mestreg — Multilevel mixed-effects parametric survival models

Description

mestreg fits a mixed-effects parametric survival-time model. The conditional distribution of the response given the random effects is assumed to be an exponential, Weibull, lognormal, loglogistic, or gamma distribution. mestreg can be used with single- or multiple-record st data.

Quick start

Without weights

Two-level Weibull survival model with covariates x1 and x2 and random intercepts by lev2 using stset data

```
mestreg x1 x2 || lev2:, distribution(weibull)
```

Mixed-effects model adding random coefficients for x1

```
mestreg x1 x2 || lev2:x1, distribution(weibull)
```

Three-level random-intercept model with lev2 nested within lev3

```
mestreg x1 x2 || lev3: || lev2:, distribution(weibull)
```

With weights

Two-level Weibull survival model with covariates x1 and x2, random intercepts by lev2, and observation-level frequency weights wvar1 using stset data

```
mestreg x1 x2 [fweight=wvar1] || lev2:, distribution(weibull)
```

Two-level random-intercept model from a two-stage sampling design with PSUs identified by psu using PSU-level and observation-level sampling weights wvar2 and wvar1

```
mestreg x1 x2 [pweight=wvar1] || psu:, pweight(wvar2)
```

Same as above, but svyset the data first

```
svyset psu, weight(wvar2) || _n, weight(wvar1)
svy: mestreg x1 x2 || psu:, distribution(weibull)
```

Note: Any supported parametric survival distribution may be specified in place of weibull above.

Menu

Statistics > Multilevel mixed-effects models > Parametric survival regression
Syntax

```
. estreg fe_equation || re_equation || re_equation ...,
    distribution(distname) [options]
```

where the syntax of `fe_equation` is

```
[ indepvars ] [ if ] [ in ] [ weight ] [ , fe_options ]
```

and the syntax of `re_equation` is one of the following:

for random coefficients and intercepts

```
levelvar: [ varlist ] [ , re_options ]
```

for random effects among the values of a factor variable

```
levelvar: R.varname
```

`levelvar` is a variable identifying the group structure for the random effects at that level or is `_all` representing one group comprising all observations.

**fe_options**

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
</tr>
<tr>
<td><code>noconstant</code></td>
</tr>
<tr>
<td>suppress constant term from the fixed-effects equation</td>
</tr>
<tr>
<td><code>offset(varname)</code></td>
</tr>
<tr>
<td>include <code>varname</code> in model with coefficient constrained to 1</td>
</tr>
</tbody>
</table>

**re_options**

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
</tr>
<tr>
<td><code>covariance(vartype)</code></td>
</tr>
<tr>
<td>variance–covariance structure of the random effects</td>
</tr>
<tr>
<td><code>noconstant</code></td>
</tr>
<tr>
<td>suppress constant term from the random-effects equation</td>
</tr>
<tr>
<td><code>fweight(varname)</code></td>
</tr>
<tr>
<td>frequency weights at higher levels</td>
</tr>
<tr>
<td><code>iweight(varname)</code></td>
</tr>
<tr>
<td>importance weights at higher levels</td>
</tr>
<tr>
<td><code>pweight(varname)</code></td>
</tr>
<tr>
<td>sampling weights at higher levels</td>
</tr>
</tbody>
</table>
### options

<table>
<thead>
<tr>
<th>Description</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>Description</td>
</tr>
<tr>
<td><code>distribution(distname)</code></td>
<td>specify survival distribution</td>
</tr>
<tr>
<td><code>time</code></td>
<td>use accelerated failure-time metric</td>
</tr>
<tr>
<td><code>constraints(constraints)</code></td>
<td>apply specified linear constraints</td>
</tr>
<tr>
<td><code>collinear</code></td>
<td>keep collinear variables</td>
</tr>
<tr>
<td>SE/Robust</td>
<td>Description</td>
</tr>
<tr>
<td><code>vce(vcetype)</code></td>
<td><code>vcetype</code> may be <code>oim</code>, <code>robust</code>, or <code>cluster clustvar</code></td>
</tr>
<tr>
<td>Reporting</td>
<td>Description</td>
</tr>
<tr>
<td><code>level(#)</code></td>
<td>set confidence level; default is <code>level(95)</code></td>
</tr>
<tr>
<td><code>nohr</code></td>
<td>do not report hazard ratios</td>
</tr>
<tr>
<td><code>tratio</code></td>
<td>report time ratios</td>
</tr>
<tr>
<td><code>noshow</code></td>
<td>do not show st setting information</td>
</tr>
<tr>
<td><code>nocnsreport</code></td>
<td>do not display constraints</td>
</tr>
<tr>
<td><code>notable</code></td>
<td>suppress coefficient table</td>
</tr>
<tr>
<td><code>noheader</code></td>
<td>suppress output header</td>
</tr>
<tr>
<td><code>nogroup</code></td>
<td>suppress table summarizing groups</td>
</tr>
<tr>
<td><code>display_options</code></td>
<td>control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling</td>
</tr>
<tr>
<td>Integration</td>
<td>Description</td>
</tr>
<tr>
<td><code>intmethod(intmethod)</code></td>
<td>integration method</td>
</tr>
<tr>
<td><code>intpoints(#)</code></td>
<td>set the number of integration (quadrature) points for all levels; default is <code>intpoints(7)</code></td>
</tr>
<tr>
<td>Maximization</td>
<td>Description</td>
</tr>
<tr>
<td><code>maximize_options</code></td>
<td>control the maximization process; seldom used</td>
</tr>
<tr>
<td><code>startvalues(svmethod)</code></td>
<td>method for obtaining starting values</td>
</tr>
<tr>
<td><code>startgrid[ (gridspec) ]</code></td>
<td>perform a grid search to improve starting values</td>
</tr>
<tr>
<td><code>noestimate</code></td>
<td>do not fit the model; show starting values instead</td>
</tr>
<tr>
<td><code>dnumerical</code></td>
<td>use numerical derivative techniques</td>
</tr>
<tr>
<td><code>coeflegend</code></td>
<td>display legend instead of statistics</td>
</tr>
</tbody>
</table>

`*distribution(distname)` is required.
**vartype**

<table>
<thead>
<tr>
<th>vartype</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>independent</td>
<td>one unique variance parameter per random effect, all covariances 0; the default unless the R. notation is used</td>
</tr>
<tr>
<td>exchangeable</td>
<td>equal variances for random effects, and one common pairwise covariance</td>
</tr>
<tr>
<td>identity</td>
<td>equal variances for random effects, all covariances 0; the default if the R. notation is used</td>
</tr>
<tr>
<td>unstructured</td>
<td>all variances and covariances to be distinctly estimated</td>
</tr>
<tr>
<td>fixed(matname)</td>
<td>user-selected variances and covariances constrained to specified values; the remaining variances and covariances unrestricted</td>
</tr>
<tr>
<td>pattern(matname)</td>
<td>user-selected variances and covariances constrained to be equal; the remaining variances and covariances unrestricted</td>
</tr>
</tbody>
</table>

