,v_2 may be regarded as more attractive (see Lindeboom & Van den Berg (1991)). Of course, in practice one may compare different specifications by using statistical tests (see e.g. Heckman & Walker (1987)).

We conclude this subsection by examining some other specific families of distributions for v_1,v_2. Whitmore & Lee (1991) examine models in which $v_1 \equiv v_2 \equiv v$ with probability one and in which the distribution of v belongs to the family of Inverse Gaussian distributions for which all positive and negative moments exist. They show that $\mathrm{CORR}(t_1,t_2|x)$ can only attain values in $(0,2/5)$. Hougaard (1986) proposes models in which $v_1 \equiv v_2 \equiv v$ with probability one and in which the distribution of v belongs to the family of Positive Stable distributions. In that case $\mathrm{V}(1/v)$ exists (though $\mathrm{E}(v)$ does not). Using results from that paper it can be shown that in such models $\mathrm{CORR}(t_1,t_2|x)$ can only attain values in $[0,1/2)$. Finally, if $1/v_1,1/v_2$ has a Filon–Isserk bivariate Beta distribution (see e.g. Mardia (1970): $\mathrm{V}(1/v_i)$ exists) then $\mathrm{CORR}(t_1,t_2|x)$ can only attain values in $(-1/3,0]$.

3. Generalizations

In this section, we will present bounds for the correlation of two endogenous variables in models that are more general than the model examined so far.

As a starting point, consider the model set up in Section 2. In empirical analysis, the assumption that $\lambda_i(t_i)=1$ (see equation (2.2)) may be restrictive, since it rules out genuine duration dependence of the hazard rates. A popular generalization specifies that $\lambda_i(t_i) = \alpha_i . t_i^{\alpha_i-1}$ for i=1,2. This means that $t_1|x,v_1$ and $t_2|x,v_2$ both have a Weibull distribution, with duration dependence parameters $\alpha_1>0$ and $\alpha_2>0$, respectively. In such a model $\mathrm{E}(t_i|x,v_i)$ and $\mathrm{V}(t_i|x,v_i)$ are proportional to v_i^{-1/α_i} and v_i^{-2/α_i}, respectively:

$$\mathrm{E}(t_i|x,v_i) = \Gamma(1+1/\alpha_i).\theta_{0i}(x)^{-1/\alpha_i}.v_i^{-1/\alpha_i}$$

$$\mathrm{V}(t_i|x,v_i) = \left[\Gamma(1+2/\alpha_i) - (\Gamma(1+1/\alpha_i))^2\right].\theta_{0i}(x)^{-2/\alpha_i}.v_i^{-2/\alpha_i}$$

Consequently, $\mathrm{COV}(t_1,t_2|x)$, $\mathrm{V}(t_i|x)$, and $\mathrm{CORR}(t_1,t_2|x)$ can be expressed in terms of moments of v_1^{-1/α_1} and v_2^{-1/α_2}. We assume that $P(0<v_1<\infty,0<v_2<\infty)=1$ and that v_1^{-1/α_1} and v_2^{-1/α_2} have positive and finite variances. Denote

$\mathrm{CORR}(v_1^{-1/\alpha_1}, v_2^{-1/\alpha_2})$ as ρ_{12}, $V(v_i^{-1/\alpha_i})$ as σ_i^2, and $E(v_i^{-1/\alpha_i})$ as μ_i. Further, define a_i as

$$a_i = \frac{\Gamma(1+2/\alpha_i)}{(\Gamma(1+1/\alpha_i))^2}$$

From the expression for $V(t_i|x,v_i)$ it follows that $a_i>1$ for every $\alpha_i>0$. Now we obtain

$$(3.1) \qquad \mathrm{CORR}(t_1,t_2|x) = \rho_{12} \cdot \left[\left[a_1 + (a_1-1).\frac{\mu_1^2}{\sigma_1^2}\right] \cdot \left[a_2 + (a_2-1).\frac{\mu_2^2}{\sigma_2^2}\right]\right]^{-1/2}$$

$$(3.2) \qquad = \frac{1}{\sqrt{2a_1-1}\,.\sqrt{2a_2-1}} \cdot \rho_{12} \cdot \left[\left[\frac{a_1}{2a_1-1} + \frac{a_1-1}{2a_1-1} \cdot \frac{\mu_1^2}{\sigma_1^2}\right] \cdot \left[\frac{a_2}{2a_2-1} + \frac{a_2-1}{2a_2-1} \cdot \frac{\mu_2^2}{\sigma_2^2}\right]\right]^{-1/2}$$

