**menl — Nonlinear mixed-effects regression**

### Description

`menl` fits nonlinear mixed-effects models in which some or all fixed and random effects enter nonlinearly. These models are also known as multilevel nonlinear models or hierarchical nonlinear models. The overall error distribution of the nonlinear mixed-effects model is assumed to be Gaussian. Different covariance structures are provided to model random effects and to model heteroskedasticity and correlations within lowest-level groups.

### Quick start

Nonlinear mixed-effects regression of $y$ on $x_1$ and $x_2$ with random intercepts $B_0$ by $id$

```
menl y = {a}*(1-exp(-({b0}+{b1}*x1+{b2}*x2+{B0}[id])))
```

As above, but using the more efficient specification of the linear combination

```
menl y = {a}*(1-exp(-{xb: x1 x2 B0[id]}))
```

As above, but using `define()` to specify the linear combination

```
menl y = {a}*(1-exp(-{xb:})), define(xb: x1 x2 B0[id])
```

As above, but perform restricted maximum-likelihood estimation instead of the default maximum-likelihood estimation

```
menl y = {a}*(1-exp(-{xb:})), define(xb: x1 x2 B0[id]) reml
```

Specify your own initial values for fixed effects, but use the default expectation-maximization (EM) method to obtain initial values for random-effects parameters

```
menl y = {a}*(1-exp(-{xb:})), define(xb: x1 x2 B0[id])
    initial({a} 1 {xb:x1} 1 {xb:x2} 0.5 {xb:_cons} 2, fixed)
```

Include random intercepts $A_0$ by $id$ to allow parameter $a$ to vary between levels of $id$, and specify the `xb` suboption to indicate that $a$ contains a linear combination rather than a scalar parameter

```
menl y = {a}*(1-exp(-{xb:})), define(xb: x1 x2 B0[id])
    define(a: A0[id], xb)
```

Include a random slope on continuous variable $x_2$ in the linear combination, and allow correlation between random slopes $B_1$ and intercepts $B_0$

```
menl y = {a}*(1-exp(-{xb:})), define(xb: x1 x2 B0[id] c.x2*B1[id])
    covariance(B0 B1, unstructured)
```

Specify a heteroskedastic within-subject error variance that varies as a power of $x_2$

```
menl y = {a}*(1-exp(-{xb:})), define(xb: x1 x2 B0[id] c.x2*B1[id])
    covariance(B0 B1, unstructured) resvariance(power x2)
```
Display random-effects and within-group error parameters as standard deviations and correlations

\texttt{menl, stddeviations}

Fit a nonlinear marginal regression of $y$ on variables $x_1$, $x_2$, and $x_3$ with an exchangeable covariance structure for the within-id errors

\begin{verbatim}
menl y = \{phi1\}*(1-exp(-0.5*(x1-\{phi2: x2 i.x3\}))), ///
    rescovariance(exchangeable, group(id))
\end{verbatim}

Three-level nonlinear regression of $y$ on variable $\text{time}$ and factor variable $f$ with random intercepts $S_0$ by $\text{lev3}$ and $W_0$ by $\text{lev2}$ nested within $\text{lev3}$, using an AR(1) correlation structure for the residuals

\begin{verbatim}
menl y = \{phi1:\}+\{phi2:\}*exp(-\{phi3\}*\text{time}), ///
    define(phi1: i.f S0[lev3]) define(phi2: i.f W0[lev3>lev2]) ///
    rescorrelation(ar 1, t(time))
\end{verbatim}

Three-level nonlinear regression of $y$ on $x_1$ with random intercepts $W_0$ and slopes $W_1$ on continuous $x_1$ by $\text{lev3}$ and with random intercepts $S_0$ and slopes $S_1$ on $x_1$ by $\text{lev2}$ nested within $\text{lev3}$, using unstructured covariance for $W_0$ and $W_1$ and exchangeable covariance for $S_0$ and $S_1$

\begin{verbatim}
menl y = \{phi1:\}+\{b1\}*cos(\{b2\}*x1), ///
    define(phi1:x1 W0[lev3] S0[lev3>lev2] ///
    c.x1#(W1[lev3] S1[lev3>lev2])) ///
    covariance(W0 W1, unstructured) ///
    covariance(S0 S1, exchangeable)
\end{verbatim}

As above, but assume that residuals are independent but have different variances for males and females

\begin{verbatim}
menl y = \{phi1:\}+\{b1\}*cos(\{b2\}*x1), ///
    define(phi1:x1 W0[lev3] S0[lev3>lev2] ///
    c.x1#(W1[lev3] S1[lev3>lev2])) ///
    covariance(W0 W1, unstructured) ///
    covariance(S0 S1, exchangeable) ///
    rescovariance(identity, by(female))
\end{verbatim}

Menu

Statistics > Multilevel mixed-effects models > Nonlinear regression
Syntax

menl  depvar = <menlexpr>  [ if ]  [ in ]  [ , options ]

<menlexpr> defines a nonlinear regression function as a substitutable expression that contains
model parameters and random effects specified in braces {}, as in \( \exp(\{b\}+U[id]) \); see
Random-effects substitutable expressions for details.

options

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>mle</td>
<td>fit model via maximum likelihood; the default</td>
</tr>
<tr>
<td>reml</td>
<td>fit model via restricted maximum likelihood</td>
</tr>
<tr>
<td>define(name:&lt;resubexpr&gt;)</td>
<td>define a function of model parameters; this option may be repeated</td>
</tr>
<tr>
<td>covariance(covspec)</td>
<td>variance–covariance structure of the random effects; this option may be repeated</td>
</tr>
<tr>
<td>initial(initial_values)</td>
<td>initial values for parameters</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Residuals</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>rescovariance(rescovspec)</td>
<td>covariance structure for within-group errors</td>
</tr>
<tr>
<td>resvariance(resvarspec)</td>
<td>heteroskedastic variance structure for within-group errors</td>
</tr>
<tr>
<td>rescorrelation(rescorrspec)</td>
<td>correlation structure for within-group errors</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time series</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>tsorder(varname)</td>
<td>specify time variable to determine the ordering for time-series operators</td>
</tr>
<tr>
<td>tsinit(name:=&lt;resubexpr&gt;)</td>
<td>specify initial conditions for lag operators used with named expressions; this option may be repeated</td>
</tr>
<tr>
<td>tsmissing</td>
<td>keep observations with missing values in depvar in computation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reporting</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>level(#)</td>
<td>set confidence level; default is level(95)</td>
</tr>
<tr>
<td>variance</td>
<td>show random-effects and within-group error parameter estimates as variances and covariances; the default</td>
</tr>
<tr>
<td>stddeviations</td>
<td>show random-effects and within-group error parameter estimates as standard deviations and correlations</td>
</tr>
<tr>
<td>noretable</td>
<td>suppress random-effects table</td>
</tr>
<tr>
<td>nofetable</td>
<td>suppress fixed-effects table</td>
</tr>
<tr>
<td>estmetric</td>
<td>show parameter estimates as stored in e(b)</td>
</tr>
<tr>
<td>nolegend</td>
<td>suppress table expression legend</td>
</tr>
<tr>
<td>noheader</td>
<td>suppress output header</td>
</tr>
<tr>
<td>nogroup</td>
<td>suppress table summarizing groups</td>
</tr>
<tr>
<td>nostderr</td>
<td>do not estimate standard errors of random-effects parameters</td>
</tr>
<tr>
<td>lrtest</td>
<td>perform a likelihood-ratio test to compare the nonlinear mixed-effects model with ordinary nonlinear regression</td>
</tr>
<tr>
<td>notsshow</td>
<td>do not show ts setting information</td>
</tr>
</tbody>
</table>

| display_options | control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling |
EM options

- `emiterate(#)`: number of EM iterations; default is `emiterate(25)`
- `emtolerance(#)`: EM convergence tolerance; default is `emtolerance(1e-10)`
- `emlog`: show EM iteration log

Maximization

- `menlmaxopts`: control the maximization process
- `coeflegend`: display legend instead of statistics
  
  `coeflegend` does not appear in the dialog box.

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

The syntax of `covspec` is

```
rename1 rename2 […] , vartype
```

<table>
<thead>
<tr>
<th>vartype</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>independent</code></td>
<td>one unique variance parameter per random effect; all covariances are 0; the default</td>
</tr>
<tr>
<td><code>exchangeable</code></td>
<td>equal variances for random effects and one common pairwise covariance</td>
</tr>
<tr>
<td><code>identity</code></td>
<td>equal variances for random effects; all covariances are 0</td>
</tr>
<tr>
<td><code>unstructured</code></td>
<td>all variances and covariances to be distinctly estimated</td>
</tr>
</tbody>
</table>

The syntax of `rescovspec` is

```
rescov [ , rescovopts ]
```

<table>
<thead>
<tr>
<th>rescov</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>identity</code></td>
<td>uncorrelated within-group errors with one common variance; the default</td>
</tr>
<tr>
<td><code>independent</code></td>
<td>uncorrelated within-group errors with distinct variances</td>
</tr>
<tr>
<td><code>exchangeable</code></td>
<td>within-group errors with equal variances and one common covariance</td>
</tr>
<tr>
<td><code>ar [#]</code></td>
<td>within-group errors with autoregressive (AR) structure of order #, AR(#)</td>
</tr>
<tr>
<td><code>ma [#]</code></td>
<td>within-group errors with moving-average (MA) structure of order #, MA(#)</td>
</tr>
<tr>
<td><code>ctar1</code></td>
<td>within-group errors with continuous-time AR(1) structure</td>
</tr>
<tr>
<td><code>toeplitz [#]</code></td>
<td>within-group errors have Toeplitz structure of order #; toeplitz implies that all matrix off-diagonals be estimated</td>
</tr>
<tr>
<td><code>banded [#]</code></td>
<td>within-group errors with distinct variances and covariances within first # off-diagonals; banded implies all matrix bands (unstructured)</td>
</tr>
<tr>
<td><code>unstructured</code></td>
<td>within-group errors with distinct variances and covariances</td>
</tr>
</tbody>
</table>
The syntax of `resvarspec` is

```
resvarfunc [, resvaropts]
```

**resvarfunc**

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>identity</strong></td>
</tr>
<tr>
<td>equal within-group error variances; the default</td>
</tr>
<tr>
<td><strong>linear varname</strong></td>
</tr>
<tr>
<td>within-group error variance varies linearly with varname</td>
</tr>
<tr>
<td>**power varname</td>
</tr>
<tr>
<td>variance function is a power of varname or of predicted mean</td>
</tr>
<tr>
<td>**exponential varname</td>
</tr>
<tr>
<td>variance function is exponential of varname or of predicted mean</td>
</tr>
<tr>
<td><strong>distinct</strong></td>
</tr>
<tr>
<td>distinct within-group error variances</td>
</tr>
</tbody>
</table>

The syntax of `rescorrspec` is

```
rescorr [, rescorropts]
```

**rescorr**

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>identity</strong></td>
</tr>
<tr>
<td>uncorrelated within-group errors; the default</td>
</tr>
<tr>
<td><strong>exchangeable</strong></td>
</tr>
<tr>
<td>within-group errors with one common correlation</td>
</tr>
<tr>
<td><strong>ar [#]</strong></td>
</tr>
<tr>
<td>within-group errors with AR(#) structure; ar 1 is implied by ar</td>
</tr>
<tr>
<td><strong>ma [#]</strong></td>
</tr>
<tr>
<td>within-group errors with MA(#) structure; ma 1 is implied by ma</td>
</tr>
<tr>
<td><strong>ctar1</strong></td>
</tr>
<tr>
<td>within-group errors with continuous-time AR(1) structure</td>
</tr>
<tr>
<td><strong>toeplitz [#]</strong></td>
</tr>
<tr>
<td>within-group errors have Toeplitz correlation structure of order #; toeplitz implies that all matrix off-diagonals be estimated</td>
</tr>
<tr>
<td><strong>banded [#]</strong></td>
</tr>
<tr>
<td>within-group errors with distinct correlations within first # off-diagonals; banded implies all matrix bands (unstructured)</td>
</tr>
<tr>
<td><strong>unstructured</strong></td>
</tr>
<tr>
<td>within-group errors with distinct correlations</td>
</tr>
</tbody>
</table>

**Options**

- **Model**
  - `mle` and `reml` specify the statistical method for fitting the model.
    - `mle`, the default, specifies that the model be fit using maximum likelihood (ML).
    - `reml` specifies that the model be fit using restricted maximum likelihood (REML), also known as residual maximum likelihood.
  - `define(name:<resubexpr>)` defines a function of model parameters, `<resubexpr>`, and labels it as `name`. This option can be repeated to define multiple functions. The `define()` option is useful for expressions that appear multiple times in the main nonlinear specification. The `menlexpr` you define the expression once and then simply refer to it by using `{name:}` in the nonlinear specification. This option can also be used for notational convenience. See Random-effects substitutable expressions for how to specify `<resubexpr>`. `<resubexpr>` within `define()` may not contain the lagged predicted mean function.
  - `covariance(rename1 rename2 [...], vartype)` specifies the structure of the covariance matrix for the random effects. `rename1, rename2, and so on, are the names of the random effects to be correlated (see Random effects), and vartype is one of the following: independent, exchangeable, identity, or unstructured. Instead of `renames`, you can specify `restub*` to refer to random effects that share the same `restub` in their names.
**independent** allows for a distinct variance for each random effect and assumes that all covariances are 0; the default.

**exchangeable** specifies one common variance for all random effects and one common pairwise covariance.

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

**unstructured** allows for all variances and covariances to be distinct. If \( p \) random effects are specified, the unstructured covariance matrix will have \( p(p+1)/2 \) unique parameters.

**initial** \((\text{initial-values})\) specifies the initial values for model parameters. You can specify a \( 1 \times k \) matrix, where \( k \) is the total number of parameters in the model, or you can specify a parameter name, its initial value, another parameter name, its initial value, and so on. For example, to initialize \{alpha\} to 1.23 and \{delta\} to 4.57, you would type

\[
\text{. menl ... , initial(alpha 1.23 delta 4.57) ...}
\]

To initialize multiple parameters that have the same group name, for example, \{y:x1\} and \{y:x2\}, with the same initial value, you can simply type

\[
\text{. menl ... , initial({y:} 1) ...}
\]

For the full specification, see *Specifying initial values.*

---

**Residuals**

menl provides two ways to model the within-group error covariance structure, sometimes also referred to as residual covariance structure in the literature. You can model the covariance directly by using the **rescovariance()** option or indirectly by using the **resvariance()** and **rescorrelation()** options.

**rescovariance** \((\text{rescov } [, \text{rescovopts}]\) specifies the within-group errors covariance structure or covariance structure of the residuals within the lowest-level group of the nonlinear mixed-effects model. For example, if you are modeling random effects for classes nested within schools, then rescovariance() refers to the residual variance–covariance structure of the observations within classes, the lowest-level groups.

**rescov** is one of the following: **identity**, **independent**, **exchangeable**, **ar [#]**, **ma [#]**, **ctar1**, **toeplitz [#]**, **banded [#]**, or **unstructured**. Below, we describe each **rescov** with its specific options **rescovopts**:

**identity** \([, \text{by(byvar)}]\), the default, specifies that all within-group errors be independent and identically distributed (i.i.d.) with one common error variance \( \sigma^2_{\epsilon} \). When combined with **by(byvar)**, independence is still assumed, but you estimate a distinct variance for each category of **byvar**.

**independent**, **index**(varname) \([\text{group(grpvar)}]\) specifies that within-group errors are independent with distinct variances for each value (index) of varname. **index**(varname) is required. **group(grpvar)** is required if there are no random effects in the model.

**exchangeable** \([, \text{by(byvar)} \text{group(grpvar)}]\) assumes that within-group errors have equal variances and a common covariance.

**ar [#]**, **t(timevar)** \([\text{by(byvar)} \text{group(grpvar)}]\) assumes that within-group errors have an AR(#) structure. If # is omitted, **ar 1** is assumed. **t(timevar)** is required. For this structure, \( # + 1 \) parameters are estimated: \# AR coefficients and one overall error variance, \( \sigma^2_{\epsilon} \).
\texttt{ma[\#], t(timevar) [by(byvar) group(grpvar)]} assumes that within-group errors have an MA(\#) structure. If \# is omitted, ma 1 is assumed. \texttt{t(timevar)} is required. For this structure, \# + 1 parameters are estimated: \# MA coefficients and one overall error variance, \( \sigma^2_\epsilon \).

\texttt{ctar1, t(timevar) [by(byvar) group(grpvar)]} assumes that within-group errors have a continuous-time AR(1) structure. This is a generalization of the AR covariance structure that allows for unequally spaced and noninteger time values. \texttt{t(timevar)} is required. For this structure, two parameters are estimated: the correlation parameter, \( \rho \), and one overall error variance, \( \sigma^2_\epsilon \). The correlation between two error terms is the parameter \( \rho \) raised to a power equal to the absolute value of the difference between the \( t() \) values for those errors.

\texttt{toeplitz[\#], t(timevar) [by(byvar) group(grpvar)]} assumes that within-group errors have a Toeplitz structure of order \#, for which correlations are constant with respect to time lags less than or equal to \# and are 0 for lags greater than \#. \# is an integer between 1 and the maximum observed lag (the default). \texttt{t(timevar)} is required. For this structure, \# + 1 parameters are estimated: \# correlations and one overall error variance, \( \sigma^2_\epsilon \).

\texttt{banded[\#], index(varname) [group(grpvar)]} is a special case of unstructured that restricts estimation to the covariances within the first \# off-diagonals and sets the covariances outside this band to 0. \texttt{index(varname)} is required. \# is an integer between 0 and \( L - 1 \), where \( L \) is the number of levels of \texttt{index()}. By default, \# is \( L - 1 \); that is, all elements of the covariance matrix are estimated. When \# is 0, only the diagonal elements of the covariance matrix are estimated. \texttt{group(grpvar)} is required if there are no random effects in the model.

\texttt{unstructured, index(varname) [group(grpvar)]} assumes that within-group errors have distinct variances and covariances. This is the most general covariance structure in that no structure is imposed on the covariance parameters. \texttt{index(varname)} is required. When you have \( L \) levels of \texttt{index()}, then \( L(L + 1)/2 \) parameters are estimated. \texttt{group(grpvar)} is required if there are no random effects in the model.

\texttt{rescovopts} are \texttt{index(varname)}, \texttt{t(timevar)}, \texttt{by(byvar)}, and \texttt{group(grpvar)}.

\texttt{index(varname)} is used within the \texttt{rescovariance()} option with \texttt{rescov} independent, \texttt{banded}, or \texttt{unstructured}. \texttt{varname} is a nonnegative-integer–valued variable that identifies the observations within the lowest-level groups (for example, \texttt{obsid}). The groups may be unbalanced in that different groups may have different \texttt{index()} values, but you may not have repeated \texttt{index()} values within any particular group.

\texttt{t(timevar)} is used within the \texttt{rescovariance()} option to specify a time variable for the \texttt{ar}, \texttt{ma}, \texttt{ctar1}, and \texttt{toeplitz} structures.

With \texttt{rescov} \texttt{ar}, \texttt{ma}, and \texttt{toeplitz}, \texttt{timevar} is an integer-valued time variable used to order the observations within the lowest-level groups and to determine the lags between successive observations. Any nonconsecutive time values will be treated as gaps.

With \texttt{rescov} \texttt{ctar1}, \texttt{timevar} is a real-valued time variable.

\texttt{by(byvar)} is for use within the \texttt{rescovariance()} option and specifies that a set of distinct within-group error covariance parameters be estimated for each category of \texttt{byvar}. In other words, you can use \texttt{by()} to model heteroskedasticity. \texttt{byvar} must be nonnegative-integer valued and constant within the lowest-level groups.

\texttt{group(grpvar)} is used to identify the lowest-level groups (panels) when modeling within-group error covariance structures. \texttt{grpvar} is a nonnegative-integer–valued group membership variable. This option lets you model within-group error covariance structures at the lowest level of your model hierarchy without having to include random effects at that level in your model. This is useful, for instance, when fitting nonlinear marginal or population-averaged
models that model the dependence between error terms directly, without introducing random
effects; see example 19. In the presence of random effects at other levels of hierarchy in
your model, grpvar is assumed to be nested within those levels.

resvariance(resvarfunc[, resvaropts]) specifies a heteroskedastic variance structure of the
within-group errors. It can be used with the rescorrelation() option to specify flexible
within-group error covariance structures. The heteroskedastic variance structure is modeled as
\[ \text{Var}(\epsilon_{ij}) = \sigma^2 \cdot g^2(\delta, v_{ij}), \]
where \( \sigma \) is an unknown scale parameter, \( g(\cdot) \) is a function that models
heteroskedasticity (also known as variance function in the literature), \( \delta \) is a vector of unknown
parameters of the variance function, and \( v_{ij} \)'s are the values of a fixed covariate \( x_{ij} \) or of the
predicted mean \( \hat{\mu}_{ij} \).

resvarfunc, omitting the arguments, is one of the following: identity, linear, power, expo-
nential, or distinct, and resvaropts are options specific to each variance function.

identity, the default, specifies a homoskedastic variance structure for the within-group errors;
\[ g(\delta, v_{ij}) = 1, \] so that \( \text{Var}(\epsilon_{ij}) = \sigma^2 = \sigma^2_{\epsilon}. \)

linear varname specifies that the within-group error variance vary linearly with varname; that
is, \[ g(\delta, v_{ij}) = \sqrt{\text{varname}_{ij}}, \] so that \( \text{Var}(\epsilon_{ij}) = \sigma^2 \cdot \text{varname}_{ij}. \) varname must be positive.

power varname|_yhat[, strata(stratavar) noconstant] specifies that the within-group error variance, or more precisely the variance function, be expressed in terms of a power of either varname or the predicted mean _yhat, plus a constant term; \[ g(\delta, v_{ij}) = |v_{ij}|^{\delta_1} + \delta_2. \] If noconstant is specified, the constant term \( \delta_2 \) is suppressed. In general, three param-
eters are estimated: a scale parameter \( \sigma \), the power \( \delta_1 \), and the constant term \( \delta_2 \). When strata(stratavar) is specified, the power and constant parameters (but not the scale) are distinctly estimated for each stratum. A total number of \( 2L + 1 \) parameters are estimated \((L\text{ power parameters, } L\text{ constant parameters, and scale }\sigma), \) where \( L \) is the number of strata
declared by variable stratavar.

exponential varname|_yhat[, strata(stratavar)] specifies that the within-group error variance vary exponentially with varname or with the predicted mean _yhat; \[ g(\gamma, v_{ij}) = \exp(\gamma v_{ij}). \] Two parameters are estimated: a scale parameter \( \sigma \) and an exponential parameter \( \gamma. \) When strata(stratavar) is specified, the exponential parameter \( \gamma \) (but not scale \( \sigma \)) is distinctly estimated for each stratum. A total number of \( L + 1 \) parameters are estimated \((L\text{ exponential parameters and scale }\sigma), \) where \( L \) is the number of strata
declared by variable stratavar.

distinct, index(varname) [group(grpvar)] specifies that the within-group errors have
distinct variances, \( \sigma_i^2, \) for each value (index), \( i, \) of varname, \( v_{ij}; \)
\[ g(\delta, v_{ij}) = \delta_{v_{ij}} \] with \( \delta_{v_{ij}} = \sigma_{v_{ij}} / \sigma_1 (\delta_1 = 1 \text{ for identifiability purposes}) \) such that \( \text{Var}(\epsilon_{ij}) = \sigma^2_{v_{ij}} = \sigma^2_1 \delta^2_{v_{ij}} \) for \( i = 1, 2, \ldots, L \) and \( v_{ij} \in \{1, 2, \ldots, L\}. \) index(varname) is required. group(grpvar) is
required if there are no random effects in the model. resvariance(distinct) in combina-
tion with rescorrelation(identity) is equivalent to rescovariance(independent).

resvaropts are strata(stratavar), noconstant, index(), and group(grpvar).

strata(stratavar) is used within the resvariance() option with resvarfunc power and
exponential. strata() specifies that the parameters of the variance function \( g(\cdot) \) be distinctly estimated for each stratum. The scale parameter \( \sigma \) remains constant across strata. In contrast, rescovariance()’s by(byvar) suboption specifies that all covariance parameters, including \( \sigma \) (whenever applicable), be estimated distinctly for each category of byvar.
stratavar must be nonnegative-integer valued and constant within the lowest-level groups.
rescorrelation is one of the following: rescorr, covariance structures.

resvariance() option to specify flexible within-group error covariance structures.

rescorr is one of the following: identity, exchangeable, ar [#], ma [#], ctar1, toeplitz [#], banded [#], or unstructured.

identity, the default, specifies that all within-group error correlations be zeros.

exchangeable [, by(byvar) group(grpvar)] assumes that within-group errors have a common correlation.

ar [#], t(timevar) [by(byvar) group(grpvar)] assumes that within-group errors have an AR(#) correlation structure. If # is omitted, ar 1 is assumed. The t(timevar) option is required. For this structure, # AR coefficients are estimated.

ma [#], t(timevar) [by(byvar) group(grpvar)] assumes that within-group errors have an MA(#) correlation structure. If # is omitted, ma 1 is assumed. The t(timevar) option is required. For this structure, # MA coefficients are estimated.

toeplitz [#], t(timevar) [by(byvar) group(grpvar)] assumes that within-group errors have a continuous-time AR(1) correlation structure. The t(timevar) option is required. The correlation between two errors is the parameter $\rho$ raised to a power equal to the absolute value of the difference between the $t()$ values for those errors.

banded [#], index(varname) [group(grpvar)] is a special case of unstructured that restricts estimation to the correlations within the first # off-diagonals and sets the correlations outside this band to 0. index(varname) is required. # is an integer between 0 and $L - 1$, where $L$ is the number of levels of index(). By default, # is $L - 1$; that is, all elements of the correlation matrix are estimated. When # is 0, the correlation matrix is assumed to be identity. group(grpvar) is required if there are no random effects in the model.

unstructured, index(varname) [group(grpvar)] assumes that within-group errors have distinct correlations. This is the most general correlation structure in that no structure is imposed on the correlation parameters. index(varname) is required. group(grpvar) is required if there are no random effects in the model.
**rescor** are **index(varname)**, **t(timevar)**, **by(byvar)**, and **group(grpvar)**.

**index(varname)** is used within the **rescorrelation()** option with **rescorr** banded or unstructured. **varname** is a nonnegative-integer–valued variable that identifies the observations within the lowest-level groups (for example, **obsid**). The groups may be unbalanced in that different groups may have different **index()** values, but you may not have repeated **index()** values within any particular group.

**t(timevar)** is used within the **rescorrelation()** option to specify a time variable for the **ar**, **ma**, **ctar1**, and **toeplitz** structures.

With **rescorr ar**, **ma**, and **toeplitz**, **timevar** is an integer-valued time variable used to order the observations within the lowest-level groups and to determine the lags between successive observations. Any nonconsecutive time values will be treated as gaps.

With **rescorr ctar1**, **timevar** is a real-valued time variable.

**by(byvar)** is used within the **rescorrelation()** option and specifies that a set of distinct within-group error correlation parameters be estimated for each category of **byvar**. **byvar** must be nonnegative-integer valued and constant within the lowest-level groups.

**group(grpvar)** is used to identify the lowest-level groups (panels) when modeling within-group error correlation structures. **grpvar** is a nonnegative-integer–valued group membership variable. This option lets you model within-group error correlation structures at the lowest level of your model hierarchy without having to include random effects at that level in your model. This is useful, for instance, when fitting nonlinear marginal or population-averaged models that model the dependence between error terms directly, without introducing random effects; see example 19. In the presence of random effects at other levels of hierarchy in your model, **grpvar** is assumed to be nested within those levels.

---

**Time series**

**tsorder(varname)** specifies the time variable that determines the time order for time-series operators used in expressions; see Time-series operators. When you use time-series operators with **menl**, you must either **tsset** your data prior to executing **menl** or specify option **tsorder()**. When you specify **tsorder()**, **menl** uses the time variable **varname** to create a new temporary variable that contains consecutive integers, which determine the sort order of observations within the lowest-level group. **menl** also creates and uses the appropriate panel variable based on the hierarchy of your model specification and the estimation sample; see example 17 and example 18.

