

# Unobserved Heterogeneity

Germán Rodríguez  
grodri@princeton.edu

Spring, 2001; revised Spring 2005

This unit considers survival models with a random effect representing unobserved heterogeneity of *frailty*, a term first introduced by Vaupel et al. (1979). We consider models without covariates and then move on to the more general case. These notes are intended to complement Rodríguez (1995).

## 1 The Statistics of Heterogeneity

Standard survival models assume homogeneity: all individuals are subject to the same risks embodied in the hazard  $\lambda(t)$  or the survivor functions  $S(t)$ . Models with covariates relax this assumption by introducing observed sources of heterogeneity. Here we consider unobserved sources of heterogeneity that are not readily captured by covariates.

### 1.1 Conditional Hazard and Survival

A popular approach to modelling such sources is the *multiplicative frailty* model, where the hazard for individual  $i$  at time  $t$  is

$$\lambda_i(t) = \lambda(t|\theta_i) = \theta_i \lambda_0(t),$$

the product of an individual-specific random effect  $\theta_i$  representing the individual's *frailty*, and a baseline hazard  $\lambda_0(t)$ . Note that this is essentially a proportional hazards model.

The individual hazard  $\lambda_i(t)$  is interpreted as a *conditional* hazard given  $\theta_i$ . Associated with it we have a conditional survival function

$$S_i(t) = S(t|\theta_i) = S_0(t)^{\theta_i},$$

representing the probability of being alive at  $t$  given the random effect  $\theta_i$ .

The twist is that the random effect  $\theta_i$  is not observed (perhaps not observable), but is assumed to have some a distribution with density  $g(\theta)$ .

## 1.2 Unconditional Hazard and Survival

To obtain the *unconditional* survival function we need to “integrate out” the unobserved random effect:

$$S(t) = \int_0^\infty S(t|\theta)g(\theta)d\theta.$$

We integrate from 0 to  $\infty$  because frailty is non-negative. If frailty was discrete, taking values  $\theta_1, \dots, \theta_k$  with probabilities  $\pi_1, \dots, \pi_k$  then the integral would be replaced by a sum

$$S(t) = \sum_i S(t|\theta_i)\pi_i.$$

In both cases  $S(t)$  is the average  $S_i(t)$ . In a demographic context  $S(t)$  is often referred to as the *population* survivor function, while  $S_i(t)$  is the *individual* survivor function.

To obtain the unconditional hazard we start from the unconditional survival and take negative logs to obtain the cumulative hazard

$$\begin{aligned} \Lambda(t) &= -\log S(t) \\ &= -\log \int_0^\infty S(t|\theta)g(\theta)d\theta \\ &= -\log \int_0^\infty S_0(t)^\theta g(\theta)d\theta. \end{aligned}$$

The next step is to take derivatives w.r.t.  $t$ . Assuming that we can take the derivative operator inside the integral we find the unconditional hazard to be

$$\lambda(t) = -\frac{\int_0^\infty \frac{d}{dt} S_0(t)^\theta g(\theta)d\theta}{\int_0^\infty S_0(t)^\theta g(\theta)d\theta} = \frac{\int_0^\infty \theta \lambda_0(t) S_0(t)^\theta g(\theta)d\theta}{\int_0^\infty S_0(t)^\theta g(\theta)d\theta},$$

where we used the fact that  $S_0(t)^\theta = e^{-\theta\Lambda_0(t)}$ , so that

$$\frac{d}{dt} S_0(t)^\theta = -e^{-\theta\Lambda_0(t)} \theta \lambda_0(t) = -\theta \lambda_0(t) e^{-\theta\Lambda_0(t)},$$

and the last exponential can be recognized as  $S_0(t)^\theta$ .

Note that the population hazard  $\lambda(t)$  is a weighted average of the individual hazards  $\lambda_i(t)$  with weights equal to the density of  $\theta$  times the probability of surviving to  $t$ :

$$S(t|\theta)g(\theta) = S_0(t)^\theta g(\theta),$$

Why can't we calculate the population hazard as a simple average of the individual hazards, the way we calculated the population survivor function?

