Section 14.6 of *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models* by Anders Skrondal and Sophia Rabe-Hesketh (2004)

Treatment of liver cirrhosis: A joint survival and marker model

1 The joint survival and marker model

1.1 Latent marker model

Let t_{ij} be the time of the ith measurement occasion for patient j. The observed marker is denoted y_{ij} at t_{ij} and is related to the latent or 'true' marker $\eta_{ij}^{(2)}$ via the measurement model

$$y_{ij} = \beta_0 + \eta_{ij}^{(2)} + \epsilon_{ij}, \tag{1}$$

where $\epsilon_{ij} \sim N(0, \theta)$.

A structural model for the latent marker is specified as

$$\eta_{ij}^{(2)} = \gamma_1 t_{ij} + \gamma_2 T_j + \eta_j^{(3)} + \zeta_{ij}^{(2)}, \quad \text{Var}(\zeta_{ij}^{(2)}) = 0, \tag{2}$$

where $\eta_j^{(3)} \sim N(0, \psi)$ is independently distributed from ϵ_{ij}.

1.2 Hazard model with latent covariate

Let t_{rj} be the rth death time survived by patient j or the patient's own time of death or censoring and let the hazard at t_{rj} be denoted h_{rj}. A proportional hazards model with the latent marker $\eta_{rj}^{(2)}$ (at time t_{rj}) as covariate is specified as:

$$\ln h_{rj} = \alpha_0 + \alpha_1 t_{rj} + \alpha_2 t_{rj}^2 + \alpha_3 t_{rj}^3 + \lambda \eta_{rj}^{(2)} + \alpha_4 T_j. \tag{3}$$

Here the latent marker $\eta_{rj}^{(2)}$ at time t_{rj} is obtained by replacing t_{ij} in (2) by t_{rj}.

2 Formulation as generalized linear latent and mixed model (GLLAMM)

The survival data are first expanded as described in Section 2.5.1 and shown in the bottom table of Display 2.1 on page 43. However, since the baseline hazard will be modeled as a smooth function of time, we cannot ignore periods of observation from the time of the last contribution to a 'risk set' until censoring as in Cox regression. For instance, the expanded data in the display should have another row with $i = 1$, $r = 4$, $d_{ir} = 0.5$ and $y_{ir} = 0$. The required data expansion can easily be achieved using Stata's `stsplit` command.

The model can now be estimated using Poisson regression with the failure or death indicator D_{ir} (y_{ir} in Display 2.1) as response variable. As explained in Section 2.5.1, we must use the log of the corresponding time intervals of observation, w_{ir} (denoted d_{ir} in Display 2.1), as an offset. We stack these responses and the marker measurements into a single response vector y_{hij}, where $h = 1$ corresponds to a marker measurement and $h = 2$ to the death indicator, i denotes the occasion (previously denoted r for the survival model) and j the patient. Let d_{1hij} be a dummy variable equal to 1 if $h = 1$ and 0 otherwise and let d_{2hij} be a dummy for $h = 2$. The linear predictor for the response model can then be written as a three-level model

$$\begin{aligned} \nu_{hij} = {} & \beta_0 d_{1hij} + \alpha_0 d_{2hij} + \alpha_1 t_{rj} d_{2hij} + \alpha_2 t_{rj}^2 d_{2hij} + \alpha_3 t_{rj}^3 d_{2hij} + \alpha_4 T_j d_{2hij} \\ & + (\ln w_{ij}) d_{2hij} + \eta_{ij}^{(2)} (d_{1hij} + \lambda d_{2hij}) + \eta_j^{(3)} \cdot 0, \end{aligned} \tag{4}$$

see (1) and (3). Note that $\eta_j^{(3)}$ does not contribute to the response model and is hence multiplied by zero. The structural model remains as in (2) and can be written as

$$\left[\begin{array}{c} \eta_{ij}^{(2)} \\ \eta_j^{(3)} \end{array} \right] = \underbrace{\left[\begin{array}{cc} \gamma_1 & \gamma_2 \\ 0 & 0 \end{array} \right]}_{\mathbf{\Gamma}} \left[\begin{array}{c} t_{ij} \\ T_j \end{array} \right] + \underbrace{\left[\begin{array}{cc} 0 & 1 \\ 0 & 0 \end{array} \right]}_{\mathbf{B}} \left[\begin{array}{c} \eta_{ij}^{(2)} \\ \eta_j^{(3)} \end{array} \right] + \left[\begin{array}{c} \zeta_{ij}^{(2)} \\ \zeta_j^{(3)} \end{array} \right].$$

3   Estimation using `gllamm`

The variable names `resp`, `treat`, `t`, `t2`, `t3`, `lnw`, `d1` and `d2` correspond to y_{hij}, T_j, t_{ij}, t_{ij}^2, t_{ij}^3, $\ln w_{ij}$, d_{1hij}, and d_{2hij}, respectively.