It follows directly from (3.1) that $\mathrm{CORR}(t_1,t_2|x)$ does not exceed $(a_1.a_2)^{-1/2}$. Lemma 3 (see Appendix 4) gives the following lower bound for the second term on the r.h.s. of (3.2) (take $p_i=a_i/(2a_i-1)$, so $p_i \in (1/2,1)$),

$$- \frac{\sqrt{(2a_1-1).(2a_2-1)}}{\sqrt{a_1 . a_2} + \sqrt{(a_1-1).(a_2-1)}}$$

Note that we apply Lemma 3 to the distribution of $v_1^{-1/\alpha_1}, v_2^{-1/\alpha_2}$. The expression above does not depend on this distribution, so it holds for all possible distributions of $v_1^{-1/\alpha_1}, v_2^{-1/\alpha_2}$ (i.e. for all distributions of v_1, v_2). From this expression, a lower bound for $\mathrm{CORR}(t_1,t_2|x)$ follows that holds for all possible distributions of v_1, v_2.

From Appendix 4 it follows that the lower bound for $\mathrm{CORR}(t_1,t_2|x)$ is again strict in the sense that it can be approximated arbitrarily well by taking certain distributions of v_1, v_2. The upper bound is also strict. (This can be shown easily by taking a bivariate distribution with two points of support for v_1, v_2 and letting certain parameters of it go to the boundary of the parameter space.) Consequently, the bounds in (3.3) are the best bounds possible in the sense that bounds closer to zero would exclude certain distributions of v_1, v_2.

In sum, we have the following result.

Proposition 2

In the bivariate duration model of this section, with the assumptions mentioned above, there holds that

$$(3.3) \qquad \frac{-1}{\sqrt{a_1 . a_2} + \sqrt{(a_1-1).(a_2-1)}} < \mathrm{CORR}(t_1,t_2|x) < \frac{1}{\sqrt{a_1 . a_2}}$$

regardless of the values of $\theta_{01}(x)$ and $\theta_{02}(x)$, and regardless of the shape of $f(v_1,v_2)$. The inequalities in (3.3) are sharp in the sense that for every $\varepsilon_1>0$ ($\varepsilon_2>0$) there are distributions $f(v_1,v_2)$ such that CORR$(t_1,t_2|x)$ is smaller than the lower bound plus ε_1 (larger than the upper bound minus ε_2)).

For $\alpha_1=\alpha_2=1$ ($a_1=a_2=2$) the results specialize to those obtained in the previous section. In fact, for $\alpha_1=\alpha_2$ (so $a_1=a_2$ and $p_1=p_2$), (3.3) could have been derived using Lemma 2 in Appendix 1.

In general, the lower bound is closer to zero than the upper bound. Further, if, for some $i\in\{1,2\}$, $\alpha_i\downarrow 0$, then $a_i\to\infty$ and CORR$(t_1,t_2|x)\downarrow 0$. On the other hand, if $\alpha_1\to\infty$ and $\alpha_2\to\infty$ then $a_1\downarrow 1$ and $a_2\downarrow 1$, and the interval of values that CORR$(t_1,t_2|x)$ can attain goes to $(-1,1)$. So, by varying α and the distribution of v_1,v_2, all values in $(-1,1)$ can be attained. (It should be noted that Lee & Gross (1989) derive results from which it follows that all values in (0,1) can be attained.) Still, for specific values of the duration dependence parameters α_1 and α_2, the value of CORR$(t_1,t_2|x)$ is restricted. In other words, the range of values that CORR$(t_1,t_2|x)$ can attain depends on the duration dependence of the hazard rate of $t_1|x,v_1$ and the duration dependence of the hazard rate of $t_2|x,v_2$.

For another generalization, consider the following bivariate nonlinear regression model.

$$(3.4) \qquad \begin{aligned} t_1 &= \zeta_{01}(x).u_1.\varepsilon_1 \\ t_2 &= \zeta_{02}(x).u_2.\varepsilon_2 \end{aligned}$$

We assume that x,u_i,ε_i are independent of each other for every i, and that ε_1 and ε_2 are independent. Further, we assume that u_1,u_2 are positive with probability one, but we allow ε_i and $\zeta_{0i}(x)$ (and therefore t_i) to be negative. Note that if ε_i and $\zeta_{0i}(x)$ are always positive, and if $\zeta_{0i}(x)$ equals $\exp(x'\beta_i)$, then the model can be written as a bivariate log-linear regression model,

$$(3.5) \qquad \begin{aligned} \log t_1 &= x'\beta_1 + \log u_1 + \log \varepsilon_1 \\ \log t_2 &= x'\beta_2 + \log u_2 + \log \varepsilon_2 \end{aligned}$$