### 1.3 Expected Frailty of Survivors

We now show that the weights in the above expression represent the conditional distribution of frailty  $\theta$  among survivors to age  $t$ . From first principles, the density of  $\theta$  among survivors is

$$g(\theta|T \geq t) = \frac{\Pr\{T \geq t|\theta\}g(\theta)}{\Pr\{T \geq t\}} = \frac{S(t|\theta)g(\theta)}{\int_0^\infty S(t|\theta)g(\theta)d\theta} = \frac{S_0(t)^\theta g(\theta)}{\int_0^\infty S_0(t)^\theta g(\theta)d\theta},$$

which are indeed the weights in the expression for  $\lambda(t)$ . The expected frailty of survivors can be calculated as

$$E(\theta|T \geq t) = \int_0^\infty \theta g(\theta|T \geq t)d\theta = \frac{\int_0^\infty \theta S_0(t)^\theta g(\theta)d\theta}{\int_0^\infty S_0(t)^\theta g(\theta)d\theta}.$$

From this result it becomes clear that

$$\lambda(t) = \lambda_0(t)E(\theta|T \geq t). \tag{1}$$

In words, the unconditional (population) hazard at  $t$  is the baseline (individual) hazard times the mean frailty of survivors to  $t$ .

Typically, the mean frailty of survivors declines over time as the more frail tend to die earlier. As a result, the population hazard declines more steeply (or increases less rapidly) than the individual hazard. This result is the source of interesting paradoxes.

## 2 Frailty Distributions

We now specialize our results considering a few alternative assumptions about the distribution of frailty.

## 2.1 Gamma Frailty

A convenient assumption used by many authors is that  $\theta$  has a gamma distribution. This distribution has the appropriate range  $(0, \infty)$  and is mathematically tractable.

The density of a gamma distribution with parameters  $\alpha$  and  $\beta$  is

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

where  $\Gamma$  is the gamma function. The mean and variance are

$$E(\theta) = \frac{\alpha}{\beta} \quad \text{and} \quad \text{var}(\theta) = \frac{\alpha}{\beta^2},$$

so the coefficient of variation  $\sigma/\mu$  is  $1/\sqrt{\alpha}$ .

It is often convenient to take  $E(\theta) = 1$  so  $\alpha = \beta = 1/\sigma^2$ . This entails no loss of generality because the average level of frailty can always be absorbed into the baseline hazard.

### 2.1.1 Unconditional Survival and Hazard

The unconditional survivor function is

$$\begin{aligned} S(t) &= \int_0^\infty S_0(t)^\theta g(\theta) d\theta \\ &= \int_0^\infty e^{-\theta\Lambda_0(t)} \theta^{\alpha-1} e^{-\beta\theta} \beta^\alpha \frac{1}{\Gamma(\alpha)} d\theta. \end{aligned}$$

The trick now is to consolidate the coefficients of  $\theta$  and complete a gamma density:

$$S(t) = \int_0^\infty \theta^{\alpha-1} e^{-(\beta+\Lambda_0(t))\theta} (\beta + \Lambda_0(t))^\alpha \frac{1}{\Gamma(\alpha)} \frac{\beta^\alpha}{(\beta + \Lambda_0(t))^\alpha} d\theta.$$

The last fraction on the right does not depend on  $\theta$  and can be taken out of the integral. What's left is a gamma density with parameters  $\alpha$  and  $\beta + \Lambda_0(t)$ , and therefore integrates to one. This gives

$$S(t) = \left( \frac{\beta}{\beta + \Lambda_0(t)} \right)^\alpha.$$

The result is known as a Pareto distribution of the second kind. If frailty has mean one and variance  $\sigma^2$ , we write  $\alpha = \beta = 1/\sigma^2$  to obtain

$$S(t) = \frac{1}{(1 + \sigma^2 \Lambda_0(t))^{1/\sigma^2}}, \quad (2)$$

the unconditional (population) survivor function under gamma frailty.

To find the unconditional (population) hazard we first take negative logs to obtain the cumulative hazard

$$\Lambda(t) = \alpha \log(\beta + \Lambda_0(t)) - \alpha \log(\beta),$$

and then take derivatives w.r.t.  $t$ , to obtain

$$\lambda(t) = \frac{\alpha \lambda_0(t)}{\beta + \Lambda_0(t)}.$$

If frailty has mean one and variance  $\sigma^2$  we obtain

$$\lambda(t) = \frac{\lambda_0(t)}{1 + \sigma^2 \Lambda_0(t)}, \quad (3)$$

the unconditional (population) hazard function under gamma frailty.

*Example:* Gamma mixtures of exponentials are a popular model for unobserved heterogeneity. If the hazard is constant for each individual but people are heterogenous and frailty has a gamma distribution then the population hazard is

$$\lambda(t) = \frac{\lambda}{1 + \sigma^2 \lambda t},$$

where  $\lambda$  is the average individual hazard and  $\sigma^2$  is the variance of frailty.

