

Complications in Event History I: Frailty Models

Basic Problem: Heterogeneity

What is it? Usually thought of as unmeasured risk factors.

Can be induced when a relevant covariate is not included in the model's specification.

Maybe these factors are not measured (unmeasurable?) or are unknown to exist.

Heterogeneity can lead to trouble insofar as parameter estimates can be inconsistent, standard errors can be wrong, and estimates of duration dependency can be misleading.

Sometimes, the problem is manifested through “negative duration dependency.” (See F.1)

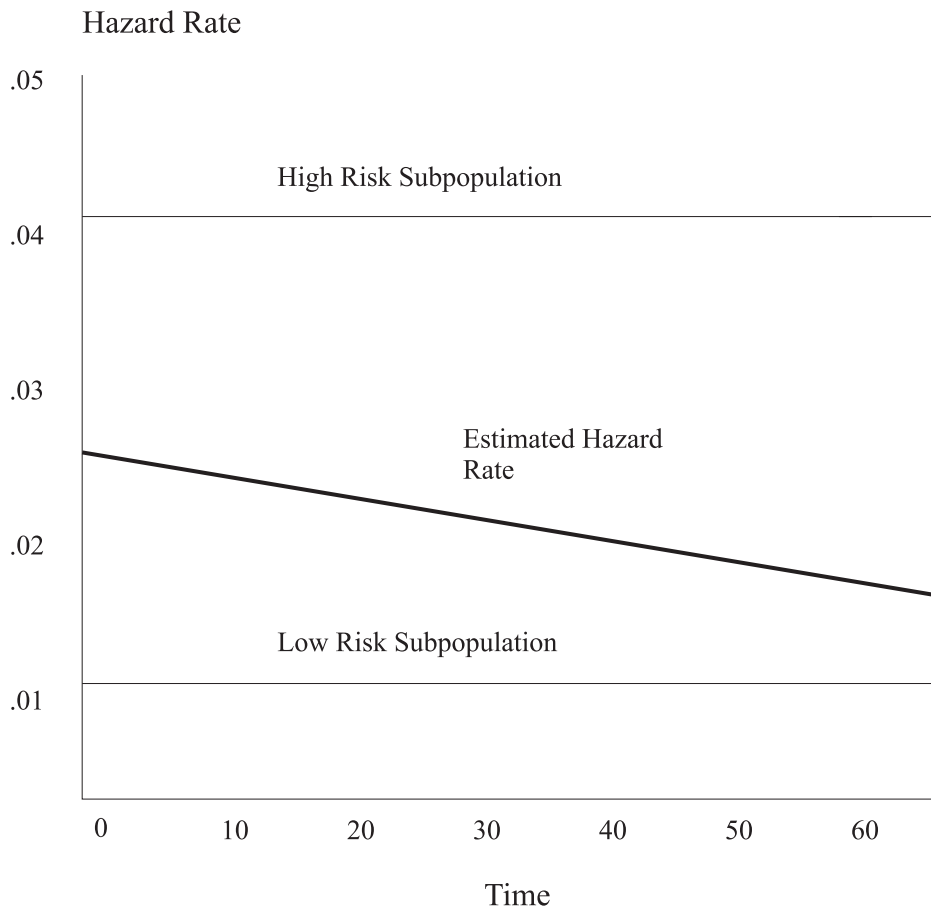


Figure 1: *This figure illustrates the implication of mixing two distinct subpopulations. The top line is an exponential hazard rate for a high risk subpopulation and the bottom line is an exponential hazard rate for a low risk subpopulation. The downward sloping line is the “estimated” hazard rate that would emerge if the two subpopulations were combined and the heterogeneity in risk factors was ignored. Figure from Blossfeld and Rohwer 1995, 241.*

Implications

At issue in the figure: If a population consists of two subgroups with distinct failure rates and the most failure-prone observations fail first, then the failure rate in the surviving population will fall over time even if an exponential duration is suitable for both subgroups individually.

Why? This is because the most failure-prone are rapidly exiting the risk set. If this kind of heterogeneity is not taken into account, then in the aggregate, the hazard rate may appear to decline, even if the group-specific hazards are flat.

All this suggests distributions may be mixed: e.g. if there are two subpopulations, the first following an exponential distribution and the second following an increasing Weibull distribution, a decreasing hazard rate for the population can still result because of frail or failure-prone observations.

Frailty Models

This class of models is a major “growth” area.

Hougaard (2000) provides an excellent and detailed presentation of these kinds of models.

