Appendix

Below we describe the derivation of a split population model for a standard parametric distribution and continuous-time duration data, and in doing so, we draw extensively on work by Schmidt and Witte (1989; see also Box-Steffensmeier and Zorn 2003). First, the density function is defined as \( f(t, \theta) \), where \( t \) is the duration of interest and \( \theta \) is a parameter vector to be estimated. The cumulative density is defined as \( F(t, \theta) = \Pr(T \leq t) \), where \( t > 0 \) and \( T \) represents the duration defined by the end of the observation period. The survival function can be written simply as \( S(t, \theta) = 1 - F(t, \theta) \). From this, we can define the hazard rate as:

\[
    h(t, \theta) = \frac{f(t, \theta)}{S(t, \theta)}
\]

The hazard rate is the conditional probability of the event of interest occurring at time \( t \) given that the event has not yet occurred.

The split population model for the duration \( t \) splits the sample into two groups: (1) a group that will eventually experience the event of interest and (2) a group that will never experience the event. Thus, define a latent variable \( Y_i \), where \( Y_i = 1 \) for those cases eventually experiencing the event of interest, and \( Y_i = 0 \) for those observations that will never experience the event. Define \( \Pr(Y_i = 1) = \delta_i \). The conditional density and distribution functions can now be defined as:

\[
    f(t \mid Y_i = 1) = g(t, \theta) \\
    F(t \mid Y_i = 1) = G(t, \theta)
\]

Note that both \( f(t \mid Y_i = 0) \) and \( F(t \mid Y_i = 1) \) are undefined since when \( Y_i = 0 \), the observation will never experience the event and the duration cannot be observed.

Next, define \( R_i \) as an observable indicator that an observation has experienced the event of interest, i.e., \( R_i = 1 \) when failure is observed, \( R_i = 0 \) otherwise. For the cases that experience the event of interest, \( R_i = 1 \), which implies that \( Y_i = 1 \). For these observations, the unconditional density is:

\[
    \Pr(Y_i = 1) \Pr(t_i \leq T_i \mid Y_i = 1) = \delta_i g(t_i, \theta)
\]

where \( T_i \) indicates censoring time. Next, we do not observe cases that experience the event of interest when \( R_i = 0 \), and this occurs for one of two reasons: (1) \( Y_i = 0 \), i.e., the observation will never fail or (2) \( t_i > T_i \), i.e., the observation is censored. For these cases, the unconditional density is:

\[
    \Pr(Y_i = 0) + \Pr(Y_i = 1) \Pr(t_i > T_i \mid Y_i = 1) = (1 - \delta_i) + \delta_i G(t_i, \theta)
\]

Combining these values for each of the two types of observation yields the following likelihood function:

\[
    L = \prod_{i=1}^{N} \delta_i g(t_i, \theta)^{R_i} [1 - \delta_i + \delta_i G(t_i, \theta)]^{(1 - R_i)}
\]
The log-likelihood is:

$$\ln L = \sum_{i=1}^{N} R_i \left[ \ln \delta_i + \ln g(t_i, \theta) \right] + (1 - R_i) \ln \left[ 1 - \delta_i + 1 - g(t_i, \theta) \right]$$

The probability $\delta_i$ is typically modeled as a logit (which we do in this paper) and can include a set of covariates either identical or not identical to those in the duration model. Thus:

$$\delta_i = \frac{\exp(Z_i \gamma)}{1 + \exp(Z_i \gamma)}$$

When $\delta_i = 1$ for all observations, i.e., when all observations will eventually experience the event of interest, the likelihood reduces to a standard duration model with censoring.