### 2.1.2 The Frailty of Survivors

In view of our earlier result connecting  $\lambda(t)$  and  $E(\theta|T \geq t)$ , the expected frailty of survivors to  $t$  under gamma frailty must be

$$E(\theta|T \geq t) = \frac{1}{1 + \sigma^2 \Lambda_0(t)}$$

In fact, we can obtain the whole distribution of frailty among survivors to  $t$ . The conditional density of  $\theta$  given  $T \geq t$  is

$$\begin{aligned} g(\theta|T \geq t) &= \frac{S(t|\theta)g(\theta)}{S(t)} \\ &= \frac{e^{-\theta\Lambda_0(t)}\theta^{\alpha-1}e^{-\beta\theta}\beta^\alpha/\Gamma(\alpha)}{\frac{\beta^\alpha}{(\beta+\Lambda_0(t))^\alpha}} \\ &= \theta^{\alpha-1}e^{-(\beta+\Lambda_0(t))\theta}(\beta + \Lambda_0(t))^\alpha/\Gamma(\alpha), \end{aligned}$$

a gamma density with parameters  $\alpha$  and  $\beta + \Lambda_0(t)$ . Thus, if frailty at birth has a gamma distribution with mean one and variance  $\sigma^2$ , so  $\alpha = \beta = 1/\sigma^2$ , then frailty of survivors to  $t$  has a gamma distribution with

$$E(\theta|T \geq t) = \frac{1}{1 + \sigma^2\Lambda_0(t)} \quad \text{and} \quad \text{var}(\theta|T \geq t) = \frac{\sigma^2}{(1 + \sigma^2\Lambda_0(t))^2}.$$

Note that  $\Lambda_0(t)$  is a monotone non-decreasing function of  $t$ . As a result, the mean frailty of survivors declines over time. The variance of frailty of survivors also declines over time, so the population becomes more homogeneous in absolute terms. However, the coefficient of variation stays constant over time, so the population does not become more homogeneous in relative terms (compared to the mean).

Note also that mean frailty will decline more rapidly over time (or selectivity will operate more quickly) when (1) the population is more heterogeneous to start with (larger  $\sigma^2$ ), or (2) the risk is higher (larger  $\Lambda_0(t)$ ).

## 2.2 Inverse Gaussian Frailty

Another distribution that can be used to represent frailty is the inverse Gaussian distribution, which arises as the first passage time in Brownian motion.

Hougaard (1984) has shown that if  $\theta$  has an inverse Gaussian distribution with mean and variance  $\sigma^2$  then the mean frailty of survivors to time  $t$  is

$$E(\theta|T \geq t) = \frac{1}{(1 + 2\sigma^2\Lambda_0(t))^{1/2}}.$$

It then follows from our general results that the unconditional (population) hazard is

$$\lambda(t) = \frac{\lambda_0(t)}{\sqrt{1 + 2\sigma^2\Lambda_0(t)}}.$$

The unconditional (population) survivor function can also be obtained explicitly, and turns out to be

$$S(t) = \exp\left\{-\frac{1}{\sigma^2}(\sqrt{1 + 2\sigma^2\Lambda_0(t)} - 1)\right\},$$

a result that can easily be verified by taking negative logs to get  $\Lambda(t)$  and differentiating w.r.t.  $t$  to obtain  $\lambda(t)$ . Can you derive this result?

### 2.2.1 Notes on the Inverse Gaussian Distribution

The so-called “ordinary” inverse Gaussian distribution has density

$$g(\theta) = \sqrt{\frac{\alpha}{\pi}} e^{\sqrt{4\alpha\theta}} \theta^{-\frac{3}{2}} e^{-\alpha\theta - \frac{\beta}{\theta}},$$

depending on parameters  $\alpha$  and  $\beta$  (called  $\psi$  and  $\theta$  by Hougaard, who uses  $z$  for our  $\theta$ ). This distribution has mean and variance

$$E(\theta) = \sqrt{\frac{\alpha}{\beta}} \quad \text{and} \quad \text{var}(\theta) = \frac{1}{2} \sqrt{\alpha} \beta^{-3/2},$$

so the coefficient of variation  $\sigma/\mu$  is  $1/\sqrt{2}(\alpha\beta)^{1/4}$ . Choosing  $\alpha = \beta$  gives a mean of one and variance  $\text{var}(\theta) = \frac{1}{2}\alpha^{\frac{1}{2}}\alpha^{-\frac{3}{2}} = \frac{1}{2\alpha}$ , so to get a distribution with variance  $\sigma^2$  we take  $\alpha = \beta = \frac{1}{2\sigma^2}$ .