**tsinit({name:}=<resubexpr>)** specifies an initial condition for the named expression **name** used with the one-period lag operator, **L.**{name:} or **L1.**{name:}, in the model specification. **name** can be the **depvar** or the name of a function of model parameters previously defined in, for instance, option **define()**. If you include the lagged predicted mean function **L.**{depvar:} or, equivalently, **L._yhat** in your model, you must specify its initial condition in **tsinit({depvar:}={...})**. The initial condition can be expressed as a random-effects substitutable expression, **<resubexpr>**. Option **tsinit()** may be repeated. Also see Time-series operators, example 17, and example 18.

**tsmissing** specifies that observations containing system missing values (.) in **depvar** be retained in the computation when a lagged named expression is used in the model specification. Extended missing values in **depvar** are excluded. Both missing and nonmissing observations are used to evaluate the predicted nonlinear mean function but only nonmissing observations are used to evaluate the likelihood. Observations containing missing values in variables used in the model other than the dependent variable are excluded. This option is often used when subjects have intermittent **depvar** measurements and the lagged predicted mean function, **L.**{depvar:} or **L._yhat**, is used
in the model specification. Such models are common in pharmacokinetics; see example 17 and example 18.

Reporting level(#); see [R] Estimation options.

variance, the default, displays the random-effects and within-group error parameter estimates as variances and covariances.

stddeviations displays the random-effects and within-group error parameter estimates as standard deviations and correlations.

noretable suppresses the random-effects table from the output.

nofetable suppresses the fixed-effects table from the output.

estmetric displays all parameter estimates in one table using the metric in which they are stored in `e(b)`. Random-effects parameter estimates are stored as log-standard deviations and hyperbolic arctangents of correlations. Within-group error parameter estimates are stored as log-standard deviations and, when applicable, as hyperbolic arctangents of correlations. Note that fixed-effects estimates are always stored and displayed in the same metric.

nolegend suppresses the expression legend that appears before the fixed-effects estimation table when functions of parameters or named substitutable expressions are specified in the main equation or in the define() options.

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.

nostderr prevents menl from calculating standard errors for the estimated random-effects parameters, although standard errors are still provided for the fixed-effects parameters. Specifying this option will speed up computation times.

lrtest specifies to fit a reference nonlinear regression model and to use this model in calculating a likelihood-ratio test, comparing the nonlinear mixed-effects model with ordinary nonlinear regression.

notsshow prevents menl from showing the key ts variables; see [TS] tsset.

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

EM options

These options control the EM iterations that occur before estimation switches to the Lindstrom–Bates method. EM is used to obtain starting values.

emiterate(#) specifies the number of EM iterations to perform. The default is emiterate(25).

emtolerance(#) specifies the convergence tolerance for the EM algorithm. The default is emtolerance(1e-10). EM iterations will be halted once the log (restricted) likelihood changes by a relative amount less than #. At that point, optimization switches to the Lindstrom–Bates method.

emlog specifies that the EM iteration log be shown. The EM iteration log is not displayed by default.
menlmaxopts: iterate(#), tolerance(#), ltolerance(#), nrtolerance(#), nonrtolerance, pnlsopts(), lmeopts(), [no] log. The convergence is declared when either tolerance() or ltolerance() is satisfied; see Stopping rules for details.

menlmaxopts control the maximization process of the Lindstrom–Bates, the generalized nonlinear least-squares (GNLS), and the nonlinear least-squares (NLS) algorithms. The Lindstrom–Bates algorithm is the main optimization algorithm used for nonlinear models containing random effects. The GNLS algorithm is used for the models without random effects but with non-i.i.d. errors. The NLS algorithm is used for the models without random effects and with i.i.d. errors. The Lindstrom–Bates and GNLS algorithms are alternating algorithms—they alternate between two optimization steps and thus support options to control the overall optimization as well as the optimization of each step. The Lindstrom–Bates algorithm alternates between the penalized least-squares (PNLS) and the linear mixed-effects (LME) optimization steps. The GNLS algorithm alternates between the GNLS and ML or, if option reml is used, REML steps. Option.pnlsopts() controls the PNLS and GNLS steps, and option lmeopts() controls the LME and ML/REML steps. The other menlmaxopts control the overall optimization of the alternating algorithms as well as the NLS optimization.

iterate(#) specifies the maximum number of iterations for the alternating algorithms and the NLS algorithm. One alternating iteration of the Lindstrom–Bates algorithm involves #_pnlsopts PNLS iterations as specified in pnlsopts()’s iterate() suboption and #_lmeopts LME iterations as specified in lmeopts()’s iterate() suboption. Similarly, one alternating iteration of the GNLS algorithm involves #_gnlsopts GNLS iterations and #_lmeopts ML/REML iterations. The default is the number set using set maxiter, which is 300 by default.

tolerance(#) specifies the tolerance for the parameter vector in the alternating algorithms and the NLS algorithm. When the relative change in the parameter vector from one (alternating) iteration to the next is less than or equal to tolerance(), the parameter convergence is satisfied. The default is tolerance(1e-6).

ltolerance(#) specifies the tolerance for the linearization log likelihood of the Lindstrom–Bates algorithm and for the log likelihood of the GNLS and NLS algorithms. The linearization log likelihood is the log likelihood from the LME optimization step in the last iteration. When the relative change in the log likelihood from one (alternating) iteration to the next is less than or equal to ltolerance(), the log-likelihood convergence is satisfied. The default is ltolerance(1e-7).

nrtolerance(#) and nonrtolerance control the tolerance for the scaled gradient.

nrtolerance(#) specifies the tolerance for the scaled gradient. Convergence is declared when \( g(-H^{-1})g' \) is less than nrtolerance(#), where \( g \) is the gradient row vector and \( H \) is the approximated Hessian matrix from the current iteration. The default is nrtolerance(1e-5).

nonrtolerance specifies that the default nrtolerance() criterion be turned off.

nrtolerance(#) and nonrtolerance are allowed only with the NLS algorithm.

pnlsopts(pnlsopts) controls the PNLS optimization step of the Lindstrom–Bates alternating algorithm and the GNLS optimization step of the GNLS alternating algorithm. pnlsopts include any of the following: iterate(#), ltolerance(#), tolerance(#), nrtolerance(#), and maximize_options. The convergence of this step within each alternating iteration is declared when nrtolerance() and one of tolerance() or ltolerance() are satisfied. This option is not allowed with the NLS algorithm.

iterate(#) specifies the maximum number of iterations for the PNLS and GNLS optimization steps of the alternating algorithms. The default is iterate(5).
ltolerance(#) specifies the tolerance for the objective function in the PNLS and GNLS optimization steps. When the relative change in the objective function from one PNLS or GNLS iteration to the next is less than or equal to ltolerance(), the objective-function convergence is satisfied. The default is ltolerance(1e-7).

tolerance(#) specifies the tolerance for the vector of fixed-effects parameters. When the relative change in the coefficient vector from one PNLS or GNLS iteration to the next is less than or equal to tolerance(), the parameter convergence criterion is satisfied. The default is tolerance(1e-6).

nrtolerance(#) specifies the tolerance for the scaled gradient in the PNLS and GNLS optimization steps. Convergence is declared when $g(-H^{-1})g'$ is less than nrtolerance(#), where $g$ is the gradient row vector and $H$ is the approximated Hessian matrix from the current iteration. The default is nrtolerance(1e-5).

maximize_options are [no]log, trace, showtolerance, nonrtolerance; see [R] Maximize.

lmeopts(lmeopts) controls the LME optimization step of the Lindstrom–Bates alternating algorithm and the ML/REML optimization step of the GNLS alternating algorithm. lmeopts include any of the following: iterate(#), ltolerance(#), tolerance(#), nrtolerance(#), and maximize_options. The convergence of this step within each alternating iteration is declared when nrtolerance() and one of tolerance() or ltolerance() are satisfied. This option is not allowed with the NLS algorithm.

iterate(#) specifies the maximum number of iterations for the LME and ML/REML optimization steps of the alternating algorithms. The default is iterate(5).

ltolerance(#) specifies the tolerance for the log likelihood in the LME and ML/REML optimization steps. When the relative change in the log likelihood from one LME or ML/REML iteration to the next is less than or equal to ltolerance(), the log-likelihood convergence is satisfied. The default is ltolerance(1e-7).

tolerance(#) specifies the tolerance for the random-effects and within-group error covariance parameters. When the relative change in the vector of parameters from one LME or ML/REML iteration to the next is less than or equal to tolerance(), the convergence criterion for covariance parameters is satisfied. The default is tolerance(1e-6).

nrtolerance(#) specifies the tolerance for the scaled gradient in the LME and ML/REML optimization steps. Convergence is declared when $g(-H^{-1})g'$ is less than nrtolerance(#), where $g$ is the gradient row vector and $H$ is the approximated Hessian matrix from the current iteration. The default is nrtolerance(1e-5).

maximize_options are [no]log, trace, gradient, showstep, hessian, showtolerance, nonrtolerance; see [R] Maximize.

[no]log; see [R] Maximize.

The following option is available with menl but is not shown in the dialog box: coeflegend; see [R] Estimation options.
Remarks and examples

Remarks are presented under the following headings:

Introduction
Random-effects substitutable expressions
Substitutable expressions
Linear combinations
Linear forms versus linear combinations
Random effects
Multilevel specifications
Time-series operators
Summary
Specifying initial values
Two-level models
Testing variance components
Random-effects covariance structures
Heteroskedastic within-group errors
Restricted maximum likelihood
Pharmacokinetic modeling
Single-dose pharmacokinetic modeling
Multiple-dose pharmacokinetic modeling
Nonlinear marginal models
Three-level models
Obtaining initial values
Linearization approach to finding initial values
Graphical approach to finding initial values
Smart regressions approach to finding initial values
Examples of specifying initial values

Introduction

Nonlinear mixed-effects (NLME) models are models containing both fixed effects and random effects where some of, or all, the fixed and random effects enter the model nonlinearly. They can be viewed as a generalization of linear mixed-effects (LME) models (see [ME] mixed), in which the conditional mean of the outcome given the random effects is a nonlinear function of the coefficients and random effects. Alternatively, they can be considered as an extension of nonlinear regression models for independent data (see [R] nl), in which coefficients may incorporate random effects, allowing them to vary across different levels of hierarchy and thus inducing correlation within observations at the same level.

Why use NLME models? Can’t we use higher-order polynomial LME models or generalized linear mixed-effects (GLME) models instead?

In principle, any smooth nonlinear function can be approximated by a higher-order polynomial. One may argue that we can use an LME (see [ME] mixed) polynomial model and increase the order of the polynomial until we get an accurate approximation of the desired nonlinear model. There are three problems with this approach. First, parameters in NLME models often have natural physical interpretations such as half-life and limiting growth. This is not the case in LME polynomial models. For example, what is the physical interpretation of the coefficient of \( \text{time}^4 \)? Second, NLME models typically use fewer parameters than the corresponding LME polynomial model, which provides a more parsimonious summarization of the data. Third, NLME models usually provide better predictions outside the range of the observed data than predictions based on LME higher-order polynomial models.

GLME models (see [ME] meglm) are also nonlinear, but in the restricted sense that the conditional mean response given random effects is a nonlinear function of the linear predictor that contains both fixed and random effects, and only indirectly nonlinear in fixed and random effects themselves. That is, the nonlinear function must be an invertible function of the linear predictor. However, many
estimation methods for GLME and NLME models are similar because random effects enter both models nonlinerly.

Population pharmacokinetics, bioassays, and studies of biological and agricultural growth processes are just a few areas that use NLME models to analyze multilevel data such as longitudinal or repeated-measures data. Comprehensive treatments of both methodology and history of NLME models may be found in Davidian and Giltinan (1995), Vonesh and Chinchilli (1997), Demidenko (2013), and Pinheiro and Bates (2000). Davidian and Giltinan (2003) provide an excellent summary.

Consider a sample of \( M \) subjects from a population of interest, where \( n_j \) measurements, \( y_{1j}, \ldots, y_{nj} \), are observed on subject \( j \) at times \( t_{1j}, \ldots, t_{nj} \). By “subject”, we mean any distinct experimental unit, individual, panel, or cluster with two or more correlated observations. The basic nonlinear two-level model can be written as follows (in our terminology, a one-level NLME is just a nonlinear regression model for independent data),

\[
y_{ij} = \mu(x_{ij}', \beta, u_j) + \epsilon_{ij} \quad i = 1, \ldots, n_j; \quad j = 1, \ldots, M
\]  

(1)

where \( \mu(\cdot) \) is a real-valued function that depends on a \( p \times 1 \) vector of fixed effects \( \beta \), a \( q \times 1 \) vector of random effects \( u_j \), which are distributed as multivariate normal with mean \( \mathbf{0} \) and variance–covariance matrix \( \Sigma \), and a covariate vector \( x_{ij} \) that contains both within-subject covariates \( x_{ij}^w \) and between-subject covariates \( x_{ij}^b \). The \( n_j \times 1 \) vector of errors \( \epsilon_j = (\epsilon_{1j}, \ldots, \epsilon_{nj})' \) is assumed to be multivariate normal with mean \( \mathbf{0} \) and variance–covariance matrix \( \sigma^2 \Lambda_j \), where depending on \( \Lambda_j \), \( \sigma^2 \) is either a within-group error variance \( \sigma^2_\epsilon \) or a squared scale parameter \( \sigma^2 \).

Parameters of NLME models often have scientifically meaningful interpretations, and research questions are formed based on them. To allow parameters to reflect phenomena of interest, (1) can be equivalently formulated as a two-stage hierarchical model as follows:

Stage 1: Individual-level model

\[
y_{ij} = m(x_{ij}^w, \phi_j) + \epsilon_{ij} \quad i = 1, \ldots, n_j
\]  

Stage 2: Group-level model

\[
\phi_j = d(x_{ij}^b, \beta, u_j) \quad j = 1, \ldots, M
\]  

(2)

In stage 1, we model the response by using a function \( m(\cdot) \), which describes within-subject behavior. This function depends on subject-specific parameters \( \phi_j \)'s, which have a natural physical interpretation, and a vector of within-subject covariates \( x_{ij}^w \). In stage 2, we use a known vector-valued function \( d(\cdot) \) to model between-subject behavior, that is, to model \( \phi_j \)'s and to explain how they vary across subjects. The \( d(\cdot) \) function incorporates random effects and, optionally, a vector of between-subject covariates \( x_{ij}^b \). The general idea is to specify a common functional form for each subject in stage 1 and then allow some parameters to vary randomly across subjects in stage 2.

To further illustrate (1) and (2), we consider the soybean plants data (Davidian and Giltinan 1995), in which we model the average leaf weight per soybean plant, \( y_{ij} \), in plot \( j \) at \( t_{ij} \) days after planting. Let’s first use (1):

\[
y_{ij} = \mu(x_{ij}', \beta, u_j) + \epsilon_{ij} = \frac{\beta_1 + u_{1j}}{1 + \exp[-(t_{ij} - (\beta_2 + u_{2j}))/(\beta_3 + u_{3j})]} + \epsilon_{ij}
\]

Here \( \beta = (\beta_1, \beta_2, \beta_3)' \), \( u_j = (u_{1j}, u_{2j}, u_{3j})' \), and \( x_{ij} \) is simply \( t_{ij} \).
Equivalently, we can use (2) to define our model,

\begin{align*}
\text{Stage 1: } y_{ij} &= m \left( x_{ij}^w, \phi_j \right) + \epsilon_{ij} \\
&= \frac{\phi_{1j}}{1 + \exp \left\{ -\left( t_{ij} - \phi_{2j} \right)/\phi_{3j} \right\}} + \epsilon_{ij}
\end{align*}

\begin{align*}
\text{Stage 2: } \phi_{1j} &= \beta_1 + u_{1j} \\
\phi_{2j} &= \beta_2 + u_{2j} \\
\phi_{3j} &= \beta_3 + u_{3j}
\end{align*}

where \( x_{ij}^w = t_{ij}, \phi_j = (\phi_{1j}, \phi_{2j}, \phi_{3j})' = d \left( x_j^b, \beta, u_j \right) = \beta + u_j \). A key advantage of (2) is the interpretability. \( \phi_j \)'s are parameters that characterize features of the trajectory. For example, \( \phi_{1j} \) can be interpreted as the asymptotic average leaf weight per soybean plant in plot \( j \) when \( t_{ij} \to \infty \) and \( \phi_{2j} \) as the time at which half of \( \phi_{1j} \) is reached; that is, if we set \( t_{ij} = \phi_{2j} \), then \( E(y_{ij}) = \phi_{1j}/2 \).

menl provides both representations.

The random effects \( u_j \) are not directly estimated (although they may be predicted) but instead are characterized by the elements of \( \Sigma \), known as variance components, which are estimated together with the parameters of the within-group error variance–covariance matrix \( \sigma^2 \Lambda_j \). Correlation among repeated measures is induced either indirectly through the subject-specific random effects \( u_j \) or directly through specification of the within-subject covariance matrix \( \sigma^2 \Lambda_j \). Several covariance structures are available for \( \Sigma \), similar to those allowed in \texttt{mixed}. In contrast to \texttt{mixed}, menl provides more flexible modeling of the within-subject variance and correlation structures.

menl uses the following decomposition of the \( \Lambda_j \) matrix,

\[ \Lambda_j = S_j C_j S_j \]

where \( S_j \) is diagonal with positive elements such that \( \text{Var}(\epsilon_{ij}) = \sigma^2 [S_j]_{ii}^2 \) and \( C_j \) is a correlation matrix such that \( \text{corr}(\epsilon_{ij}, \epsilon_{kj}) = [C_j]_{ij} \). \([A]_{ij}\) denotes the \( ij \)th element of matrix \( A \). Decomposition (3) of \( \Lambda_j \) allows us to separately model the variance structure (heteroskedasticity) and the correlation structure by using disjoint sets of parameters for \( C_j \) and \( S_j \). This is different from how \texttt{mixed} handles within-subject correlation, where heteroskedasticity and correlation are determined by the type of the chosen residual covariance structure. For convenience, menl accommodates the behavior of the \texttt{mixed} command for specifying residual covariance structures via the \texttt{rescovariance()} option. The more flexible modeling of the residual structures according to (3) is available via the \texttt{resvariance()} and \texttt{rescorrelation()} options.

For LME models, because the random effects \( u_j \)'s are unobserved, inference about \( \beta \) and the covariance parameters are based on the marginal likelihood obtained after integrating out the random effects. Unlike LME models, no closed-form solution is available because the random effects enter the model nonlinearly, making the integration analytically intractable in all but the simplest situations. There are two principal methods proposed in the literature for fitting NLME models. One is to use an adaptive Gauss–Hermite (AGH) quadrature to approximate the integral that appears in the expression of the marginal likelihood. The other one is to use the linearization method of Lindstrom and Bates (1990), also known as a conditional first-order linearization method, which is based on a first-order Taylor-series approximation of the mean function and essentially linearizes the mean function with respect to fixed and random effects. With the AGH method, the level of accuracy increases as the number of quadrature points increases but at the expense of increasing computational burden. The linearization method is computationally efficient because it avoids the intractable integration, but the approximation cannot be made arbitrarily accurate. Despite its potential limiting accuracy, the linearization method has proven the most popular in practice (Fitzmaurice et al. 2009, sec. 5.4.8).

The linearization method uses a first-order Taylor-series expansion of the specified nonlinear mean function to approximate it with a linear function of fixed and random effects. Thus an NLME model is approximated by an LME model, in which the fixed-effects and random-effects design matrices involve derivatives of the nonlinear mean function with respect to fixed effects (coefficients) and random effects, respectively. As such, inference after the linearization method uses the computational machinery of the LME models. For example, estimates of random effects are computed as best linear unbiased predictors (BLUPs) of random effects from the approximating LME model. The accuracy of the inferential results will depend on the accuracy of the linearization method in approximating the original NLME model. In general, asymptotic inference for the NLME models based on the linearization method is only “approximately asymptotic”, making it less accurate than the corresponding asymptotic inference for true LME models. In practice, however, the linearization method was found to perform well in many situations (for example, Pinheiro and Bates [1995]; Wolfinger and Lin [1997]; Plan et al. [2012]; and Harring and Liu [2016]).

Both ML and REML estimation are supported by `menl`. The ML estimates are based on the usual application of likelihood theory, given the distributional assumptions of the model. In small samples, ML estimation generally leads to small-sample bias in the estimated variance components. The REML method (Thompson 1962) reduces this bias by forming a set of linear contrasts of the response that do not depend on the fixed effects \( \beta \) but instead depend only on the variance components to be estimated. The likelihood is then formed based on the distribution of the linear contrasts, and standard ML methods are applied.

The next section describes how to specify nonlinear expressions containing random effects in `menl`.

Random-effects substitutable expressions

You define the nonlinear model to be fit by `menl` by using a random-effects substitutable expression, a substitutable expression that contains random effects. For example, \( \exp({b}+{U[id]}) \), \( \frac{{b1}}{{{b2}+{b3}*x+{U[id]}}} \), and \( \frac{{(b1)+{U1[id]}}}{(1+{b2}*x+{c.x#U2[id]})} \) are a few examples of such expressions. We describe them in more detail below.

Substitutable expressions

Let’s first consider substitutable expressions without random effects. Substitutable expressions are just like any other mathematical expressions involving scalars and variables, such as those you would use with Stata’s `generate` command, except that the parameters to be estimated are bound in braces. See [U] 13.2 Operators and [U] 13.3 Functions for more information on expressions.

For teaching purposes, we will start with simpler substitutable expressions that do not contain random effects. Suppose that we wish to fit the model

\[
y_{ij} = \alpha \left( 1 - e^{-(\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij})} \right) + \epsilon_{ij}
\]

where \( \alpha, \beta_0, \beta_1, \) and \( \beta_2 \) are the parameters to be estimated and \( \epsilon_{ij} \) is an error term. We could simply type

```
* menl y = {a}*(1 - exp(-({b0}+{b1}*x1+{b2}*x2)))
```

Because \( a, b0, b1, \) and \( b2 \) are enclosed in braces, `menl` knows that they are parameters in the model.
You can group several parameters together by assigning a group name (or equation name) to them. Parameters with the same group name, \( lc \) in the example below, will be grouped together in the output table:

\[
\texttt{menl } y = \{a\}\ast(1 - \exp(-{(\{lc:b0\} + \{lc:b1\} \ast x1 + \{lc:b2\} \ast x2)}))
\]

That is, parameters \( b0 \), \( b1 \), and \( b2 \) will appear together in the output table in the equation labeled \( lc \). Parameters without equation names will appear at the bottom of the output table.

Sometimes, it may be convenient to define subexpressions within the main expression. This can be done inside the expression itself or by using the \texttt{define()} option. For example,

\[
\texttt{menl } y = \{a\}\ast(1 - \exp(-{\{xb:\} })), \texttt{define(xb: } {lc:b0}+{lc:b1}\ast x1+{lc:b2}\ast x2)
\]

defines the linear predictor of the exponent in the \texttt{define()} option with label \( xb \) and then refers to it inside the exponent as \{\texttt{xb:}\}. You can define as many subexpressions as you like by using the \texttt{define()} option repeatedly. Defining subexpressions is also useful for later predictions; see, for instance, example 13.

The above is equivalent to

\[
\texttt{menl } y = \{a\}\ast(1 - \exp(-{\{xb:\} } {lc:b0}+{lc:b1}\ast x1+{lc:b2}\ast x2})
\]

Parameters \{a\}, \{lc:b0\}, \{lc:b1\}, and \{lc:b2\} are what we call “free parameters”, meaning that they are not defined by a linear form, which we describe in the next section. Free parameters are displayed with a forward slash in front of their names or their group names.

The general syntax for a free parameter is

\[
\{ \texttt{eqname:} \ \texttt{name} \}
\]

Linear combinations

Nonlinear functions will often contain linear combinations of variables. Recall our nonlinear function from \textit{Substitutable expressions}:

\[
y_{ij} = \alpha \left( 1 - e^{-(\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij})} \right) + \epsilon_{ij}
\]

Instead of explicitly specifying the linear combination that appears in the exponent, as we did in the previous section, we can use \texttt{menl}'s shorthand notation

\[
\texttt{menl } y = \{a\}\ast(1 - \exp(-{(\{lc: x1 \ x2\}))})
\]

By specifying \{lc::x1 x2\}, you are telling \texttt{menl} that you are declaring a linear combination named \( lc \) that is a function of two variables, \( x1 \) and \( x2 \). \texttt{menl} will create three parameters, named \{lc::cons\}, \{lc::x1\}, and \{lc::x2\}.

Although both specifications produce the same results, the shorthand specification is more convenient.

The general syntax for defining a linear combination is

\[
\{ \texttt{eqname: varspec} , \texttt{xb noconstant} \}
\]

where \texttt{varspec} includes a list of variables (\texttt{varlist}), a list of \texttt{random-effects terms}, or both.
The `xb` option is used to distinguish between the linear combination that contains one variable and a free parameter that has the same name as the variable and the same group name as the linear combination. For example, `{lc: x1, xb}` is equivalent to `{lc:_cons} + {lc:x1}*x1`, whereas `{lc:x1}` refers to either a free parameter `x1` with a group name `lc` or the coefficient of the `x1` variable, if `{lc:}` has been previously defined in the expression as a linear combination that involves variable `x1`; see examples below. Thus the `xb` option indicates that the specification is a linear combination rather than a single parameter to be estimated.

When you define a linear combination, a constant term is included by default (a mathematician would argue that “affine combination” is the correct terminology!). The `noconstant` option suppresses the constant.

Having defined a linear combination such as `{lc1: x1 x2}`, you can refer to its individual coefficients by using `{lc1:x1}` and `{lc1:x2}` or, more generally, `{eqname:varname}`. For example, suppose that we want to fit the following model, where the coefficient of `x1`, $\beta_1$, appears in two places in the expression:

$$ y_{ij} = \frac{1}{1 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij}} \exp \left\{ - \left( \alpha_0 + \alpha_1 z_{ij} \right) / \left( 1 + \beta_1 x_{4ij} \right) \right\} + \epsilon_{ij} $$

We use `{lc1: x1 x2 x3, noconstant}` to specify the first linear combination, which appears in the denominator outside the exponentiated expression, and then use `{lc1:x1}` to refer to $\beta_1$ in the denominator inside the exponentiated expression. We also use the `xb` option, when we specify the second linear combination that contains only one covariate $z$. Below is the full specification:

```
. menl y = 1/(1+{lc1: x1 x2 x3, nocons})*exp(-{lc2: z, xb}/(1+{lc1:x1}*x4))
```

You may also refer to a “subset” of a previously defined linear combination. For example, let’s modify our previous expression by substituting $\beta_1 x_{4ij}$ in the denominator in the exponent with the subset $\beta_1 x_{1ij} + \beta_3 x_{3ij}$ of the first linear combination:

$$ y_{ij} = \frac{1}{1 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij}} \exp \left\{ - \left( \alpha_0 + \alpha_1 z_{ij} \right) / \left( 1 + \beta_1 x_{1ij} + \beta_3 x_{3ij} \right) \right\} + \epsilon_{ij} $$

The coefficients for variables `x1` and `x3` are the same in the denominators inside and outside the exponent. We fit this model by typing

```
. menl y = 1/(1+{lc1: x1 x2 x3, nocons})*exp(-{lc2: z, xb}/1+{lc1:})
```

Just like subexpressions, linear combinations can be defined in the `define()` option. For example, the above is equivalent to

```
. menl y = 1/(1+{lc1:})*exp(-{lc2:}/1+{lc1:}), define(lc1: x1 x2 x3, nocons) //
    define(lc2: z, xb)
```
Linear forms versus linear combinations

As we mentioned in Linear combinations, the linear-combination specification is syntactically convenient. It can also be more computationally efficient when a linear combination is a linear form.

A linear combination is what we call a linear form as long as you do not refer to its coefficients or any subset of the linear combination anywhere in the expression. Linear forms are beneficial for some nonlinear commands such as nl because they make derivative computation faster and more accurate. Although menl does not fully utilize the linear-form specification in its computations, it is still important to learn to distinguish between linear forms and linear combinations.

For example, in Linear combinations, the first linear combination \{lc:\}, the linear combination \{lc2:\}, and the linear combination \{lc1:\} in the last example are all linear forms. The linear combination \{lc1:\} in the examples where we referred to \{lc1:x1\} and \{lc1:x1 x3\} is not a linear form.

In contrast to free parameters, parameters of a linear form are displayed without forward slashes in the output. Rather, they are displayed as parameters within an equation whose name is the linear combination name. Parameters of linear combinations that are not linear forms are considered free parameters.