**distname**

<table>
<thead>
<tr>
<th>distname</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>exponential</td>
<td>exponential survival distribution</td>
</tr>
<tr>
<td>loglogistic</td>
<td>loglogistic survival distribution</td>
</tr>
<tr>
<td>loglogistic</td>
<td>synonym for loglogistic</td>
</tr>
<tr>
<td>weibull</td>
<td>Weibull survival distribution</td>
</tr>
<tr>
<td>lognormal</td>
<td>lognormal survival distribution</td>
</tr>
<tr>
<td>lognormal</td>
<td>synonym for lognormal</td>
</tr>
<tr>
<td>gamma</td>
<td>gamma survival distribution</td>
</tr>
</tbody>
</table>

**intmethod**

<table>
<thead>
<tr>
<th>intmethod</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>mvaghermite</td>
<td>mean–variance adaptive Gauss–Hermite quadrature; the default unless a crossed random-effects model is fit</td>
</tr>
<tr>
<td>mcaghermite</td>
<td>mode-curvature adaptive Gauss–Hermite quadrature</td>
</tr>
<tr>
<td>ghermite</td>
<td>nonadaptive Gauss–Hermite quadrature</td>
</tr>
<tr>
<td>laplace</td>
<td>Laplacian approximation; the default for crossed random-effects models</td>
</tr>
</tbody>
</table>

You must *stset* your data before using *mestreg*; see [ST] *stset*.

*indepvars* may contain factor variables; see [U] 11.4.3 Factor variables.

*bayes*, by, and *svy* are allowed; see [U] 11.10 Prefix commands. For more details, see [BAYES] *bayes: mestreg*. *vce()* and weights are not allowed with the *svy* prefix; see [SVY] *svy*.

*fw* weights, *iweights*, and *pweights* are allowed; see [U] 11.1.6 weight. Only one type of weight may be specified.

Weights are not supported under the Laplacian approximation or for crossed models.

*startvalues()*., *startgrid*, *noestimate*, *dnumerical*, and *coeflegend* do not appear in the dialog box.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.

**Options**

- **noconstant** suppresses the constant (intercept) term and may be specified for the fixed-effects equation and for any of or all the random-effects equations.
offset(varname) specifies that varname be included in the fixed-effects portion of the model with the coefficient constrained to be 1.

covariance(vartype) specifies the structure of the covariance matrix for the random effects and may be specified for each random-effects equation. vartype is one of the following: independent, exchangeable, identity, unstructured, fixed(matname), or pattern(matname).

covariance(independent) covariance structure allows for a distinct variance for each random effect within a random-effects equation and assumes that all covariances are 0. The default is covariance(independent) unless a crossed random-effects model is fit, in which case the default is covariance(identity).

covariance(exchangeable) structure specifies one common variance for all random effects and one common pairwise covariance.

covariance(identity) is short for “multiple of the identity”; that is, all variances are equal and all covariances are 0.

covariance(unstructured) allows for all variances and covariances to be distinct. If an equation consists of p random-effects terms, the unstructured covariance matrix will have p(p + 1)/2 unique parameters.

covariance(fixed(matname)) and covariance(pattern(matname)) covariance structures provide a convenient way to impose constraints on variances and covariances of random effects. Each specification requires a matname that defines the restrictions placed on variances and covariances. Only elements in the lower triangle of matname are used, and row and column names of matname are ignored. A missing value in matname means that a given element is unrestricted. In a fixed(matname) covariance structure, (co)variance (i,j) is constrained to equal the value specified in the i,jth entry of matname. In a pattern(matname) covariance structure, (co)variances (i,j) and (k,l) are constrained to be equal if matname[i,j] = matname[k,l].

fweight(varname) specifies frequency weights at higher levels in a multilevel model, whereas frequency weights at the first level (the observation level) are specified in the usual manner, for example, [fw=fwtvar1]. varname can be any valid Stata variable name, and you can specify fweight() at levels two and higher of a multilevel model. For example, in the two-level model

```
. mecmd fixed_portion [fw = wt1] || school: ... , fweight(wt2) ...
```

the variable wt1 would hold the first-level (the observation-level) frequency weights, and wt2 would hold the second-level (the school-level) frequency weights.

iweight(varname) specifies importance weights at higher levels in a multilevel model, whereas importance weights at the first level (the observation level) are specified in the usual manner, for example, [iw=iwtvar1]. varname can be any valid Stata variable name, and you can specify iweight() at levels two and higher of a multilevel model. For example, in the two-level model

```
. mecmd fixed_portion [iw = wt1] || school: ... , iweight(wt2) ...
```

the variable wt1 would hold the first-level (the observation-level) importance weights, and wt2 would hold the second-level (the school-level) importance weights.

pweight(varname) specifies sampling weights at higher levels in a multilevel model, whereas sampling weights at the first level (the observation level) are specified in the usual manner, for example, [pw=pwtvar1]. varname can be any valid Stata variable name, and you can specify pweight() at levels two and higher of a multilevel model. For example, in the two-level model

```
. mecmd fixed_portion [pw = wt1] || school: ... , pweight(wt2) ...
```

variable wt1 would hold the first-level (the observation-level) sampling weights, and wt2 would hold the second-level (the school-level) sampling weights.
distribution(*distname*) specifies the survival model to be fit. *distname* is one of the following: exponential, loglogistic, llogistic, weibull, lognormal, lnormal, or gamma. This option is required.

time specifies that the model be fit in the accelerated failure-time metric rather than in the log relative-hazard metric. This option is valid only for the exponential and Weibull models because these are the only models that have both a proportional-hazards and an accelerated failure-time parameterization. Regardless of metric, the likelihood function is the same, and models are equally appropriate in either metric; it is just a matter of changing interpretation.

time must be specified at estimation.

constraints(*constraints*), collinear; see [R] estimation options.

---

**vce(vcetype)** specifies the type of standard error reported, which includes types that are derived from asymptotic theory (oim), that are robust to some kinds of misspecification (robust), and that allow for intragroup correlation (cluster clustvar); see [R] vce_option. If vce(robust) is specified, robust variances are clustered at the highest level in the multilevel model.

---

**level(#)**; see [R] estimation options.

**nohr**, which may be specified at estimation or upon redisplaying results, specifies that coefficients rather than exponentiated coefficients be displayed, that is, that coefficients rather than hazard ratios be displayed. This option affects only how coefficients are displayed, not how they are estimated. This option is valid only for the exponential and Weibull models because they have a natural proportional-hazards parameterization. These two models, by default, report hazards ratios (exponentiated coefficients).