Hougaard shows that under the multiplicative frailty model the distribution of  $\theta$  among survivors to  $t$  is also inverse Gaussian, with parameters  $\alpha$  and  $\beta + \Lambda_0(t)$ . In particular, the mean frailty of survivors is

$$E(\theta|T \geq t) = \sqrt{\frac{\alpha}{\beta + \Lambda_0(t)}} = \frac{1}{(1 + 2\sigma^2\Lambda_0(t))^{1/2}}.$$

Interestingly, the distribution of frailty among those who die at  $t$  is a “generalized” inverse Gaussian.

### 2.3 A More General Family

If you look again at the expressions for  $E(\theta|T \geq t)$ , the mean frailty of survivors to  $t$ , under gamma and inverse Gaussian frailty, you will notice a certain resemblance. In fact, you could write

$$E(\theta|T \geq t) = \frac{1}{(1 + \frac{\sigma^2}{k}\Lambda_0(t))^k}, \quad (4)$$

where  $k = 1$  gives the result for gamma frailty and  $k = 1/2$  gives the result for inverse Gaussian frailty.

One naturally wonders whether this result makes sense more generally. Does this formula represent expected frailty of survivors under some distribution for other values of  $k$ ?

Hougaard (1986) shows that Equation 4 makes sense for any  $k < 1$ , yielding a family based on the *stable distributions*, which includes the inverse Gaussian as a special case.

A distribution is called “stable” if the distribution of the sum of  $n$  i.i.d. r.v.’s is the same as the distribution of  $n^{1/\alpha}$  times one of them for some  $\alpha \in [0, 2]$ . In symbols,

$$\mathcal{D}(X_1 + X_2 + \dots + X_n) = \mathcal{D}(n^{1/\alpha} X_1),$$

where  $\mathcal{D}$  denotes distribution. For example the normal distribution  $N(\mu, \sigma^2)$  is stable with  $\alpha = 1$ , because  $\sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2)$ .

Aalen (1988) showed that Equation 4 also makes sense for  $k > 1$ , showing that all the remaining cases could be obtained by assuming that  $\theta$  has a *compound Poisson distribution*.

To construct this distribution suppose  $N$  is distributed Poisson and  $X_1, X_2, \dots$  are i.i.d. gamma r.v.’s, and define

$$\theta = \begin{cases} 0 & \text{if } N = 0 \\ X_1 + \dots + X_N & \text{if } N > 0 \end{cases}$$

One way to think about this distribution is to imagine a population that has infinitely many strata, one with no frailty, one where frailty is gamma, one where frailty is the sum of two gammas, and so on, with relative stratum sizes given by a Poisson distribution.

Note that this distribution leads to improper survival functions, because for some people  $\theta = 0$  and the event of interest has no risk of ever occurring.

## 2.4 Frailty Transforms

A very useful tool in frailty analysis is the *Laplace transform*. Given a function  $f(x)$ , the Laplace transform, considered as a function of a real argument  $s$  is defined as

$$\mathcal{L}(s) = \int_0^\infty e^{-sx} f(x) dx.$$

The reason why this is useful in our context is that the Laplace transform has exactly the same form as the unconditional survival function. Think of  $f(x)$  as the frailty distribution  $g(\theta)$  and  $s$  as the cumulative baseline hazard  $\Lambda_0(t)$  and you obtain

$$\begin{aligned} S(t) &= \int_0^\infty e^{-\Lambda_0(t)\theta} g(\theta) d\theta \\ &= \mathcal{L}(\Lambda_0(t)), \end{aligned}$$

where  $\mathcal{L}$  denotes the Laplace transform.



Because Laplace transforms are well-known, and many are tabulated, our task is easier. For example the Laplace transform of the gamma distribution with parameters  $\alpha$  and  $\beta$  is

$$\mathcal{L}(s) = \left( \frac{\beta}{\beta + s} \right)^\alpha.$$

Evaluating this at  $s = \Lambda_0(t)$  we obtain the same result as before, but with a lot less work.

Vaupel (1990) has defined the *frailty transform* as the function

$$\mathcal{F}(m, s) = \int_0^\infty \theta^m e^{-s\theta} g(\theta) d\theta.$$

Note that  $\mathcal{F}(0, s)$  is the good old Laplace transform, and  $\mathcal{F}(m, s)$  gives the  $m$ -th moment of the distribution of  $\theta$  at birth. For the gamma distribution the frailty transform is

$$\mathcal{F}(m, s) = \frac{\Gamma(\alpha + m)}{\Gamma(m)} \frac{\beta^\alpha}{(\beta + s)^{\alpha+m}}.$$

The connection with Laplace transforms has practical as well as theoretical importance.