The models have a funny name. Why?

They get their name because they attempt to account for unobserved heterogeneity that occurs because some observations are more failure-prone—and hence, more “frail”—than other observations in a data set.

The basic idea is to introduce into the hazard rate, an additional random parameter that accounts for the random frailties.

These frailties may be individual-specific or group-specific thus giving rise to the nomenclature “individual frailty” or “shared frailty” models.

Individual Frailty

Suppose we have a sample of j observations where some observations are more failure prone due to reasons unknown (or unmeasured) but go ahead and estimate a garden variety model like this one:

$$h(t)_j = h_0(t) \exp(\beta' \mathbf{x}_j).$$

In this model (a PH model) the hazard is increasing or decreasing as a function of \mathbf{x} .

The Problem: If there are unmeasured or unobserved “frailties,” the hazard rate will not only be a function of the covariates, but *also* a function of the frailties:

$$h(t_j) = h_0(t) \exp(\beta' \mathbf{x}_j + \psi' w_j), \quad (1)$$

where w_j are the frailties and are assumed to be an independent sample from a distribution with mean 0 and variance 1 (Klein and Moeschberger 1997). (That is, they follow some distribution function).

Note a couple of important things here: 1) if $\psi = 0$, then the standard proportional hazards model is obtained; 2) if the relevant factors comprising w_j could be measured, then ψ would go to 0.

A Model

A tractable model to account for heterogeneity can be derived if one is willing to make some assumptions regarding the distribution of the frailty. To see this, let's rewrite our model to show how the frailties act multiplicatively on the hazard:

$$h_j(t) \mid \beta' \mathbf{x}_j, \nu_j) = h_0(t) \nu_j \exp(\beta' \mathbf{x}_j). \quad (2)$$

(Note that $\nu_j = \exp(\psi' w_j)$.)

For identification purposes, it is conventionally assumed that the mean of ν is 1 and the variance is unknown and equal to some parameter θ .

Note that we always make assumptions about ν : in standard non-frailty models, we assume ν to be 1 with probability 1! (Frailty may exist; we choose to ignore it.)

If the hazard is a function of the frailties, the survivor function must also be conditional on both the covariates and on the frailty term.

The conditional survivor function (omitting subscripts) is given by

$$\begin{aligned} S(t \mid \beta' \mathbf{x}, \nu) &= \exp \left(- \int_0^t h(u \mid \nu) du \right) \\ &= \exp \left(- \nu \int_0^t h(u) du \right). \end{aligned} \quad (3)$$

and the *marginal* survivor function is given by

$$\begin{aligned} S(t) &= E[S(t \mid \beta' \mathbf{x}, \nu)] \\ &= E \left[\exp \left(- \nu \int_0^t h(u) du \right) \right] \\ &= L \left[\exp \left(\int_0^t h(u) du \right) \right], \end{aligned} \quad (4)$$

where L is the Laplace transformation. Hougaard 2000 refers to this distribution as the “marginal survivor function” because it is the observed survivor function after ν has been integrated out.

To derive the expected value of the survivor function, we need to specify a probability distribution for ν , call this $g(\nu)$.

Lots of choices: Any continuous distribution with positive support, a unit mean, and finite variance θ can be used for $g(\nu)$.

These include: gamma, inverse Gaussian, log-normal, and power variance model. The gamma has most readily been adopted in applied research.

Bottom Line: If we assume ν has a probability distribution, then a tractable model is obtained.

How? By taking the expected value of the survivor function through integrating out the frailty, the problem reduces to one of estimating the frailty variance term, θ and evaluating the null.

Getting There from Here

All this looks rather complicated.

(It is!)

The goal of it all is easy to understand. Let me summarize:

- With heterogeneity, you are likely to have a mixture of hazards.
- If so, you might consider a frailty model.
- Frailty models, as should be clear now, are essentially random effects survival models.
- Frailty “terms” seek to explicitly account for the extra variance associated with unmeasured risk factors.
- To obtain a model, we need to make the “usual” assumptions about which model to pursue and then make the added assumption about $g(\nu)$.
- The problem *with ignoring* frailties is seen in the hazard. In the PH models, the hazard is a multiplicative function of the measured covariates.
- With frailty, the hazard is *also* a function of ν .
- To make the problem more tractable, we integrate out ν and so we’re left with the problem of estimating the variance, θ .