Random effects

So far, we have restricted our discussion to substitutable expressions that do not contain random effects. Examples of random effects specified within the menl syntax are \{U1[id]\}, \{U2[id1>id2]\}, \{c.x1#U3[id]\}, and \{2.f1#U4[id]\}. These represent a random intercept at the id level, a random intercept at the id2-within-id1 level, a random slope for the continuous variable x1, and a random slope associated with the second level of the factor variable f1, respectively.

The general syntax for specifying random effects, respec, is provided below.

<table>
<thead>
<tr>
<th>Description</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>{rename[levelspec]}</td>
<td>Random intercepts rename at hierarchy levelspec</td>
</tr>
<tr>
<td>{c.varname#rename[levelspec]}</td>
<td>Random coefficients rename for continuous variable varname</td>
</tr>
<tr>
<td>{#.fvvarname#rename[levelspec]}</td>
<td>Random coefficients rename for the #th level of factor variable fvvarname</td>
</tr>
</tbody>
</table>

rename is a random-effects name. It is a Stata name that starts with a capital letter. levelspec defines the level of hierarchy and is described below.

<table>
<thead>
<tr>
<th>Description</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>levelspec</td>
<td>variable identifying the group structure for the random effect at that level</td>
</tr>
<tr>
<td>lv2 &gt; lv1</td>
<td>two-level nesting: levels of variable lv1 are nested within lv2</td>
</tr>
<tr>
<td>lv3 &gt; lv2 &gt; lv1</td>
<td>three-level nesting: levels of variable lv1 are nested within lv2,</td>
</tr>
<tr>
<td></td>
<td>which is nested within lv3</td>
</tr>
<tr>
<td>... &gt; lv3 &gt; lv2 &gt; lv1</td>
<td>higher-level nesting</td>
</tr>
</tbody>
</table>

You can equivalently specify levels in the opposite order, from the lowest level to the highest; for example, lv1 < lv2 < lv3, but they will be displayed in the canonical order, from the highest level to the lowest.

Random effects can be specified within a linear-combination specification such as \{lc_u: x1 x2 U1[id1] U2[id2>id1]\}. In this case, the curly braces around each random effect are not needed.
Let us illustrate several random-effects specifications with \texttt{menl}. In this section, we concentrate on two-level nonlinear models; see \textit{Multilevel specifications} for higher-level models.

Suppose that we want to fit the following model:

\[
y_{ij} = \frac{\alpha z_{ij} + u_{0j}}{1 + \exp\left\{-(\beta_0 + \beta_1 x_{1ij})\right\}} + \epsilon_{ij}
\]

Compared with models we considered in previous sections, this model includes random effects or, specifically, random intercepts $u_{0j}$. Suppose that these random intercepts correspond to the levels of the \text{id} variable. Then, we can include them in our model by using \{U0[id]\}, where U0 will be their name.

\[
\texttt{menl y = (\{a\}*z+\{U0[id]\})/(1+exp(-\{b0\}+\{b1\}*x1)))}
\]

A more efficient specification is to use the linear-combination notation:

\[
\texttt{menl y = \{lc1: z U0[id], nocons\}/(1+exp(-\{lc2: x1, xb\}))}
\]

The curly braces around \text{U0[id]} are removed when it is specified within a linear-combination specification.

If you need to refer to the random-effects term again in the expression, you can simply use its name. For example, suppose that our model includes the same random intercepts in both the numerator and the denominator.

\[
y_{ij} = \frac{\alpha z_{ij} + u_{0j} + u_{1j} z_{ij}}{1 + \exp\left\{-(\beta_0 + \beta_1 x_{1ij} + u_{0j})\right\}} + \epsilon_{ij}
\]

We include random intercepts $u_{0j}$’s in the second linear combination by simply referring to their name, U0:

\[
\texttt{menl y = \{lc1: z U0[id], nocons\}/(1+exp(-\{lc2: x1 U0\}))}
\]

If instead of $u_{0j}$’s, we had a different set of random intercepts, $v_{0j}$’s, in the denominator, we would need to specify a new set of random intercepts, say, \text{V0[id]}, with \texttt{menl}:

\[
\texttt{menl y = \{lc1: z U0[id], nocons\}/(1+exp(-\{lc2: x1 V0[id]\}))}
\]

The shorthand notation for referring to random effects only by name, that is, without the brackets and the \textsf{levelspar}, is also useful when specifying the \texttt{covariance()} option, especially for multilevel random effects with long-level specifications; see \textit{Multilevel specifications}.

Let’s now see how to include random slopes. Consider the following extension of the first, simpler model in this subsection:

\[
y_{ij} = \frac{\alpha z_{ij} + u_{0j} + u_{1j} z_{ij}}{1 + \exp\left\{-(\beta_0 + \beta_1 x_{1ij})\right\}} + \epsilon_{ij}
\]

Here $u_{1j}$ is a random slope for a continuous variable \text{z} and is specified as \{c.z#U1[id]\} directly or without curly braces within a linear-combination specification.

\[
\texttt{menl y = \{lc1: z U0[id] c.z#U1[id], nocons\}/(1+exp(-\{lc2: x1, xb\}))}
\]
We can also include random slopes for factor variables. To demonstrate this, let’s consider a different nonlinear model for variety. Consider the model below, where binary variables $x_{1ij}$ and $x_{2ij}$ correspond to the factor levels 1 and 2 of a factor variable $x$ that takes on values 0, 1, and 2, with 0 being the base level.

$$y_{ij} = \alpha_0 + \alpha_1 z_{1ij} - \sqrt{u_{ij}^2 + \exp(\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + u_{0j} + u_{1j} x_{1ij} + u_{2j} x_{2ij})} + \epsilon_{ij}$$

There are three random-effects terms in this model: random intercepts $u_{0j}$, random slopes $u_{1j}$ for $x_{1ij}$ (level 1 of $x$), and random slopes $u_{2j}$ for $x_{2ij}$ (level 2 of $x$). In Stata, for a factor variable $x$, we can use the factor-variable notation ([U] 11.4.3 Factor variables) to refer to its levels, 1.x for level 1 and 2.x for level 2. So, to include the three random-effects terms in `menl`, we will use `U0[id]`, 1.x#U1[id], and 2.x#U2[id], respectively.

```
.menl y = {lc1: z1, xb} - sqrt(c.w#c.w + \
    exp({lc2: i.x U0[id] 1.x#U1[id] 2.x#U2[id]}))
```

In the above specification, we used the factor-variable notations i.x to include fixed effects for all levels of $x$, except the base level, and c.w#c.w to include a square of $w$; see [U] 11.4.3 Factor variables for details. The factor-variable specification i. or any other specification that refers to multiple levels of a factor variable is not allowed when specifying random coefficients, because each level will typically require a different set of random effects. For example, if we had specified i.x#U[id] in the above example, we would have received an error.

### Multilevel specifications

In Random effects, we focused on specifying substitutable expressions containing random effects for two-level nonlinear mixed-effects models. Here we will consider higher-level models.

Suppose that we want to fit the following three-level nonlinear mixed-effects model,

$$y_{ijk} = \beta_0 + u_{0k}^{(3)} + u_{0jk}^{(2)} + \cos\left(\left(\beta_1 + u_{1k}^{(3)}\right)x_{1ijk}\right) + \epsilon_{ijk}$$

where first-level observations, indexed by $i$, are nested within second-level groups, indexed by $j$, which are nested within third-level groups, indexed by $k$.

There are three random-effects terms in this model: random intercepts $u_{0k}^{(3)}$, and random slopes for $x_1$, $u_{1k}^{(3)}$, at the third level (idk) and random intercepts $u_{0jk}^{(2)}$ at the second level (idj-nested-within-idk). We specify random intercepts and random slopes for $x_1$ at the highest hierarchical level just like we did in Random effects for two-level models. Specifically, we can use U0[idk] and c.x1#U1[idk], respectively. To specify random intercepts $u_{0jk}^{(2)}$ at the idj-nested-within-idk level, we need to use one of the levelspec specifications for two nested levels. For example, we can use U0[idk>idj]. Below is the full specification:

```
.menl y = {lc1: U0[idk] UU0[idk>idj]} + cos({lc2: x1 c.x1#U1[idk], noconstant})
```

We can also include a random slope of the $x_1$ variable at the idj-nested-within-idk level in the cosine function by specifying c.x1#U1[idk>idj] inside the `cos()` function.

```
.menl y = {lc1: U0[idk] UU0[idk>idj]} + cos({lc2: x1 c.x1#U1[idk], noconstant})
```

We can shorten the above specification by writing c.x1#U1[idk] c.x1#U1[idk>idj] more compactly as c.x1#(U1[idk] UU1[idk>idj]),

```
.menl y = {lc1: U0[idk] UU0[idk>idj]} + cos({lc2: x1 c.x1#(U1[idk] UU1[idk>idj]), noconstant})
```
Similarly, if we had a four-level model with, say, a random intercept at the idj-within-idk-within-idl level, we could specify it as \( W[idl>idk>idj] \); see levelspec for other specifications.

### Time-series operators

You can use time-series operators in the specification of your nonlinear model (see [U] 11.4.4 Time-series varlists) but with some exceptions described next. You can use time-series operators in the main nonlinear specification <menlexpr> or any random-effects substitutable expression <resubexpr>. The supported time-series operators include L. and L#, F. and F#, and D. and D#. You cannot combine time-series operators or use them with a list of variables. Also, you cannot combine time-series operators with factor variables.

You can also include the lagged predicted mean function and lagged functions of model parameters in your expressions. For brevity, we will refer to both types of lagged functions as lagged named expressions. Lagged named expressions are useful, for instance, for fitting certain pharmacokinetic models; see example 17 and example 18.

To include the lagged predicted mean function, you can use the specification L.\{depvar:} or, equivalently, L._yhat. (Do not confuse this with the lagged dependent variable specification L.depvar.) You can specify the lagged predicted mean function only in the main nonlinear specification menlexpr. To include a lagged function of model parameters, you can use the specification L.\{name:}, where name is the name of the previously defined function of model parameters. Such functions are typically defined in the define() options. Only the one-period lag operator, L. or L1., is supported with named expressions.

To use time-series operators, you must either tsset your data prior to executing menl or specify the tsorder() option with menl. You must specify time and panel variables with tsset. When you use the tsorder(varname) option, menl uses the time variable varname to determine the ordering for time-series operators. menl creates a new temporary time variable that takes on values 1, 2, ... in each panel for the estimation sample. menl also creates the appropriate panel variable and uses the newly generated variables with tsset. For two-level models, menl uses the specified level variable as the panel variable. With more than two levels, menl creates the panel variable as a variable that takes on values 1, 2, ... for the groups formed by all level variables in the estimation sample. The generated panel and time variables are labeled as <panel> and <time> in the output of tsset as displayed by menl.

When you use time-series operators with variables in the dataset, some of the observations are used to initialize the series for those variables. For example, if you include a lagged variable varname_{t-1} (L.varname) in your model, the value of varname in the first observation in each panel is used to initialize the series; see [TS] tsset. But what happens when you include a lagged named expression for which there is no existing variable in the dataset? If your named expression is a function of existing variables, the values of those variables in the first observation (in each panel) will be used to compute an initial value for the lagged named expression. For some models, a named expression can depend on its own lag; see example 17 and example 18. In this case, you must specify the initial condition for it in the tsinit() option. Note that you will always need to specify the tsinit() option for the lagged predicted mean function. The tsinit() option may be repeated and may contain functions of variables and model parameters. When you specify the tsinit() option, menl uses its value (or values in the first observation of each panel) to initialize the corresponding lagged named expression. Just like with regular time-series variables, the first observation in each panel will be excluded from the estimation sample whenever you use lagged named expressions in the model.
Summary

To summarize, here are a few rules to keep in mind when defining substitutable expressions.

1. Model parameters and random effects are bound in braces if specified directly in the expression: \{b0\}, \{U0[id]\}, etc.
2. Model parameters can be assigned group names: \{slopes:x1\}, \{slopes:x2\}, etc.
3. Random-effects names must start with a capital letter as in \{U0[id]\}, \{c.x#U1[id]\}, \{V0[id2>id1]\}, \{1.z#V1[id2>id1]\}, etc.
4. The factor-variable specification \i\., as in \{i.z#V1[id2>id1]\}, or any other specification that refers to multiple levels of a factor variable, as in \{i(1/4).z#V1[id2>id1]\}, is not allowed when specifying random coefficients.
5. Linear combinations of variables can be included using the specification
   \{eqname:varlist[, xb noconstant]\}
   For example, \{price: mpg weight i.rep78\} and \{lc: x1 x2, noconstant\}.
6. Random effects can be specified within a linear combination, in which case they should be included without curly braces, for example, \{lc_u: x1 x2 U[id]\}.
7. To specify a linear combination that contains only one variable, use the \xb\ option, for example, \{lc: x1, xb\}.
8. To refer to the previously defined linear combination again in the expression, simply use its name \{eqname:\}, for example, \{lc:\} and \{lc_u:\}.
9. You can refer to individual parameters of the linear combination by using \{eqname:_cons\} and \{eqname:varname\}, for example, \{price:_cons\} and \{price:weight\}.
10. You can refer to a “subset” of the previously defined linear combination by using \{eqname:subset\}, where \subset\ is a subset of the variables from \varlist\ used to define \eqname, as in \{price: mpg weight\}. To refer to the subset containing only one variable, use the \xb\ option, as in \{price: weight, xb\}. If a linear combination contains only one random-effects term, the \xb\ option is implied.
11. To refer to the previously defined random effects again in the expression or in the \covariance()\ option, simply use their names, such as \{U0\} and \{U1\}.
12. You can define subexpressions, including linear combinations, inside the main expression or in the \define()\ option, which can be repeated. For example,
   ```
   . menl y = \{numer:}/{denom:}, define(numer: z U0[id]) ///
   define(denom:1+exp(-{lc: x1, xb}))
   ```
13. Specify linear forms whenever possible for faster and more accurate computation of derivatives; see Linear forms versus linear combinations.
14. Model parameters that are not defined by linear forms are considered free parameters. They are included in the output with a forward slash in front of their names or group names and displayed after linear forms in the estimation table.

Specifying initial values

By default, menl uses the EM algorithm to obtain initial values, but you may often need to specify your own. You specify your own initial values in the \initial()\ option. For example, specifying the \initial(a 1.1 b -2)\ option with menl initializes parameter \{a\} to 1.1 and parameter \{b\} to -2.
When you specify your own initial values, they are used for initialization, and the EM algorithm is not performed. When you specify initial values for only a subset of model parameters, the remaining parameters are initialized with some predetermined values such as zeros for fixed-effects parameters and correlations and variances. You can specify the iterate(0) option to see the initial values that will be used by menl in the optimization.

Often, you may have good initial values for fixed-effects parameters but not for random-effects parameters. In this case, you can specify initial()’s fixed suboption to supply your own fixed-effects parameters, but use the EM algorithm to obtain initial values for the random-effects parameters.

There are three ways in which you can use the initial(initial_values) option: you can specify a vector of values, a list of values, or values for individual parameters and groups of parameters.

Specifically, initial_values is one of the following:

\[ \text{vectorname} [, \text{skip copy fixed}] \]

\[ # \ldots \], \text{copy} \]

\[ \text{paramlist} = # [\text{paramlist} = # \ldots] \], \text{fixed} \]

skip specifies that any parameters found in the specified initialization vector, vectorname, that are not also found in the model be ignored. The default action is to issue an error message.

copy specifies that the initial values be copied into the initialization vector without checking for valid column names. copy must be specified when initial values are supplied as a list of numbers.

fixed specifies that initial estimates are being supplied for the fixed effects only and that menl should still perform the EM algorithm to refine initial values for variance components. The specified initial values are used for fixed-effects parameters during the EM algorithm. If you omit fixed, menl presumes that you are specifying starting values for all parameters in e(b), and the EM algorithm will not be performed.

Examples of paramlist are param, {param}, {param1} {param2}, {param1 param2}, {grp:param1} {grp:param2} {grp:param3}, {grp:param1 param2}, and {grp:}.

Let’s describe each specification in more detail. You can specify the name of a vector containing the initial values, say, initial(b0). Vector b0 should be properly labeled with labels found in column names of e(b), unless you specify the copy option. A properly labeled vector can have fewer elements than e(b), or, if skip is specified, even more elements. A vector without labels must be of the same dimension as e(b).

Alternatively, you can supply a list of numbers to initial(), in which case copy must be specified. The list of numbers should be of length equal to the dimension of e(b). For example, if e(b) has four parameters and you type initial(1.1 0 3 -2, copy), then the four coefficients in e(b) will be initialized to 1.1, 0, 3, and -2, respectively. If instead you specify, for example, only three initial values in your list, an error will be issued.

Finally, you can initialize parameters by referring to their names. You can specify a parameter name, its initial value, another parameter name, its initial value, and so on, for example, initial(a 0.5 b -2). You can also assign the same initial value to a group of parameters. For example, initial({a b c} 1) will initialize parameters {a}, {b}, and {c} to 1 and initial({lc:x1 x2 _cons} 0) will initialize {lc:x1}, {lc:x2}, and {lc:_cons} to 0. You can assign the same initial value to all parameters with the same group name. For example, we can shorten the previous specification to initial({lc:} 0).

Depending on the situation, it may also be beneficial to specify initial values for the NLS algorithm used by menl to obtain starting values for the EM algorithm. These initial values can be specified in the parameter definition such as {a=0.5}, in which case the NLS algorithm used during the initialization
will use 0.5 as the starting value for parameter $a$ instead of the default 0. Such initialization is particularly useful for parameters used in the denominators for which zero values may lead to an undefined value of the mean function.

See Examples of specifying initial values and Obtaining initial values for examples.

Two-level models

The sole purpose of this section and its examples is to highlight the syntax of \texttt{menl} and make you familiar with how to specify substitutable expressions in \texttt{menl} and with its output. Also see an introductory example in Nonlinear models in \cite{ME me}.

We will use the data from the Longitudinal Study of Unicorn Health in Zootopia, which contain the brain weight (weight) of 30 newborn male unicorns and 30 newborn female unicorns. Measurements were collected at 13 occasions every 2 months over the first 2 years after birth (time). Based on previous studies, a model for unicorn brain shrinkage is believed to be

$$
\text{weight}_{ij} = \beta_1 + (\beta_2 - \beta_1) \exp(-\beta_3 \text{time}_{ij}) + \epsilon_{ij} \quad i = 1, 2, \ldots, 13; \ j = 1, 2, \ldots, 60
$$

Parameter $\beta_1$ represents the average brain weight of unicorns as time$_{ij}$ increases to infinity. Parameter $\beta_2$ is the average brain weight at birth (at time$_{ij} = 0$), and $\beta_3$ is a scale parameter that determines the rate at which the average brain weight of unicorns approaches the asymptotic weight $\beta_1$ (decay rate). This model can be fit with the \texttt{nl} command; see \cite{R nl}.

We will start with a simple two-level model in which we allow the asymptote parameter $\beta_1$ to vary between unicorns by replacing $\beta_1$ in the above equation with $\beta_1 + u_{0j}$,

$$
\text{weight}_{ij} = \beta_1 + u_{0j} + (\beta_2 - \beta_1 - u_{0j}) \exp(-\beta_3 \text{time}_{ij}) + \epsilon_{ij}
$$

(4)

where $\beta_1$, $\beta_2$, and $\beta_3$ are fixed-effects parameters to be estimated and $u_{0j}$ is a random intercept at the unicorn, id, level that follows a normal distribution with mean 0 and variance $\sigma^2_u$.

Equivalently, the model defined by (4) can be written as a two-stage model,

$$
\text{weight}_{ij} = \phi_{1j} + (\phi_{2j} - \phi_{1j}) \exp(-\phi_{3j} \text{time}_{ij}) + \epsilon_{ij}
$$

(5)

with the following stage 2 specification:

$$
\phi_{1j} = \beta_1 + u_{0j} \\
\phi_{2j} = \beta_2 \\
\phi_{3j} = \beta_3
$$

(6)

Parameters $\phi_{1j}$, $\phi_{2j}$, and $\phi_{3j}$ now describe the behavior of the $j$th unicorn. For example, $\phi_{1j}$ represents the brain weight of the $j$th unicorn as time$_{ij}$ increases to infinity.
> Example 1: Simple two-level model

Let’s use `menl` to first fit a single-equation model defined by (4), described above.

```
use https://www.stata-press.com/data/r16/unicorn
(Brain shrinkage of unicorns in the land of Zootopia)
mensl weight = {b1}+{U0[id]}+({b2}-{b1}-{U0[id]})*exp(-{b3}*time)
```

Obtaining starting values by EM:
Alternating PNLS/LME algorithm:
Iteration 1: linearization log likelihood = -56.9757597

Computing standard errors:
Mixed-effects ML nonlinear regression
Number of obs = 780
Group variable: id
Number of groups = 60
Obs per group:
  min = 13
  avg = 13.0
  max = 13

Linearization log likelihood = -56.97576

| weight | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] |
|--------|-------|-----------|---|-----|---------------------|
| /b1    | 4.707954 | 0.1414511 | 33.28 | 0.000 | 4.430715 - 4.985193 |
| /b2    | 8.089432 | 0.0260845 | 310.12 | 0.000 | 8.038307 - 8.140556 |
| /b3    | 4.13201  | 0.0697547 | 59.24 | 0.000 | 3.995293 - 4.268726 |

Random-effects Parameters
<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
</table>
| id: Identity
  var(U0)  | 1.189809  | 0.2180058  - 1.703887 |
  var(Residual) | .0439199  | 0.0023148 - 0.0486995 |

Notes:
1. The response variable `weight` is specified on the left-hand side of the equality sign, and parameters to be estimated are enclosed in curly braces `{b1}`, `{b2}`, and `{b3}` on the right-hand side.
2. By typing `{U0[id]}`, we specified a random intercept at the level identified by the group variable `id`, that is, the unicorn level (level two).
3. The estimation log consists of three parts:
   a. A set of EM iterations used to refine starting values. By default, the iterations themselves are not displayed, but you can display them by using the `emlog` option. NLME models may often have multiple solutions and converge to a local maximum. It is thus important to try different initial values to investigate the existence of multiple solutions and the convergence to a global maximum; see `Obtaining initial values`.
   b. A set of iterations displaying the value of the linearization log likelihood from the Lindstrom–Bates algorithm or alternating algorithm. The term “linearization” reflects the fact that the reported log likelihood corresponds to the linear mixed-effects model obtained after linearization of the specified nonlinear mean function with respect to fixed and random effects. See `Inference based on linearization` and `Stopping rules` for details about the algorithm.
   c. The message “Computing standard errors”. This is just to inform you that `menl` has finished its iterative maximization and is now reparameterizing the variance components (see `Methods and formulas`) to their natural metric and computing their standard errors. If you are interested only in the fixed-effects estimates, you can use the `nostderr` option to bypass this step.
4. The output title, “Mixed-effects ML nonlinear regression”, informs us that our model was fit using ML, the default. For REML estimates, use the reml option.

5. The header information is similar to that of the mixed command, but unlike mixed, menl in general does not report a model $\chi^2$ statistic in the header because a test of the joint significance of all fixed-effects parameters (except the constant term) may not be relevant in a nonlinear model.

6. The first estimation table reports the fixed effects. We estimate $\beta_1 = 4.71$, $\beta_2 = 8.09$, and $\beta_3 = 4.13$. Although $z$ tests against zeros are reported automatically for all fixed-effects parameters, as part of standard estimation output, they may not always be of interest or even appropriate for parameters of nonlinear models. You can use the test command ([R] test) to test hypotheses of interest or reparameterize your model so that the tests of parameters against zeros are meaningful.

7. The second estimation table shows the estimated variance components. The first section of the table is labeled id: Identity. In general, this means that our model includes random effects at the id (unicorn) level and that their variance–covariance matrix, $\Sigma$, is the identity matrix (all random effects have the same variance). In our example, because we have only one random effect, $u_{0j}$, the random-effect covariance structure is irrelevant, and the variance of the random intercept, $\sigma_u^2$, labeled as var(U0) in the output, is estimated as 1.19 with standard error 0.22.

8. The row labeled var(Residual) displays the estimated overall error variance or variance of the error term; that is, $\widehat{\text{Var}}(\epsilon_{ij}) = \widehat{\sigma}_\epsilon^2 = 0.044$.

Example 2: Two-level model as a two-stage model, using the define() option

The model from example 1 can also be specified as a two-stage model, as defined by (5) and (6), by using the define() option. The define() option is particularly useful when you have a complicated nonlinear expression, and you would like to break it down into smaller pieces. Parameters of interest that are functions of other parameters can be defined using the define() option. This will make it easier to predict them for each subject after estimation; see [ME] menl postestimation.

Below we specify the asymptote parameter, $\phi_{1j}$, by using define(). The colon (:) in \{phi1:} instructs menl that phi1 will be specified as a substitutable expression within the define() option. Parameters \{phi2\} and \{phi3\} are simple free parameters and thus do not need to be specified in define().
\[ \text{menl weight} = \phi_1: + (\phi_2 - \phi_1:) \cdot \exp(-\phi_3 \cdot \text{time}), \]

\[ \text{define}(\phi_1: = \{b1\} + \{U0[id]\}) \]

Obtaining starting values by EM:

Alternating PNLS/LME algorithm:

Iteration 1: linearization log likelihood = -56.9757597

Computing standard errors:

Mixed-effects ML nonlinear regression

<table>
<thead>
<tr>
<th>Number of obs</th>
<th>Number of groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>780</td>
<td>60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obs per group:</th>
</tr>
</thead>
<tbody>
<tr>
<td>min = 13</td>
</tr>
<tr>
<td>avg = 13.0</td>
</tr>
<tr>
<td>max = 13</td>
</tr>
</tbody>
</table>

Linearization log likelihood = -56.97576

\[ \phi_1: = \{b1\} + \{U0[id]\} \]

| weight | Coef. | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|--------|-------|-----------|------|-------|----------------------|
| \(\phi_1\) | 4.707954 | .1414511 | 33.28 | 0.000 | 4.430715 - 4.985193 |
| \(\phi_2\) | 8.089432 | .0260845 | 310.12 | 0.000 | 8.038307 - 8.140556 |
| \(\phi_3\) | 4.13201 | .0697547 | 59.24 | 0.000 | 3.995293 - 4.268726 |

<table>
<thead>
<tr>
<th>Random-effects Parameters</th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>id: Identity</td>
<td>var(U0)</td>
<td>1.189809</td>
<td>.2180053</td>
</tr>
<tr>
<td></td>
<td>var(Residual)</td>
<td>.0439199</td>
<td>.0023148</td>
</tr>
</tbody>
</table>

The results are identical to those obtained in example 1, but the estimation table now has a legend that lists the expression \(\phi_1\) defined in the model. We can suppress this legend by specifying the `nolegend` option.

We could have defined \(\phi_1\) directly in the main expression instead of in the `define()` option,

\[ \text{menl weight} = \phi_1: + \{U0[id]\} + (\phi_2 - \phi_1:) \cdot \exp(-\phi_3 \cdot \text{time}) \]

(output omitted)

but by using the `define()` option, we simplified the main expression.

### Example 3: Two-level model containing covariates

Reducing brain weight loss has been an active research area in Zootopia for the past two decades, and scientists believe that consuming rainbow cupcakes right after birth may help slow down brain shrinkage. Recall that the scale parameter \(\phi_3j\) determines the rate at which the brain weight of the \(j\)th unicorn decreases to its asymptotic value \(\phi_1j\). Hence, covariate cupcake, which represents the number of rainbow cupcakes consumed right after birth, is added to the equation of \(\phi_3j\). Also, we would like to investigate whether the asymptote, \(\phi_1j\), is gender specific, so we include the factor variable `female` in the equation for \(\phi_1j\). `femalej` is 1 if the \(j\)th unicorn is a female and 0 otherwise.