- Given a function  $f(x)$ , computation of the Laplace transform is a well understood problem with efficient algorithms.
- Given the Laplace transform  $\mathcal{L}(s)$ , recovery of the function  $f(x)$  by inversion is an ill-conditioned problem, in the sense that slight changes in  $\mathcal{L}(s)$  can induce huge fluctuations in  $f(x)$ .

### 3 The Inversion Formula

So far we have gone from conditional to unconditional (or if you wish from individual to population) hazard and survival by a process of “mixing”. Can we go the other way? Can we infer the unconditional (individual) hazard and survival from the conditional (population) counterparts by a process of “unmixing”?

The answer is yes, provided we know the distribution of frailty (or how the mixing was done). In the next two subsections we provide inversion formulas for gamma and inverse Gaussian frailty. These results can be used to express population survival functions as gamma or inverse Gaussian mixtures of individual survival functions.

### 3.1 Gamma Mixtures

We have shown that under gamma frailty the unconditional hazard can be written as

$$\lambda(t) = \frac{\lambda_0(t)}{1 + \sigma^2 \Lambda_0(t)}.$$

We will integrate the left-hand side to obtain the cumulative hazard  $\Lambda(t)$ . In order to do this it helps to rewrite the previous equation as a derivative

$$\lambda(t) = \frac{1}{\sigma^2} \frac{d}{dt} \log(1 + \sigma^2 \Lambda_0(t)),$$

because then we can integrate to obtain

$$\Lambda(t) = \frac{1}{\sigma^2} \log(1 + \sigma^2 \Lambda_0(t)),$$

where we used the boundary condition  $\Lambda_0(t) = 0$ . This gives

$$1 + \sigma^2 \Lambda_0(t) = e^{\sigma^2 \Lambda(t)},$$

or

$$\Lambda_0(t) = \frac{1}{\sigma^2} (e^{\sigma^2 \Lambda(t)} - 1).$$

Taking derivatives w.r.t.  $t$  we obtain the conditional (individual) baseline hazard as a function of the unconditional (population) hazard

$$\lambda_0(t) = \lambda(t) e^{\sigma^2 \Lambda(t)}. \tag{5}$$

*Example:* We noted earlier the popularity of gamma mixtures of exponentials. We now show that the exponential distribution itself can be viewed as a gamma mixture of something else. If the population survival function is exponential then

$$\lambda(t) = \lambda \quad \text{and} \quad \Lambda(t) = \lambda t.$$

Plugging these functions into our inversion formula we find the conditional (individual) hazard to be

$$\lambda_0(t) = \lambda e^{\sigma^2 \lambda t},$$

which we recognize as a Gompertz or extreme value hazard, where the log of the hazard is a linear function of  $t$ .

Thus, we have the remarkable result that a population that shows a constant hazard over time may result from individuals with gamma-distributed heterogeneity and Gompertz hazards that increase exponentially with time.

You may begin to suspect that we have a bit of an identification problem here, because a flat population hazard could also result from a homogeneous population where the hazard for each individual is constant over time.

No amount of data can help us distinguish between these two models because they have identical observable consequences.

### 3.2 Inverse Gaussian Mixtures

We can also obtain an inversion formula for inverse Gaussian frailty. Recall that the unconditional hazard was

$$\lambda(t) = \frac{\lambda_0(t)}{(1 + 2\sigma^2\Lambda_0(t))^{1/2}}.$$

As before, we write the right-hand side as a derivative, so integrating is simpler:

$$\lambda(t) = \frac{1}{\sigma^2} \frac{d}{dt} (1 + 2\sigma^2\Lambda_0(t))^{1/2}.$$

To integrate from 0 to  $t$  we impose the boundary condition  $\Lambda(t) = 0$  and obtain

$$\Lambda(t) = \frac{1}{\sigma^2} ((1 + 2\sigma^2\Lambda_0(t))^{1/2} - 1),$$

which incidentally answers the question posed earlier, on how to derive the unconditional survival for inverse Gaussian frailty (see page 2.2). Now we use this result to solve for the baseline integrated hazard:

$$\begin{aligned} (1 + 2\sigma^2\Lambda_0(t))^{1/2} &= 1 + \sigma^2\Lambda(t) \\ 1 + 2\sigma^2\Lambda_0(t) &= (1 + \sigma^2\Lambda(t))^2 \\ \Lambda_0(t) &= \frac{(1 + \sigma^2\Lambda(t))^2 - 1}{2\sigma^2}. \end{aligned}$$

Now take derivatives w.r.t.  $t$  to obtain

$$\begin{aligned} \lambda_0(t) &= \frac{1}{2\sigma^2} 2(1 + \sigma^2\Lambda(t))\sigma^2\lambda(t) \\ &= \lambda(t)(1 + \sigma^2\Lambda(t)). \end{aligned}$$

This result gives us a baseline hazard  $\lambda_0(t)$  that can be mixed using an inverse Gaussian distribution to obtain any given population hazard  $\lambda(t)$ .