**tratio** specifies that exponentiated coefficients, which are interpreted as time ratios, be displayed. **tratio** is appropriate only for the loglogistic, lognormal, and gamma models or for the exponential and Weibull models when fit in the accelerated failure-time metric.

**tratio** may be specified at estimation or upon replay.

**noshow** prevents mestreg from showing the key st variables. This option is rarely used because most users type stset, show or stset, noshow to set once and for all whether they want to see these variables mentioned at the top of the output of every st command; see [ST] stset.

**nocnsreport**; see [R] estimation options.

**notable** suppresses the estimation table, either at estimation or upon replay.

**noheader** suppresses the output header, either at estimation or upon replay.

**nogroup** suppresses the display of group summary information (number of groups, average group size, minimum, and maximum) from the output header.

**display_options**: noci, nopvalues, noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fvwrap(#), fvwrapon(style), cformat(%,fmt), pformat(%,fmt), sformat(%,fmt), and nolstretch; see [R] estimation options.
Integration

`intmethod(intmethod)` specifies the integration method to be used for the random-effects model. `mvaghermite` performs mean–variance adaptive Gauss–Hermite quadrature; `mcaghermite` performs mode-curvature adaptive Gauss–Hermite quadrature; `ghermite` performs nonadaptive Gauss–Hermite quadrature; and `laplace` performs the Laplacian approximation, equivalent to mode-curvature adaptive Gaussian quadrature with one integration point.

The default integration method is `mvaghermite` unless a crossed random-effects model is fit, in which case the default integration method is `laplace`. The Laplacian approximation has been known to produce biased parameter estimates; however, the bias tends to be more prominent in the estimates of the variance components rather than in the estimates of the fixed effects.

For crossed random-effects models, estimation with more than one quadrature point may be prohibitively intensive even for a small number of levels. For this reason, the integration method defaults to the Laplacian approximation. You may override this behavior by specifying a different integration method.

`intpoints(#)` sets the number of integration points for quadrature. The default is `intpoints(7)`, which means that seven quadrature points are used for each level of random effects. This option is not allowed with `intmethod(laplace)`.

The more integration points, the more accurate the approximation to the log likelihood. However, computation time increases as a function of the number of quadrature points raised to a power equaling the dimension of the random-effects specification. In crossed random-effects models and in models with many levels or many random coefficients, this increase can be substantial.

Maximization

`maximize_options: difficult, technique(algorithm_spec), iterate(#), [no]log, trace, gradient, showstep, hessian, showtolerance, tolerance(#), itolerance(#), nrtolerance(#), nonrtolerance, and from(init_specs);` see `[R] maximize. Those that require special mention for `mestreg` are listed below.

`from()` accepts a properly labeled vector of initial values or a list of coefficient names with values. A list of values is not allowed.

The following options are available with `mestreg` but are not shown in the dialog box:

`startvalues(svmethod), startgrid[(gridspec)], noestimate, and dnumerical;` see `[ME] meglm.

`coeflegend;` see `[R] estimation options.

Remarks and examples

For a general introduction to `me` commands, see `[ME] me.

Remarks are presented under the following headings:

- Introduction
- Two-level models
- Three-level models
**Introduction**

Mixed-effects survival models contain both fixed effects and random effects. In longitudinal data and panel data, random effects are useful for modeling intracluster correlation; that is, observations in the same cluster are correlated because they share common cluster-level random effects.

`mestreg` allows for many levels of random effects. However, for simplicity, we now consider two-level models, where we have a series of $M$ independent clusters and a set of random effects $u_j$ corresponding to those clusters. Two often-used models for adjusting survivor functions for the effects of covariates are the accelerated failure-time (AFT) model and the multiplicative or proportional hazards (PH) model.

In the AFT model, the natural logarithm of the survival time, $\log t$, is expressed as a linear function of the covariates; when we incorporate random-effects, this yields the model

$$\log t_{ji} = x_{ji}^T \beta + z_{ji}^T u_j + v_{ji}$$

for $j = 1, \ldots, M$ clusters, with cluster $j$ consisting of $i = 1, \ldots, n_j$ observations. The $1 \times p$ row vector $x_{ji}$ contains the covariates for the fixed effects, with regression coefficients (fixed effects) $\beta$.

The $1 \times q$ vector $z_{ji}$ contains the covariates corresponding to the random effects and can be used to represent both random intercepts and random coefficients. For example, in a random-intercept model, $z_{ji}$ is simply the scalar 1. The random effects $u_j$ are $M$ realizations from a multivariate normal distribution with mean $0$ and $q \times q$ variance matrix $\Sigma$. The random effects are not directly estimated as model parameters but are instead summarized according to the unique elements of $\Sigma$, known as variance components.

Finally, $v_{ji}$ are the observation-level errors with density $\varphi(\cdot)$. The distributional form of the error term determines the regression model. Five regression models are implemented in `mestreg` using the AFT parameterization: exponential, gamma, loglogistic, lognormal, and Weibull. The lognormal regression model is obtained by letting $\varphi(\cdot)$ be the normal density. Similarly, by letting $\varphi(\cdot)$ be the logistic density, one obtains the loglogistic regression. Setting $\varphi(\cdot)$ equal to the extreme-value density yields the exponential and the Weibull regression models.

In the PH models fit by `mestreg`, the covariates have a multiplicative effect on the hazard function

$$h(t_{ji}) = h_0(t_{ji}) \exp(x_{ji}^T \beta + z_{ji}^T u_j)$$

for some baseline hazard function $h_0(t)$. For the `mestreg` command, $h_0(t)$ is assumed to be parametric. The exponential and Weibull models are implemented in `mestreg` for the PH parameterization. These two models are implemented using both the AFT and PH parameterizations.

`mestreg` is suitable only for data that have been `stset`. By using `stset` on your data, you define the variables `_t0`, `_t`, and `_d`, which serve as the trivariate response variable $(t_0, t, d)$. Each response corresponds to a period under observation, $(t_0, t]$, resulting in either failure ($d = 1$) or right-censoring ($d = 0$) at time $t$.

`mestreg` does not allow delayed entry or gaps. However, `mestreg` is appropriate for data exhibiting multiple records per subject and time-varying covariates. `mestreg` requires subjects to be nested within clusters.