The stage 2 specification of the model defined by (5) becomes

\[ \begin{align*}
\phi_{1j} &= \beta_{10} + \beta_{11} \text{female}_j + u_{0j} \\
\phi_{2j} &= \beta_2 \\
\phi_{3j} &= \beta_{30} + \beta_{31} \text{cupcake}_j
\end{align*} \]
The define() option can be repeated, so we specify a separate define() option for $\phi_1j$, $\phi_2j$, and $\phi_3j$. We could have left $\phi_2j$ as a free parameter \{phi2\} in our specification, but we wanted to closely match the stage 2 specification (7).

```
  . menl weight = {phi1:}+({phi2:}-{phi1:})*exp(-{phi3:}*time),
  > define(phi1: {b10}+{b11}*1.female+{U0[id]})
  > define(phi2: {b2})
  > define(phi3: {b30}+{b31}*cupcake)
```

Obtaining starting values by EM:

```
Alternating PNLS/LME algorithm:
Iteration 1: linearization log likelihood = -29.0149875
```

Computing standard errors:

```
Mixed-effects ML nonlinear regression Number of obs = 780
Group variable: id Number of groups = 60
Obs per group:
  min = 13
  avg = 13.0
  max = 13
```

```
Linearization log likelihood = -29.014988
phi1: {b10}+{b11}*1.female+{U0[id]}
phi2: {b2}
phi3: {b30}+{b31}*cupcake
```

```
weight Coef. Std. Err. z P>|z| [95% Conf. Interval]
/b10 4.072752  .1627414 25.03 0.000  3.753785  4.39172
/b11 1.264407  .2299723  5.50 0.000  0.8136694  1.715144
/b2  8.088102  .0255465 316.60 0.000  8.038032  8.138172
/b30 4.706926  .1325714 35.50 0.000  4.44709   4.966761
/b31 -.2007309 .0356814 -5.63 0.000 -.2706651 -.1307966
```

```
Random-effects Parameters Estimate Std. Err. [95% Conf. Interval]
id: Identity
  var(U0) .7840578  .1438928  .5471833  1.123475
  var(Residual) .0420763  .0022176  .0379468  .0466551
```

In the table legend, /b10 and /b11 correspond, respectively, to the constant term and coefficient of 1.female in the equation for $\phi_1j$. /b2 is $\phi_2j$, and /b30 and /b31 correspond, respectively, to the constant term and coefficient for cupcake in the equation for $\phi_3j$.

Based on our results, consuming rainbow cupcakes after birth indeed slows down brain shrinkage: /b31 is roughly $-0.2$ with a 95% CI of $[-0.271, -0.131]$.

Example 4: Specifying linear combinations

A more convenient way to specify the model in example 3 is to use linear-combination specifications; see Random-effects substitutable expressions.

For example, define(phi1: {b10}+{b11}*1.female+{U0[id]}) can be replaced with define(phi1: i.female U0[id]). menl knows that we are defining $\phi_{1j}$ as a linear combination of i.female and U0[id] and thus will create a constant term and a coefficient for each level of factor
variable female and will use a coefficient of 1 for any random effect. Because female has only two levels, menl will create two coefficients for 0b.female and 1.female, respectively, but will constrain the coefficient of the base level, level 0, to be 0.

We now fit our model by using linear-combination specifications within the define() options. We explain the use of the second and third define() specifications following estimation.

```
. menl weight = {phi1:}+({phi2:}-{phi1:})*exp(-{phi3:}*time),
> define(phi1: i.female U0[id])
> define(phi2: _cons, xb)
> define(phi3: cupcake, xb)

Obtaining starting values by EM:
Alternating PNLS/LME algorithm:
Iteration 1:  linearization log likelihood = -29.0149875
Computing standard errors:
Mixed-effects ML nonlinear regression  Number of obs = 780
Group variable: id                Number of groups =  60
Obs per group:
       min =  13
       avg = 13.0
       max =  13
Wald chi2(2)  =  61.78
Linearization log likelihood = -29.014988  Prob > chi2 = 0.0000

weight
Coef. Std. Err. z  P>|z|  [95% Conf. Interval]
phi1
female
    female  1.264407   .2299723  5.50 0.000   .8136695   1.715144
    _cons  4.072752   .1627414 25.03 0.000   3.753785   4.391722
phi2
    _cons  8.088102   .0255465 316.60 0.000   8.038032   8.138172
phi3
cupcake
    _cons -.2007309  .0356814  -5.63 0.000  -.2706651  -.1307966
    _cons  4.706926   .1325714 35.50 0.000   4.44709     4.966761

Random-effects Parameters  Estimate  Std. Err.  [95% Conf. Interval]
id: Identity
    var(U0)  .7840578  .1438924   .5471839  1.123473
    var(Residual)  .0420763  .0022176  .0379468   .0466551
```

By using linear-combination specifications within the define() options, we improved the readability of the coefficient table. For example, it is now clear that _cons in the equation labeled phi3 corresponds to the constant term in the equation for $\phi_{3j}$. This term was labeled /b30 previously.

Notes:

1. The define() option interprets its specification as a random-effects substitutable expression, so you do not need to specify the curly braces (()) around the specification.
2. All rules for random-effects substitutable expressions apply to the specifications within define().
3. Following one of the rules for random-effects substitutable expressions, we used the xb option within define()s for phi2 and phi3, because their linear combinations contained only one term: _cons for phi2 and cupcake for phi3.

4. Specification define(phi2: _cons, xb) is the same as define(phi2:, xb). In other words, _cons is implied with any linear combination, unless the noconstant option is specified. We chose to include _cons directly for clarity.

5. We could have used a free parameter {phi2} instead of the linear combination {phi2: _cons, xb}, but we wanted to preserve the order in which phi1, phi2, and phi3 appear in the estimation table. See example 5, where we specify \( \phi_{2j} \) as a free parameter {phi2}.

6. In the presence of linear combinations, menl reports a joint test of significance of all coefficients (except the constant term) across all linear combinations.

7. Linear combinations containing only a constant such as {phi2:} are not listed in the table expression legend for brevity.

Example 5: Including random coefficients

In previous examples, we only illustrated how to specify random intercepts such as \{U0[id]\}, and it is bad karma to end a unicorn story without showing how to specify random coefficients or random slopes.

Continuing with our model as defined by (5) and (7), let’s suppose that the equation for the brain-weight scale parameter, \( \phi_{3j} \), is as follows:

\[
\phi_{3j} = \beta_{30} + (\beta_{31} + u_{1j})\text{cupcake}_j
\]

We incorporated a unicorn-specific random slope for variable cupcake. The random slope, \( u_{1j} \), for a continuous variable cupcake can be specified in menl as c.cupcake#U1[id], and by default, menl assumes that it is independent of the random intercept, \( u_{0j} \). (See example 9 for specifying other random-effects covariance structures.)
. menl weight = {phi1:}+({phi2}-{phi1:})*exp(-{phi3:}*time),
> define(phi1: i.female U0[id])
> define(phi3: cupcake c.cupcake#U1[id])

Obtaining starting values by EM:

Alternating PNLS/LME algorithm:
Iteration 1:  linearization log likelihood = 165.417513
Iteration 2:  linearization log likelihood = 165.420082
Iteration 3:  linearization log likelihood = 165.420112
Iteration 4:  linearization log likelihood = 165.420101

Computing standard errors:

Mixed-effects ML nonlinear regression  Number of obs = 780
Group variable: id  Number of groups = 60
Obs per group:
  min = 13
  avg = 13.0
  max = 13

Linearization log likelihood = 165.4201  Prob > chi2 = 0.0000

  phi1: i.female U0[id]
  phi3: cupcake c.cupcake#U1[id]

| weight     | Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|------------|--------|-----------|-------|------|----------------------|
| phi1       |        |           |       |      |                      |
| female     | 1.320623 | 0.2215707 | 5.96  | 0.000| 0.8863522 1.754894   |
| _cons      | 4.006823 | 0.1568268 | 25.55 | 0.000| 3.699448  4.314197   |
| phi3       |        |           |       |      |                      |
| cupcake    | -0.219661 | 0.0659984 | -3.33 | 0.001| -0.3490155 -0.0903066|
| _cons      | 4.771466  | 0.1128421 | 42.28 | 0.000| 4.5503    4.992633   |
| /phi2      | 8.087655  | 0.0179406 | 450.80| 0.000| 8.052492  8.122818   |

Random-effects Parameters

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>id: Independent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>var(U0)</td>
<td>.727464</td>
<td>.1337143</td>
</tr>
<tr>
<td>var(U1)</td>
<td>.1258914</td>
<td>.0309569</td>
</tr>
<tr>
<td>var(Residual)</td>
<td>.0208202</td>
<td>.0011403</td>
</tr>
</tbody>
</table>

In addition to the overall error variance and the random-intercept variance, we now have a random-slope variance, which is labeled var(U1) in the output. In this example, we also specified parameter $\phi_{2j}$ as a free parameter $\{\phi2\}$ instead of a linear combination as in example 4. As we mentioned in Summary, free parameters are displayed after linear combinations, so $\phi_{12}$ is listed last in the estimation table.

Previous studies of unicorns considered a model that also incorporated gender-specific variation between unicorns in asymptotic weight $\phi_{1j}$,

$$\phi_{1j} = \beta_{10} + u_{0j} + (\beta_{11} + u_{2j})_{\text{female}_j}$$

but found no statistical evidence of such variation.
If we wanted to verify this for our data, we could have fit the following model:

```
    . menl weight = {phi1:}+({phi2}-{phi1:})*exp(-{phi3:}*time), ///
    define(phi1: i.female U0[id] 1.female#U2[id]) ///
    define(phi3: cupcake c.cupcake#U1[id])
```

Compared with our previous specification, we included a new term in the `define()` option for `phi1`—a random slope for level 1 of the factor variable `female`, `1.female#U2[id]`. To include random slopes for a factor variable, we must specify random effects for each level, except the base level, of the factor variable. The specification `i.fvvarname` for referring to all levels of a factor variable is not allowed in the context of random effects, because a different set of random effects must be defined for each level. For example, if we specified `i.female#U2[id]` in our example, we would have received an error.

To summarize:

1. Use `{name}` to define free parameters such as `{b1}`.
2. Use, for example, `{U0[id]}` to define random intercepts at the id level, `{c.varname#U1[id]}` to define random slopes for continuous variable `varname` at the id level, and `{#.fvvarname#U1[id]}` for each level #, except the base level, of variable `fvvarname` to include random slopes for factor variable `fvvarname`. The specification `{i.fvvarname#U1[id]}` is not allowed.
3. Use linear-combination specifications whenever possible. Do not use `{}` around random effects when they are specified within a linear combination.
4. Use multiple `define()` options to specify parameters of interest that are functions of other parameters, and use linear-combination specifications within `define()` whenever possible.
5. Use the `xb` option within a linear combination or within `define()` whenever you specify one variable such as `define(phi1: cupcake, xb)`, one random effect such as `{phi2: U0[id], xb}`, or a constant-only linear combination such as `{phi2: _cons, xb}` or `{phi2: , xb}`. When you specify the `xb` option, the above specifications are interpreted by `menl`, respectively, as a linear combination `{phi1:cons}+{phi1:cupcake}*cupcake`, a linear combination `{phi:cons}+{U0[id]}`, and a constant term `{phi2:_cons}`.
6. Unicorns do exist in our world, they are just gray, fat, and called rhinos.

### Testing variance components

Consider data on the intensity of 23 large earthquakes in western North America between 1940 and 1980 analyzed originally by Joyner and Boore (1981) and then also by Davidian and Giltinan (1995, sec. 11.4). The objective of the study was to model the maximum horizontal acceleration (in g units), `accel`, taken at the $i$th measuring station for the $j$th earthquake, as a function of the magnitude of the quake on the Richter scale, `richter`, and the distance (in km) of the measuring station from the quake epicenter, `distance`. We are also interested in the possible effect of the soil type `soil`, soil versus rock, at a given measuring station on acceleration. The results of this study are useful to understand the nature of the damage caused by a particular earthquake and to determine the location for sensitive installations such as nuclear facilities.

Let’s consider one of the models studied by Davidian and Giltinan (1995) for these data,

\[
\log_{10}(accel_{ij}) = \phi_{1j} - \log_{10}\sqrt{distance_{ij}^2 + \exp(\phi_{2j}) - \phi_{3ij}\sqrt{distance_{ij}^2 + \exp(\phi_{2j})} + \epsilon_{ij}}
\]

(8)
where
\[\phi_{1j} = \beta_0 + \beta_1 \text{richter}_{j} + u_{1j}\]
\[\phi_{2j} = \beta_2\]
\[\phi_{3i} = \beta_3 + u_{3j}\]

and
\[u_j = \begin{bmatrix} u_{1j} \\ u_{3j} \end{bmatrix} \sim N(0, \Sigma), \text{ diagonal } \Sigma = \begin{bmatrix} \sigma_{u_1}^2 & 0 \\ 0 & \sigma_{u_3}^2 \end{bmatrix}, \text{ and } \epsilon_{ij} \sim N(0, \sigma_\epsilon^2)\]

> Example 6: Fitting an NLME model for the earthquake data

Let's fit the model defined by (8), (9), and (10) by using menl.

. use https://www.stata-press.com/data/r16/earthquake
(Earthquake intensity (Joyner and Boore, 1981))
. menl laccel = {phi1:}-log10(sqrt(c.distance#c.distance+exp({phi2})))
> -{phi3:}*sqrt(c.distance#c.distance+exp({phi2})),
> define(phi1: richter U1[quake]) define(phi3: U3[quake], xb)

Obtaining starting values by EM:

Alternating PNLS/LME algorithm:
Iteration 1: linearization log likelihood = 2.41158111
Iteration 2: linearization log likelihood = 2.40751409
Iteration 3: linearization log likelihood = 2.407347
Iteration 4: linearization log likelihood = 2.40734239
Iteration 5: linearization log likelihood = 2.40734121
Iteration 6: linearization log likelihood = 2.40734113

Computing standard errors:

Mixed-effects ML nonlinear regression
Number of obs = 182
Group variable: quake
Number of groups = 23
Obs per group:
  min = 1
  avg = 7.9
  max = 38
Wald chi2(1) = 26.26
Prob > chi2 = 0.0000

phi1: richter U1[quake]
phi3: U3[quake], xb

|            | Coef.     | Std. Err. | z    | P>|z|  | [95% Conf. Interval] |
|------------|-----------|-----------|------|-------|----------------------|
| phi1       |           |           |      |       |                      |
| richter    | 0.2310021 | 0.0450804 | 5.12 | 0.000 | 0.1426461 - 0.319358 |
| _cons      | -0.8836537 | 0.2826255 | -3.13 | 0.002 | -1.437589 - -0.329718 |
| phi3       |           |           |      |       |                      |
| _cons      | 0.004575  | 0.0014192 | 3.22 | 0.001 | 0.0017935 - 0.0073566 |
| /phi2      | 4.063075  | 0.4023386 | 10.10| 0.000 | 3.274506 - 4.851644  |
### Random-effects Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>quake: Independent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>var(U1)</td>
<td>0.0056676</td>
<td>0.0073404</td>
<td>0.0004477 .071752</td>
</tr>
<tr>
<td>var(U3)</td>
<td>0.000013</td>
<td>8.42e-06</td>
<td>3.66e-06 .0000463</td>
</tr>
<tr>
<td>var(Residual)</td>
<td>0.0461647</td>
<td>0.0054421</td>
<td>0.0366409 .0581639</td>
</tr>
</tbody>
</table>

We also store our estimates for later use:

```stata
. estimates store E1
```

By default, `menl` assumes that the random effects $u_{1j}$ and $u_{3j}$ are independent, so there is no need to specify the `covariance()` option in this case. In other words, omitting the `covariance()` option is equivalent to specifying `covariance(U1 U3, independent)`.

### Example 7: Likelihood-ratio test for variance components

Davidian and Giltinan (1995) did not include any random effects in the model for the $\phi_{2j}$ parameters. Let’s check whether the random effects are needed in the equations for $\phi_{1j}$ and $\phi_{3j}$ parameters in (9).

One simple way to assess whether a random effect associated with a certain $\phi_{j}$ can be omitted, is to examine its coefficient of variation (CV), the ratio of the standard deviation to the mean. Let’s compute the CV for $\phi_{3j}$. For convenience, let’s redisplay the results from example 6 as standard deviations for variance components.
. menl, stddeviations

Mixed-effects ML nonlinear regression
Number of obs = 182
Group variable: quake
Number of groups = 23
Obs per group:
min = 1
avg = 7.9
max = 38

Linearization log likelihood = 2.4073411
Prob > chi2 = 0.0000

phi1: richter U1[quake]
phi3: U3[quake], xb

|       | Coef.    | Std. Err. | z    | P>|z|   | [95% Conf. Interval] |
|-------|----------|-----------|------|-------|----------------------|
| phi1  |          |           |      |       |                      |
| richter | .2310021 | .0450804  | 5.12 | 0.000 | .1426461 .319358     |
| _cons  | -.8836537| .2826255  | -3.13| 0.002 | -1.437589 -.329718   |
| phi3   |          |           |      |       |                      |
| _cons  | .004575  | .0014192  | 3.22 | 0.001 | .0017935 .0073566    |

/phi2

|       | Coef.    | Std. Err. | z    | P>|z|   | [95% Conf. Interval] |
|-------|----------|-----------|------|-------|----------------------|
|       |          |           |      |       |                      |
| 4.063075 | .4023386 | 10.10     | 0.000|       | 3.274506 4.851644   |

Random-effects Parameters

<table>
<thead>
<tr>
<th>Quake: Independent</th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>sd(U1)</td>
<td>.0752832</td>
<td>.0487517</td>
<td>.0211582 .2678656</td>
</tr>
<tr>
<td>sd(U3)</td>
<td>.0036085</td>
<td>.0011673</td>
<td>.0019142 .0068026</td>
</tr>
<tr>
<td>sd(Residual)</td>
<td>.2148596</td>
<td>.0126644</td>
<td>.1914181 .241172</td>
</tr>
</tbody>
</table>

The `stddeviations` option specifies that `menl` display random-effects and error standard deviations instead of variances. It will also display correlations instead of covariances whenever they are in the model. Because random-effects variances for these data are very small, we will use this option in all subsequent examples to display results in the standard deviation metric.

The interquake random variation in the $\phi_{3j}$ values about their mean is $CV = \text{sd}(U3)/\{\text{phi3:}_\text{cons}\} = 0.0036/0.0046 \approx 78\%$, and it appears reasonable to keep it in the model. You can perform a formal likelihood-ratio (LR) test of $H_0: \sigma^2_{u3} = 0$ to verify this, as we show below for the test of $H_0: \sigma^2_{u1} = 0$.

Let’s check whether we need random intercept $u_{1j}$ to model $\phi_{1j}$. Computing CV in this case to get an initial assessment is not simple because the mean of $\phi_{1j}$ depends on the $j$th quake through variable `richter`. Given the same main equation (8), we will use the LR test to compare the restricted model, with $u_{1j}$ excluded, which is defined by (11) and (12) below, with the full model defined by (9) and (10).

The stage 2 specification of the restricted model is

$$
\phi_{1j} = \beta_0 + \beta_1 \text{richter}_j \\
\phi_{2j} = \beta_2 \\
\phi_{3j} = \beta_3 + u_{3j}
$$

(11)

where

$$
\begin{align*}
u_{3j} & \sim N(0, \sigma^2_{u3}) \\
\epsilon_{ij} & \sim N(0, \sigma^2_{\epsilon})
\end{align*}
$$

(12)
We now fit the restricted model:

```plaintext
. menl laccel = {phi1:}-log10(sqrt(c.distance#c.distance+exp({phi2})))
> -{phi3:}*sqrt(c.distance#c.distance+exp({phi2})),
> define(phi1: richter, xb) define(phi3: U3[quake], xb)
> stddeviations
```

Obtaining starting values by EM:

Alternating PNLS/LME algorithm:

- Iteration 1: linearization log likelihood = 2.12628622
- Iteration 2: linearization log likelihood = 2.126043
- Iteration 3: linearization log likelihood = 2.12603276
- Iteration 4: linearization log likelihood = 2.12603003
- Iteration 5: linearization log likelihood = 2.1260297

Computing standard errors:

Mixed-effects ML nonlinear regression

Number of obs = 182
Group variable: quake
Number of groups = 23
Obs per group:
  min = 1
  avg = 7.9
  max = 38

Linearization log likelihood = 2.1260297
Wald chisq(1) = 32.22
Prob > chi2 = 0.0000

<table>
<thead>
<tr>
<th>phi1: richter, xb</th>
<th>phi3: U3[quake], xb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coef.</td>
<td>Std. Err.</td>
</tr>
<tr>
<td>richter</td>
<td>0.2208878</td>
</tr>
<tr>
<td>_cons</td>
<td>-0.7863293</td>
</tr>
<tr>
<td>phi3</td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>0.0054348</td>
</tr>
<tr>
<td>/phi2</td>
<td>4.228431</td>
</tr>
</tbody>
</table>

Random-effects Parameters

<table>
<thead>
<tr>
<th>quake: Identity</th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>sd(U3)</td>
<td>0.0042144</td>
<td>0.0011309</td>
<td>0.0024907 - 0.0071309</td>
</tr>
<tr>
<td>sd(Residual)</td>
<td>0.217084</td>
<td>0.0122821</td>
<td>0.194231 - 0.2424668</td>
</tr>
</tbody>
</table>

. estimates store E2

Next, we use `lrtest` to perform an LR test of the hypothesis:

\[ H_0: \sigma^2_{u_1} = 0 \quad \text{versus} \quad H_1: \sigma^2_{u_1} \neq 0 \]
Because testing of $H_0: \sigma^2_{u_1} = 0$ is on the boundary of the parameter space, \texttt{lrtest} reports a note that the provided LR test is conservative; that is, the actual $p$-value is smaller than the one reported. For a test of $H_0: \sigma^2_{u_1} = 0$ in a two-level model, the true asymptotic distribution is not $\chi^2(1)$ but a mixture of $\chi^2(0)$ and $\chi^2(1)$ with equal weights, $0.5\chi^2(0) + 0.5\chi^2(1)$; thus the $p$-value is actually $0.4532/2 = 0.2266$ (see Rabe-Hesketh and Skrondal 2012, sec 8.8). We do not have sufficient evidence to reject the null hypothesis, so we can omit random effect $u_{1j}$ from the full model. \texttt{AIC} and \texttt{BIC} also favor a simpler, reduced model.

Example 8: Including within-subject covariates

One of the questions of interest in the earthquake study was the potential effect of the soil type on acceleration. Variable $\text{soil}$ is a within-subject covariate because the values $\text{soil}_{ij}$ may vary within a subject (earthquake). We include variable $\text{soil}$ in the equation for $\phi_{3ij}$ in (11),

\[
\begin{align*}
\phi_{1j} &= \beta_0 + \beta_1 \text{richter}_j \\
\phi_{2j} &= \beta_2 \\
\phi_{3ij} &= \beta_3 + \beta_4 \text{soil}_{ij} + u_{3j}
\end{align*}
\]
and fit the corresponding model:

```plaintext
.laccel = {phi1:}-log10(sqrt(c.distance#c.distance+exp({phi2})))
> -{phi3:}*sqrt(c.distance#c.distance+exp({phi2})),
> define(phi1: richter, xb) define(phi3: i.soil U3[quake]) stddeviations
```

Obtaining starting values by EM:

Alternating PNLS/LME algorithm:

Iteration 1:  linearization log likelihood = 3.56347788
Iteration 2:  linearization log likelihood = 3.56324717
Iteration 3:  linearization log likelihood = 3.56323391
Iteration 4:  linearization log likelihood = 3.56323036
Iteration 5:  linearization log likelihood = 3.56322978

Computing standard errors:

Mixed-effects ML nonlinear regression

<table>
<thead>
<tr>
<th>Number of obs = 182</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group variable: quake</td>
</tr>
<tr>
<td>Number of groups = 23</td>
</tr>
<tr>
<td>Obs per group:</td>
</tr>
<tr>
<td>min = 1</td>
</tr>
<tr>
<td>avg = 7.9</td>
</tr>
<tr>
<td>max = 38</td>
</tr>
<tr>
<td>Wald chi2(2) = 34.20</td>
</tr>
</tbody>
</table>

Linearization log likelihood = 3.5632298

<table>
<thead>
<tr>
<th>phi1: richter, xb</th>
</tr>
</thead>
<tbody>
<tr>
<td>phi3: i.soil U3[quake]</td>
</tr>
</tbody>
</table>

| laccel | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] |
|--------|-------|-----------|---|------|----------------------|
| phi1   |       |           |   |       |                      |
| richter| .2275944 | .0395549 | 5.75 | 0.000 | .1500683 .3051206 |
| _cons  | -.8079826 | .2548833 | -3.17 | 0.002 | -1.307545 -.3084205 |

| phi3   |       |           |   |       |                      |
| soil   |       |           |   |       |                      |
| soil   | -.0011041 | .0006441 | -1.71 | 0.087 | -.0023665 .0001583 |
| _cons  | .0067347 | .0017416 | 3.87 | 0.000 | .0033213 .0101481 |
| _/phi2 | 4.3212 | .3653809 | 11.83 | 0.000 | 3.605067 5.037334 |

Random-effects Parameters

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>quake: Identity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sd(U3)</td>
<td>.0043088</td>
<td>.0011285 .0025788 .0071992</td>
</tr>
<tr>
<td>sd(Residual)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.2147101</td>
<td>.0121424</td>
<td>.1921829 .2398779</td>
</tr>
</tbody>
</table>

The estimated coefficient for the soil type is $-0.0011$ with a 95% CI of $[-0.0024, 0.0002]$. The knowledge of the soil type at a particular site does not appear to add explanatory power to our model.

Random-effects covariance structures

menl supports various covariance structures to model the random-effects covariance matrix. They are specified using the covariance() option. The covariance() option may be repeated. This is necessary to accommodate multilevel NLME models, where you may need to specify different covariance matrices for the random effects at different levels. Repeating this option may also be useful if you want to specify a block-diagonal covariance structure. See example 23 for details.
Davidian and Giltinan (1995, sec. 1.1.3 and 11.2) discuss a study of soybean plants that started in 1988 and spanned over three growing seasons, year. The central objective of the study was to compare the growth patterns of two genotypes of soybean plants, variety: a commercial variety of soybean, denoted by F, and an experimental variety, denoted by P. In each season, eight plots were planted using F variety and eight using P variety. To assess growth, researchers sampled each plot 8 to 10 times \((8 \leq n_j \leq 10)\) at approximately weekly intervals, time. At each sampling time, six plants were taken from each plot at random. Leaves from the plants were weighed, and the resulting total weight was divided by six to yield a measure of the average leaf weight per plant (in g) for the plot for that week, weight. Plots are identified by the plot variable.

Let’s plot the data first.

```stata
use https://www.stata-press.com/data/r16/soybean
(Growth of soybean plants (Davidian and Giltinan, 1995))
.twoway connected weight time if year==2, connect(L) by(variety)
```

The graph shows the average leaf weights per plant over time for the eight plots with plants of each genotype in the 1989 growing season. Longitudinal growth measures for each plot are connected with solid lines. Apart from some intraplot variation, the growth profile of each plot follows roughly an S shape, according to which growth begins slowly, then shows a linear trend during the middle of the growing season, and then “levels off” at the end. Such pattern is typical for many growth studies.

The main goal of the study was to compare growth patterns over the growing season for the two soybean genotypes. Because the three growing seasons differed markedly in terms of precipitation—1988 was unusually dry, 1989 was wet, and 1990 was normal—contrasting these growth patterns across years was also of interest. The results of this study are useful, for example, for harvesting purposes.

A popular model for individual profiles that resemble an S shape is the logistic growth model:

\[
weight_{ij} = \frac{\phi_{1j}}{1 + \exp\left\{-(time_{ij} - \phi_{2j})/\phi_{3j}\right\}} + \epsilon_{ij}
\]  

(13)

\(\phi_{1j}\) is the asymptotic average leaf weight per soybean plant in plot \(j\) as \(time_{ij} \to \infty\). \(\phi_{2j}\) is the time at which half of \(\phi_{1j}\) is reached; that is, if \(time_{ij} = \phi_{2j}\), then \(E(weight_{ij}) = 0.5\phi_{1j}\). \(\phi_{1j}\)
and $\phi_{2j}$ will henceforth be referred to as “the limiting growth” and “half-life”, respectively. $\phi_{3j}$ is a scale parameter, and it represents the number of days it takes for average leaf weight to grow from 50% (half-life) to about 73% of its limiting growth. That is, if we set $t_{0.73} = \phi_{2j} + \phi_{3j}$, the right-hand side of (13), ignoring the error term, reduces to $\phi_{1j}/\{1 + \exp(-1)\} = 0.73\phi_{1j}$, and then $\phi_{3j} = t_{0.73} - \phi_{2j}$.