*Example:* Using this result we should be able to produce an exponential distribution as an inverse Gaussian mixture of something else. Let's try. If the population survival function is exponential then

$$\lambda(t) = \lambda \quad \text{and} \quad \Lambda(t) = \lambda t,$$

and plugging these into our general result we obtain

$$\lambda_0(t) = \lambda(1 + \sigma^2 \lambda t) = \lambda + \sigma^2 \lambda^2 t,$$

a linear hazard.

Thus, a population with a constant hazard could consist of an inverse Gaussian mix of individuals with linearly rising hazards. Note that  $\sigma^2$  is not specified, so the steepness of the individual hazards is arbitrary.

My conclusion from these results is that models with unobserved heterogeneity are not identified in the sense that we can not distinguish between competing models that have identical observable consequences.

However, it is very important to know that the data we observe could have been generated by different mechanisms. For example in the analysis of waiting time to conception the hazard typically declines over time. This could be due to the fact that the hazard actually declines for each individual. But it is also possible that the individual hazard is constant and the observed decline reflects a selection effect.

## 4 Models with Covariates

Many of the ideas discussed so far extend to models with covariates. Here we will summarize some of the key ideas.

### 4.1 The Omitted Variable Bias

We know from linear models that omitting a variable from a model introduces a bias unless the omitted variable is uncorrelated with the other predictors in the model. In hazard models it turns out that we obtain a bias *even* if the the omitted variable is uncorrelated with the predictors.

To see this point consider a simple problem with two dummy variables,  $x_1$  and  $x_2$ . Suppose these variables are independent and  $\frac{1}{4}$  of the population falls in each of the four categories defined by them.

Suppose further that both variables affect survival time. When both are zero the hazard is constant at  $\lambda_0(t) = 1$ ,  $x_1$  doubles the risk, so  $e^{\beta_1} = 2$ , and  $x_2$  triples the risk, so  $e^{\beta_2} = 3$ .

Under these assumptions we have four exponentials with a proportional hazards structure. If  $i$  denotes the value of  $x_1$  and  $j$  the value of  $x_2$ , the hazard is  $\lambda_{ij}(t) = 2^i 3^j$  for  $i = 0, 1; j = 0, 1$ .

Now suppose we do not observe  $x_2$ . The survival functions we observe for  $x_1 = i$  are a mixture of the curves for  $x_2 = 0$  and  $x_2 = 1$  with equal

weights, so we can write

$$\begin{aligned} S_i(t) &= \frac{1}{2}S_{i0}(t) + \frac{1}{2}S_{i1}(t) \\ &= \frac{1}{2}e^{-2^i t} + \frac{1}{2}e^{-2^{i+3} t} \end{aligned}$$

We can take negative logs to obtain the cumulative hazards and then differentiate to obtain the hazards. (If you do this numerically you can just difference the cumulative hazards.) Figure 1 shows the resulting hazards.

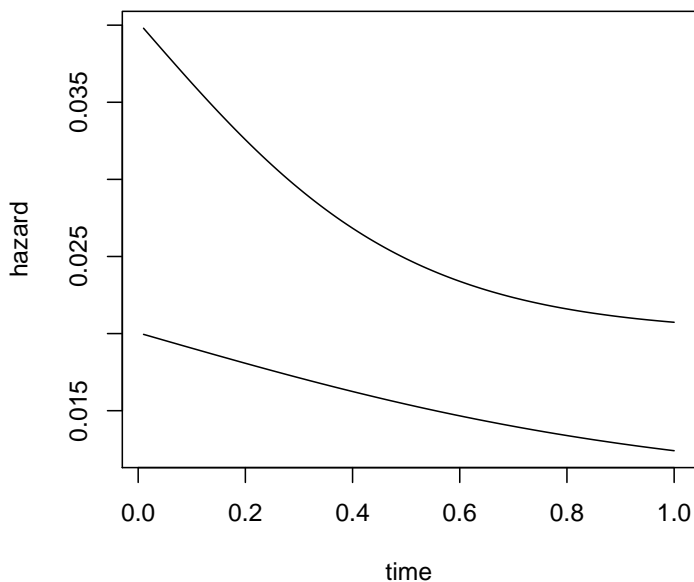


Figure 1: Proportional Hazards with Unobserved Heterogeneity

First, the hazards are *not constant*, even though we started with exponentials. This shows the effect of selection. Individuals with  $x_2 = 1$  die more quickly than those with  $x_2 = 0$  and are selected out of the risk set, so the observed hazard declines over time.