`stset` weights are not used; instead, weights must be specified at estimation. Weights are not allowed with crossed models or the Laplacian approximation. See *Survey estimation* in *Methods and formulas* for details.
Two-level models

Example 1: Two-level random-intercept PH model

In example 11 of [ST] streg, we fit a Weibull model with an inverse-Gaussian shared frailty to the recurrence times for catheter-insertion point infection for 38 kidney dialysis patients. In this example, the subjects are the catheter insertions, not the patients themselves. This is a function of how the data were recorded—the onset of risk occurs at the time the catheter is inserted and not, say, at the time of admission of the patient into the study. Thus we have two subjects (insertions) within each group (patient). Each catheter insertion results in either infection (infect==1) or right-censoring (infect==0). The stset results are shown below.

```
. use http://www.stata-press.com/data/r15/catheter
   (Kidney data, McGilchrist and Aisbett, Biometrics, 1991)
. stset
   -> stset time, failure(infect)
     failure event:  infect != 0 & infect < .
     obs. time interval:  (0, time]
     exit on or before:  failure
```

76 total observations
0 exclusions
76 observations remaining, representing
58 failures in single-record/single-failure data
7,424 total analysis time at risk and under observation
   at risk from t = 0
   earliest observed entry t = 0
   last observed exit t = 562

While it is reasonable to assume independence of patients, we would not want to assume that recurrence times within each patient are independent. The model used in [ST] streg allowed us to model the correlation by assuming that it was the result of a latent patient-level effect, or frailty.

The random-effects approach used by mestreg is more flexible because it allows you to experiment with several levels of random effects, including random coefficients, or both. You can then choose the model that best suits your data. Here we use mestreg to fit a random-effects Weibull model with normally distributed random effects. This model can be viewed as a shared frailty model with lognormal frailty.
...mestreg age female || patient:, distribution(weibull)
    failure _d: infect
    analysis time _t: time

Fitting fixed-effects model:
Iteration 0:  log likelihood = -1700989.9
Iteration 1:  log likelihood = -440.1998
Iteration 2:  log likelihood = -336.62162
Iteration 3:  log likelihood = -334.64937
Iteration 4:  log likelihood = -334.57959
Iteration 5:  log likelihood = -334.57944
Iteration 6:  log likelihood = -334.57944

Refining starting values:
Grid node 0:  log likelihood = -336.03604

Fitting full model:
Iteration 0:  log likelihood = -336.03604 (not concave)
Iteration 1:  log likelihood = -333.14043
Iteration 2:  log likelihood = -330.40952
Iteration 3:  log likelihood = -329.89242
Iteration 4:  log likelihood = -329.87847
Iteration 5:  log likelihood = -329.87832
Iteration 6:  log likelihood = -329.87832

Mixed-effects Weibull PH regression
Group variable:  patient
Obs per group:
    min = 2
    avg = 2.0
    max = 2
Integration method: mvaghermite
Integration pts. = 7
Wald ch2(2) = 10.12
Number of obs = 76
Number of groups = 38
Log likelihood = -329.87832
Prob > ch2 = 0.0063

| _t | Haz. Ratio | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|----|------------|-----------|-------|------|----------------------|
|    |            |           |       |      |                      |
| age| 1.007348   | 0.013788  | 0.53  | 0.593| .9806828 1.034737    |
| female| .1904727 | .099992   | -3.16 | 0.002| .0680737 .5329493   |
| _cons| .0072901  | .0072274  | -4.96 | 0.000| .0010444 .0508881   |
| /ln_p| .2243233  | .1402795  |       |      | -.0506195 .4992661  |
| patient| var(_cons)| .8308495  | .4978425 |      | .256735 2.688808    |

Note: Estimates are transformed only in the first equation.
Note: _cons estimates baseline hazard (conditional on zero random effects).
LR test vs. Weibull model: chibar2(01) = 9.40 Prob >= chibar2 = 0.0011

The results are similar to those in [ST] streg. The likelihood-ratio test compares the random-effects model with a survival model with fixed-effects only. The results support the random-effects model.

By default, when fitting a model with the PH parameterization, mestreg displays exponentiated coefficients, labeled as hazard ratios. These hazard ratios should be interpreted as “conditional hazard ratios”, that is, conditional on the random effects.

For example, the hazard ratio for age is 1.01. This means that according to the model, for a given patient, the hazard would increase 1% with each year of age. However, at the population level, marginal hazards corresponding to different levels of the covariates are not necessarily proportional. Example 5 in [ME] mestreg postestimation illustrates this point with simulated data.
The exponentiated coefficients of covariates that usually remain constant within a group do not have a natural interpretation as conditional hazard ratios. However, the magnitude of the exponentiated coefficients always gives an idea of the effect of the covariates. In this example, female is constant within the group. The estimated hazard ratio for female is 0.19, which indicates that hazard functions for females tend to be smaller than hazard functions for males. Both conditional and unconditional predictions can be obtained with predict. Unconditional predictions can be visualized by using stcurve. Unconditional effects can be tested and visualized by using margins and marginsplot. See example 1 in [ME] mestreg postestimation for an example using predict, margins, and marginsplot.

Example 2: Two-level random-intercept AFT model

Although the PH parameterization is more popular in the literature because the output is easier to interpret, the AFT parameterization is useful when we need to make comparisons with other models that have only an AFT parameterization. For example, we might want to compare the Weibull results from example 1 with the results from a gamma model.

Let’s redisplay the results of a Weibull PH model from example 1 as coefficients:

```
mestreg, nohr
Mixed-effects Weibull PH regression
Group variable: patient
Obs per group: min = 2 avg = 2.0 max = 2
Integration method: mvaghermite
Integration pts. = 7
Log likelihood = -329.87832 Wald chi2(2) = 10.12 Prob > chi2 = 0.0063
_t
Coef. Std. Err. z P>|z| [95% Conf. Interval]
age .0073207 .0136874 0.53 0.593 -.0195062 .0341476
female -1.658247 .5249676 -3.16 0.002 -2.687164 -.629329
_cons -4.921236 .9914009 -4.96 0.000 -6.864346 -2.978126
/ln_p .2243233 .1402795 -1.65 0.096 -.4988642 .9475108
patient var(_cons) .8308495 .4978425 .256735 2.688808
LR test vs. Weibull model: chibar2(01) = 9.40 Prob >= chibar2 = 0.0011
```
We can refit the Weibull model using the AFT parameterization by specifying option `time`.

```stata
. mestreg age female || patient:, distribution(weibull) time
   failure _d: infect
   analysis time _t: time
```

Fitting fixed-effects model:

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Log Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-346.46486</td>
</tr>
<tr>
<td>1</td>
<td>-343.29515</td>
</tr>
<tr>
<td>2</td>
<td>-335.0513</td>
</tr>
<tr>
<td>3</td>
<td>-334.58308</td>
</tr>
<tr>
<td>4</td>
<td>-334.57944</td>
</tr>
<tr>
<td>5</td>
<td>-334.57944</td>
</tr>
</tbody>
</table>