Let c_i denote the square of the coefficient of variation of ε_i, i.e., $c_i = V(\varepsilon_i)/E^2(\varepsilon_i)$, if it exists. One can show along the lines of the first part of this section that

$$(3.6) \qquad \frac{-1}{\sqrt{c_1 . c_2} + \sqrt{(c_1+1).(c_2+1)}} < \mathrm{CORR}(t_1,t_2|x) < \frac{1}{\sqrt{(c_1+1).(c_2+1)}}$$

for every distribution of u_1,u_2, provided that this correlation exists. Consequently, the range of values that $\mathrm{CORR}(t_1,t_2|x)$ can attain depends on the coefficients of variation of ε_1 and ε_2 (or the coefficients of variation of $t_1|x,u_1$ and $t_2|x,u_2$).

In fact, the model examined in the first part of this section is a special case of the nonlinear regression model (3.4) (take $c_i=a_i-1$, take ε_i such that $\varepsilon_i^{\alpha_i}$ has an exponential distribution with parameter one, and take $u_i=v_i^{-1/\alpha_i}$ and $\zeta_{0i}(x)=\theta_{0i}(x)^{-1/\alpha_i}$).

4. Conclusion

In this paper we have examined the correlation of duration variables in the context of multivariate mixed proportional hazard models. We showed that such models restrict the range of values that this correlation can attain. If the baseline hazards are constant (so the durations have marginal distributions that are mixtures of exponentials), then the correlation necessarily lies between −1/3 and 1/2. This holds regardless of the actual values of the parameters and functions in the model, and regardless of the values of the observed explanatory variables. Moreover, these bounds are sharp in the sense that, for either one of these bounds, there are distributions for the unobserved heterogeneity terms for which the correlation of the durations is arbitrarily close to it.

We also derived bounds for this correlation in case the distribution of the unobserved heterogeneity terms belongs to a specific family of distributions. If it is assumed that these heterogeneity terms follow a (multivariate) discrete distribution, then all values between −1/3 and 1/2 can be attained. However, if the unobserved heterogeneity terms are restricted to have a (multivariate) normal distribution, then the range of attainable values is smaller. Similarly, if discrete distributions with a priori fixed values of the points of support are taken, then not all values between −1/3 and 1/2 can be attained. All of these families of distributions have been used to parametrically model unobserved heterogeneity in empirical analyses of multiple durations. Our results suggest that, in terms of flexibility, using a discrete distribution with unspecified points of support is to be preferred over using one of the other families.

Most of the qualitative results carry over to more general models. Moreover, if the baseline hazards are not constant, then the upper and lower bound for the correlation of the duration variables depend on the duration dependence of the hazard rates.

Appendix

A.1. Some moment inequalities for two random variables with positive means

In this appendix we present and discuss some lemmas that will be used in the main text. The lemmas provide moment inequalities for two random variables X and Y. These inequalities involve the correlation coefficient ρ_{xy} of X and Y and the means and variances μ_x, μ_y, σ_x^2 and σ_y^2 of X and Y, respectively, and are derived under the assumption that μ_x, μ_y and E(XY) are positive. Obviously, sufficient for the latter is that $P(X>0,Y>0)=1$. To ensure that ρ_{xy} and the first and second moments of X and Y exist, we will assume throughout the appendix that the variances σ_x^2 and σ_y^2 of X and Y are positive and finite.

Lemma 1

Let X,Y have a bivariate distribution with positive finite variances. If E(XY)>0 then $\sigma_{xy} > -\mu_x.\mu_y$. Consequently,

(A.1) $$\rho_{xy} > -\frac{\mu_x}{\sigma_x}\cdot\frac{\mu_y}{\sigma_y}$$

Proof. $\sigma_{xy} \equiv E(XY)-\mu_x.\mu_y > -\mu_x.\mu_y$.

This simple result formalizes and extends the idea that for nonnegative random variables having a negative linear relation, the variances can not be arbitrarily large relative to the means. As a very simple example, suppose that $X\sim LN(\mu,\log 3)$ and $Y\sim LN(\mu,\log 3)$, so log X and log Y have the same normal distribution with unspecified mean μ and variance equal to log 3. Then it follows from Lemma 1 that ρ_{xy} necessarily exceeds $-1/2$, regardless the joint distribution of X,Y.

