competing risks model

A competing risks model is a model for multiple durations that start at the same point in time for a given subject, where the subject is observed until the first duration is completed and one also observes which of the multiple durations is completed first.

The term ‘competing risks’ originates in the interpretation that a subject faces different risks $i$ of leaving the state it is in, each risk giving rise to its own exit destination, which can also be denoted by $i$. One may then define random variables $T_i$ describing the duration until risk $i$ is materialized. Only the smallest of all these durations $Y := \min_i T_i$ and the corresponding actual exit destination, which can be expressed as $Z := \arg\min_i T_i$, are observed. The other durations are censored in the sense that all is known is that their realizations exceed $Y$. Often those other durations are latent or counterfactual, for example if $T_i$ denotes the time until death due to cause $i$.

In economics, the most common application concerns individual unemployment durations. One may envisage two durations for each individual: one until a transition into employment occurs, and one until a transition into non-participation occurs. We observe only one transition, namely, the one that occurs first. Other applications include the duration of treatments, where the exit destinations are relapse and recovery, and the duration of marriage, where one risk is divorce and the other is death of one of the spouses. More generally, the duration until an event of interest may be right-censored due to the occurrence of another event, or due to the data sampling design. The duration until the censoring is then one of the variables $T_i$.

Sometimes one is interested only in the distribution of $Y$. For example, an unemployment insurance (UI) agency may be concerned only about the expenses on UI and not in the exit destinations of recipients. In such cases one may employ standard statistical duration analysis for empirical inference with register data on the duration of UI receipt. However, in studies on individual behaviour one is typically interested in one or more of the marginal distributions of the $T_i$. If these variables are known to be independent, then again one may employ standard duration analysis for each of the $T_i$ separately, treating the other variables $T_j (j \neq i)$ as independent right-censoring variables. But often it is not clear whether the $T_i$ are independent. Indeed, economic theory often predicts that they are dependent, in particular if they can be affected by the individual’s behaviour and individuals are heterogeneous. It may even be sensible from the individual’s point of view to use their privately observed exogenous exit rates into destinations $j$ as inputs for the optimal strategy affecting the exit rate into destination $i (i \neq j)$ (see, for example, van den Berg, 1990). Erroneously assuming independence leads to incorrect inference, and in fact the issue of whether the durations $T_i$ are related is often an important question in its own right.
Unfortunately, the joint distribution of all $T_i$ is not identified from the joint distribution of $Y, Z$, a result that goes back to Cox (1959). In particular, given any specific joint distribution, there is a joint distribution with independent durations $T_i$ that generates the same distribution of the observable variables $Y, Z$. In other words, without additional structure, each dependent competing risks model is observationally equivalent to an independent competing risks model. The marginal distributions in the latter can be very different from the true distributions.

Of course, some properties of the joint distribution are identified. To describe these it is useful to introduce the concept of the hazard rate of a continuous duration variable, say $W$. Formally, the hazard rate at time $t$ is $\theta(t) := \lim_{dt \to 0} \Pr(W \in [t, t + dt])/dt$. Informally, this is the rate at which the duration $W$ is completed at $t$ given that it has not been completed before $t$. The hazard rate is the basic building block of duration analysis in social sciences because it can be directly related to individual behaviour at $t$. The data on $Y, Z$ allow for identification of the hazard rates of $T_i$ at $t$ given that $T \geq t$. These are called the ‘crude’ hazard rates. If the $T_i$ are independent, then these equal the ‘net’ hazard rates of the marginal distributions of the $T_i$.

We now turn to a number of approaches that overcome the general non-identification result for competing risks models. In econometrics, one is typically interested in covariate or regressor effects. The main approach has therefore been to specify semi-parametric models that include observed regressors $X$ and unobserved heterogeneity terms $V$. With a single risk, the most popular duration model is the mixed proportional hazard (MPH) model, which specifies that $\theta(t|X = x, V) = \psi(t) \exp(x^\prime \beta)V$ for some function $\psi(.)$. $V$ is unobserved, and the composition of the survivors changes selectively as time proceeds, so identification from the observable distributions of $T|X$ is non-trivial. However, it holds under the assumptions that $X \perp V$ and $\text{var}(X) > 0$ and some regularity assumptions (see van den Berg, 2001, for an overview of results). With competing risks, the analogue of the MPH model is the multivariate MPH (MMPH) model. With two risks, $\theta_1(t|x, V) = \psi_1(t) \exp(x^\prime \beta_1)V_1$ and $\theta_2(t|x, V) = \psi_2(t) \exp(x^\prime \beta_2)V_2$. where $T_1, T_2|X, V$ are assumed independent, so that a dependence of the durations given $X$ is modelled by way of their unobserved determinants $V_1$ and $V_2$ being dependent. Many empirical studies have estimated parametric versions of this model, using maximum likelihood estimation.

The semi-parametric model has been shown to be identified, under only slightly stronger conditions than those for the MPH model (Abbring and van den Berg, 2003). Specifically, $\text{Var}(X) > 0$ is strengthened to the condition that the vector $X$ includes two continuous variables with the properties that (a) their joint support contains a non-empty open set in $\mathbb{R}^2$, and (b) the vectors $\tilde{\beta}_1, \tilde{\beta}_2$ of the corresponding elements of $\beta_1$ and $\beta_2$ form a matrix $(\tilde{\beta}_1, \tilde{\beta}_2)$ of full rank. Somewhat loosely, $X$ has two continuous