We will start with a simple stage 2 specification that does not contain any covariates. Also, because the number of soybean plots, 48, is large compared with the number of random effects, 3, we consider a general positive-definite, unstructured, random-effects covariance matrix:

$$
\phi_j = \begin{bmatrix}
\phi_{1j} \\
\phi_{2j} \\
\phi_{3j}
\end{bmatrix} = \begin{bmatrix}
\beta_1 \\
\beta_2 \\
\beta_3
\end{bmatrix} + \begin{bmatrix}
u_{1j} \\
\nu_{2j} \\
\nu_{3j}
\end{bmatrix}
$$

(14)

$$
u_j = \begin{bmatrix}
u_{1j} \\
\nu_{2j} \\
\nu_{3j}
\end{bmatrix} \sim N(\mathbf{0}, \Sigma), \quad \Sigma = \begin{bmatrix}
\sigma_{11} & \sigma_{12} & \sigma_{13} \\
\sigma_{12} & \sigma_{22} & \sigma_{23} \\
\sigma_{13} & \sigma_{23} & \sigma_{33}
\end{bmatrix}, \quad \epsilon_{ij} \sim N(0, \sigma^2_\epsilon)
$$

To specify this covariance structure in menl, we specify unstructured in the covariance() option. The covariance() option also requires that we list the names of random effects to be correlated.

```
. menl weight = {phi1:}/(1+exp(-(time-{phi2:})/{phi3:})),
> define(phi1: U1[plot], xb) define(phi2: U2[plot], xb) define(phi3: U3[plot], xb)
> covariance(U1 U2 U3, unstructured)
```

Obtaining starting values by EM:

```
Alternating PNLS/LME algorithm:
Iteration 1:  linearization log likelihood = -739.901421
Iteration 2:  linearization log likelihood = -739.84929
(Iteration log omitted)
Iteration 39: linearization log likelihood = -739.834521
Iteration 40: linearization log likelihood = -739.834448
```

Computing standard errors:

Mixed-effects ML nonlinear regression

|                | Coef.  | Std. Err. | z      | P>|z|  | [95% Conf. Interval] |
|----------------|--------|-----------|--------|------|----------------------|
| phi1_cons      | 19.25314 | 0.8031811 | 23.97  | 0.000 | 17.67893 20.82734    |
| phi2_cons      | 55.01999 | 0.7272491 | 75.65  | 0.000 | 53.59461 56.44537    |
| phi3_cons      | 8.403468  | 0.3152551 | 26.66  | 0.000 | 7.78558 9.021357     |
The expected limiting growth or expected maximum average weight, $\beta_1 = E(\phi_{1j})$, of soybean leaves is estimated to be around 19.25 grams. The expected half-life or the time at which the leaves reach half of their expected maximum average weight, $\beta_2 = E(\phi_{2j})$, is estimated to be around 55 days after planting. The expected time needed for the average leaf weight per plant to grow from 50% to 73% of the limiting growth, $\beta_3 = E(\phi_{3j})$, is about 8.4 days.

The estimates of the six random-effects variance–covariance parameters $\sigma_{11}$, $\sigma_{22}$, $\sigma_{33}$, $\sigma_{12}$, $\sigma_{13}$, and $\sigma_{23}$ are displayed in the upper part of the random-effects parameters table. There is a plot-to-plot variation in the estimates of all three parameters of interest: $\beta_1$, $\beta_2$, and $\beta_3$. Also, the plot-specific effects associated with the parameters of interest are positively correlated. For example, based on the estimate of 5.19 of $\text{cov}(U1,U3)$, plants with larger maximum weights tend to grow faster.

We store our estimates for later use:

```
. estimates store S1
```

Example 10: Residuals-vs-fitted plot to check for heteroskedasticity

A popular tool for investigating within-cluster heteroskedasticity is the plot of residuals against the predicted values and other candidate variance covariates. For growth models, variance is often a function of the mean (predicted values). Below we construct the plot of residuals versus predicted values to evaluate the assumption of homoskedastic errors in example 9.
. predict fitweight, yhat
. predict res, residuals
. scatter res fitweight

The plot reveals increasing variability with the predicted average leaf weights, which indicates that our within-cluster variance model is misspecified. In *Heteroskedastic within-group errors*, we will show how to account for within-cluster heteroskedasticity by using the `resvariance()` option.

**Heteroskedastic within-group errors**

Until now, we assumed that the within-group errors—the $\epsilon$’s in the considered models—are i.i.d. Gaussian with common variance $\sigma^2$, labeled as `var(Residual)` by `menl` in the output.

To relax the assumptions of homoskedasticity and the independence of errors, `menl` provides two alternatives. You can model the within-group error variance–covariance matrix, $\sigma^2 \Lambda_j$, directly by using the `rescovariance()` option. If you used the `mixed` command and its `residuals()` option before, you should be familiar with this approach. Alternatively, you can model the error variance–covariance matrix indirectly by modeling the heteroskedasticity structure with the `resvariance()` option and the correlation structure with the `rescorrelation()` option; see *Variance-components parameters*. The latter approach offers more flexibility, particularly in modeling the heteroskedasticity structure. For example, many NLME models exhibit within-subject heterogeneity that is a power function of the mean. The `rescovariance()` option cannot model this, but `resvariance(power _yhat)` can.

If your error structure is simple and is similar to those encountered in `mixed`, you can use the `rescovariance()` option. Otherwise, use `resvariance()`, `rescorrelation()`, or both to model more flexible within-group error covariance structures.

**Example 11: Heteroskedastic power structure**

Continuing with example 9, for these types of growth data, we find it is common for the intraplot variance to increase systematically with the average leaf weight, as we saw in example 10 from the residuals-versus-fitted plot. Davidian and Giltinan (1995) proposed a variance structure that models the within-group error variance as a power function of the mean to account for the intraplot variability.
To reduce the number of parameters to be estimated, the authors assume that the random effects are independent.

Stage 2 specification of the model defined by (13) becomes

\[
\phi_j = \begin{bmatrix} \phi_{1j} \\ \phi_{2j} \\ \phi_{3j} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} u_{1j} \\ u_{2j} \\ u_{3j} \end{bmatrix}
\]

(15)

where

\[
u_j = \begin{bmatrix} u_{1j} \\ u_{2j} \\ u_{3j} \end{bmatrix} \sim N(0, \Sigma), \text{ diagonal } \Sigma = \begin{bmatrix} \sigma_{u_1}^2 & 0 & 0 \\ 0 & \sigma_{u_2}^2 & 0 \\ 0 & 0 & \sigma_{u_3}^2 \end{bmatrix}
\]

and

\[
\text{Var}(\epsilon_{ij}) = \sigma^2(\hat{\text{weight}}_{ij})^{2\delta}
\]

Parameter \(\sigma^2\) in the above is no longer an overall error variance \(\sigma^2_e\) but a common multiplier or a (squared) scale parameter.

In menl, this type of heteroskedasticity is modeled by specifying resvariance(power \_yhat, noconstant). \_yhat designates that the variance should be modeled as a function of predicted values, \(\text{weight}_{ij}\). By default, variance function power includes a constant, which we suppress by specifying the noconstant option.

. menl weight = {phi1:}/(1+exp(-{time-}{phi2:})/{phi3:}), > define(phi1: U1[plot], xb) define(phi2: U2[plot], xb) define(phi3: U3[plot], xb) > resvariance(power _yhat, noconstant)

Obtaining starting values by EM:

Alternating PNLS/LME algorithm:

Iteration 1:  linearization log likelihood = -364.022492
Iteration 2:  linearization log likelihood = -364.228375
Iteration 3:  linearization log likelihood = -364.431678
Iteration 4:  linearization log likelihood = -364.383194
Iteration 5:  linearization log likelihood = -364.389643
Iteration 6:  linearization log likelihood = -364.389152
Iteration 7:  linearization log likelihood = -364.389201

Computing standard errors:

Mixed-effects ML nonlinear regression Number of obs = 412
Group variable: plot Number of groups = 48
Obs per group:
  min = 8
  avg = 8.6
  max = 10
Linearization log likelihood = -364.3892

phi1: U1[plot], xb
phi2: U2[plot], xb
phi3: U3[plot], xb

| weight      | Coef.  | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|-------------|--------|-----------|-----|-----|-----------------------|
| phi1 _cons  | 16.822989 | .6030531  | 27.90 | 0.000 | 15.64093 18.00485     |
| phi2 _cons  | 51.74669 | .4579632  | 112.99 | 0.000 | 50.8491 52.64429     |
| phi3 _cons  | 7.545371 | .0856321  | 88.11 | 0.000 | 7.377535 7.713206    |

Random-effects Parameters

<table>
<thead>
<tr>
<th>plot: Independent</th>
</tr>
</thead>
<tbody>
<tr>
<td>var(U1)</td>
</tr>
<tr>
<td>var(U2)</td>
</tr>
<tr>
<td>var(U3)</td>
</tr>
</tbody>
</table>

Residual variance: Power _yhat

| sigma2  | .0509223 | .004422 | .0429527 | .0603706 |
| delta   | .9339856 | .0244477 | .886069 | .9819023 |

The near-zero estimate of the variance component of $u_{3j}$, var(U3), suggests that the random-effects model is overparameterized. The within-group heteroskedasticity structure appears to explain enough variability in our data, and we no longer need random effects specific to $\phi_{3j}$. This is quite common in mixed-effects models: the random-effects covariance structure and the within-group error covariance structure compete with each other, in the sense that fewer random effects are needed when the within-group error covariance structure is present, and vice versa.

Let’s omit $u_{3j}$ from (15) but now assume an unstructured covariance matrix for $u_{1j}$ and $u_{2j}$. The EM algorithm used by `menl` to obtain initial values produces the starting values for variance components that are, in general, close to the final estimates upon convergence. Thus it can be used as a tool to help us detect potential convergence problems because of an overparameterized random-effects structure at an earlier stage. For example, we can check whether an unstructured covariance matrix is a reasonable choice for the random effects $u_{1j}$ and $u_{2j}$ for these data by displaying estimates after a few iterations. This can be done by specifying the `iterate(#)` option, where # is a small number of iterations, say, between 1 and 4. Below we specify `iterate(3)` to perform only three iterations and the `stddeviations` option to obtain standard deviations and correlations instead of variances and covariances for easier interpretability:
. menl weight = {phi1:}/(1+exp(-(time-{phi2:})/{phi3})),
> define(phi1: U1[plot], xb) define(phi2: U2[plot], xb)
> covariance(U*, unstructured) resvariance(power _yhat, noconstant)
> iterate(3) stddeviations

Obtaining starting values by EM:

Alternating PNLS/LME algorithm:

Iteration 1: linearization log likelihood = -379.663433
Iteration 2: linearization log likelihood = -362.909207
Iteration 3: linearization log likelihood = -361.92309

Computing standard errors:

Mixed-effects ML nonlinear regression
Group variable: plot

| Coef.     | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|-----------|-----------|------|------|----------------------|
| phi1_cons | 16.92772  | .567725 | 29.82 | 0.000 | 15.815 | 18.04044 |
| phi2_cons | 51.81715  | .4484432 | 115.55 | 0.000 | 50.93821 | 52.69608 |
| /phi3     | 7.54089   | .0869075 | 86.77 | 0.000 | 7.370554 | 7.711226 |

Random-effects Parameters

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>plot: Unstructured</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sd(U1)</td>
<td>2.904818</td>
<td>.4070691</td>
</tr>
<tr>
<td>sd(U2)</td>
<td>1.282297</td>
<td>.2555105</td>
</tr>
<tr>
<td>corr(U1,U2)</td>
<td>-.99999</td>
<td>.0033744</td>
</tr>
</tbody>
</table>

Residual variance:

| sigma    | .225507 | .0095097 | .2076177 | .2449377 |
| delta    | .9552958 | .0230657 | .9100879 | 1.000504 |

Warning: convergence not achieved

The U* in covariance(U*, unstructured) is a shorthand notation to reference all random effects starting with U, that is, U1 and U2 in this example. The correlation between u1j and u2j is near −1 with a 95% CI of [−1, 1], which indicates that the random-effects model may still be overparameterized. If you try to fit this model without the iteration(3) option, it would keep iterating without convergence.

Therefore, we further simplify the random-effects covariance structure by assuming independence between u1j and u2j. Stage 2 specification of the model defined by (13) is now

$$\phi_j = \begin{bmatrix} \phi_{1j} \\ \phi_{2j} \\ \phi_{3j} \end{bmatrix} = \begin{bmatrix} \beta_1 + u_{1j} \\ \beta_2 + u_{2j} \\ \beta_3 \end{bmatrix}$$

(16)
where
\[
\mathbf{u}_j = \begin{bmatrix} u_{1j} \\ u_{2j} \end{bmatrix} \sim N(0, \Sigma), \quad \Sigma = \begin{bmatrix} \sigma_{u_1}^2 & 0 \\ 0 & \sigma_{u_2}^2 \end{bmatrix}
\]

and
\[
\text{Var}(\epsilon_{ij}) = \sigma^2(\text{weight}_{ij})^{2\delta}
\]

We fit this model and store its results as S2:

```plaintext
. menl weight = {phi1:}/(1+exp(-{phi2:}/{phi3})), > define(phi1: U1[plot], xb) define(phi2: U2[plot], xb) > resvariance(power _yhat, noconstant)
```

Obtaining starting values by EM:

Alternating PNLS/LME algorithm:

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Linearization log likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-402.761818</td>
</tr>
<tr>
<td>2</td>
<td>-372.07611</td>
</tr>
<tr>
<td>3</td>
<td>-364.151316</td>
</tr>
<tr>
<td>4</td>
<td>-364.340214</td>
</tr>
<tr>
<td>5</td>
<td>-364.387415</td>
</tr>
<tr>
<td>6</td>
<td>-364.388891</td>
</tr>
<tr>
<td>7</td>
<td>-364.389191</td>
</tr>
<tr>
<td>8</td>
<td>-364.389199</td>
</tr>
</tbody>
</table>

Computing standard errors:

Mixed-effects ML nonlinear regression

|                           | Estimate | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|---------------------------|----------|-----------|-------|-------|---------------------|
| weight phil1 _cons        | 16.82289 | .6030522  | 27.90 | 0.000 | 15.64093 18.00485   |
| weight phil2 _cons        | 51.74669 | .4579626  | 112.99| 0.000 | 50.8491 52.64428    |
| /phi3                     | 7.54537  | .085632   | 88.11 | 0.000 | 7.377534 7.713205   |

Random-effects Parameters

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>plot: Independent var(U1)</td>
<td>11.32134</td>
<td>2.831139</td>
<td>6.934847 18.48241</td>
</tr>
<tr>
<td>var(U2)</td>
<td>2.689111</td>
<td>.9344031</td>
<td>1.360931 5.313508</td>
</tr>
</tbody>
</table>

Residual variance:

<table>
<thead>
<tr>
<th>Power _yhat</th>
<th>sigma2</th>
<th>delta</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.0509223</td>
<td>.9339856</td>
</tr>
<tr>
<td></td>
<td>.004422</td>
<td>.0244477</td>
</tr>
<tr>
<td></td>
<td>.0429527</td>
<td>.886069</td>
</tr>
<tr>
<td></td>
<td>.0603706</td>
<td>.9819023</td>
</tr>
</tbody>
</table>

. estimates store S2
Because (16) is not nested in (14), we assess the adequacy of the heteroskedastic model by using information criteria. We use \texttt{estimates stats} to display the AIC and BIC values for the three models.

\begin{verbatim}
. estimates stats S1 S2
Akaike’s information criterion and Bayesian information criterion

<table>
<thead>
<tr>
<th>Model</th>
<th>N</th>
<th>ll(null)</th>
<th>ll(model)</th>
<th>df</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>412</td>
<td>-739.8344</td>
<td>10</td>
<td>1499.669</td>
<td>1539.879</td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>412</td>
<td>-364.3892</td>
<td>7</td>
<td>742.7784</td>
<td>770.9256</td>
<td></td>
</tr>
</tbody>
</table>
\end{verbatim}

Note: BIC uses \texttt{N = number of observations}. See \texttt{[R] BIC note}.

The heteroskedastic model defined by (16) has smaller AIC and BIC values and thus provides a much better representation of the data than (14).

\section*{Example 12: Heteroskedastic model with interactions}

The main goal of the soybean study was to compare growth patterns of the two genotypes of soybean over the three growing seasons, represented by calendar years 1988 through 1990. More specifically, we would like to compare the limiting growth, the half-life, and the growth rate of soybeans across growing seasons and genotypes.

Let \( P_j = I(\text{variety}_j = \text{P}) \) be the indicator for genotype variety \( \text{P} \), \( S_{89,j} = I(\text{year}_j = 1989) \) be the indicator for growing season 1989, and \( S_{90,j} = I(\text{year}_j = 1990) \) be the indicator for growing season 1990. Genotype variety \( \text{F} \) and growing season 1988 are baselines.

Consider an extension of the model defined by (13) and (16), where, in addition to random effects, \( \phi_{1j} \) includes main and interaction effects of growing seasons and genotype variety, \( \phi_{2j} \) includes main effects of growing seasons and genotype variety, and \( \phi_{3j} \) contains main effects of growing seasons only.

\[
\phi_j = \begin{bmatrix}
\phi_{1j} \\
\phi_{2j} \\
\phi_{3j}
\end{bmatrix} = \begin{bmatrix}
\beta_{11} + \beta_{12}S_{89,j} + \beta_{13}S_{90,j} + \beta_{14}P_j + \beta_{15}S_{89,j} \times P_j + \beta_{16}S_{90,j} \times P_j + u_{1j} \\
\beta_{21} + \beta_{22}S_{89,j} + \beta_{23}S_{90,j} + \beta_{24}P_j + u_{2j} \\
\beta_{31} + \beta_{32}S_{89,j} + \beta_{33}S_{90,j}
\end{bmatrix}
\]

(17)

To fit the model defined by (13) and (17) by using \texttt{menl}, we extend \texttt{menl}’s specification from example 11 by including the full-factorial interaction \texttt{i.year##i.variety} in the expression \{\texttt{phi1:}\}, main effects \texttt{i.year} and \texttt{i.variety} in the expression \{\texttt{phi2:}\}, and main effects \texttt{i.year} in the expression \{\texttt{phi3:}\}.

\begin{verbatim}
. menl weight = {phi1:}/(1+exp(-(time-{phi2:})/{phi3:})),
> define(phi1: i.year##i.variety U1[plot])
> define(phi2: i.year i.variety U2[plot])
> define(phi3: i.year, xb)
> resvariance(power _yhat, noconstant)

Obtaining starting values by EM:
Alternating PNLS/LME algorithm:
Iteration 1:  linearization log likelihood = -292.62615
Iteration 2:  linearization log likelihood = -290.243889
(iteration log omitted)
Iteration 10: linearization log likelihood = -290.90729
Iteration 11: linearization log likelihood = -290.907297
\end{verbatim}
### Nonlinear mixed-effects regression

#### Computing standard errors:

<table>
<thead>
<tr>
<th>Model</th>
<th>Observations</th>
<th>Groups</th>
<th>Obs per group</th>
<th>Wald chi²(10)</th>
<th>Prob &gt; chi²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixed-effects ML nonlinear regression</td>
<td>412</td>
<td>48</td>
<td></td>
<td>413.88</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Linearization log likelihood = -290.9073

#### Model Parameters

- **phi1**: i.year i.variety i.year#i.variety U1[plot]
- **phi2**: i.year i.variety U2[plot]
- **phi3**: i.year

#### Coefficients

| Weight | Coef.   | Std. Err. | z      | P>|z| | [95% Conf. Interval] |
|--------|---------|-----------|--------|-----|---------------------|
| phi1   |         |           |        |     |                     |
| year   |         |           |        |     |                     |
| 1989   | -8.837933 | 1.056113  | -8.37  | 0.000 | -10.90788 -6.76799      |
| 1990   | -3.666206 | 1.165969  | -3.14  | 0.002 | -5.951463 -1.38095      |
| variety|         |           |        |     |                     |
| P      | 1.648139 | 1.033433  | 1.59   | 0.111 | -0.377353 3.673631      |
| year#variety | 5.563008 | 1.167782  | 4.76 | 0.000 | 3.274196 7.851819         |
| 1989#P | .0974815 | 1.178054  | 0.08 | 0.934 | -2.211462 2.406425        |
| _cons  | 19.42734 | .9445749  | 20.57 | 0.000 | 17.57601 21.27867         |
| phi2   |         |           |        |     |                     |
| year   |         |           |        |     |                     |
| 1989   | -2.253227 | .9746495 | -2.31 | 0.021 | -4.163505 -.3429494     |
| 1990   | -4.970736 | .9778317 | -5.08 | 0.000 | -6.887251 -3.054221      |
| variety|         |           |        |     |                     |
| P      | -1.294058 | .4255317 | -3.04 | 0.002 | -2.128085-.4600314       |
| _cons  | 54.81257 | .7587239 | 72.24 | 0.000 | 53.3255 56.29964          |
| phi3   |         |           |        |     |                     |
| year   |         |           |        |     |                     |
| 1989   | -.9023768 | .1992358 | -4.53 | 0.000 | -1.292872 -.5118818      |
| 1990   | -.6805314 | .2100799 | -3.24 | 0.001 | -1.09228 -.2687823       |
| _cons  | 8.060677 | .1459662 | 55.22 | 0.000 | 7.774588 8.346765         |

#### Random-effects Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>plot: Independent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>var(U1)</td>
<td>.8643052</td>
<td>.5272014</td>
<td>.2614921 2.856773</td>
</tr>
<tr>
<td>var(U2)</td>
<td>.1341755</td>
<td>.2307058</td>
<td>.0046141 3.901728</td>
</tr>
<tr>
<td>Residual variance:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power _yhat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sigma2</td>
<td>.0467091</td>
<td>.0039176</td>
<td>.0396286 .0550546</td>
</tr>
<tr>
<td>delta</td>
<td>.9451193</td>
<td>.0227608</td>
<td>.9005089 .9897297</td>
</tr>
</tbody>
</table>

.estimates store S3
By including more fixed effects in the model, which explain some of the variability in the average leaf weight, we substantially reduced the estimates of variance components. Compared with example 11, \( \text{var}(U_1) \) decreased from 11.32 to 0.86, and \( \text{var}(U_2) \) decreased from 2.69 to 0.13. It often happens that specifying a better-fitting model for the fixed effects reduces the need for random effects in the model.

We can compare model S3 or the model defined by (17) with model S2 or the one defined by (16) by using, for example, information criteria.

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
\text{Model} & N & \text{ll(null)} & \text{ll(model)} & \text{df} & \text{AIC} & \text{BIC} \\
\hline
S2 & 412 & -364.3892 & 7 & 742.7784 & 770.9256 \\
S3 & 412 & -290.9073 & 17 & 615.8146 & 684.172 \\
\hline
\end{array}
\]

Note: BIC uses \( N = \) number of observations. See [R] BIC note.

Even though S3 has many more parameters, it fits the soybean data better than S2.

By inspecting the fixed-effects estimates from the output of model S3, we see that both the type of year and genotype variety affect all three parameters: the expected maximum leaf weight, half-life, and scale. For example, all three parameters achieve their highest values in the dry year, baseline year 1988, because coefficient estimates for the other years are negative. Also, the genotype variety F reaches its half-life roughly a day later \( (\beta_{24} = -1.29) \) than genotype variety P.

Example 13: Obtaining predictions

After estimation, we may want to obtain predicted values for the outcome or for the parameters of interest. Continuing with example 12, we want to predict the asymptotic average leaf weight per soybean plant in each plot, \( \hat{\phi}_{1j} \). The \( \phi_{1j} \) parameter is not constant but varies for each plot, growing season, and genotype variety. We can use \texttt{predict} after \texttt{menl} to obtain predicted values for \( \phi_{1j} \); see [ME] \texttt{menl postestimation}.

First, we create a new grouping variable for growing seasons, genotype variety, and plot types. We also create the \texttt{tolist} variable to mark the first observation in each group.

\[
\begin{align*}
&\text{. egen group = group(year variety plot)} \\
&\text{. by group, sort: generate byte tolist=(_n==1)}
\end{align*}
\]

Next, we use \texttt{predict} to compute predicted values for the expression \{\phi1:} and store them in the new variable \texttt{phi1}. We store only unique values in \texttt{phi1}, one for each \texttt{group}; the remaining observations are replaced with missing values.

\[
\begin{align*}
&\text{. predict double (phi1 = \{phi1:\})} \\
&\text{. qui replace phi1 = . if tolist!=1}
\end{align*}
\]
We now list the five smallest and the five largest values of the asymptotic average leaf weight.

```
. sort phil
. list plot year variety phi1 if (_n<=5 | _n>43) & phi1<., sep(5)
```

<table>
<thead>
<tr>
<th>plot</th>
<th>year</th>
<th>variety</th>
<th>phi1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>1989</td>
<td>F</td>
<td>8.8421451</td>
</tr>
<tr>
<td>2.</td>
<td>1989</td>
<td>F</td>
<td>10.449521</td>
</tr>
<tr>
<td>3.</td>
<td>1989</td>
<td>F</td>
<td>10.473849</td>
</tr>
<tr>
<td>4.</td>
<td>1989</td>
<td>F</td>
<td>10.721364</td>
</tr>
<tr>
<td>5.</td>
<td>1989</td>
<td>F</td>
<td>10.810197</td>
</tr>
<tr>
<td>44.</td>
<td>1988</td>
<td>P</td>
<td>20.86739</td>
</tr>
<tr>
<td>45.</td>
<td>1988</td>
<td>P</td>
<td>21.237692</td>
</tr>
<tr>
<td>46.</td>
<td>1988</td>
<td>P</td>
<td>21.310511</td>
</tr>
<tr>
<td>47.</td>
<td>1988</td>
<td>P</td>
<td>21.506007</td>
</tr>
<tr>
<td>48.</td>
<td>1988</td>
<td>P</td>
<td>21.581873</td>
</tr>
</tbody>
</table>

Soybean plants with genotype variety P have substantially larger asymptotic average leaf weight in the dry year, 1988, than soybean plants with genotype variety F in the wet year, 1989.

Example 14: Within-group error correlation structure

Pinheiro and Bates (2000, chap. 8) analyzed data from a study of the estrus cycles of mares. Originally analyzed in Pierson and Ginther (1987), the data contain daily records of the number of ovarian follicles larger than 10 mm over a period ranging from 3 days before ovulation to 3 days after the subsequent ovulation. The measurement times for each mare are scaled so that the ovulations for each mare occur at times 0 and 1 and are recorded in stime.

The considered model is

$$\text{follicles}_{ij} = \phi_{1j} + \phi_{2j} \sin(2\pi \phi_{3j}\text{stime}_{ij}) + \phi_{4j} \cos(2\pi \phi_{3j}\text{stime}_{ij}) + \epsilon_{ij}$$

where $\phi_{1j}$ is an intercept, $\phi_{3j}$ is the frequency of the sine wave for the $j$th mare, and $\phi_{2j}$ and $\phi_{4j}$ are terms determining the amplitude and phase of the sine wave for the $j$th mare. If $a_j$ and $p_j$ are the amplitude and phase for mare $j$, then $\phi_{2j} = a_j \cos(p_j)$ and $\phi_{4j} = a_j \sin(p_j)$.

This model was fit in example 8 of [ME] mixed in the context of a linear mixed-effects model, where the number of ovarian follicles was a periodic function of time with known frequency $\phi_{3j}$ equal to 1. If we want to estimate frequency, we cannot use the mixed command, because $\phi_{3j}$ enters the model nonlinearly.

Pinheiro and Bates (2000) suggested an AR(1) correlation structure for modeling the within-group error correlation. This structure can be specified by using the rescorrelation() option as

```
rescorrelation(ar 1, t(time))
```

where $t$ is an integer-valued time variable used to order the observations within mares and to determine the lags between successive observations.