Lemma 2

Let X,Y have a bivariate distribution with positive finite variances. If $\mu_x>0$, $\mu_y>0$ and E(XY)>0, then

(A.2) $$\sigma_{xy} > -\sqrt{p.\sigma_x^2 + (1-p).\mu_x^2}\,.\sqrt{p.\sigma_y^2 + (1-p).\mu_y^2} \qquad \text{for all } p\in[0,1)$$

Consequently,

(A.3) $$\rho_{xy} > -\sqrt{p + (1-p).\frac{\mu_x^2}{\sigma_x^2}}\ .\sqrt{p + (1-p).\frac{\mu_y^2}{\sigma_y^2}} \qquad \text{for all } p\in[0,1)$$

Proof. The result trivially holds if $\rho_{xy}\geq 0$. Suppose $\rho_{xy}<0$. Let $\eta_i=\mu_i/\sigma_i$. From Lemma 1 we know that $\rho_{xy} > -\eta_x\eta_y$. This is equivalent to stating that the following matrix M is positive definite.

(A.4) $$M = \begin{bmatrix} \eta_x^2 & \rho_{xy} \\ \rho_{xy} & \eta_y^2 \end{bmatrix}$$

since the determinant of M equals $(\eta_x\eta_y-\rho_{xy}).(\eta_x\eta_y+\rho_{xy})$. and $\eta_x>0$, $\eta_y>0$ and $\rho_{xy}<0$. There also holds that $\rho_{xy} \geq -1$. This is equivalent to stating that the following matrix S is positive semi-definite.

(A.5) $$S = \begin{bmatrix} 1 & \rho_{xy} \\ \rho_{xy} & 1 \end{bmatrix}$$

This implies that for every scalar $p\in[0,1)$ the matrix $pS+(1-p)M$ is positive definite. Therefore, its determinant is positive.

$$(p + (1-p).\eta_x^2)\ .(p + (1-p).\eta_y^2) - \rho_{xy}^2 > 0$$

and the lemma follows.

The bound for ρ_{xy} in (A.3) can be interpreted as a (by p) weighted average of the two bounds $-\eta_x\eta_y$ and -1 for ρ_{xy} that follow from $\rho_{xy} > -\eta_x\eta_y$ and $\rho_{xy} \geq -1$, respectively. For $p=0$, (A.3) reduces to $\rho_{xy} > -\eta_x\eta_y$ while for $p\uparrow 1$ the bound in (A.3) goes to -1. It can be shown that for values of p between 0 and 1, the bound in (A.3) is smaller than $max(-\eta_x\eta_y,-1)$ (unless $\eta_x=\eta_y=1$). In general, therefore, $max(-\eta_x\eta_y,-1)$ provides a sharper lower bound for ρ_{xy} than the bound in (A.3). This means that for $p\in(0,1)$, equation (A.3) is not relevant in case a bound for ρ_{xy} in terms of μ_x, μ_y, σ_x^2 and σ_y^2 is needed that has to be as sharp as possible. Rather, it may be relevant for cases in which the value of p is given; that is, it may be relevant in cases in which bounds are needed for ρ_{xy} divided by the r.h.s. of (A.3) with given p. As will be shown below, such bounds are the sharpest possible in the sense that there are distributions for X,Y for which ρ_{xy} divided by the r.h.s. of (A.3) with given p is almost equal to 1.

A.2. Cases in which the inequalities almost hold with an equality sign

We will now examine whether there are distributions of X,Y satisfying the assumptions of Lemma 2 for which the moment inequalities derived almost hold with an equality sign. Clearly, to find a distribution for which equation (A.3) holds with an equality sign, we have to relax one or more of the assumptions made. Suppose we replace the assumption that $E(XY)>0$ by the assumption that $E(XY)\geq 0$. This implies that equation (A.1) may hold with an equality sign, or, equivalently, that the matrix M defined in equation (A.4) may be singular with rank one.

From the proof of Lemma 2 it follows that equation (A.3) holds with an equality sign if and only if $pS+(1-p)M$ is singular. Consider cases in which $p\in(0,1)$. Under the sustained assumptions, necessary and sufficient conditions for $pS+(1-p)M$ to be singular are (i) S and M are singular and (ii) S and M have the same set of eigenvectors corresponding to their zero eigenvalues.