Refining starting values:

<table>
<thead>
<tr>
<th>Grid node 0</th>
<th>Log Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-335.10428</td>
</tr>
</tbody>
</table>

Fitting full model:

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Log Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-335.10428</td>
</tr>
<tr>
<td>1</td>
<td>-332.13546</td>
</tr>
<tr>
<td>2</td>
<td>-330.01623</td>
</tr>
<tr>
<td>3</td>
<td>-329.88013</td>
</tr>
<tr>
<td>4</td>
<td>-329.87832</td>
</tr>
<tr>
<td>5</td>
<td>-329.87832</td>
</tr>
</tbody>
</table>

Mixed-effects Weibull AFT regression

| Coef.       | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|-------------|-----------|-------|------|----------------------|
| age         | -0.0058496| 0.010872 | -0.54 | 0.591    | -0.0271585 - 0.0154592 |
| female      | 1.325034  | 0.3719102 | 3.56  | 0.000    | 0.596103 - 2.053964   |
| _cons       | 3.932346  | 0.5663757 | 6.94  | 0.000    | 2.82227 - 5.042422    |

Integration method: mvaghermite

<table>
<thead>
<tr>
<th>Wald chi2(2)</th>
<th>Prob &gt; chi2</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.00</td>
<td>0.0015</td>
</tr>
</tbody>
</table>

Log likelihood = -329.87832

The estimates of coefficients and variance components are different between the two models. In fact, the coefficients have the opposite signs. This is expected because the two models have different parameterizations. The relationship between the coefficients and variances of the two parameterizations for the Weibull model is

\[
\beta_{PH} = -p \times \beta_{AFT}
\]

\[
\text{var}_{PH} = p^2 \times \text{var}_{AFT}
\]

where \( p \) denotes the ancillary parameter. It is estimated in the logarithmic metric and is displayed in the output as \(/ln_p\).

For example, we could calculate \( \beta_{PH} \) for `female` as approximately \(- \exp(0.22) \times 1.33 = -1.66\). If we exponentiate this to obtain the hazard ratio that was reported in example 1, we obtain the same reported result, 0.19.
For a discussion of the differences between the PH and AFT parameterizations, see, for example, Cleves, Gould, and Marchenko (2016).

Now, we can compare the results from our Weibull specification with the results from a gamma specification.

```
.mestreg age female || patient:, distribution(gamma)
   failure _d: infect
   analysis time _t: time
```

Fitting fixed-effects model:
Iteration 0:  log likelihood = -351.17349
Iteration 1:  log likelihood = -337.04571
Iteration 2:  log likelihood = -335.10167
Iteration 3:  log likelihood = -335.09115
Iteration 4:  log likelihood = -335.09115
Refining starting values:
Grid node 0:  log likelihood = -334.49759
Fitting full model:
Iteration 0:  log likelihood = -334.49759
Iteration 1:  log likelihood = -331.87827
Iteration 2:  log likelihood = -329.64795
Iteration 3:  log likelihood = -329.52682
Iteration 4:  log likelihood = -329.52635
Iteration 5:  log likelihood = -329.52634
Mixed-effects gamma AFT regression Number of obs = 76
Group variable:  patient Number of groups = 38
Obs per group:
   min =  2
   avg =  2.0
   max =  2
Integration method: mvaghermite Integration pts. = 7
Wald chi2(2) = 13.23
Log likelihood = -329.52634 Prob > chi2 = 0.0013

|     _t | Coef.  | Std. Err. | z    | P>|z|  | [95% Conf. Interval] |
|-------|--------|-----------|------|------|---------------------|
| age   | -.0060276 | .0108267 | -0.56| 0.578| -.0272475 - .0151924 |
| female| 1.324745  | .3685132 | 3.59 | 0.000| .6024726  2.047018  |
| _cons | 3.873854  | .5628993 | 6.88 | 0.000| 2.770592  4.977117  |

/logs  -.1835075  .1008892 -.3812467  .0142317

patient
var(_cons)  .5071823  .2241959  .213254  1.206232

LR test vs. gamma model:  chibar2(01) = 11.13 Prob >= chibar2 = 0.0004

The coefficients and the random-effects variance are very similar for the two AFT models.

We can compare the marginal distributions or hazard functions for the two models by using stcurve; see example 2 in [ME] mestreg postestimation.
Example 3: Two-level random-slope model

In this example, we use a modified form of the dataset from Rabe-Hesketh and Skrondal (2012, chap. 15.7), previously published in Danahy et al. (1977) and analyzed by Pickles and Crouchley (1994,1995) and Rabe-Hesketh, Skrondal, and Pickles (2004).

angina.dta includes data on 21 patients with coronary heart disease who participated in a randomized crossover trial comparing a drug to prevent angina (chest pain) with a placebo. The participants are identified by pid.

Before receiving the drug (or placebo), participants were asked to exercise on exercise bikes to the onset of angina or, if angina did not occur, to exhaustion. The exercise time, seconds, and the result of the exercise, angina—angina (angina=1) or exhaustion (angina=0)—were recorded. The drug (treat=1) or placebo (treat=0) was then taken orally, and the exercise test was repeated one, three, and five hours (variable occasion) after drug or placebo administration. Because each exercise test can have a failure (the occurrence of angina), the test is the subject. Each test is identified by tid. Failure is indicated by the variable angina. In this case, we have eight repeated measures per study participant.

Before fitting the model, we stset our data:

```
. use http://www.stata-press.com/data/r15/angina
(Angina drug data, Rabe-Hesketh and Skrondal, 2012, ch 15.7)
. stset seconds, failure(angina) id(tid)
   id: tid
   failure event: angina != 0 & angina < .
   obs. time interval: (seconds[_n-1], seconds]
   exit on or before: failure

168  total observations
    0  exclusions

168  observations remaining, representing
168  subjects
155  failures in single-failure-per-subject data
47,267  total analysis time at risk and under observation

at risk from t = 0
earliest observed entry t = 0
last observed exit t = 743
```

To reiterate, we specify seconds as the time variable, angina as the failure variable, and tid as the variable identifying multiple observations per test.