Weibull Example

Consider a Weibull mixture. The conditional survivor function is given by

$$S(t | \nu) = \exp^{-(\nu\lambda t)^p}. \quad (5)$$

With the exception of the frailty term, ν , this expression is identical to equation a standard Weibull.

Now suppose that the gamma distribution is specified for $g(\nu)$. We can define the gamma distribution as $g(\nu, \alpha, \beta)$ where $\alpha = 1/\theta$ and $\beta = \theta$. The density function for the gamma is then given by

$$g(\nu, \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \nu^{\alpha-1} e^{-\nu/\beta},$$

where $\Gamma(\alpha)$ is the gamma integral ($\int_0^\infty \nu^{\alpha-1} e^{-\nu}$). With gamma frailty, the marginal Weibull survivor function is equal to

$$S(t) = [1 + \theta(\lambda t)^p]^{-1/\theta}$$

and the Weibull hazard with gamma frailty is

$$h(t) = \lambda p (\lambda t)^{p-1} [S(t)]^\theta.$$

When the variance of the frailty, ν , is 0, the model reduces to the standard Weibull.

What makes this approach attractive is that it is possible to evaluate the hypothesis that $\theta = 0$.

Note again the following: The frailty approach produces a *mixture model* in that the conditional distribution is described by the Weibull, while the mixture distribution is described by the gamma. We now turn to a brief illustration.

Shared-Frailty Models

Main difference between “shared” and “unshared” frailty models is the assumption of how the frailty is “distributed” in the data.

Shared- or Group-frailty models assumes that similar observations share the same frailty, even as that frailty may vary from group-to-group.

Example: some countries may be more prone for war than others; some U.S. states may be more prone to adopt certain kinds of policies than others.

A “multilevel” problem: shared frailty models are akin to multi-level models. In multilevel modeling, one usually has cases (“level-1 units”) nested within some higher-order unit (“level-2 units”) . . . people nested within countries; students within classrooms...so forth and so on.

Duration data are “multilevel” but in a weird way: we have “time nested within cases.” In this sense, observations on t are “level-1” units and the individual (state, country, regime, etc.) is the “level-2” unit.

It is natural to ask about extra variance associated with your “level-2” units. In event history language, this means asking about frailties associated with your cases (again, individuals, patients, countries, regimes, etc.).

The Basic Model

Suppose we have j observations and i subgroups (for repeated measures data, the j observations will simply be the period-specific records of data for the individual). The hazard rate for the j th individual in the i th subgroup (with frailties) is

$$h(t_{ij}) = h_0(t) \exp(\beta' \mathbf{x}_{ij} + \psi' w_i), \quad (6)$$

where w_i are the *subgroup* frailties, which as before, are assumed to be an independent sample from a distribution with mean 0 and variance 1.

Note again: if $\psi = 0$, then the standard proportional hazards model is obtained, thus implying the absence of group-level heterogeneity.

Also like before, we can reexpress equation (6) as

$$h_{ij}(t) \mid \beta' \mathbf{x}_{ij}, \nu_i = h_0(t) \nu_i \exp(\beta' \mathbf{x}_{ij}), \quad (7)$$

where $\nu_i = \exp(\psi' w_i)$.

These are the *shared* frailties.

Note the difference between this expression and that given in equation (2). Here, the frailty is shared among the j observations in the i th group.

This is why they are called *shared* frailty models!

To get to an estimable model, we essentially work through the same steps as before (having to make assumptions about $g(\nu)$).

Uses of Frailty Models

The use of the frailty approach is attractive as a means to account for possible unobserved heterogeneity; however, the models do have some important caveats to be aware of.

First, the models can be sensitive to the distribution posited for ν . That is to say, estimates of the frailty variance can vary, sometimes substantially, from distribution to distribution.

Second, estimation of the marginal survivor function can be highly nontrivial for some distribution functions that are defined on ν (Hougaard 2000). (You won't get estimates sometimes!)

Third, even though alternatives to the gamma frailty model have been proposed, the gamma is usually the default choice; however, as Hougaard (2000) notes, the gamma has some undesirable properties. Chiefly, the gamma frailty model may give "strange results" because the non-proportional hazards may have a larger influence on the estimates than the actual degree of dependence found in the data (Hougaard 2000, p. 256).

Third (part II): Interpretation of results are contingent on the frailty effect. (Life is easy when $\theta = 0$!)

Fourth, if dependence among observations is primarily regarded as a nuisance, then simpler alternatives to the frailty approach may be desirable like robust variance estimation using either Huber's (1967) method or Lin and Wei's (1989) method.