Consider condition (i). The matrix S is singular if and only if $\rho_{xy}=-1$. This means that there has to be a negative linear relationship between X and Y. If $\rho_{xy}\leq 0$, then the matrix M is singular if and only if $E(XY)=0$. It is straightforward to construct distributions for which $\mu_x>0$, $\mu_y>0$, $\sigma_x>0$, $\sigma_y>0$, $E(XY)=0$, and $\rho_{xy}=-1$ (take e.g. $P(X=Y=1) = P(X=-1/2,Y=2) = 1/2$). If we restrict the attention to distributions for which $P(X\geq 0,Y\geq 0)=1$, then $E(XY)=0$ is equivalent to stating that all probability mass is concentrated on the x-axis and on the y-axis. Consequently, in that case S and M are singular if and only if the distribution of X,Y belongs to the following family of discrete distributions with two points of support:

$$\text{(A.6)} \qquad \begin{aligned} P(X=c_1,Y=0) &= \pi \qquad && c_1>0,\ 0<\pi<1 \\ P(X=0,Y=c_2) &= 1-\pi && c_2>0 \end{aligned}$$

However, such distributions do not necessarily satisfy the second condition for $pS+(1-p)M$ to be singular. One can show that in the family described by (A.6), the second condition holds if and only if $\pi=1/2$. As a result, the class of distributions satisfying $P(X\geq 0,Y\geq 0)=1$, $\sigma_x>0$, and $\sigma_y>0$ for which equation (A.3) with $p\in(0,1)$ holds with an equality sign, is described by (A.6) with $\pi=1/2$. Note that p itself does not enter the description of this class of distributions.

Until now we have considered cases in which $p\in(0,1)$. If $p=0$ then equation

(A.3) holds with an equality sign if and only if M is singular. If $\rho_{xy} \leq 0$, then M is singular if and only if E(XY)=0.

Summarizing, if we allow for E(XY)=0, then for every $p \in [0,1)$ there are distributions of X,Y such that equation (A.3) holds with an equality sign. For reasons of continuity, this implies that for every $p \in [0,1)$ there are distributions with E(XY)>0 for which equation (A.3) 'almost' holds with an equality sign, in the sense that ρ_{xy} is arbitrarily close to its bound on the r.h.s. of (A.3). In other words, for every $p \in [0,1)$ there are distributions with E(XY)>0 for which ρ_{xy} divided by the r.h.s. of equation (A.3) is arbitrarily close to 1. Consider for example the following family of distributions.

$$\text{(A.7)} \qquad \begin{aligned} P(X=c_1, Y=\varepsilon_1) &= 1/2 \qquad & c_1 > \varepsilon_2 > 0 \\ P(X=\varepsilon_2, Y=c_2) &= 1/2 \qquad & c_2 > \varepsilon_1 > 0 \end{aligned}$$

Such distributions satisfy the assumptions of Lemma 2. Note that $\rho_{xy}=-1$ for all $\varepsilon_1 \in (0,c_2)$ and all $\varepsilon_2 \in (0,c_1)$. The bound for ρ_{xy} in equation (A.3) is a continuous function of $\varepsilon_1,\varepsilon_2$ on $(0,c_2)\times(0,c_1)$. Let $\varepsilon_1 \downarrow 0$ and $\varepsilon_2 \downarrow 0$. In the limit, the distribution of X,Y becomes the distribution in (A.6) with $\pi=1/2$, and the bound for ρ_{xy} goes to -1. The limiting distribution does not satisfy all assumptions in Lemma 2. However, because of the continuity in ε_1 and ε_2, the value of bound for ρ_{xy} can be made arbitrarily close to -1 by taking ε_1 and ε_2 sufficiently small.

A.3. Discrete distributions with fixed points of support for v_1 and v_2

Suppose that v_1,v_2 has a bivariate discrete distribution with points of support v_1^l and v_1^u for v_1, and v_2^l and v_2^u for v_2, with $0<v_i^l \leq v_i^u < \infty$ for $i=1,2$. The probabilities associated with different points of support of v_1 and v_2 are defined in the following way:

$$\pi_1 = P(v_1=v_1^u,\ v_2=v_2^u) \qquad \pi_3 = P(v_1=v_1^u,\ v_2=v_2^l)$$
$$\pi_2 = P(v_1=v_1^l,\ v_2=v_2^u) \qquad \pi_4 = P(v_1=v_1^l,\ v_2=v_2^l)$$