Rabe-Hesketh and Skrondal (2012) apply several models to this dataset, including a lognormal model and a Cox model with random effects. We fit a Weibull model with covariates occasion and treat and interaction between occasion and treat. We include a random effect at the subject level.
. mestreg occasion#treat || pid:, distribution(weibull)
  failure _d: angina
  analysis time _t: seconds
  id:  tid

note: 1.occasion#1.treat identifies no observations in the sample
note: 4.occasion#1.treat omitted because of collinearity
(output omitted)

Mixed-effects Weibull PH regression
Number of obs = 168
Group variable:  pid
Number of groups = 21
Obs per group:
  min = 8
  avg = 8.0
  max = 8

Integration method: mvaghermite
Integration pts. = 7

Wald chi2(6) = 78.14
Log likelihood = -885.67135  Prob > chi2 = 0.0000

|     | Haz. Ratio | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|-----|------------|-----------|------|------|----------------------|
| _t     |            |           |      |      |                      |
| occasion |            |           |      |      |                      |
| 2       | .719456    | .2031744  | -1.17 | 0.244 | 0.4136423 1.251364 |
| 3       | .902988    | .2542476  | -0.36 | 0.717 | 0.5200146 1.568009 |
| 4       | 1.264262   | .3516347  | 0.84  | 0.399 | .7329746 2.180648 |
| treat |            |           |      |      |                      |
| Drug    | .3825531   | .128784   | -2.85 | 0.004 | .1977608 .7400195 |
| occasion#treat |           |           |      |      |                      |
| 1#Drug | 1 (empty)  |           |      |      |                      |
| 2#Drug | .1576401   | .0804767  | -3.62 | 0.000 | .0579589 .4287586 |
| 3#Drug | .4512793   | .2127706  | -1.69 | 0.091 | .1791093 1.137032 |
| 4#Drug | 1 (omitted) |           |      |      |                      |
| _cons  | 4.90e-13   | 9.98e-13  | -13.91 | 0.000 | 9.03e-15 2.66e-11 |
| /ln_p  | 1.640297   | .0689544  | 1.505149 | 1.775445 |
| pid    |            |           |      |      |                      |
| var(_cons) | 4.529641  | 1.544175  | 2.322124 | 8.835725 |

Note: Estimates are transformed only in the first equation.
Note: _cons estimates baseline hazard (conditional on zero random effects).

Because individuals were exercising without the administration of a placebo or treatment at the first occasion (occasion==1), the category for interaction between occasion==1 and treat==1 is empty.

The estimated variance at the individual level (that is, the variance between individuals) is equal to 4.53. The likelihood-ratio test shows evidence in favor of the random-effects model versus the fixed-effects model.

The parameter $p$ is $\exp(1.640297) = 5.16$, which is larger than 1. This means that the estimated hazard (conditional on the covariates and on the random effects) is a monotonically increasing function if we assume a Weibull distribution.
The model contains interaction terms for occasion and treat. Interpretation of interaction terms is usually less straightforward. Briefly, to interpret the exponentiated coefficients as conditional hazard ratios, we need to examine all the covariates in the interaction. The hazard function for pid = j, when we set occasion = k and treat = l, will be

\[ h(t) = h_0(t) \times \exp(\beta_{\text{occ}_k} + \beta_{\text{treat}_l} + \beta_{\text{occ}_k \times \text{treat}_l} + \_\text{cons} + u_j) \]

where \( \beta_{\text{occ}_k}, \beta_{\text{treat}_l}, \) and \( \beta_{\text{occ}_k \times \text{treat}_l} \) are, respectively, the coefficients for the dummies for occasion = k and treat = l and the interaction (occasion = k x treatment = l).

For example, when treat = 0, the hazard function is

\[ h(t|\text{treat} = 0, \text{occasion} = k, \text{pid} = j) = h_0(t) \times \exp(\beta_{\text{occ}_k} + \_\text{cons} + u_j) \]

where \( \beta_{\text{occ}_1} \) is equal to 0 because occasion = 1 is the base category. This means that for a given pid,

\[ \frac{h(t|\text{treat} = 0, \text{occ} = k, \text{pid} = j)}{h(t|\text{treat} = 0, \text{occ} = 1, \text{pid} = j)} = \exp(\beta_{\text{occ}_k}) \]

Notice that this is only true within pid, because different participants have different \( u_j \)s.

The coefficients have already been exponentiated, so we can see clearly that according to this model, when there is no treatment, the hazard for occasion 2 is smaller than the hazard for occasion 1. The increasing ratios indicate that the hazard increases with the occasion. Similar calculations could be performed for other interaction terms.

The easiest way to interpret models with interactions is by using `margins` and `marginsplot`, which allow us to compute and then visualize unconditional predictions and marginal effects. See [R] `margins` for more information.

Above we assumed a constant treatment effect for all individuals for each occasion. However, we may instead believe that the treatment effect varies also with individuals. This can be modeled by adding a random coefficient for the treatment, treat, at the individual level; we also include the `covariance(unstructured)` option to estimate a covariance term between the random intercept and the random slope for treat.
. mestreg occasion#treat || pid: treat, distribution(weibull) 
  > covariance(unstructured) 
    failure _d: angina 
    analysis time _t: seconds 
    id: tid 

note: 1.occasion#1.treat identifies no observations in the sample 
note: 4.occasion#1.treat omitted because of collinearity 
(output omitted)

Mixed-effects Weibull PH regression  Number of obs = 168
Group variable:  pid  Number of groups = 21
Obs per group: 
  min =  8 
  avg =  8.0 
  max =  8 
Integration method: mvaghermite  Integration pts. =  7
Log likelihood = -859.50038  Wald chi2(6) =  50.18
Prob > chi2 =  0.0000

|      | Haz. Ratio | Std. Err. |     z  | P>|z| | [95% Conf. Interval] |
|------|------------|-----------|-------|-------|----------------------|
| _t   |            |           |       |       |                      |
| occasion |          |           |       |       |                      |
| 2    |  0.5993591 |  0.1861745| -1.65 | 0.099 |  0.3260503  1.1017661 |
| 3    |  0.8643306 |  0.2560242| -0.49 | 0.623 |  0.4836650  1.5445970 |
| 4    |  1.333201  |  0.3843218|  1.00 | 0.318 |  0.7577392  2.3456944 |
| treat |            |           |       |       |                      |
| Drug |  0.2147751 |  0.1280091| -2.58 | 0.010 |  0.0667814  0.6907365 |
| occasion#treat |         |           |       |       |                      |
| 1#Drug | (empty) |           |       |       |                      |
| 2#Drug |  0.1594337 |  0.0885644| -3.31 | 0.001 |  0.0536714  0.4736058 |
| 3#Drug |  0.4632936 |  0.2273925| -1.57 | 0.117 |  0.1770402  1.2123850 |
| 4#Drug | (omitted) |           |       |       |                      |
| _cons |  6.21e-17  |  1.75e-16 | -13.20| 0.000 |  2.44e-19  1.58e-14 |

/ln_p  1.91931  0.0736166  1.775024  2.063596

pid  | var(treat) | 4.682507  1.956897  2.064178  10.62208 
| var(_cons) | 6.939041  2.372975  3.549852  13.56403 

pid  | cov(treat, _cons) | 1.73782  1.313054  1.32  0.186  -0.8357182  4.311357

Note: Estimates are transformed only in the first equation. 
Note: _cons estimates baseline hazard (conditional on zero random effects). 
LR test vs. Weibull model: chi2(3) = 229.74  Prob > chi2 = 0.0000 
Note: LR test is conservative and provided only for reference.
We obtain somewhat different estimates of hazard ratios, but our inferential conclusions remain the same. We now observe two variances in the output, the variance for the intercept at the individual level and the variance for the coefficient for treatment at the individual level. The variance for the intercept is smaller because some of the variability is now explained by varying coefficients for treatment. The covariance is positive, meaning that the random slope tends to be larger for individuals who have a larger random intercept. See example 4 in [ME] mestreg postestimation for an application of predict that presents a graphical analysis of this relationship.