We also considered several random-effects structures and found that we need only one random intercept to model $\phi_{1j}$. 
The full specification for the stage 2 model is

\[
\phi_j = \begin{bmatrix}
\phi_{1j} \\
\phi_{2j} \\
\phi_{3j} \\
\phi_{4j}
\end{bmatrix} = \begin{bmatrix}
\beta_1 + u_{1j} \\
\beta_2 \\
\beta_3 \\
\beta_4
\end{bmatrix}
\]

where

\[
u_j = u_{1j} \sim N(0, \sigma_u^2), \quad \epsilon_j \sim N(0, \sigma^2 \Lambda_j)
\]

and

\[
\sigma^2 \Lambda_j = \sigma^2 \begin{bmatrix}
1 & \rho & \rho^2 & \ldots & \rho^{n_j-1} \\
\rho & 1 & \rho & \ldots & \rho^{n_j-2} \\
\rho^2 & \rho & 1 & \ldots & \rho^{n_j-3} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho^{n_j-1} & \rho^{n_j-2} & \rho^{n_j-3} & \ldots & 1
\end{bmatrix}
\]

We fit this model by using \texttt{menl} as follows:

\[
\text{. use https://www.stata-press.com/data/r16/ovary, clear (Ovarian follicles in mares)}
\]

\[
\text{. menl follicles = \{phi1: U1[mare], xb\} + \{phi2\}*sin(2*_pi*stime*{phi3}) +}
\]

\[
\text{\quad \{phi4\}*cos(2*_pi*stime*{phi3}), rescorrelation(ar 1, t(time))}
\]

Obtaining starting values by EM:

Alternating PNLS/LME algorithm:

Iteration 1: linearization log likelihood = -789.434153
Iteration 2: linearization log likelihood = -789.434391
Iteration 3: linearization log likelihood = -789.434391

Computing standard errors:

Mixed-effects ML nonlinear regression

| Coef. | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|-------|-----------|------|------|---------------------|
| follicles |       |      |      |                     |
| phil    |       |      |      |                     |
| _cons  | 11.98929 | .9055946 | 13.24 | 0.000 | 10.21436 | 13.76422 |
| /phi2  | .2226033 | .3290159 | 0.68  | 0.499 | -0.4222559 | 0.8674626 |
| /phi3  | 4.18747  | .2746499 | 15.25 | 0.000 | 3.649166  | 4.725774  |
| /phi4  | .279653  | .3223277 | 0.87  | 0.386 | -0.3520977 | 0.9114036 |
By using estimates of $\phi_{2j}$ and $\phi_{4j}$, we can compute the amplitude and phase for the sine wave for mare $j$. The amplitude and the phase are the same for all the mares because $\phi_{2j}$ and $\phi_{4j}$ are constant and not mare specific.

For example, the amplitude $a_j$ can be computed as $\sqrt{\phi_{2j}^2 + \phi_{4j}^2}$ by using the relationship $\phi_{2j}^2 + \phi_{4j}^2 = a_j^2 \{\sin^2(p_j) + \cos^2(p_j)\} = a_j^2$. The phase $p_j$ can be computed as $p_j = \text{atan}(\phi_{4j}/\phi_{2j})$ by using the relationship $\phi_{4j}/\phi_{2j} = \{a_j \sin(p_j)\} / \{a_j \cos(p_j)\} = \tan(p_j)$.

We can use `nlcom` to compute the amplitude and the phase.

```
.nlcom (amplitude: sqrt(_b[/phi2]^2 + _b[/phi4]^2))
> (phase: atan(_b[/phi4]/_b[/phi2]))
```

| follicles | Coef. | Std. Err. | z    | P>|z|  | [95% Conf. Interval] |
|-----------|-------|-----------|------|------|---------------------|
| amplitude | .3574325 | .2451183  | 1.46 | 0.145 | -.1229904 to .8378555 |
| phase     | .8985001 | 1.090985  | 0.82 | 0.410 | -1.23979 to 3.03679  |

As we mentioned in example 1, it is important to try different initial values when fitting NLME models to investigate potential convergence to a local maximum, especially for models containing periodic functions, as in our example. We explore different initial values for this model in `Linearization approach to finding initial values` by considering the functional form of the mean function and arrive at a different solution with a larger log likelihood.

---

**Restricted maximum likelihood**

Like `mixed`, `menl` provides estimation by using ML or REML. The difference between the two approaches is described in detail in `Likelihood versus restricted likelihood` in [ME] `mixed`. Briefly, REML is preferable when you have a small number of groups because it produces unbiased, at least for balanced data, estimates of variance components. In large samples, there is little difference between ML and REML. One disadvantage of REML, however, is that LR tests based on REML are inappropriate for comparing models with different fixed-effects specifications. See example 15 for an example of REML estimation.
Pharmacokinetic modeling

Pharmacokinetics (PKs) is the study of drug absorption, distribution, metabolism, and excretion. It is often referred to as the study of “what the body does with a drug”. The goal of PK modeling is to summarize the concentration-time measurements using a model that relates drug input to drug response, to relate the parameters of this model to patient characteristics, and to provide individual dose–response predictions to optimize individual doses. In other words, by understanding between-subject variation in drug disposition, we can individualize the dosage regimen for a particular patient based on relevant physiological information identified by our PK model.

Single-dose pharmacokinetic modeling

Example 15: Single-oral-dose model

Consider a PK study of the antiasthmatic agent theophylline that was reported by Boeckmann, Sheiner, and Beal (2011) and analyzed by Davidian and Giltinan (1995). The drug was administered orally to 12 subjects, where dosage $d(j)$ was given on a per weight basis. Serum concentrations (in mg/L) were obtained at 11 time points per subject over 25 hours following administration. The graph below shows the resulting concentration-time profiles for four subjects.

In PKs, the pattern of rapid rise to a peak concentration followed by an apparent exponential decay may be described by a so-called one-compartment open model with first-order absorption and elimination. The model corresponds roughly to viewing the body as one “blood compartment” and is particularly useful for the PK analysis of drugs that distribute relatively rapidly throughout the body, which makes it a reasonable model for the kinetics of theophylline after oral administration. Further details about compartmental modeling may be found in Gibaldi and Perrier (1982). The one-compartment open model for theophylline kinetics may be expressed as

$$
	ext{conc}_{ij} = \frac{d(j)k_{ej}k_{aj}}{Cl_{j}(k_{aj} - k_{ej})} \left\{ \exp(-k_{ej}\text{time}_{ij}) - \exp(-k_{aj}\text{time}_{ij}) \right\} + \epsilon_{ij}
$$

(18)
for $i = 1, \ldots, 11$ and $j = 1, \ldots, 12$. Model parameters are the elimination rate constant $k_{e,j}$, the absorption rate constant $k_{a,j}$, and the clearance $Cl_j$ for each subject $j$.

Because each of the model parameters must be positive to be meaningful, we write

\begin{align*}
Cl_j &= \exp (\beta_0 + u_{0j}) \\
    k_{a,j} &= \exp (\beta_1 + u_{1j}) \\
    k_{e,j} &= \exp (\beta_2)
\end{align*}

where $u_{0j}$ and $u_{1j}$ are assumed independent and normally distributed with means zero and variance $\sigma^2_{u_0}$ and $\sigma^2_{u_1}$, respectively.

The model defined by (18) implies that the predicted value for the concentration at time $time_{ij} = 0$ is $\hat{conc}_{ij} = 0$. Therefore, a power variance function, a natural candidate for this type of heteroskedastic pattern, cannot be used in this example because error variance will be 0 at $time_{ij} = 0$. So the constant plus power variance function, which adds a constant to the power term, is used instead to model the within-group error variance:

$$\text{Var}(\epsilon_{ij}) = \sigma^2\{(\hat{conc}_{ij})^\delta + c\}^2$$
In `menl`, we use the `resvariance(power _yhat)` option to specify the constant plus power variance function and the following model specification:

```
> . menl conc = (dose*{ke:}*{ka:}/({cl:}*({ka:}-{ke:}))*
> (exp(-{ke:}*time)-exp(-{ka:}*time)), define(cl: exp({b0}+{U0[subject]}))
> define(ka: exp({b1}+{U1[subject]})) define(ke: exp({b2}))
> resvariance(power _yhat)
```

Obtaining starting values by EM:

Alternating PNLS/LME algorithm:

```
Iteration 1:  linearization log likelihood = -167.519533
Iteration 2:  linearization log likelihood = -167.65729
(Iteration log omitted)
Iteration 26: linearization log likelihood = -167.679657
Iteration 27: linearization log likelihood = -167.679641
```

Computing standard errors:

```
Mixed-effects ML nonlinear regression                              Number of obs = 132
Group variable: subject                                           Number of groups = 12
Obs per group:  
  min =  11
  avg = 11.0
  max =  11

Linearization log likelihood = -167.67964
```

```
cl: exp({b0}+{U0[subject]})
ka: exp({b1}+{U1[subject]})
ke: exp({b2})
```

|        | Coef. | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|--------|-------|-----------|------|------|----------------------|
| /b0    | -3.227479 | .0598389  | -53.94 | 0.000 | -3.344761 to -3.110197 |
| /b1    | .432931  | .1980835  | 2.19  | 0.029 | .0446945 to .8211674  |
| /b2    | -2.453742 | .0514567  | -47.69 | 0.000 | -2.554595 to -2.352889 |

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>subject: Independent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>var(U0)</td>
<td>.0288787</td>
<td>.0127763</td>
<td>.0121337 to .0687323</td>
</tr>
<tr>
<td>var(U1)</td>
<td>.4075667</td>
<td>.1948713</td>
<td>.1596654 to 1.040367</td>
</tr>
</tbody>
</table>

Residual variance:    
  Power _yhat

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>sigma2</td>
<td>.0976905</td>
<td>.0833033</td>
<td>.0183659 to .5196273</td>
</tr>
<tr>
<td>delta</td>
<td>.3187133</td>
<td>.2469517</td>
<td>-.1653031 to .8027297</td>
</tr>
<tr>
<td>_cons</td>
<td>.7288982</td>
<td>.3822969</td>
<td>.2607496 to 2.037559</td>
</tr>
</tbody>
</table>
The number of groups, 12, is fairly small in these data, so we now refit the model by using REML estimation.

```
. menl conc = (dose*{ke:}*{ka:}/({cl:}*({ka:}-{ke:}))) *
>  (exp(-{ke:}*time)-exp(-{ka:}*time)), define(cl: exp({b0}+{U0[subject]}))
>  define(ka: exp({b1}+{U1[subject]})) define(ke: exp({b2}))
>  resvariance(power _yhat) reml
```

Obtaining starting values by EM:

Alternating PNLS/LME algorithm:

Iteration 1:  linearization log restricted-likelihood = -172.31734
Iteration 2:  linearization log restricted-likelihood = -172.423246

(Iteration log omitted)

Iteration 23: linearization log restricted-likelihood = -172.443827
Iteration 24: linearization log restricted-likelihood = -172.443845

Computing standard errors:

Mixed-effects REML nonlinear regression

Number of obs = 132
Group variable: subject
Number of groups = 12

Obs per group:
min = 11
avg = 11.0
max = 11

Linear. log restricted-likelihood = -172.44384

| cl: exp({b0}+{U0[subject]}) |
| ka: exp({b1}+{U1[subject]}) |
| ke: exp({b2}) |

```
\begin{tabular}{lcccr}
\textbf{conc} & \textbf{Coef.} & \textbf{Std. Err.} & \textbf{z} & \textbf{P>|z|} & \textbf{[95\% Conf. Interval]} \\
\hline
/b0 & -3.227295 & .0619113 & -52.13 & 0.000 & -3.348639 -3.105951 \\
/b1 & .4354519 & .2072387 & 2.10 & 0.036 & .0292716 .8416322 \\
/b2 & -2.453743 & .0517991 & -47.37 & 0.000 & -2.555267 -2.352218 \\
\hline
\end{tabular}
```

<table>
<thead>
<tr>
<th>Random-effects Parameters</th>
<th>\textbf{Estimate}</th>
<th>\textbf{Std. Err.}</th>
<th>\textbf{[95% Conf. Interval]}</th>
</tr>
</thead>
</table>
| subject: Independent | \textbf{var(U0)} & .0316416 & .014531 & .0128634 & .0778326 \\
| & \textbf{var(U1)} & .4500585 & .2228207 & .1705475 & 1.187662 \\
| Residual variance: | \textbf{sigma2} & .1015759 & .0865353 & .0191262 & .5394517 \\
| & \textbf{delta} & .3106636 & .2466555 & -.1727723 & .7940996 \\
| & \textbf{_cons} & .7150935 & .3745261 & .2561833 & 1.996065 \\

As expected, the estimates of the random-effects variances are slightly larger than the corresponding ML estimates, but we arrive at similar inferential conclusions based on our REML estimates.

Example 16: Nonlinear functions of parameters

A distinctive feature of example 15 is that parameters of interest are nonlinear functions of the estimated parameters and random effects. To interpret parameters that depend on random effects, we can either integrate random effects out of the parameter expression or condition on them. The former parameter estimates are often referred to as population-based estimates. The latter parameter estimates
are referred to as conditional estimates and, when conditioning on zero random effects, \( u_j = 0 \), as estimates for an “average” or typical subject. For linear functions, the population-based estimates coincide with the conditional estimates. This is no longer true for nonlinear functions.

In PK modeling, the parameters of interest are clearance, elimination rate, and absorption rate. These are nonlinear functions of the estimated parameters \( \beta_0, \beta_1, \beta_2 \), and subject-specific random effects. Depending on the context, we may be interested in their population-based estimates or in their conditional estimates.

In general, obtaining population-based estimates would require numerical integration to integrate the subject-specific random effects out of the expression. In our example, we can compute population-based estimates directly by using the fact that \( \exp(u_{0j})'s \) and \( \exp(u_{1j})'s \) are lognormally distributed. Thus the population-based clearance \( Cl_P \) can be computed as

\[
E(Cl_j) = E\{ \exp(\beta_0 + u_{0j}) \} = \exp(\beta_0 + \sigma^2_{u_0}/2)
\]

and the population-based absorption rate \( k_{aP} \) as

\[
E(\exp(\beta_1 + u_{1j})) = \exp(\beta_1 + \sigma^2_{u_1}/2).
\]

The elimination rate \( k_e \) does not depend on subject-specific effects and can thus be computed simply as

\[
k_{eP} = k_e = \exp(\beta_2).
\]

Alternatively, if we want parameters to represent a typical subject, we can simply set \( u_{0j} = 0 \) and \( u_{1j} = 0 \) in their expressions. Thus we can compute clearance and absorption rate for a typical subject simply as

\[
Cl = \exp(\beta_0)
\]

and

\[
k_a = \exp(\beta_1).
\]

If we compare the formulas for, say, \( Cl_P \) and \( Cl \), the former considers variation in clearances across subjects, whereas the latter ignores such variation and instead reflects what the clearance would be for a typical subject with \( u_{0j} = 0 \).

Both approaches have merit, and here we will compute, for example, \( Cl_P = \exp(\beta_0 + \sigma^2_{u_0}/2) = \exp(-3.23+0.032/2) = 0.04 \). That is, 0.04 liters of serum concentration are cleared of the theophylline drug per hour per kg body weight in the considered population. In other words, for the population of subjects that weigh 75 kg, an average of \( 75 \times 0.04 \approx 3 \) liters of serum concentration are cleared of theophylline every hour.

We can also use \texttt{nlcom} to compute the estimates of \( Cl_P \) and \( Cl \). To use \texttt{nlcom}, we need to know how parameters are labeled by \texttt{menl} for postestimation. We can use \texttt{menl}'s option \texttt{coeflegend} to display parameter names. We also specify \texttt{noheader} to suppress the table header.

```
. menl, coeflegend noheader
```

<table>
<thead>
<tr>
<th>conc</th>
<th>Coef. Legend</th>
</tr>
</thead>
<tbody>
<tr>
<td>/b0</td>
<td>-3.227295 _b[/b0]</td>
</tr>
<tr>
<td>/b1</td>
<td>.4354519 _b[/b1]</td>
</tr>
<tr>
<td>/b2</td>
<td>-2.453743 _b[/b2]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>/subject</th>
<th>Coef. Legend</th>
</tr>
</thead>
<tbody>
<tr>
<td>lnsd(U0)</td>
<td>-1.726641 _b[/subject:lnsd(U0)]</td>
</tr>
<tr>
<td>lnsd(U1)</td>
<td>-.3991888 _b[/subject:lnsd(U1)]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>/Residual</th>
<th>Coef. Legend</th>
</tr>
</thead>
<tbody>
<tr>
<td>lnsigma</td>
<td>-1.143475 _b[/Residual:lnsigma]</td>
</tr>
<tr>
<td>delta</td>
<td>.3106636 _b[/Residual:delta]</td>
</tr>
<tr>
<td>ln_cons</td>
<td>-.335342 _b[/Residual:ln_cons]</td>
</tr>
</tbody>
</table>

If we examine the output carefully, we will notice that \texttt{menl}, \texttt{coeflegend} displayed results in the estimation metric—as log-standard deviations instead of variances. Although by default \texttt{menl} displays
parameters in their original metric, it stores them in the estimation metric, the metric that was used
during optimization; see Examples of specifying initial values and Methods and formulas for more
details about the estimation metric.

The parameters we need to compute \( C_l^P \) and \( C_l \) are coefficient \( b[0] \) and the variance of \( U_0 \),
which can be obtained as \( \exp(2* b[/ subject:lnsd(U0)]) \) based on the stored estimate of the
log-standard deviation of \( U_0 \). We now use \texttt{nlcom} to compute our nonlinear estimates.

\[
\texttt{. nlcom (C_l^P: exp(_b[/b0]+0.5*exp(2*_b[/subject:lnsd(U0)]))) (C_l: exp(_b[/b0]))}
\]

| conc | Coef.     | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|------|-----------|-----------|------|-----|---------------------|
| C_l  | 0.0396646 | 0.0024557 | 16.15| 0.000| 0.0348516 .0444777  |

Working with parameters in the estimation metric can be tedious, especially when nonlinear
expressions contain multiple variance components. In that case, you may consider using \texttt{estat sd}
after \texttt{menl} to obtain results in the standard deviation metric or, if you also specify the variance
option, in the variance metric; see [ME] menl postestimation. If you specify the post option with \texttt{estat sd},
the results will also be stored in the standard deviation or variance metrics, which you
can use for further postestimation analysis.

\[
\texttt{. estat sd, post variance coeflegend}
\]

<table>
<thead>
<tr>
<th>conc</th>
<th>Coef. Legend</th>
</tr>
</thead>
<tbody>
<tr>
<td>/b0</td>
<td>-3.227295 _b/[b0]</td>
</tr>
<tr>
<td>/b1</td>
<td>0.4354519 _b/[b1]</td>
</tr>
<tr>
<td>/b2</td>
<td>-2.453743 _b/[b2]</td>
</tr>
</tbody>
</table>

Random-effects Parameters | Estimate | Legend
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>subject: Independent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>var(U0)</td>
<td>0.0316416 _b[/subject:var(U0)]</td>
<td></td>
</tr>
<tr>
<td>var(U1)</td>
<td>0.4500585 _b[/subject:var(U1)]</td>
<td></td>
</tr>
</tbody>
</table>

| Residual variance:       |          |                     |
| Power _yhat              |          |                     |
| sigma2                   | 1.0157595 _b/[Residual:sigma2] | |
| delta                    | 0.3106636 _b/[Residual:delta] | |
| _cons                    | 0.7150935 _b/[Residual:_cons] | |

In addition to results being displayed in the variance metric, because of the \texttt{post} option, they are
stored in that metric. We also specified the \texttt{coeflegend} option with \texttt{estat sd} to see how parameters
are labeled so that we could refer to them in other postestimation commands such as \texttt{nlcom}.

Now, we can simply refer to the variance of \( U_0 \) as \( _b[/subject:var(U0)] \) in our \texttt{nlcom}
command.
. nlcom (Cl_P: exp(_b[/b0]+0.5*_b[/subject:var(U0)]))
    Cl_P: exp(_b[/b0]+0.5*_b[/subject:var(U0)])

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cl_P</td>
<td>0.0402972</td>
<td>0.002512</td>
<td>16.04</td>
<td>0.000</td>
</tr>
</tbody>
</table>

_estat sd_’s _post_ option should be used with caution because it clears all estimation results except
the parameter estimates in _e(b)_ and their VCE in _e(V)_ Thus the only postestimation features that
will work after _estat sd, post_ are those that need only _e(b)_ and _e(V)_ such as _lincom_ and
_nlcom_. Other postestimation features will not be available, and you will need to refit your model to
run them. To avoid refitting your model, you may consider storing your estimation results in memory
(see [R] _estimates store_) or saving them on disk (see [R] _estimates save_) before using _estat sd, post_. We no longer needed the estimation results from _menl_, so we did not mind clearing them.

Multiple-dose pharmacokinetic modeling

In example 15, a single dose of the analgesic theophylline was administered to each subject followed
by multiple serum concentration measurements per subject. For long-duration illnesses, multiple doses
are often given to each subject, with multiple serum concentration measurements interspersed. After a
single-dose drug administration, the plasma drug level rises above and then falls below the minimum
effective concentration, resulting in a decline in therapeutic effect. To treat chronic diseases, multiple-
dosage or intravenous infusion regimens are used to maintain the plasma drug levels within the narrow
limits of the therapeutic window to achieve optimal clinical effectiveness.

Example 17: Multiple-intravenous-doses model

Grasela and Donn (1985) report a study of the neonatal PKs of phenobarbital. Data were collected
on 59 preterm infants given phenobarbital for prevention of seizures during the first 16 days after
birth. Each infant received one or more intravenous doses, _dose_ (mg/kg). One to six blood serum
phenobarbital concentration measurements, _conc_ (mg/L), were obtained from each infant, _subject_,
for a total of 155 measurements. The birthweight, in kilograms, and a five-minute Apgar score, a
measure of the physical condition, were also obtained on each infant. The Apgar score is obtained
by adding points (2, 1, or 0) for heart rate, respiratory effort, muscle tone, response to stimulation,
and skin coloration; a score of 10 represents the best possible condition. _time_ is measured in hours.
Davidian and Giltinan (1995) and Pinheiro and Bates (2000) also analyze this dataset.

A one-compartment open model with intravenous administration and first-order elimination was
used to model the PKs of this phenobarbital study

\[ \text{conc}_{ij} = \sum_{t \leq i} \frac{\text{dose}_{ik}}{V_j} \exp \left\{ -\frac{\text{Cl}_j}{V_j} (\text{time}_{ij} - \text{time}_{tj}) \right\} + \epsilon_{ij} \]  

(19)

for \( i = 1, \ldots, n_j \) and \( j = 1, \ldots, 59 \). Model parameters are the clearance \( \text{Cl}_j \) (L/h) and volume of
distribution \( V_j \) (L) for each subject \( j \). Clearance is the volume of blood or plasma that is totally cleared
of its content of drug per unit time. It is the proportionality factor between the rate of elimination
and concentration, \( \frac{dC}{dt} = -k_e C = -\left(\frac{\text{Cl}}{V}\right) C \), where \( C \) is the plasma concentration and \( k_e \)
is the elimination rate (h\(^{-1}\)). The volume of distribution, \( V \), is defined as the apparent space or volume
into which a drug distributes.
To fit this model using \texttt{menl}, we consider an alternative recursive formulation of model (19)

\[
\text{conc}_{ij} = \mu(x'_{ij}, \beta, u_j) = \frac{\text{dose}_{ij}}{V_j} + \mu(x'_{i-1,j}, \beta, u_j) \exp \left\{ -\frac{\text{Cl}_j}{V_j} (\text{time}_{ij} - \text{time}_{i-1,j}) \right\} + \epsilon_{ij}
\]

Here, \(x'_{ij} = (\text{time}_{ij}, \text{dose}_{ij}, \text{fapgar}_j, \text{weight}_j)\) is the vector of covariates corresponding to subject \(j\) at \(\text{time}_{ij}\). Notice that concentration \(\text{conc}_{ij} = \mu(x'_{ij}, \beta, u_j)\) depends on its previous expected value, \(\mu(x'_{i-1,j}, \beta, u_j)\), and on the time difference, \(\text{time}_{ij} - \text{time}_{i-1,j}\). In Stata, we can use the lag operator, \(L\), to refer to previous values and the difference operator, \(D\), to refer to the difference between the two successive values. \texttt{menl} supports time-series operators in the model specification; see \textit{Time-series operators}. We can use \(D.\text{time}\) to include the time difference in the model. However, we cannot simply use \(L.\text{conc}\), because this would include the previous observed value of \(\text{conc}\) in the model, and we need the previous (predicted) value of the mean function. \texttt{menl} provides a special syntax \(L.\_\text{yhat}\) to include lagged predicted values or, equivalently, a special syntax \(L.\{\text{conc}:\}\) to include the lagged predicted mean function. \{conc:\} refers to the nonlinear expression for the mean function of the \text{conc} variable. Thus, our \texttt{menl} main specification of the recursive model would be

\[
. \texttt{menl conc = dose}/\{V:\} + L.\{\text{conc}:\}*\exp(-\{\text{Cl}:\}/\{V:\}*D.\text{time}), \ldots
\]

where expressions for \{V:\} and \{Cl:\} will be defined later.

Because we are using time-series operators in the expression, we need to declare our data to be time-series data. There are two ways to do this: you can specify \texttt{tsset} prior to calling \texttt{menl} or you can specify the time variable in \texttt{menl}'s option \texttt{tsorder()}; see \textit{Time-series operators} for details. In this example, we will use the \texttt{tsorder()} option; see the technical note \textit{below} for an example using \texttt{tsset}.

\[
. \texttt{menl conc = dose}/\{V:\} + L.\{\text{conc}:\}*\exp(-\{\text{Cl}:\}/\{V:\}*D.\text{time}), \ldots \texttt{tsorder(time)}
\]

Let's take a quick look at our data by listing the observations for the first subject.

\[
. \texttt{use https://www.stata-press.com/data/r16/phenobarb} \\
(\text{Pharmacokinetics study of phenobarbital in neonatal infants}) \\
. \texttt{list if subject==1, sepby(subject)}
\]

<table>
<thead>
<tr>
<th>subject</th>
<th>weight</th>
<th>apgar</th>
<th>time</th>
<th>dose</th>
<th>conc</th>
<th>fapgar</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1.4</td>
<td>7</td>
<td>0</td>
<td>25</td>
<td>&gt;= 5</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1.4</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>17.3</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1.4</td>
<td>7</td>
<td>12.5</td>
<td>3.5</td>
<td>&gt;= 5</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1.4</td>
<td>7</td>
<td>24.5</td>
<td>3.5</td>
<td>&gt;= 5</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1.4</td>
<td>7</td>
<td>37</td>
<td>3.5</td>
<td>&gt;= 5</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1.4</td>
<td>7</td>
<td>48</td>
<td>3.5</td>
<td>&gt;= 5</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1.4</td>
<td>7</td>
<td>60.5</td>
<td>3.5</td>
<td>&gt;= 5</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1.4</td>
<td>7</td>
<td>72.5</td>
<td>3.5</td>
<td>&gt;= 5</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1.4</td>
<td>7</td>
<td>85.3</td>
<td>3.5</td>
<td>&gt;= 5</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1.4</td>
<td>7</td>
<td>96.5</td>
<td>3.5</td>
<td>&gt;= 5</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1.4</td>
<td>7</td>
<td>108.5</td>
<td>3.5</td>
<td>&gt;= 5</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1.4</td>
<td>7</td>
<td>112.5</td>
<td>0</td>
<td>31</td>
</tr>
</tbody>
</table>

The most noticeable feature of our PK data is the presence of many missing values for the concentration. In fact, this is a common structure of PK data in the presence of multiple doses. Notice that the \text{conc} variable contains missing values for each nonzero \text{dose}. It is typical to measure concentration only after a dose or multiple doses are administered, which gives rise to missing concentration at some time points. By default, Stata commands omit all observations containing missing values in variables used with the command. In this example, we need to retain missing \text{conc} observations. We can use \texttt{menl}'s option \texttt{tsmissing} to do so.
When you specify the `tsmissing` option, `menl` uses predicted values in place of system missing `conc` values in the computation. (Observations with extended missing values .a, .b, and so on in `conc`, if there were any, would have been omitted from the computation.) These predicted values are used to compute predicted values for the observed concentrations but are not used to compute the log likelihood. Only observed concentrations contribute to the log-likelihood calculation.

Another aspect of our data is that they are time-series data. Thus, the first observation in each panel provides starting values for the time-series operators. For example, from the data, the initial time value used by `D.time` for the first subject is $t_{i-1,j} = t_{0,1} = 0$. But how do we initialize $\{conc:\}$ given that $\{conc:\}$ does not exist as a variable in our dataset? We use `menl`’s option `tsinit()`.

```
  . menl conc = dose/{V:} + L.{conc:}*exp(-{Cl:}/{V:}*D.time), ...
  > tsinit({conc:}=dose/{V:}) tsmissing tsorder(time)
```

The `tsinit()` option allows us to specify initial conditions for the lagged predicted mean functions as expressions. In our example, the initial condition for the mean concentration for each subject $j$ at time 0 is $dose_{0,j}/V_j$, which we specified in `tsinit()`.