Of course, $0 \leq \pi_1,\pi_2,\pi_3,\pi_4 \leq 1$ and $\pi_1+\pi_2+\pi_3+\pi_4=1$. We now assume that $v_1^l/v_1^u = v_2^l/v_2^u$ and we define c as $c = v_1^l/v_1^u$. It follows that $0<c\leq 1$. If c=1 then CORR$(t_1,t_2|x)=0$. If $0<c<1$ then, by elaborating on equation (2.4), we obtain (see also Lindeboom & Van den Berg (1991))

$$\text{(A.8)} \qquad \mathrm{CORR}(t_1,t_2|x) = (\pi_1.\pi_4 - \pi_2.\pi_3) \Big/ \Bigg[\Big[(\pi_1+\pi_2).(\pi_3+\pi_4) + \frac{c^2.(\pi_1+\pi_2)+(\pi_3+\pi_4)}{(1-c)^2}\Big] . \Big[(\pi_1+\pi_3).(\pi_2+\pi_4) + \frac{c^2.(\pi_1+\pi_3)+(\pi_2+\pi_4)}{(1-c)^2}\Big] \Bigg]^{1/2}$$

We now have to find values of π_1,π_2,π_3,π_4 for which this expression attains its maximum and minimum, given c. Fisrt, substitute $\pi_4=1-\pi_1-\pi_2-\pi_3$ into (A.8) and rewrite this equation using the following one-to-one mapping of π_1, π_2, and π_3 on γ_1, γ_2, and γ_3: $\gamma_1=\pi_1$, $\gamma_2=\pi_1+\pi_2$, and $\gamma_3=\pi_1+\pi_3$. As a result,

$$\text{(A.9)} \qquad \mathrm{CORR}(t_1,t_2|x) = (\gamma_1 - \gamma_2.\gamma_3) . \left[-\gamma_2^2 - \gamma_2.\frac{2c}{1-c} + \frac{1}{(1-c)^2}\right]^{-1/2} . \left[-\gamma_3^2 - \gamma_3.\frac{2c}{1-c} + \frac{1}{(1-c)^2}\right]^{-1/2}$$

Because of the definitions of γ_2 and γ_3, there holds that the derivative of $\mathrm{CORR}(t_1,t_2|x)$ w.r.t. π_2 (π_3) equals the derivative w.r.t. γ_2 (γ_3). It is straightforward but tedious to show that the derivative w.r.t. γ_2 has the same sign as the following expression.

$$\text{(A.10)} \qquad c^2.(\gamma_1\gamma_2 - \gamma_1 - \gamma_2\gamma_3) + c.(-2\gamma_1\gamma_2 + \gamma_2\gamma_3 + \gamma_1) + (\gamma_1\gamma_2 - \gamma_3)$$

Consider the equation that follows by equating this expression to zero. The discriminant of this quadratic equation in c can be written as

$$(\gamma_1 + 3\gamma_2\gamma_3).(\gamma_1 - \gamma_3) + \gamma_3.(\gamma_2\gamma_3 + 3\gamma_1).(\gamma_2 - 1)$$

Now note that $\gamma_1-\gamma_3\le 0$ and $\gamma_2-1\le 0$, so the discriminant is smaller than or equal to zero. Therefore, either (A.10) cannot be positive, or it cannot be negative, for all possible values of c, γ_1, γ_2, and γ_3. The value of (A.10) as $c\uparrow 1$ equals $-\gamma_3$. As a result, the derivative of $\mathrm{CORR}(t_1,t_2|x)$ w.r.t. γ_2 is always smaller than or equal to zero. Because of the symmetry of the r.h.s. of (A.9) as a function of γ_2 and γ_3, this implies that the derivative of $\mathrm{CORR}(t_1,t_2|x)$ w.r.t. γ_3 is smaller than or equal to zero as well. Consequently, the derivatives of $\mathrm{CORR}(t_1,t_2|x)$ w.r.t. π_2 and π_3 are smaller than or equal to zero. This means that the maximum of $\mathrm{CORR}(t_1,t_2|x)$ is attained at $\pi_2=\pi_3=0$ and the minimum at $\pi_1+\pi_2+\pi_3=1$.

Consider the hyperplane $\pi_1+\pi_2+\pi_3=1$. After substituting $\pi_1+\pi_2+\pi_3=1$ into (A.8) and differentiating w.r.t. π_2 (π_3), it is easy to show that, in this hyperplane, the derivative of $\mathrm{CORR}(t_1,t_2|x)$ w.r.t. π_2 (π_3) is smaller than or equal to zero. Consequently, the minimum of $\mathrm{CORR}(t_1,t_2|x)$ is attained at $\pi_2+\pi_3=1$.