Second, the hazards are *not proportional*, even though we started with a proportional hazards structure. Recall that the hazard for  $x_1 = 1$  was twice the hazard for  $x_1 = 0$ , holding everything else constant. This is now true

only at  $t = 0$ . After that the curves come closer together and the effect of  $x_1 = 1$  is less than 2. If you fitted a proportional hazards model to data generated from this model you would underestimate the effect of  $x_1$ .

The reason for this result is that selection happens more quickly when the hazard is higher (as shown before). The group with  $x_1 = 1$  has a higher hazard (in fact, twice the hazard)—and therefore can select out the frail (those with  $x_2 = 1$ ) more quickly—than the group with  $x_1 = 0$ .

This happened even though  $x_1$  and  $x_2$  were independent. The key to understanding these results is to realize that they were independent at  $t = 0$  but they are no longer independent at  $t > 0$ . The proportion “frail” (i.e. with  $x_2 = 1$ ) is 50% for  $x_1 = 1$  and for  $x_1 = 0$  at the outset, but is 12% for  $x_1 = 1$  compared to 2% for  $x_1 = 0$  at  $t = 1$ . Can you reproduce these percents? As you can imagine, the situation is worse if  $x_1$  and  $x_2$  are correlated.

## 4.2 Models with Unobserved Heterogeneity

One possible solution to the problem of unobserved heterogeneity is to introduce a random effect  $\theta$  in the hope that it will capture the effects of omitted variables that are independent of the  $X$ 's in the model.

The general model we will entertain is a proportional hazards model with a frailty term, where the hazard at time  $t$  for an individual with covariates  $x$  and frailty  $\theta$  is

$$\lambda(t, x, \theta) = \theta \lambda_0(t) e^{x'\beta},$$

where  $\theta$  is a random effect with mean zero and a distribution that does not depend on the observed covariates.

Estimation of this model can be done by maximum likelihood using standard techniques if you can assume

- a parametric form for the baseline hazard  $\lambda_0(t)$ , and
- a distribution for the random effect  $\theta$ .

For example Newman and McCulloch analyzed birth intervals using gamma frailty (here representing fecundability). An alternative tractable functional form is the inverse Gaussian.

It is possible to relax one of these two assumptions, but not both.

## 4.3 Heckman-Singer

Heckman and Singer (1984), in a very influential paper, noted some instability of parameter estimates depending on the type of assumption made about the distribution of frailty.

As a solution, they proposed using a non-parametric maximum likelihood estimator (NPMLE) of the distribution of frailty. Following on earlier work by Laird and others, they show that the NPMLE is a discrete mixing distribution that assigns positive mass to a finite (usually small) set of points of support.

Specifically, the non-parametric estimate takes values  $\theta_1, \theta_2, \dots, \theta_k$  with probabilities  $\pi_1, \pi_2, \dots, \pi_k$  for some value of  $k$ . The distribution has  $2(k-1)$  parameters if one restricts  $\theta$  to have mean one.

Usually one fits a model with  $k = 2$  and increases  $k$  by adding an additional point of support until the likelihood fails to improve, at which point two of the points often coalesce. When one of the points has negligible risk ( $\theta \approx 0$ ) the result can be interpreted as a mover-stayer model.

Flexibility in estimation of the frailty distribution requires parametric assumptions about the hazard. A common choice used in the program CTM (Continuous Time Models) developed by Heckman and associates is the Box-Cox specification

$$\lambda_0(t) = \lambda + \beta_1 \frac{t^{\lambda_1} - 1}{\lambda_1} + \beta_2 \frac{t^{\lambda_2} - 1}{\lambda_2},$$

where  $t^\lambda$  is interpreted as  $\log t$  for  $\lambda = 0$ . This includes as special cases the exponential, Weibull, Gompertz and a log-quadratic hazard.

#### 4.4 Trussell-Richards

Trussell and Richards (1985) wondered whether models estimated using the Heckman-Singer technique were sensitive to the choice of baseline hazard. They found that the results were indeed sensitive, a conclusion confirmed in further work by Trussell and Montgomery.

In fact, it seems clear that the results should be more sensitive to the choice of the baseline hazard than to the choice of the distribution of the unobservable. Why? Recall that the unconditional survival function  $S(t)$ , the only piece of the puzzle that we can actually estimate, has the structure of a Laplace transform

$$S(t, x) = \mathcal{L}_{g(\theta)}(\Lambda_0(t)e^{x'\beta}),$$

so that large variations in  $g(\theta)$  tend to be “smoothed” out and result in small variations in  $S(t, x)$ .