Three-level models

Example 4: Three-level random-slope model

Blossfeld, Golsch, and Rohwer (2007) analyze a dataset based on the German Life History Study of Mayer and Brückner (1989), collected in the years 1981–1983. The jobhistory dataset contains a modified version of Blossfeld, Golsch, and Rohwer’s anonymization of a random sample of 201 respondents from the original data. Each of the 600 observations in the dataset corresponds to a job episode. Variable id contains identification of the individual, tstart contains the starting point of the job (in months from the beginning of the century), tend is the end of the job episode, and failure indicates whether the date in tend corresponds to the actual end of the employment in a certain job or whether it is a censored observation.

We first stset the data. As explained in Cleves (1999) and Therneau and Grambsch (2000), when analyzing multiple-failure data, we can consider two main approaches. One approach is to define the study time from the first time that an individual starts being at risk. The second approach is to define the study time from the last failure. We will take the second approach, which means that we treat each job episode as the subject.

Therefore, the origin is defined as the start of each job episode, and the study time will be the time from the start of each episode until the jobs end or the episode is censored.

```
. use http://www.stata-press.com/data/r15/jobhistory
(Job history data, Blossfeld et al 2007)
. stset tend, origin(tstart) failure(failure)
    failure event: failure != 0 & failure < .
    obs. time interval: (origin, tend]
    exit on or before: failure
    t for analysis: (time-origin)
    origin: time tstart

600  total observations
    0  exclusions
600  observations remaining, representing
458  failures in single-record/single-failure data
40,782  total analysis time at risk and under observation
769  at risk from t =
    earliest observed entry t = 0
    last observed exit t = 428
```

We want to fit a Weibull model using the education level, the number of previous jobs, the prestige of the current job, and gender as explanatory variables. education records the highest education level before entering the labor market, njobs contains the number of previous jobs for each individual, and prestige is an index for the prestige of the current job. The birthyear variable indicates the
The estimated variance of the random intercept is equal to 1.02

According to this model, an increase in the number of previous jobs is negatively associated with job mobility; the same is true for an increase in the prestige of the current job. By contrast, an increase in the years of education is positively associated with job mobility. Also, women seem to be more mobile than men.
We now store our estimates for later use:

```
  . estimates store randint
```

The dataset has only two natural levels. However, for illustration purposes, let’s consider the following situation. Assume that we want to account for unobserved variables associated with the date of birth, such as life experience, level of familiarity with new technologies, and family situation. We therefore add a random effect for the year of birth. Now, individuals will be nested within birth years.

```
  . mestreg education njobs prestige i.female || birthyear: || id:,
       > distribution(weibull)
       > failure _d: failure
       > analysis time _t: (tend-origin)
       > origin: time tstart

(output omitted)
```

Mixed-effects Weibull PH regression

<table>
<thead>
<tr>
<th>Group Variable</th>
<th>No. of Observations per Group</th>
<th>Number of obs = 600</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Groups Minimum Average Maximum</td>
<td></td>
</tr>
<tr>
<td>birthyear</td>
<td>12  3  50.0  99</td>
<td></td>
</tr>
<tr>
<td>id</td>
<td>201  1  3.0  9</td>
<td></td>
</tr>
</tbody>
</table>

Integration method: mvaghermite Integration pts. = 7

Log likelihood = -2439.9066 Wald chi2(4) = 83.20

_\_t | Haz. Ratio | Std. Err. | z | P>|z| | [95% Conf. Interval] |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>education</td>
<td>1.120373</td>
<td>.045203</td>
<td>2.82</td>
<td>0.005</td>
<td>1.035189 1.212566</td>
</tr>
<tr>
<td>njobs</td>
<td>.7181197</td>
<td>.0372039</td>
<td>-6.39</td>
<td>0.000</td>
<td>.6487813 .7948686</td>
</tr>
<tr>
<td>prestige</td>
<td>.966567</td>
<td>.0069189</td>
<td>-4.75</td>
<td>0.000</td>
<td>.9531009 .9802234</td>
</tr>
<tr>
<td>1.female</td>
<td>1.734236</td>
<td>.3022479</td>
<td>3.16</td>
<td>0.002</td>
<td>1.232419 2.440384</td>
</tr>
<tr>
<td>_cons</td>
<td>.0059091</td>
<td>.0031758</td>
<td>-9.55</td>
<td>0.000</td>
<td>.0020609 .0169429</td>
</tr>
</tbody>
</table>

_/ln_p | .1685641 | .0454824 | .0794203 | .257708 |

birthyear
var(_cons) | .0950371 | .0741445 | .0205976 | .4385006 |

birthyear>id
var(_cons) | .8728384 | .2020938 | .5544339 | 1.374099 |

The results for the fixed part of the model are similar to the ones in the previous model.

Now, we have two estimated variances—one estimate for the random intercept at the individual level and one estimate for the random intercept at the birth-year level.

The variance component for the individual level is smaller for this model, and it looks as if the first model might have been trying to explain a variance component at the birth-year level by incorporating it into the individual-level variance. We can perform a likelihood-ratio test to compare the stored model randint with the current model:
The test is significant at the 0.05 level. It supports the three-level model with the additional variance component at the birth-year level.