Let’s now return to our nonlinear model specification and provide expressions for $\{V:\}$ and $\{Cl:\}$. One of the model parameterizations that Davidian and Giltinan (1995) consider for these data use `weight` as a covariate for clearance and volume. They also include a dichotomized Apgar score, factor variable `fapgar` in our dataset, to model volume. They express clearance and volume as

\[
Cl_j = \beta_1 weight_j \times \exp(u_{1j})
\]
\[
V_j = \beta_2 weight_j (1 + \beta_3 fapgar_j) \exp(u_{2j})
\]

where $u_{1j}$’s and $u_{2j}$’s are two independent sets of random effects that follow $N(0, \sigma^2_{u1})$ and $N(0, \sigma^2_{u2})$, respectively.
We specify the above expressions for subject-specific volume and clearance in *menl* using the `define()` options and fit the model:

```
   . menl conc = dose/{V:} + L.{conc:}*exp(-{Cl:}/{V:}*D.time),
   > define(Cl: {cl:weight}*weight*exp({U1[subject]}))
   > define(V: {v:weight}*weight*(1+{v:apgar}*1.fapgar)*exp({U2[subject]}))
   > tsinit({conc:} = dose/{V:})
   > tsmissing tsorder(time)
```

Obtaining starting values by EM:

Alternating PNLS/LME algorithm:

Iteration 1:  linearization log likelihood = -432.588874
Iteration 2:  linearization log likelihood = -436.355248
Iteration 3:  linearization log likelihood = -436.367352
Iteration 4:  linearization log likelihood = -436.368940
Iteration 5:  linearization log likelihood = -436.368999
Iteration 6:  linearization log likelihood = -436.368960

Computing standard errors:

Mixed-effects ML nonlinear regression

| Path | No. of Groups | Observations per Group |  |  |  |
|------|---------------|------------------------|  |  |  |
| subject | 59 | 1 | 11.6 | 19 |
| conc   | 59 | 1 | 2.6  | 6  |

Linearization log likelihood = -436.36896

| Conc  | Coef.  | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|-------|--------|-----------|-------|------|----------------------|
| /cl   |        |           |       |      |                      |
| weight| .004705 | .0002219  | 21.20 | 0.000| .0042701 .00514 |
| /v    |        |           |       |      |                      |
| weight| .9657032| .0294438  | 32.80 | 0.000| .9079945 1.023412 |
| apgar | .1749755| .0845767  | 2.07  | 0.039| .0092082 .3407429 |

Random-effects Parameters

<table>
<thead>
<tr>
<th>Subject: Independent</th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>var(U1)</td>
<td>.0404098</td>
<td>.0187133</td>
<td>.0163044 .1001537</td>
</tr>
<tr>
<td>var(U2)</td>
<td>.030259</td>
<td>.0078857</td>
<td>.0181562 .0504295</td>
</tr>
<tr>
<td>var(Residual)</td>
<td>7.469354</td>
<td>1.280411</td>
<td>5.337875 10.45196</td>
</tr>
</tbody>
</table>

Note: Lagged predicted mean function `L.{conc:}` is used in the model.

From the coefficient table, we see that heavier babies have a higher clearance and volume of distribution. There is a positive association between the volume of distribution and the Apgar score: healthier babies have a better ability to eliminate the drug.
Because we specified the \texttt{tsmissing} option, the header reported the number of missing and nonmissing concentration values used in the computation. Also, the table containing the information about the number of groups has an additional entry for \texttt{conc} providing the group information for nonmissing observations of \texttt{conc}.

When we specified the time variable in the \texttt{tsorder()} option, \texttt{menl} generated the corresponding consecutive integer-valued time variable and used it with \texttt{tsset}. From the output of \texttt{tsset}, as displayed by \texttt{menl}, we see that \texttt{menl} also identified the panel variable, \texttt{subject}, from our model specification and used it with \texttt{tsset}. The generated time variable used with \texttt{tsset} is labeled as <\texttt{time}> in the output.

\textbf{Technical note}

In example 17, we used the \texttt{tsorder()} option to specify the ordering for time-series operators. We could have used \texttt{tsset} instead, but we would need to create the appropriate time variable first. Here, we demonstrate how to do this.

We must specify the panel and time variables with \texttt{tsset}. Intuitively, we would want to type

\begin{verbatim}
    . tsset subject time
\end{verbatim}

but that would not produce the intended results. First, \texttt{tsset} requires an integer time variable, which the \texttt{time} variable is not. Second, even if \texttt{time} contained integers, it is not equally spaced, which would lead to gaps in the time series and thus missing values for time-series operators.

In our example, we are concerned only with the ordering of observations within a subject with respect to the \texttt{time} variable for the purpose of time-series operators. So, we create a new variable, \texttt{tsorder}, to contain consecutive integers based on \texttt{time} and use it with \texttt{tsset}.

\begin{verbatim}
    . sort subject time
    . by subject (time): generate long tsorder = _n
    . tsset subject tsorder
\end{verbatim}

You can verify that the following specification of \texttt{menl} will produce the same results as in example 17.

\begin{verbatim}
    . menl conc = dose/\{V:\} + L.\{conc:\}*exp(-\{Cl:\}/\{V:\}*D.time),
        > define(Cl: \{cl:weight\}*weight*exp(U1\{subject\}))
        > define(V: \{v:weight\}*weight*(1+\{v:apgar\}*1.fapgar)*exp(U2\{subject\}))
        > tsinit(\{conc:\} = dose/\{V:\})
        > tsmissing
\end{verbatim}

(output omitted)

Note that we still use the \texttt{time} variable with the difference operator, \texttt{D.}, in the model specification.

\textbf{Example 18: Multiple-oral-doses model}

\textcite{Verme et al. (1992)} evaluated the \textit{PK} behavior of quinidine, a pharmaceutical agent used to prevent cardiac arrhythmias, in a study of 136 subjects receiving oral quinidine therapy. A total of 361 serum quinidine concentrations (variable \texttt{conc}, mg/L) were measured over time (variable \texttt{time}, hours), ranging from 1 to 11 observations per subject. Multiple doses (variable \texttt{dose}, mg) of quinidine,
in two different forms, were administrated to each subject. The doses were adjusted for differences in salt content by conversion into milligrams of quinidine base. These data are also presented as examples in Davidian and Giltinan (1995) and Pinheiro and Bates (2000).

A one-compartment open model with first-order absorption and elimination is assumed for serum quinidine concentrations. This model, expressed in a compact recursive form, is

$$\text{conc}_{ij} = \mu_1 (x'_{ij}, \beta, u_j) = \mu_1 (x'_{i-1,j}, \beta, u_j) Q_{e_{ij}} + \frac{k_{a_j}}{k_{a_j} - k_{e_j}} (Q_{e_{ij}} - Q_{a_{ij}}) + \epsilon_{ij} \tag{20}$$

where

$$\begin{align*}
\text{Ca}_{ij} &= \mu_2 (z'_{ij}, \beta, u_j) = \mu_2 (z'_{i-1,j}, \beta, u_j) Q_{a_{ij}} + \frac{\text{dose}_{ij}}{V_j} \\
Q_{e_{ij}} &= \exp \left\{ -k_{e_j} (\text{time}_{ij} - \text{time}_{i-1,j}) \right\} \\
Q_{a_{ij}} &= \exp \left\{ -k_{a_j} (\text{time}_{ij} - \text{time}_{i-1,j}) \right\}
\end{align*}$$

for subject $j = 1, \ldots, 136$ and subject observation $i = 1, \ldots, n_j, n_j \in [1, 11]$. The quantities $Q_{a_{ij}}$ and $Q_{e_{ij}}$ are defined for notational convenience to simplify the model expression. $z'_{ij} = (\text{time}_{ij}, \text{dose}_{ij})$ and $x'_{ij} = (z'_{ij}, \text{glyco}_{ij}, \text{creatinine}_j, \text{weight}_j)$ are vectors of covariates, which we describe later, corresponding to subject $j$ at time $\text{time}_{ij}$. Because the drug administration is extravascular, the quinidine concentration in the body over time is a function of both the absorption rate, $k_{a_j}$, and the elimination rate, $k_{e_j}$, for subject $j$. The function $\text{Ca}_{a_j}$ is the apparent concentration of quinidine in the absorption depot over time (indexed by $i$) for subject $j$.

From example 17, we know that $k_{e_j} = \frac{\text{Cl}_j}{V_j}$, where $\text{Cl}_j$ is the clearance, defined as the volume of plasma or blood that is totally cleared from its content of drug per unit time, and $V_j$ is the apparent volume of distribution, defined as theoretical volume that would be necessary to contain the total amount of an administered drug at the same concentration that is observed in the blood plasma.

The menl specification corresponding to model (20) is

```plaintext
. menl conc = L.{conc:}*{Qe:}*{Ca:}*{ka:}*{ke:}*{Qa:},
> define(Ca: L.{Ca:}*{Qa:}+dose/{V:})
> define(Qe: exp(-{ke:}*D.time))
> define(Qa: exp(-{ka:}*D.time))
> define(ke: {Cl:}/{V:})
> define(ka: exp({lka}))
> ...
```

where expressions for {Cl:} and {V:} will be defined later. Similarly to example 17, we use D.time to specify differences between two successive time values and L.{conc:} to specify the lagged predicted mean function; also see Time-series operators. New in this specification is the inclusion of the lagged function of model parameters or lagged named expression L.{Ca:}. Expression Ca is defined in the define() option and is a function of its own lag, L.{Ca:}. Finally, parameter {ka} is reparameterized as $\exp({lka})$ to ensure that it is positive.

When a patient receives the same dosage at regular time intervals (variable interval), model (20) simplifies to the steady-state model

$$\text{conc}_{ij}^{ss} = \frac{\text{dose}_{ij} k_{a_j}}{V_j (k_{a_j} - k_{e_j})} (Q_{e_{ij}}^{ss} - Q_{a_{ij}}^{ss}) \tag{21}$$
and

$$Ca_{ij}^{ss} = \frac{dose_{ij}}{V_j} Q_{a_{ij}}^{ss}$$

where

$$Q_{e_{ij}}^{ss} = \frac{1}{1 - \exp(-k_{e_j} \text{interval}_{ij})}$$

$$Q_{a_{ij}}^{ss} = \frac{1}{1 - \exp(-k_{a_j} \text{interval}_{ij})}$$

The quantities $Q_{e_{ij}}^{ss}$ and $Q_{a_{ij}}^{ss}$ are also defined for notational convenience.

The `menl` specification corresponding to model (21) is

```
. menl conc = dose*{ka:}/({V:}*({ka:}-{ke:}))*({Qe_ss:} - {Qa_ss:}),
> define(Qe_ss: 1/(1-exp(-{ke:}*interval))
> define(Qa_ss: 1/(1-exp(-{ka:}*interval))
> define(ke: {Cl:}/{V:})
> define(ka: exp({lka}))
> ...
```

For the quinidine model, the steady-state model (21) is assumed whenever $\text{interval}_{ij}$ is nonzero and the nonsteady-state model (20) is assumed otherwise. Thus, we need to switch back and forth between these two models in our `menl` specification. We can use the Stata function `cond(condition, expr_if_condition_true, expr_if_condition_false)`.

For example, the `menl` specification becomes

```
. menl conc = cond(interval==0,
> L.{conc:}*{Qe:}+L.{Ca:}*({ka:}/({ka:}-{ke:}))*({Qe:}-{Qa:}),
> dose*{ka:}/({V:}*({ka:}-{ke:}))*({Qe:}-{Qa:}))
> define(Ca: cond(interval==0,L.{Ca:}*{Qa:}+dose/{V:}, dose/{V:}*{Qa:}))
> ...
```

where other expressions such as $\{Qe:\}$ and $\{Qa_{ss}:\}$ are as defined earlier. We used `cond()` for the main `menl` specification and for the definition of the $\{Ca:\}$ function.

Recall from example 17 that when we specify the lagged predicted mean function, we need to specify an initial condition for it in the `tsinit()` option. Just like the main nonlinear specification, the initial condition for $L.\{conc:\}$ will depend on the value of `interval`. The mean concentration at time 0 will be 0 for observations with zero `interval` values and will be equal to the expression for the steady-state model otherwise: `tsinit(L.\{conc:\}=cond(interval==0,0,dose*{ka:}/({V:}*({ka:}-{ke:}))*({Qe:}-{Qa:})))`. Similarly, we need to provide an initial condition for the lagged function of model parameters $L.\{Ca:\}$. It also depends on `interval`: `tsinit(L.\{Ca:\}=cond(interval==0,dose/{V:},dose/{V:}*{Qa:}))`. Because we are using the same expressions in the function definitions and the initial conditions, we can define additional functions to minimize typing:

```
. menl conc = cond(interval==0,
> L.\{conc:\}*\{Qe:\}+L.\{Ca:\}*({ka:}/({ka:}-{ke:}))*\{Qe:\}-{Qa:}),
> \{Ca:},
> define(Ca: cond(interval==0,L.\{Ca:\}*\{Qa:\}+dose/{V:}, \{Ca_{ss}:\})
> define(Css: cond(interval==0,0,dose*{ka:}/({V:}*({ka:}-{ke:}))*\{Qe:}-{Qa:}))
> define(Ca_{ss}: dose/{V:})*\{Qa:})
> ...
> tsinit(\{conc:\} = cond(interval==0, 0, \{Css:\}
> tsinit(\{Ca:\} = cond(interval==0, dose/{V:}, \{Ca_{ss}:\)
```

```
\{\text{Css}\} \text{ contains the expression for the steady-state model (or 0 for observations in a nonsteady state), and } \{\text{Ca\_ss}\} \text{ contains the expression for the Ca function in the steady state.}

Let’s now finalize our \text{menl} specification by defining expressions for \{\text{Cl}\} and \{\text{V}\}. The goal of the study from Verme et al. (1992) was to examine the relationship between quinidine PKs and several potential covariates: body weight (kg); age (years); height (in); glyco, \(\alpha_1\)-acid glycoprotein concentration (mg/dL); creatinine, creatinine clearance (\(\geq 50\) or \(< 50\) ml/min); race (Caucasian, Latin, black); smoke, smoking status (yes, no); ethanol, alcohol abuse (former, none, current); and heart, congestive heart failure (no or mild, moderate, severe). We provide more details about covariates creatinine and glyco below.

Creatinine is a waste product from the normal breakdown of muscle tissue. As creatinine is produced, it is filtered through the kidneys and excreted in urine. Doctors use creatinine and creatinine clearance tests to check renal function (kidney function). Testing the rate of creatinine clearance shows the kidneys’ ability to filter the blood. As renal function declines, creatinine clearance also goes down. Creatinine clearance in a healthy young person is about 95 ml/min for women and 120 ml/min for men.

\(\alpha_1\)-acid glycoprotein (also known as AAG) is an important plasma protein involved in the binding and transport of many drugs, including quinidine. A healthy range is 50–120 mg/dl. Changes in AAG concentration could potentially alter the free fraction of drugs in plasma or at their target sites and eventually affect their PK disposition and pharmacological action. Because AAG levels are increased in response to stress, serum levels of total quinidine may be greatly increased in settings such as acute myocardial infarction. Protein binding is also increased in chronic renal failure. There tends to be a small increase in AAG with age.

For the purpose of illustration, we fit a modified version of model 2 from pages 248–249 of Davidian and Giltinan (1995). The clearance, \(\text{Cl}_j\), is modeled on the log scale as a linear combination \{1\text{Cl}\} of glyco, ib1.creatinine, weight, and a random intercept, \(U_1\), at the subject level. The apparent volume, \(V_j\), is modeled on the log scale using a fixed-effect intercept and weight. The absorption rate, \(k_a\), is modeled on the log scale as a free parameter \{1\text{ka}\}, and is assumed fixed for all subjects. The full second-stage specification is as follows:

\[
\begin{align*}
\text{Cl}_{ij} &= \exp \left( \beta_1 + \beta_2 \text{glyco}_{ij} + \beta_3 \text{creatinine}_{ij} + \beta_4 \text{weight}_j + u_{1j} \right) \\
V_j &= \exp \left( \beta_5 + \beta_6 \text{weight}_j \right) \\
k_{a_j} &= \exp (\beta_7) \\
k_{e_{ij}} &= \frac{\text{Cl}_{ij}}{V_j}
\end{align*}
\]

where \(u_{1j}\)'s are random effects that follow \(N \left( 0, \sigma^2_{u1} \right)\).
Similarly to the phenobarbital data from example 17, the quinidine data also contain missing concentration values, so we specify the \texttt{tsmissing} option to retain them in the computation. Again, we will specify the time variable in the \texttt{tsorder()} option and let \texttt{menl tsset} the data for us.

. use https://www.stata-press.com/data/r16/quinidine
. menl conc = cond(interval==0,
   > L.{conc:}+L.{Ca:}*({ka:}-{ke:})*({Qe:}-{Qa:}),
   > {Css:}),
. define(Ca: cond(interval==0, L.{Ca:}*{Qa:}+dose/{V:}, {Ca_ss:}))
. define(Qe: exp(-{ke:}*D.time))
. define(Qa: exp(-{ka:}*D.time))
. define(Css: cond(interval==0,0,{ka:}*dose/{V:}*({ka:}-{ke:})*({Qe:}-{Qa:})))
. define(Ca_ss: cond(interval==0,0,dose/{V:}*{Qa_ss:}))
. define(Qe_ss: 1/(1-exp(-{ke:}*interval)))
. define(Qa_ss: 1/(1-exp(-{ka:}*interval)))
. define(ke: {Cl:}/{V:})
. define(ka: exp({lka}))
. define(Cl: exp({lCl:glyco ib1.creatinine weight U1[subject], xb}))
. define(V: exp({lV: weight, xb}))
. tsinit({conc:} = cond(interval==0, 0, {Css:}))
. tsinit({Ca:} = cond(interval==0, dose/{V:}, {Ca_ss:}))
. tsorder(time) tsmissing
   panel variable: subject (unbalanced)
   time variable: <time>, 1 to 47
   delta: 1 unit

Obtaining starting values by EM:

Alternating PNLS/LME algorithm:

Iteration 1: linearization log likelihood = -423.266877
Iteration 2: linearization log likelihood = -425.823117
Iteration 3: linearization log likelihood = -425.811238
Iteration 4: linearization log likelihood = -425.811904
Iteration 5: linearization log likelihood = -425.812407
Iteration 6: linearization log likelihood = -425.812232
Iteration 7: linearization log likelihood = -425.812329
Iteration 8: linearization log likelihood = -425.812284
Iteration 9: linearization log likelihood = -425.812305

Computing standard errors:

Mixed-effects ML nonlinear regression

<table>
<thead>
<tr>
<th>Path</th>
<th>No. of Groups</th>
<th>Observations per Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimum</td>
<td>Average</td>
</tr>
<tr>
<td>subject</td>
<td>136</td>
<td>1</td>
</tr>
<tr>
<td>conc</td>
<td>136</td>
<td>1</td>
</tr>
</tbody>
</table>

Number of obs = 1,335
Nonmissing = 361
Missing = 974
Nonlinear mixed-effects regression

Wald chi2(4) = 169.94
Prob > chi2 = 0.0000

Linearization log likelihood = -425.81231

Ca: cond(interval==0,L.{Ca:}*{Qa:}+dose/{V:},{Ca_ss:})
Ca_ss: cond(interval==0,0,dose/{V:}*{Qa_ss:})
Cl: exp({lCl:})
Css: cond(interval==0,0,{ka:}*dose/({V:}*({ka:}-{ke:}))*({Qe_ss:}-{Qa_ss:}))
Qa: exp(-{ka:}*D.time)
Qa_ss: 1/(1-exp(-{ka:}*interval))
Qe: exp(-{ke:}*D.time)
Qe_ss: 1/(1-exp(-{ke:}*interval))
V: exp({lV:})
ka: exp({lka})
ke: {Cl:}/{V:}
lCl: glyco ib1.creatinine weight U1[subject], xb
lV: weight, xb

| conc     | Coef.  | Std. Err. | z      | P>|z|   | [95% Conf. Interval] |
|----------|--------|-----------|--------|-------|---------------------|
| lCl      |        |           |        |       |                     |
| glyco    | -.4689097 | .0416876 | -11.25 | 0.000 | -.5506159 -.3872035 |
| creatinine >= 50 | .1851334 | .0464825 | 3.98  | 0.000 | .0940294 .2762373  |
| weight   | .0036181 | .0018213 | 1.99  | 0.047 | .0000485 .0071877  |
| _cons    | 2.668191 | .1524726 | 17.50 | 0.000 | 2.36935 2.967031   |
| lV       |        |           |        |       |                     |
| weight   | .0087346 | .0058603 | 1.49  | 0.136 | -.0027514 .0202206 |
| _cons    | 4.572762 | .47765   | 9.57  | 0.000 | 3.636585 5.508939  |
| /lka     | -.8956278 | .301     | -2.98 | 0.003 | -1.485577 -.3056787 |

Random-effects Parameters

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>var(U1)</td>
<td>.0589024</td>
<td>.0108271 .0410838 .0844921</td>
</tr>
<tr>
<td>var(Residual)</td>
<td>.4122599</td>
<td>.0364831 .346612 .4903413</td>
</tr>
</tbody>
</table>

Note: Lagged predicted mean function L.{conc} is used in the model.
Note: Lagged named expression L.{Ca:} is used in the model.

From the coefficient table, we see that the clearance decreases with increase of AAG (glyco) as would be expected with the greater protein binding. The clearance is greater for creatinine clearance ≥ 50 as would be expected with better renal function. Both clearance and volume increase with weight; although, the effect of weight on volume is not statistically significant at the 5% level. The subject variability for clearance contributes to the model as seen by the confidence interval for the random-effects variance var(U1).
Nonlinear marginal models

The variance–covariance matrix of the response vector \( y_j = (y_{1j}, \ldots, y_{nj}) \) involves two components to model heteroskedasticity and correlation: A random-effects component \( \Sigma \) and a within-group error component \( \Lambda_j \). In some applications, one may wish to directly model the covariance structure of the response by choosing the appropriate within-group error component \( \Lambda_j \) without introducing random effects. This results in the so-called nonlinear marginal model (for example, Pinheiro and Bates [2000, sec. 7.5.1]):

Stage 1: Individual-level model
\[
y_j = m(x^w_j, \phi_j) + \epsilon_j \quad \epsilon_j \sim N(0, \sigma^2 \Lambda_j)
\]

Stage 2: Group-level model
\[
\phi_j = d(x^b_j, \beta) \quad j = 1, \ldots, M
\]

The above is essentially a vector representation of (2) after excluding the random effects \( u_j \). Random effects are used in NLME models to explain the between-subject or between-group variation, but they are not used in the specification of nonlinear marginal models. This key difference implies that mixed-effects models allow for subject-specific inference, whereas marginal models do not. For this reason, mixed-effects models are often called subject-specific models, while marginal models are called population-averaged models.

\texttt{menl} provides the \texttt{group()} suboption within the \texttt{rescovariance()} and \texttt{rescorrelation()} options to model the dependence between within-group observations without introducing random effects. Below, we show an example of fitting a nonlinear marginal model, without random effects, using the \texttt{group()} suboption. See example 22 for the usage of the \texttt{group()} suboption in the presence of random effects.

Example 19: Nonlinear marginal model

Vonesh and Carter (1992) analyzed data on 20 high-flux hemodialyzers to assess their in-vitro ultrafiltration performance. Dialyzers are used in hemodialysis, a treatment that replaces the work of kidneys, to filter harmful wastes out of blood for patients with kidney failure. High-flux dialyzers do this more efficiently than conventional dialyzers—they are composed of membranes with larger pores, which allows them to remove larger molecules and water during blood filtration. A dialyzer’s ultrafiltration performance, or ability to filter blood, is controlled by so-called transmembrane pressure and also depends on the blood flow rate used during hemodialysis. In these data, the response variable, \( \text{rate} \), is the dialyzer’s ultrafiltration rate in mL/hr measured at 7 different transmembrane pressures, \( \text{pressure} \), in dmHg. Ten dialyzers were evaluated using bovine blood at a blood flow rate, \( q_b \), of 200 mL/min, whereas the other 10 dialyzers were evaluated at 300 mL/min.

The ultrafiltration rate, \( \text{rate}_{ij} \), at the \( i \)th transmembrane pressure, \( \text{pressure}_{ij} \), for the \( j \)th subject is represented by the nonlinear model

\[
\text{rate}_{ij} = \phi_{1j} \left[ 1 - \exp\{-\exp(\phi_{2j}) (\text{pressure}_{ij} - \phi_3)\} \right] + \epsilon_{ij}
\]

The parameters \( \phi_1 \), \( \phi_2 \), and \( \phi_3 \) have physiological interpretation: \( \phi_1 \) is the maximum attainable ultrafiltration rate, \( \phi_2 \) is the logarithm of the hydraulic permeability transport rate of the membrane (rate at which water and molecules pass through the dialyzer membrane), and \( \phi_3 \) is the transmembrane pressure required to offset the oncotic pressure (the transmembrane pressure at which the ultrafiltration rate is 0).

One of the models proposed in Vonesh and Carter (1992) included no random effects and used an exchangeable (also known as compound symmetry) covariance structure to model the within-dialyzer error covariance structure. The full description of the second stage of the model is
\[ \phi_{1j} = \beta_{10} + \beta_{11} q_b^j \]
\[ \phi_{2j} = \beta_{20} + \beta_{21} q_b^j \]
\[ \phi_{3j} = \beta_3 \]

and

\[ \epsilon_j \sim N(0, \sigma^2 \Lambda_j), \quad \Lambda_j = \begin{bmatrix} 1 & \rho & \ldots & \rho \\ 1 & \rho & \ldots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ 1 & & & 1 \end{bmatrix} \]

Below, we use `rescovariance(exchangeable, group(dialyzer))` to request an exchangeable within-group error covariance structure where groups are identified by the dialyzer variable.

```
. use https://www.stata-press.com/data/r16/dialyzer (High-flux hemodialyzers (Vonesh and Carter, 1992))
. menl rate = {phi1:}(1-exp(-exp({phi2:})*(pressure - {phi3}))),
> define(phi1: i.qb, xb)
> define(phi2: i.qb, xb)
> rescovariance(exchangeable, group(dialyzer)) stddev
```

Obtaining starting values:
Alternating GNLS/ML algorithm:
Iteration 1: log likelihood = -365.34244
Iteration 2: log likelihood = -365.32697
Iteration 3: log likelihood = -365.32697
Iteration 4: log likelihood = -365.32697
Iteration 5: log likelihood = -365.32697
Iteration 6: log likelihood = -365.32697

Computing standard errors:
Mixed-effects ML nonlinear regression
Number of obs = 140
Group variable: dialyzer
Number of groups = 20
Obs per group:
  min = 7
  avg = 7.0
  max = 7
Wald chi2(2) = 194.77
Prob > chi2 = 0.0000

<table>
<thead>
<tr>
<th>phi1: i.qb</th>
<th>phi2: i.qb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

```
|           | Coef. | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|-----------|-------|-----------|-------|-------|----------------------|
| rate      |       |           |       |       |                      |
| phi1      |       |           |       |       |                      |
| qb        | 17.23062 | 1.24589 | 13.83 | 0.000 | 14.78872 - 19.67252  |
| cons      | 44.95795 | 0.8841506 | 50.85 | 0.000 | 43.22505 - 46.69086  |
| phi2      |       |           |       |       |                      |
| qb        | -.5034708 | .0763513 | -6.59 | 0.000 | -.6531166 - -.353825  |
| cons      | .7626986 | .0630914 | 12.09 | 0.000 | .6390417 - .8863555  |
| /phi3     | .2249104 | .0102113 | 22.03 | 0.000 | .2048965 - .2449243  |
```
Random-effects Parameters | Estimate | Std. Err. | [95% Conf. Interval]
--- | --- | --- | ---
Residual: Exchangeable | sd | 3.722521 | .3064517 | 3.167839 | 4.374326 |
corr | .3867847 | .0993617 | .1771206 | .5628662 |

The estimated values of \( \rho \) and \( \sigma_\epsilon \) are \( \hat{\rho} = 0.39 \) and \( \hat{\sigma}_\epsilon = 3.72 \), respectively. The 95% confidence interval [0.18, 0.56] for \( \rho \) suggests a positive correlation within dialyzer measurements. The maximum ultrafiltration rate, \( \phi_1 \), and the logarithm of the hydraulic permeability transport rate, \( \phi_2 \), appear to be affected by the blood flow rate.

Three-level models

Representation of (1) can be extended to, for example, two-nested levels of clustering, to form the following three-level model, with observations composing the first level,

\[
y_{jk} = \mu \left( X_{jk}, \beta, u^{(3)}_k, u^{(2)}_{jk} \right) + \epsilon_{jk}
\]

where the first-level observations \( i = 1, \ldots, n_{jk} \) are nested within the second-level groups \( j = 1, \ldots, M_k \), which are nested within the third-level groups \( k = 1, \ldots, M \). Group \( j \) nested within group \( k \) consists of \( n_{jk} \) observations, so \( y_{jk}, X_{jk} \), and \( \epsilon_{jk} \) each have row dimension \( n_{jk} \).

Also, assume that

\[
\begin{align*}
  u^{(3)}_k &\sim N(0, \Sigma_3) \\
  u^{(2)}_{jk} &\sim N(0, \Sigma_2) \\
  \epsilon_{jk} &\sim N(0, \sigma^2 \Lambda_{jk})
\end{align*}
\]

and that \( u^{(3)}_k, u^{(2)}_{jk}, \) and \( \epsilon_{jk} \) are independent.

Example 20: Three-level model

Hand and Crowder (1996, 118–120) analyzed a study where the blood glucose levels of 7 volunteers, subject, who took alcohol at time 0 were measured 14 times, time, over a period of 5 hours after alcohol consumption. The same experiment was repeated at a later date with the same subjects but with a dietary additive, guar, used for all subjects. Variable guar is a binary variable that identifies whether a subject received a dietary additive. It also identifies each experiment, with 0 corresponding to the experiment without guar and 1 corresponding to the experiment with guar. Thus we will use the guar variable both as the level indicator and, later, as a fixed-effects variable.
Here is a plot of the whole dataset.