Now consider the line $\pi_2+\pi_3=1$. Let us substitute $\pi_3=1-\pi_2$ into (A.8) and differentiate the resulting expression w.r.t. π_2. It is tedious to show that the derivative has the same sign as

$$-2.(1-c)^2.(c^2-c+1).\pi_2^3 \;+\; (1-c)^2.(3c^2+c+3).\pi_2^2 \;+$$

$$+\;(-c^4+c^3+2c^2+c-1).\pi_2 \;-\; c^2$$

This expression can be rewritten as follows

$$\text{(A.11)}\qquad 2\,.(\pi_2-\tfrac{1}{2})\,.\,\left[\,(1-c)^2.(c^2-c+1).\pi_2^2 + (1-c)^2.(c^2+c+1).\pi_2 + c^2\,\right]$$

The quadratic equation that results from equating the expression in square brackets in (A.11) to zero has two real roots. However, it is easily shown that one of them is negative while the other exceeds one, for each value of c in (0,1). Therefore, $\pi_2=1/2$ is the only value of π_2 in (0,1) for which (A.11) equals zero. Again, it is easily shown that (A.11) is negative for $\pi_2\in[0,1/2)$ and positive for $\pi_2\in(1/2,1]$, for each value of c in (0,1). As a result, $\mathrm{CORR}(t_1,t_2|x)$ attains its minimum at $\pi_2=\pi_3=1/2$, for every $c\in(0,1)$. By substituting $\pi_2=\pi_3=1/2$ into (A.8) we obtain the lower bound of $\mathrm{CORR}(t_1,t_2|x)$ as a function of c on (0,1).

Now let us turn to the upper bound for $\mathrm{CORR}(t_1,t_2|x)$. It is necessary that $\pi_4=1-\pi_1$. If we substitute this into equation (A.8), and differentiate the resulting expression w.r.t. π_1, we obtain an expression that has the same sign as

$$\text{(A.12)}\qquad (1-c)^2.\pi_1^2 - 2.\pi_1 + 1$$

It is easy to show that $\pi_1 = 1/(1+c)$ is the only value of π_1 in [0,1] for which (A.12) is zero. Moreover, since (A.12) is negative for $\pi_1\in[0,1/(1+c))$ and positive for $\pi_1\in(1/(1+c),1]$, it follows that $\mathrm{CORR}(t_1,t_2|x)$ attains its maximum at $\pi_1 = 1-\pi_4 = 1/(1+c)$, for every $c\in(0,1)$. By substituting $\pi_1 = 1-\pi_4 = 1/(1+c)$ into (A.8) we obtain the upper bound of $\mathrm{CORR}(t_1,t_2|x)$ as a function of

c on $(0,1)$. Note that the bounds in equation (2.8) in the main text also capture the case $c=1$.

A.4. Generalizations

This appendix contains a generalization of the results derived in the first two appendices above.

In equation (A.3), the same constant p appears under both roots. The following lemma generalizes Lemma 2 by providing a lower bound for ρ_{xy} in which the constants under these roots may differ.

Lemma 3

Let X,Y have a bivariate distribution with positive finite variances. If $\mu_x>0$, $\mu_y>0$ and $E(XY)>0$, then

$$\text{(A.13)} \qquad \left[\sqrt{p_1 p_2}+\sqrt{(1-p_1)(1-p_2)}\right].\sigma_{xy} > -\sqrt{p_1.\sigma_x^2+(1-p_1).\mu_x^2}\;.\sqrt{p_2.\sigma_y^2+(1-p_2).\mu_y^2}$$

for all $p_1\in[0,1)$ and $p_2\in[0,1)$

Consequently,

$$\text{(A.14)} \qquad \rho_{xy} > -\frac{1}{\sqrt{p_1 p_2}+\sqrt{(1-p_1)(1-p_2)}}\cdot\sqrt{p_1+(1-p_1).\frac{\mu_x^2}{\sigma_x^2}}\cdot\sqrt{p_2+(1-p_2).\frac{\mu_y^2}{\sigma_y^2}}$$

for all $p_1\in[0,1)$ and $p_2\in[0,1)$

Proof. We use the same notation as in the proof of Lemma 2. In addition, let the matrices Q and R be defined as follows,

$$\text{(A.15)} \qquad Q = \begin{bmatrix} p_1^{1/2} & 0 \\ 0 & p_2^{1/2} \end{bmatrix} \qquad\qquad R = \begin{bmatrix} (1-p_1)^{1/2} & 0 \\ 0 & (1-p_2)^{1/2} \end{bmatrix}$$

Then, for every $p_1,p_2\in[0,1]$, the matrix QSQ is positive semi-definite. Also, for every $p_1,p_2\in[0,1)$, the matrix RMR is positive definite. As a result, for every $p_1,p_2\in[0,1)$, the matrix $QSQ + RMR$ is positive definite. Thus, its determinant is positive, and the lemma directly follows.