As a result, I think that one is usually be better off using a flexible specification of the baseline hazard combined with a parametric assumption for the distribution of frailty.

## 4.5 The Identification Problem

One difficulty with these models is that the underlying assumption of proportionality of hazards is confounded with unobserved heterogeneity. Consider again Figure 1. We know that the underlying hazards are proportional, but look non-proportional because we are missing  $x_2$ . But I could have generated the same hazards without any omitted variables by assuming that the baseline hazard declines over time and the effect of  $x_1$  is non-proportional.

To further explore these issues we extend our earlier results on unobserved heterogeneity to the case where we have covariates. We start from a proportional hazards model where the conditional or subject-specific hazard is

$$\lambda(t, x, \theta) = \theta \lambda_0(t) e^{x'\beta},$$

and  $\theta$  has density  $g(\theta)$ . To obtain the unconditional or population-average hazard we integrate out  $\theta$  using the appropriate conditional density

$$\lambda(t, x) = \int_0^\infty \lambda(t, x, \theta) g(\theta|T \geq t, x) d\theta.$$

Using the proportional hazards structure we can write this as

$$\lambda(t, x) = \lambda_0(t) e^{x'\beta} \int_0^\infty \theta g(\theta|T \geq t, x) d\theta.$$

The integral can be recognized as the expected frailty of survivors, so we have our first result:

$$\lambda(t, x) = \lambda_0(t) e^{x'\beta} E(\theta|T \geq t, x). \quad (6)$$

The form of the expectation can be worked out for specific distributions. From our earlier results, we can write

$$E(\theta|T \geq t, x) = \frac{1}{(1 + \frac{1}{k} \sigma^2 \lambda_0(t) e^{x'\beta})^k},$$

with  $k = 1$  for gamma frailty and  $k = \frac{1}{2}$  for inverse Gaussian frailty. If we substitute this result on the formula for the hazard and take logs we can write the model as

$$\log \lambda(t, x) = \alpha(t) + x'\beta + \gamma(t, x),$$

where  $\alpha(t) = \log \lambda_0(t)$  is the log-baseline hazard, representing the main effect of duration,  $x'\beta$  is the log-relative risk, representing the main effects



of the covariates, and  $\gamma(t, x) = \log E(\theta|T \geq t, x)$ , the log of the expected frailty of survivors, representing a form of interaction between duration and the covariates.

In other words, the presence of unobserved heterogeneity in a subject-specific proportional hazards model results in a population-average model where the hazards are no longer proportional.

As a result, we conclude that unobserved heterogeneity is indeed confounded with proportionality of hazards. We can't test for one without assuming the other.

*Example 1:* My 1995 paper shows an example of a proportional hazards models that, combined with gamma heterogeneity, leads to declining non-proportional hazards.

But it also shows that exactly the same population hazards could have been generated from a model with inverse Gaussian heterogeneity where the individual hazards are non-proportional.

And of course, it is possible (though unlikely) that there is no unobserved heterogeneity and the individual hazards look just like the population hazards that we observe.

*Example 2:* Suppose you have found that the following proportional hazards model (with a constant baseline!) fits your data well:

$$\lambda(t, x) = \exp\{\alpha + x'\beta\} \tag{7}$$

Before you conclude that the hazard is indeed constant for each individual, consider the alternative subject-specific model

$$\lambda(t, x, \theta) = \theta \exp\{\alpha + x'\beta + \sigma^2 t e^{\alpha + x'\beta}\},$$

where heterogeneity has a gamma distribution with mean one and any variance  $\sigma^2$  that you like. This is an accelerated life model with a Gompertz baseline. You should be able to verify that this model leads to exactly the same population-average hazard as Equation 7.

But there is more. Consider the following subject-specific model

$$\lambda(t, x, \theta) = \theta \exp\{\alpha + x'\beta\}(1 + \sigma^2 \exp\{\alpha + x'\beta\}t),$$

where  $\theta$  has an inverse Gaussian distribution with mean one and variance  $\sigma^2$ . In this model the hazard is a linear function of  $t$ . Again, you should be able to verify that this model leads to the same population-average hazard as Equation 7.

Thus, the hazards in Equation 7 could represent

- homogeneous populations with constant risk,
- gamma frailty with Gompertz accelerated life, or
- inverse Gaussian frailty with linear risks.

The choice between these interpretations cannot be made on statistical grounds.

In our next unit we will see that these models are in fact identified when we have multiple observations, as we do in multivariate survival and event-history models.