Stored results

mestreg stores the following in e():

Scalars
- \( e(N) \) number of observations
- \( e(k) \) number of parameters
- \( e(k_{-eq}) \) number of equations in \( e(b) \)
- \( e(k_{-eq\_model}) \) number of equations in overall model test
- \( e(k_{-dv}) \) number of dependent variables
- \( e(k_{-f}) \) number of fixed-effects parameters
- \( e(k_{-_r}) \) number of random-effects parameters
- \( e(k_{-rs}) \) number of variances
- \( e(k_{-rc}) \) number of covariances
- \( e(df_{_m}) \) model degrees of freedom
- \( e(ll) \) log likelihood
- \( e(chi2) \) \( \chi^2 \)
- \( e(p) \) \( p \)-value for model test
- \( e(ll_{-_c}) \) log likelihood, comparison model
- \( e(chi2_{-_c}) \) \( \chi^2 \), comparison test
- \( e(df_{-_c}) \) degrees of freedom, comparison test
- \( e(p_{-_c}) \) \( p \)-value for comparison test
- \( e(N_{-_clust}) \) number of clusters
- \( e(rank) \) rank of \( e(V) \)
- \( e(ic) \) number of iterations
- \( e(rc) \) return code
- \( e(converged) \) 1 if converged, 0 otherwise

Macros
- \( e(cmd) \) \texttt{gsem}
- \( e(cmd2) \) \texttt{mestreg}
- \( e(cmdline) \) command as typed
- \( e(depvar) \) name of dependent variable
- \( e(type) \) weight type
- \( e(wtype) \) weight expression (first-level weights)
- \( e(fweight{k}) \) \( fweight \) variable for \( k \)th highest level, if specified
- \( e(iweight{k}) \) \( iweight \) variable for \( k \)th highest level, if specified
- \( e(pweight{k}) \) \( pweight \) variable for \( k \)th highest level, if specified
- \( e(covariates) \) list of covariates
- \( e(ivars) \) grouping variables
- \( e(model) \) model name
- \( e(title) \) title in estimation output
- \( e(distribution) \) distribution
Methods and formulas

Methods and formulas are presented under the following headings:

Survival models
Survey data

Survival models

Survival models have a trivariate response \((t_0, t, d)\):

- \(t_0\) is the starting time under observation \(t_0 \geq 0\);
- \(t\) is the ending time under observation \(t \geq t_0\); and
- \(d\) is an indicator for failure \(d \in \{0, 1\}\).

The survival function for a given family is the complement of the cumulative distribution function, \(S(t) = 1 - F(t)\). The unconditional density for a failure at time \(t\) is given by

\[
g(t) = \frac{\partial F(t)}{\partial t} = -\frac{\partial S(t)}{\partial t}
\]

Some distributions contain ancillary parameters that are not denoted here.
The conditional density for a failure at time \( t \) is
\[
g(t | t \geq t_0, d = 1) = \frac{g(t)}{S(t_0)}
\]
and the conditional probability of survival without failure up to time \( t \) is
\[
P(T \geq t | t \geq t_0, d = 0) = \frac{S(t)}{S(t_0)}
\]
The conditional likelihood is given by
\[
L(t, t_0, d) = \left\{ \frac{g(t)}{S(t_0)} \right\}^d \left\{ \frac{S(t)}{S(t_0)} \right\}^{1-d}
\]
See Survival distributions in [SEM] methods and formulas for gsem for the specific density function corresponding to each distribution.

Given a set of cluster-level random effects \( u_j \) for \( j = 1, \ldots, M \), the conditional distribution of \( t_j = (t_j1, \ldots, t_jn_j)' \) on \( \eta_j = X_j \beta + Z_j u_j = (x_{j1} \beta + z_{j1} u_j, \ldots, x_{jn_j} \beta + z_{jn_j} u_j) \) for cluster \( j \) is
\[
f(t_j | \eta_j) = \prod_{i=1}^{n_j} f(t_{ji} | \eta_{ji})
\]
where \( f(t_{ji} | \eta_{ji}) \) is the contribution to the likelihood from observation \( ji \); that is,
\[
f(t_{ji} | \eta_{ji}) = \left\{ \frac{g(t_{ji} | x_{ji} \beta + z_{ji} u_j)}{S(t_{0ji} | x_{ji} \beta + z_{ji} u_j)} \right\}^{d_{ji}} \left\{ \frac{S(t_{ji} | x_{ji} \beta + z_{ji} u_j)}{S(t_{0ji} | x_{ji} \beta + z_{ji} u_j)} \right\}^{1-d_{ji}}
\]
where \( g(t | \eta) \) and \( S(t | \eta) \) are, respectively, the density and the survivor function conditional on the linear prediction \( \eta \).

As mentioned in Introduction under Remarks and examples, \texttt{mestreg} does not allow delayed entry or gaps. Therefore, the first observation for a given subject will have a value of \( t_0 = 0 \), and subsequent spells for the subject must start at the end of the previous spell. That is, if observations \( ji \) and \( j, i + 1 \) belong to the same subject, then \( t_{0ji,i+1} = t_{ji} \).

Because the prior distribution of \( u_j \) is multivariate normal with mean \( 0 \) and \( q \times q \) variance matrix \( \Sigma \), the likelihood contribution for the \( j \)th cluster is obtained by integrating \( u_j \) out of the joint density \( f(t_j, u_j) \),
\[
L_j(\beta, \Sigma) = (2\pi)^{-q/2} |\Sigma|^{-1/2} \int f(t_j | X_j \beta + Z_j u_j) \exp \left( -u_j' \Sigma^{-1} u_j / 2 \right) du_j
\]
The integration in (2) has no closed form and thus must be approximated; see Methods and formulas in [ME] meglm for details.
Survey data

In the presence of sampling weights, following Rabe-Hesketh and Skrondal (2006), the weighted log pseudolikelihood for a two-level model is given as

\[
\mathcal{L}(\beta, \Sigma) = \sum_{j=1}^{M} w_j \log \int_{-\infty}^{\infty} \exp \left\{ \sum_{i=1}^{n_j} w_{i|j} \log f(t_{ji}|\eta_{ji}) \right\} \phi(v_{j1}) \, dv_{j1}
\]

where \( w_j \) is the inverse of the probability of selection for the \( j \)th cluster; \( w_{i|j} \) is the inverse of the conditional probability of selection of individual \( i \), given the selection of cluster \( j \); \( f(t_{ji}|\eta_{ji}) \) is as in (1); and \( \eta_{ji}, \phi(\cdot), v_{j1} \) are defined as in Methods and formulas in [ME] meglm.

Weighted estimation is achieved through the direct application of \( w_j \) and \( w_{i|j} \) into the likelihood calculations as detailed above to reflect replicated clusters for \( w_j \) and replicated observations within clusters for \( w_{i|j} \). Because this estimation is based on replicated clusters and observations, frequency weights are handled similarly.

References


Also see

[ME] mestreg postestimation — Postestimation tools for mestreg
[ME] me — Introduction to multilevel mixed-effects models
[BAYES] bayes: mestreg — Bayesian multilevel parametric survival model
[ST] streg — Parametric survival models
[ST] Glossary
[SVY] svy estimation — Estimation commands for survey data
[XT] xtstreg — Random-effects parametric survival models
[U] 20 Estimation and postestimation commands