By taking $p_1=p_2$, Lemma 2 is obtained as a special case of Lemma 3. It is clear from the proof above that Lemma 3 can be generalized by allowing the elements of the matrices Q and R to be unrelated to each other. For example, one can easily derive an equation analogous to (A.14) in which the terms $(1-p_i)$ are replaced by constants $q_i \in (0,1]$.

Using the same argument as in Appendix 2, it can be shown that for every $p_1,p_2 \in [0,1)$ there are distributions of X,Y for which the bound for ρ_{xy} in Lemma 3 almost holds with an equality sign. The class of distributions satisfying $P(X \geq 0, Y \geq 0)=1$, $\sigma_x>0$ and $\sigma_y>0$, for which equation (A.14) with $p_1,p_2 \in (0,1)$ holds with an equality sign, is as follows.

(A.16) $\quad P(X=c_1, Y=0) = \pi \qquad c_1>0$

$\qquad\quad\; P(X=0, Y=c_2) = 1-\pi \qquad c_2>0$

$$\text{with } \pi = \left[1 + \sqrt{\frac{p_1(1-p_2)}{p_2(1-p_1)}} \right]^{-1}$$

Note that if $p_1=p_2$ then $\pi=1/2$, which is in accordance to what we found in Appendix 2.

References

Butler, J.S., K.H. Anderson and R.V. Burkhauser (1986), Testing the relationship between work and health, *Economics Letters* 20, 383–386.

Butler, J.S., K.H. Anderson and R.V. Burkhauser (1989), Work and health after retirement, *Review of Economics and Statistics* 71, 46–53.

Cantor, A.B. and R.C. Knapp (1985), A test of the equality of survival distributions based on paired observations from conditionally independent exponential distributions, *IEEE Transactions on Reliability* 34, 342–346.

Flinn, C.J. and J.J. Heckman (1982), Models for the analysis of labor force dynamics, in: R. Basmann and G. Rhodes, eds., *Advances in econometrics* Vol. 1 (JAI Press, Greenwich).

Flinn, C.J. and J.J. Heckman (1983), Are unemployment and out of the labor force behaviorally distinct labor force states?, *Journal of Labor Economics* 1, 28–42.

Ham, J.C. and R. Lalonde (1990), Estimating the effect of training on the incidence and duration of unemployment, Mimeo, University of Pittsburgh.

Heckman, J.J., and B.E. Honoré (1989), The identifiability of the competing risks model, *Biometrika* 76, 325–330.

Heckman, J.J., V.J. Hotz and J.R. Walker (1985), New evidence on the timing and spacing of births, *American Economic Review* 75, 179–184.

Heckman, J.J. and J.R. Walker (1987), Using goodness of fit and other criteria to choose among competing duration models: a case study of the Hutterite data, in: C. Clogg, ed., *Sociological Methodology* 1987 (American Sociological Association, Washington).

Heckman, J.J. and J.R. Walker (1990), The relationship between wages and income and the timing and spacing of births, *Econometrica* 58, 1411–1441.

Honoré, B. (1991), Identification results for duration models with multiple spells, Research Memorandum, Northwestern University.

Hougaard, P. (1986), A class of multivariate failure time distributions, *Biometrika* 73, 671–678.

Hougaard, P. (1987), Modelling multivariate survival, *Scandinavian Journal of Statistics* 14, 291–304.

Kalbfleisch, J.D. and R.L. Prentice (1980), *The statistical analysis of failure time data* (Wiley, New York).

Lancaster, T. (1990), *The econometric analysis of transition data* (Cambridge University Press, Cambridge).

Lee, M.T. and A.J. Gross (1989), Properties of conditionally independent generalized gamma distributions, *Probability in the Engineering and Informational Sciences* 3, 289–297.

Lindeboom. M. and G.J. van den Berg (1991), Heterogeneity in bivariate duration models: the importance of the mixing distribution, Research Memorandum, Leiden University.

Mardia, K.V. (1970), *Families of bivariate distributions* (Griffin, London).

Newman, J.L. and C.E. McCulloch (1984), A hazard rate approach to the timing of births, *Econometrica* 52, 939–962.

Van den Berg, G.J., M. Lindeboom and G. Ridder (1991), Attrition in longitudinal panel data, and the empirical analysis of dynamic labour market behaviour, Research Memorandum, Groningen University.

Visser, M. (1990), Multiple duration spells with unobserved heterogeneity, Research Memorandum, ENSAE CREST Paris.

Whitmore, G.A. and M.T. Lee (1991), A multivariate survival distribution generated by an inverse gaussian mixture of exponentials, *Technometrics* 33, 39–50.