To specify the response model in (4), we must first create all the necessary interactions with `d2`:

```
gen d2_t = d2*t
gen d2_t2 = d2*t2
gen d2_t3 = d2*t3
gen d2_treat = d2*treat

gen d2_lnw = cond(d2==1,lnw,0)  /* lnw was missing if d2=0 */
```

We can now define equations for the linear combinations of variables multiplying the level-2 and level-3 latent variables $\eta_{ij}^{(2)}$ and $\eta_j^{(3)}$ in the response model. For $\eta_{ij}^{(2)}$, we require $(d_{1hij} + \lambda d_{2hij})$,

```
eq lev2: d1 d2  /* coefficient of first variable set to 1 */
```

whereas we require 0 for $\eta_j^{(3)}$:

```
gen zero = 0
eq lev3: zero
```

The fixed part of the structural model (first row of $\mathbf{\Gamma}$) can be specified using an equation for the first latent variable. To let `gllamm` know that this is the equation for the first latent variable, the second character of the equation name must be a '1':

```
eq f1:  t treat
```

Finally, we can specify the $\mathbf{B}$ matrix by defining a matrix with elements equal to '1' corresponding to coefficients we wish to estimate, all other elements being set to 0:

```
matrix B = (0,1\ 0,0)
```

However, we do not wish to estimate b_{12} which should be fixed at 1. We therefore define a constraint

```
cons def 1 [b1_2]_cons = 1
```

and another constraint to set the variance of the level-2 disturbance in the structural model to zero

```
cons def 2 [occ1_1]d1 = 0
```

We are now ready to specify the `gllamm` command:

```
gllamm resp d1 d2 d2_t d2_t2 d2_t3 d2_treat, i(occ id) link(ident log)    /*
  */ lv(var) fam(gauss poiss) fv(var) nocons offset(d2_lnw) eqs(lev2 lev3) /*
  */ nip(1 20) constr(1 2) geqs(f1) bmat(B)
```

`occ` and `id` are identifiers for the individual observations and subjects respectively and `var` is a variable equal to 1 for marker responses and equal to 2 for death indicator responses. The `link()` and `family()` options, together with `lv(var)` and `fv(var)` specify a linear model with gaussian errors for the marker responses and a log-linear Poisson model for the failure indicator responses.

Since the variance of the level-2 disturbance is zero, we avoid integrating over the level-2 disturbance by specifying only one integration point for level 2 using the `nip(1 20)` option. The output is:

```
number of level 1 units = 73046
number of level 2 units = 73046
number of level 3 units = 488

Condition Number = 409.06292

gllamm model with constraints:
 ( 1) [b1_2]_cons = 1
 ( 2) [occ1_1]d1 = 0

log likelihood = -8422.867933859132
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
d1	7.88475	.1157757	68.10	0.000	7.657834	8.111666
d2	-7.95779	.1190091	-66.87	0.000	-8.191044	-7.724537
d2_t	-.1528312	.0508363	-3.01	0.003	-.2524685	-.0531939
d2_t2	.0952418	.0232594	4.09	0.000	.0496543	.1408294
d2_t3	-.007631	.0024718	-3.09	0.002	-.0124756	-.0027865
d2_treat	-.1791224	.1242597	-1.44	0.149	-.4226669	.064422
d2_lnw	(offset)					

```
Variance at level 1
--------------------------------------------------------------------------------

  3.4301326 (.09655311)

Variances and covariances of random effects
--------------------------------------------------------------------------------


***level 2 (occ)

    var(1): 0 (0)

    loadings for random effect 1
    d1: 1 (fixed)
    d2: -.38306696 (.03896087)

***level 3 (id)

    var(1): 3.7132765 (.25822641)

B-matrix:
--------------------------------------------------------------------------------


    B(1,2): 1 (0)

Regressions of latent variables on covariates
--------------------------------------------------------------------------------


    random effect 1 has 2 covariates:
    t: .14781708 (.01683579)
    treat: -.6387828 (.18535363)
--------------------------------------------------------------------------------
```

so that $\widehat{\beta}_0 = 7.88$, $\widehat{\gamma}_1 = 0.15$, $\widehat{\gamma}_2 = -0.64$, $\widehat{\psi} = 3.71$, $\widehat{\theta} = 3.43$ $\widehat{\alpha}_0 = -7.96$, $\widehat{\alpha}_1 = -0.15$, $\widehat{\alpha}_2 = 0.10$, $\widehat{\alpha}_3 = -0.01$, $\widehat{\alpha}_4 = -0.18$ and $\widehat{\lambda} = -0.38$. The estimates in Table 14.8 were obtained using more quadrature points, namely, `nip(1 30)`.

For more information on gllamm, see http://www.gllamm.org