```
. use https://www.stata-press.com/data/r16/glucose
   (Glucose levels following alcohol ingestion (Hand and Crowder, 1996))
. twoway connected glucose time if guar==0 ||
    connected glucose time if guar==1 ||, by(subject, rows(2))
    legend(order(1 "without guar" 2 "with guar"))
```

Our preliminary assessment based on the above graph is that, except for subject 6, the effect of the dietary additive guar on the temporal trajectory of the blood glucose levels does not seem to be important. The effect of guar will be formally tested in example 21.

Hand and Crowder (1996) proposed the following empirical model relating the expected glucose level to time,

\[
glucose_{ijk} = \phi_1 + \phi_2 time^3 \exp(-\phi_3 time) + \epsilon_{ijk}
\]  

(22)

where \( k = 1, \ldots, 7, j = 1, 2, \) and \( i = 1, \ldots, 14 \). The blood glucose level is \( \phi_1 \) at time \( 0 \) and as \( time \to \infty \). This is intentional, so that \( \phi_1 \) can be interpreted as both the blood glucose level before ingesting alcohol and the blood glucose level after the effect of alcohol ingestion has washed out.

Pinheiro and Bates (2000, exercise 3, 412) analyzed this dataset in the context of a three-level NLME model. They initially proposed the following stage 2 specification,

\[
\begin{align*}
\phi_{1jk} &= \beta_1 + u_{1j,k}^{(3)} + u_{1j,k}^{(2)} \\
\phi_{2jk} &= \beta_2 + u_{2j,k}^{(3)} + u_{2j,k}^{(2)} \\
\phi_{3jk} &= \beta_3
\end{align*}
\]

(23)

\[
\begin{bmatrix}
u_{1j,k}^{(3)} \\
u_{2j,k}^{(3)}
\end{bmatrix} \sim N(0, \Sigma_3) \\
\begin{bmatrix}
u_{1j,k}^{(2)} \\
u_{2j,k}^{(2)}
\end{bmatrix} \sim N(0, \Sigma_2) \\
\epsilon_{ijk} \sim N(0, \sigma^2_\epsilon)
\]

where \( \Sigma_2 \) and \( \Sigma_3 \) are general symmetric covariance matrices. \( u_{1j,k}^{(2)} \) and \( u_{2j,k}^{(2)} \) are random intercepts at the guar-within-subject level and can be specified in `menl` as `UU1[subject>guar]` and `UU2[subject>guar]`. 
The full model defined by (22) and (23) contains many parameters. We will follow our own advice from example 11 and specify the iterate() option to check how reasonable our model is for the data we have.

```
.menl glucose = {phi1:} + {phi2:}*c.time#c.time#c.time*exp(-{phi3}*time),
> define(phi1: U1[subject] UU1[subject>guar])
> define(phi2: U2[subject] UU2[subject>guar])
> covariance(U1 U2, unstructured) covariance(UU*, unstructured)
> stddeviations iterate(3)
```

Obtaining starting values by EM:

Alternating PNLS/LME algorithm:

Iteration 1:  linearization log likelihood = -189.447112
Iteration 2:  linearization log likelihood = -189.441164
Iteration 3:  linearization log likelihood = -189.441125

Computing standard errors:

Mixed-effects ML nonlinear regression  Number of obs = 196

<table>
<thead>
<tr>
<th>Path</th>
<th>No. of Groups</th>
<th>Observations per Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>subject</td>
<td>7</td>
<td>28</td>
</tr>
<tr>
<td>subject&gt;guar</td>
<td>14</td>
<td>14</td>
</tr>
</tbody>
</table>

Linearization log likelihood = -189.44113

phi1:  U1[subject] UU1[subject>guar]
phi2:  U2[subject] UU2[subject>guar]

| glucose        | Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|----------------|--------|-----------|-------|------|---------------------|
| phi1_cons      | 3.661565 | 0.1160346 | 31.56 | 0.000 | 3.434142 3.888989   |
| phi2_cons      | 0.4283297 | 0.0530027 | 8.08  | 0.000 | 0.3244463 0.5322132 |
| /phi3          | 0.5896813 | 0.013861  | 42.54 | 0.000 | 0.5625143 0.6168483 |

Random-effects Parameters

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>subject: Unstructured</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sd(U1)</td>
<td>0.2624563</td>
<td>0.0926846</td>
<td>0.1313594 0.524388</td>
</tr>
<tr>
<td>sd(U2)</td>
<td>0.0598425</td>
<td>0.0724595</td>
<td>0.0055763 0.6422089</td>
</tr>
<tr>
<td>corr(U1,U2)</td>
<td>-0.1489692</td>
<td>0.9201993</td>
<td>-0.9636424 0.9347058</td>
</tr>
<tr>
<td>subject&gt;guar: Unstructured</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sd(UU1)</td>
<td>0.0919525</td>
<td>0.0764238</td>
<td>0.0180348 0.4688308</td>
</tr>
<tr>
<td>sd(UU2)</td>
<td>0.1227069</td>
<td>0.0412879</td>
<td>0.0634542 0.237289</td>
</tr>
<tr>
<td>corr(UU1,UU2)</td>
<td>0.99999</td>
<td>0.0047045</td>
<td>0.9636424 0.9347058</td>
</tr>
</tbody>
</table>

sd(Residual) 0.5712262 | 0.0305339 | 0.514409 0.6343189

Warning: convergence not achieved

The estimated correlation corr(UU1,UU2) is near one with the confidence interval spanning the entire range for the correlation parameter, which indicates that the random-effects structure is overparameterized. The confidence interval for corr(U1,U2) contains zero, which suggests that this term does not contribute much to explaining between-subject variability. If we try to fit this model without the iterate() option, it will continue iterating without convergence.
We simplify our model by assuming independence between random effects; that is, we assume that random-effects covariance matrices $\Sigma_2$ and $\Sigma_3$ are diagonal.

Recall that covariance(, independent) is assumed by default, so we do not need to explicitly specify the covariance() option:

```
. menl glucose = {phi1:} + {phi2:}*c.time#c.time#c.time*exp(-{phi3}*time),
  > define(phi1: U1[subject] UU1[subject>guar])
  > define(phi2: U2[subject] UU2[subject>guar]) stddeviations
```

Obtaining starting values by EM:

Alternating PNLS/LME algorithm:

```
Iteration 1:  linearization log likelihood = -190.355295
Iteration 2:  linearization log likelihood = -190.36034
Iteration 3:  linearization log likelihood = -190.363301
Iteration 4:  linearization log likelihood = -190.364176
Iteration 5:  linearization log likelihood = -190.363753
Iteration 6:  linearization log likelihood = -190.363969
Iteration 7:  linearization log likelihood = -190.363859
Iteration 8:  linearization log likelihood = -190.363914
Iteration 9:  linearization log likelihood = -190.363886
```

Computing standard errors:

Mixed-effects ML nonlinear regression

<table>
<thead>
<tr>
<th>Path</th>
<th>No. of Groups</th>
<th>Observations per Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>subject</td>
<td>7</td>
<td>28</td>
</tr>
<tr>
<td>subject&gt;guar</td>
<td>14</td>
<td>14</td>
</tr>
</tbody>
</table>

Linearization log likelihood = -190.36389

| glucose            | Coef.    | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|--------------------|----------|-----------|-------|--------|---------------------|
| phi1 _cons         | 3.658712 | .1168642  | 31.31 | 0.000  | 3.429662 3.887762  |
| phi2 _cons         | .4239173 | .0526333  | 8.05  | 0.000  | .320758 .5270766  |
| /phi3              | .5876636 | .0137214  | 42.83 | 0.000  | .5607701 .6145571 |

Random-effects Parameters

<table>
<thead>
<tr>
<th>Path</th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
</table>
| subject: Independent
| sd(U1)          | .2685609  | .092104   | .137126 .5259758    |
| sd(U2)          | .0422075  | .1078502  | .0002821 .6315585   |
| subject>guar: Independent
| sd(UU1)         | .0666034  | .1527525  | .0007435 5.966202   |
| sd(UU2)         | .1362263  | .0433548  | .0730065 .2541913   |
| sd(Residual)    | .5732488  | .0309928  | .5156118 .6373288   |

The random-effects structure may still be overparameterized, given small estimates for sd(U2) and sd(UU1). If we were to perform an LR test of the corresponding variance components being zero,
we would have no statistical evidence to reject this null hypothesis; see example 7 for an instance of performing an LR test.

Example 21: Three-level model with continuous-time AR(1) error structure

The main objective of the study from example 20 was to determine whether the use of the dietary additive guar significantly affected time profiles of the blood glucose levels of subjects.

We continue with the model without random effects $U_2[\text{subject}]$ and $UU_1[\text{subject}>\text{guar}]$ and include covariate $\text{guar}$ for all $\phi_{jk}$'s. Hand and Crowder (1996) also suggested to use a continuous-time AR(1) correlation structure for the guar-within-subject errors, which is specified in menl as `rescorrelation(ctar1, t(time))`:

\[
\text{menl glucose = \{phi1:\} + \{phi2:\}c.time#c.time#exp(-\{phi3:\}time),}
\]
\[
> \text{define(phi1: i.guar U1[subject]) define(phi2: i.guar UU2[subject>guar])}
\]
\[
> \text{define(phi3: i.guar, xb) rescorrelation(ctar1, t(time)) stddeviations}
\]

Obtaining starting values by EM:
Alternating PNLS/LME algorithm:
Iteration 1: linearization log likelihood = -180.623038
Iteration 2: linearization log likelihood = -181.270273
(iteration log omitted)
Iteration 24: linearization log likelihood = -181.187002
Iteration 25: linearization log likelihood = -181.186986
Computing standard errors:
Mixed-effects ML nonlinear regression

<table>
<thead>
<tr>
<th>Path</th>
<th>No. of Groups</th>
<th>Minimum</th>
<th>Average</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>subject</td>
<td>7</td>
<td>28</td>
<td>28.0</td>
<td>28</td>
</tr>
<tr>
<td>subject&gt;guar</td>
<td>14</td>
<td>14</td>
<td>14.0</td>
<td>14</td>
</tr>
</tbody>
</table>

Linearization log likelihood = -181.18699
Wald chi2(3) = 0.66
Prob > chi2 = 0.8814

| glucose       | Coef.  | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|---------------|--------|-----------|-------|-------|---------------------|
| phi1 guar     | -.0814355 | .1532736 | -0.53 | 0.595 | -.381846            |
| _cons         | 3.685365  | .1433368  | 25.71 | 0.000 | 3.40443            |
| phi2 guar     | .0109469 | .0883807  | 0.12  | 0.901 | -.162276           |
| _cons         | .344372   | .0609614  | 5.67  | 0.000 | .2254191           |
| phi3 guar     | .0103743  | .0330196  | 0.31  | 0.753 | -.054343           |
| _cons         | .5514012  | .022009   | 25.05 | 0.000 | .5082642           |
The dietary additive guar does not seem to affect the blood-glucose-level profiles over time. This actually conforms with the plot of the data from example 20, where, except for subject 6, the profiles with and without guar are similar.

Example 22: Using group() in the presence of random effects

The actual NLME model presented in Hand and Crowder (1996) for these glucose data included random effects for $\phi_1$ and $\phi_2$ only at the subject level and used a continuous-time AR(1) correlation structure on time for the guar-within-subject errors, with errors from different guar-within-subject clusters assumed to be independent. This model can be specified in menl using rescorrelation()'s group() suboption:

```
  . menl glucose = {phi1:} + {phi2:}*c.time#c.time#c.time*exp(-{phi3:}*time),
  >    define(phi1: i.guar U1[subject])
  >    define(phi2: i.guar U2[subject])
  >    define(phi3: i.guar, xb)
  >    rescorrelation(ctar1, t(time) group(guar)) stddeviations
```

Obtaining starting values by EM:

Alternating PNLS/LME algorithm:

Iteration 1:  linearization log likelihood = -183.720799
Iteration 2:  linearization log likelihood = -183.916983
            (iteration log omitted)
Iteration 13: linearization log likelihood = -183.90513
Iteration 14: linearization log likelihood = -183.905112

Computing standard errors:

Mixed-effects ML nonlinear regression  Number of obs   =   196

<table>
<thead>
<tr>
<th>Path</th>
<th>No. of Groups</th>
<th>Observations per Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Minimum    Average   Maximum</td>
</tr>
<tr>
<td>subject</td>
<td>7</td>
<td>28         28.0      28</td>
</tr>
<tr>
<td>guar</td>
<td>14</td>
<td>14         14.0      14</td>
</tr>
</tbody>
</table>
The fixed-effects estimates are similar to those in example 21, and the same conclusion is reached regarding the effect of the dietary additive guar on the blood-glucose-levels profiles over time. AIC and BIC may be used to decide on which model is better.

Notice the note displayed by menl following the command specification about the group variable guar being nested within variable subject. When you specify group(grpvar) within the rescorrelation() (or rescovariance()) option in the presence of random effects, grpvar is assumed to represent the lowest level of hierarchy and is thus assumed to be nested within other hierarchical levels.

Example 23: Three-level model with block-diagonal covariance matrix

Pinheiro and Bates (2000) report the data from the experiment conducted by Microelectronics Division of Lucent Technologies to study the variability in the manufacturing of analog MOS circuits. The intensities of the current (in mA) were collected on n-channel devices at five ascending voltages: 0.8, 1.2, 1.6, 2.0, and 2.4 V. Measurements were made on 8 sites of each of 10 wafers. The main objective of the study was to build an empirical model to simulate the behavior of similar circuits.
The intensity of the current at the $i$th level of voltage in the $j$th site within the $k$th wafer is expressed as

$$\text{current}_{ijk} = \phi_{1jk} + \phi_{2jk} \cos(\phi_{3jk} \text{voltage}_i + \pi/4) + \epsilon_{ijk}$$

where

$$\phi_{1jk} = \beta_0 + u_{0k}^{(3)} + u_{1j,k}^{(2)} + (\beta_1 + u_{1k}^{(3)} + u_{1j,k}^{(2)}) \text{voltage}_i + (\beta_2 + u_{2k}^{(3)} + u_{2j,k}^{(2)}) \text{voltage}_i^2$$

$$\phi_{2jk} = \beta_3 + u_{3k}^{(3)} + u_{3j,k}^{(2)}$$

$$\phi_{3jk} = \beta_4 + u_{4k}^{(3)}$$

$$\mathbf{u}_k^{(3)} = \left[ \begin{array}{c} u_{0k}^{(3)} \\ u_{1k}^{(3)} \\ u_{2k}^{(3)} \\ u_{3k}^{(3)} \\ u_{4k}^{(3)} \end{array} \right] \sim N(\mathbf{0}, \Sigma_3) \quad \mathbf{u}_{j,k}^{(2)} = \left[ \begin{array}{c} u_{0j,k}^{(2)} \\ u_{1j,k}^{(2)} \\ u_{2j,k}^{(2)} \\ u_{3j,k}^{(2)} \end{array} \right] \sim N(\mathbf{0}, \Sigma_2) \quad \epsilon_{ijk} \sim N(0, \sigma^2_\epsilon)$$

Parameters $\beta_0$, $\beta_1$, and $\beta_2$ characterize the quadratic component of the model, and amplitude $\beta_3$ and frequency $\beta_4$ characterize the periodic component represented by the cosine wave.

For illustration, consider the following random-effects covariance structures:

$$\Sigma_3 = \begin{bmatrix} \sigma_{11}^{(3)} & \sigma_{12}^{(3)} & \sigma_{13}^{(3)} & \sigma_{14}^{(3)} & \sigma_{15}^{(3)} \\ \sigma_{12}^{(3)} & \sigma_{22}^{(3)} & 0 & 0 & 0 \\ \sigma_{13}^{(3)} & 0 & \sigma_{33}^{(3)} & 0 & 0 \\ \sigma_{14}^{(3)} & 0 & 0 & \sigma_{44}^{(3)} & 0 \\ \sigma_{15}^{(3)} & 0 & 0 & 0 & \sigma_{55}^{(3)} \end{bmatrix} \quad \Sigma_2 = \begin{bmatrix} \sigma_{11}^{(2)} & \sigma_{12}^{(2)} & 0 & 0 & 0 \\ \sigma_{12}^{(2)} & \sigma_{22}^{(2)} & 0 & 0 & 0 \\ 0 & 0 & \sigma_{33}^{(2)} & \sigma_{34}^{(2)} & 0 \\ 0 & 0 & 0 & \sigma_{44}^{(2)} & 0 \\ 0 & 0 & 0 & 0 & \sigma_{55}^{(2)} \end{bmatrix}$$

If we were to fit this model by using \texttt{menl}, we would type

```
. use https://www.stata-press.com/data/r16/wafer
(Modeling of analog MOS circuits)
. menl current = {phi1:}+{phi2:}*cos({phi3:}*voltage + _pi/4),
> define(phi1: voltage c.voltage#c.voltage W0[wafer] S0[wafer>site])
> c.voltage#(W1[wafer] S1[wafer>site])
> c.voltage#c.voltage#(W2[wafer] S2[wafer>site]))
> define(phi2: W3[wafer] S3[wafer>site]) define(phi3: W4[wafer], xb)
> covariance(S0 S1, unstructured) covariance(S2 S3, unstructured)
> covariance(W*, independent) stddeviations
```

In the specification above, $\Sigma_3$ is specified as \texttt{covariance(W*, independent)}, although this specification could have been omitted because \texttt{independent} is \texttt{menl}'s default random-effects covariance structure. The block-diagonal matrix $\Sigma_2$ is specified by using repeated \texttt{covariance()} options: \texttt{covariance(S0 S1, unstructured)} and \texttt{covariance(S2 S3, unstructured)}. If we tried to run this model, we would find out that it is overparameterized.
Because of the large number of random effects at each grouping level, to avoid numerically unstable estimates, we will further simplify our model by assuming independence between $u_{2j,k}^{(2)}$ and $u_{3j,k}^{(2)}$, which implies that $\sigma_{34}^{(2)} = 0$:

$$\Sigma_3 = \begin{bmatrix}
\sigma_{11}^{(3)} & \sigma_{12}^{(3)} & \sigma_{13}^{(3)} & \sigma_{14}^{(3)} \\
\sigma_{21}^{(3)} & \sigma_{22}^{(3)} & 0 & 0 \\
\sigma_{31}^{(3)} & 0 & \sigma_{33}^{(3)} & 0 \\
\sigma_{41}^{(3)} & 0 & 0 & \sigma_{44}^{(3)} \\
\end{bmatrix} \quad \Sigma_2 = \begin{bmatrix}
\sigma_{11}^{(2)} & \sigma_{12}^{(2)} & 0 & 0 \\
\sigma_{12}^{(2)} & \sigma_{22}^{(2)} & 0 & 0 \\
0 & 0 & \sigma_{33}^{(2)} & 0 \\
0 & 0 & 0 & \sigma_{44}^{(2)} \\
\end{bmatrix}$$

We now try to fit the above simpler model. Note that given the complexity of this model, it takes some time to execute.

```
use https://www.stata-press.com/data/r16/wafer
(Modeling of analog MOS circuits)
menl current = {phi1:}+{phi2:}*cos({phi3:}*voltage + _pi/4),
> define(phi1: voltage c.voltage#c.voltage W0[wafer] S0[wafer>site]
>           c.voltage#(W1[wafer] S1[wafer>site])
>           c.voltage#c.voltage#(W2[wafer] S2[wafer>site]))
> define(phi2: W3[wafer] S3[wafer>site]) define(phi3: W4[wafer], xb)
> covariance(S0 S1, unstructured) covariance(S2 S3, independent)
> covariance(W*, independent) stddeviations
Obtaining starting values by EM:
Alternating PNLS/LME algorithm:
Iteration 1: linearization log likelihood = 735.215957
Iteration 2: linearization log likelihood = 804.989061
Iteration 3: linearization log likelihood = 825.917169
Iteration 4: linearization log likelihood = 825.917107
Computing standard errors:
Mixed-effects ML nonlinear regression  Number of obs =  400

<table>
<thead>
<tr>
<th>Path</th>
<th>No. of Groups</th>
<th>Observations per Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimum</td>
<td>Average</td>
</tr>
<tr>
<td>wafer</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>wafer&gt;site</td>
<td>80</td>
<td>5</td>
</tr>
</tbody>
</table>
Nonlinear mixed-effects regression

Wald chi2(2) = 8764.00
Linearization log likelihood = 825.91711 Prob > chi2 = 0.0000

phi1: voltage c.voltage#c.voltage W0[wafer] S0[wafer>site]
c.voltage#W1[wafer] c.voltage#S1[wafer>site]
c.voltage#c.voltage#W2[wafer]
c.voltage#c.voltage#S2[wafer>site]
phi2: W3[wafer] S3[wafer>site]
phi3: W4[wafer], xb

| current | Coef. | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|---------|-------|-----------|-----|------|----------------------|
| phi1   | voltage | 6.046937 | .1022628 | 59.13 | 0.000 | 5.846505 6.247368 |
|     | c.voltage#c.voltage | 1.158782 | .0159668 | 72.57 | 0.000 | 1.127487 1.190076 |
|     | _cons | -4.658034 | .0361762 | -128.76 | 0.000 | -4.728938 -4.58713 |
| phi2   | _cons | .1684428 | .002054 | 82.01 | 0.000 | .1644171 .1724685 |
| phi3   | _cons | 6.449391 | .001963 | 3285.40 | 0.000 | 6.445543 6.453238 |

Random-effects Parameters

<table>
<thead>
<tr>
<th>wafer: Independent</th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>sd(W0)</td>
<td>.1107108</td>
<td>.0262518</td>
<td>.0695589 .1762087</td>
</tr>
<tr>
<td>sd(W1)</td>
<td>.3041975</td>
<td>.0764653</td>
<td>.1858623 .4978746</td>
</tr>
<tr>
<td>sd(W2)</td>
<td>.0449994</td>
<td>.0125441</td>
<td>.026057 .077121</td>
</tr>
<tr>
<td>sd(W3)</td>
<td>.0057862</td>
<td>.0016144</td>
<td>.0033489 .0099974</td>
</tr>
<tr>
<td>sd(W4)</td>
<td>.0061349</td>
<td>.0013878</td>
<td>.0039377 .0095579</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>wafer&gt;site: Unstructured</th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>sd(S0)</td>
<td>.0729495</td>
<td>.006297</td>
<td>.0615952 .0863969</td>
</tr>
<tr>
<td>sd(S1)</td>
<td>.2930062</td>
<td>.0252425</td>
<td>.2474831 .3469029</td>
</tr>
<tr>
<td>corr(S0,S1)</td>
<td>-.8113227</td>
<td>.0413367</td>
<td>-.8782248 -.7132763</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>wafer&gt;site: Independent</th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>sd(S2)</td>
<td>.0627587</td>
<td>.0053067</td>
<td>.0531738 .0740712</td>
</tr>
<tr>
<td>sd(S3)</td>
<td>.0080612</td>
<td>.0006861</td>
<td>.0068227 .0095244</td>
</tr>
</tbody>
</table>

| sd(Residual) | .0008407 | .0000711 | .0007122 .0009922 |
In this example, our primary focus was to demonstrate how to use \texttt{menl} to fit a block-diagonal random-effects covariance structure. But if we were to interpret our fixed-effects estimates, the average frequency of the cosine wave, $\beta_4 = E(\phi_{3jk})$, for example, is estimated to be $6.45V^{-1}$, with a corresponding estimated period of $2\pi/\hat{\beta}_4 \approx 0.97V$. Also, some of the estimates of standard deviations such as $\text{sd}(W_2)$, $\text{sd}(W_3)$, and $\text{sd}(W_4)$ are very small, which suggests that this model may still be too rich for the observed data. If we proceeded to further analyze these data, we would consider simpler models. For example, at the very least, we would have omitted the term $W_3[\text{wafer}]$ from this model.

### Obtaining initial values

Obtaining good starting or initial values is important for the estimation of many statistical models, but it is often crucial for the estimation of NLME models. NLME models are known to be sensitive to the initial values and to have difficulty converging. Highly nonlinear mean specification or complicated variance–covariance structures for random effects and errors can often lead to multiple solutions, which requires considering different sets of initial values.

By default, \texttt{menl} uses the EM algorithm to obtain initial values. This default routine works well in many cases but cannot be guaranteed to provide good initial values in all situations. Sometimes, you may need to specify your own initial values. Trying different initial values can also be useful to investigate the existence of multiple solutions and to verify convergence to a global maximum.

So far we have been “lucky” that all the examples worked without us having to specify initial estimates. You may not be that lucky with your data and model. So, in this section, we provide some guidance on how to find good initial values when the default initial values do not work well.

We present three approaches that you may choose to explore to find good initial estimates for the fixed effects. In some cases, you may also be able to obtain initial estimates for covariance parameters; see \textit{Linearization approach to finding initial values}.

### Linearization approach to finding initial values

Sometimes, we can use an LME model to obtain initial values of the NLME model by holding some of the parameters fixed at specific values. We can then fit the resulting LME model by using the \texttt{mixed} command and use the corresponding estimates as initial values for the NLME model. We refer to this initialization method as the linearization method.

We could have used this method in example 14 and example 23, if the default EM method did not provide reasonable initial estimates. In any case, it is good practice to specify different initial values to investigate potential convergence of the algorithm to a local maximum.

For instance, in example 14, we fit

$$\text{follicles}_{ij} = \phi_{1j} + \phi_{2j} \sin(2\pi\phi_{3j}\text{stime}_{ij}) + \phi_{4j} \cos(2\pi\phi_{3j}\text{stime}_{ij}) + \epsilon_{ij}$$

where

$$\phi_j = \begin{bmatrix} \phi_{1j} \\ \phi_{2j} \\ \phi_{3j} \\ \phi_{4j} \end{bmatrix} = \begin{bmatrix} \beta_1 + u_{1j} \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}$$
This model is nonlinear because of the parameter $\phi_{3j}$. To obtain initial values, we can hold $\phi_{3j}$ (or $\beta_3$) fixed at a specific value, say, $\beta_3 = 1$, thus making the above model linear,

$$\text{follicles}_{ij} = \phi_{1j} + \phi_{2j} \sin(2\pi \phi_{3j}\text{stime}_{ij}) + \phi_{4j} \cos(2\pi \phi_{3j}\text{stime}_{ij}) + \epsilon_{ij}$$

where

$$\phi_j = \begin{bmatrix} \phi_{1j} \\ \phi_{2j} \\ \phi_{3j} \\ \phi_{4j} \end{bmatrix} = \begin{bmatrix} \beta_1 + u_{1j} \\ \beta_2 \\ 1 \\ \beta_4 \end{bmatrix}$$

Or, more compactly,

$$\text{follicles}_{ij} = \beta_1 + u_{1j} + \beta_2 \sin(2\pi \text{stime}_{ij}) + \beta_4 \cos(2\pi \text{stime}_{ij}) + \epsilon_{ij}$$

Now that the model is linear, we can use the mixed command to obtain initial values for $\beta_1$, $\beta_2$, and $\beta_4$ to be used in menl. In the code below, variables sin1 and cos1 are $\sin(2\pi \text{stime}_{ij})$ and $\cos(2\pi \text{stime}_{ij})$, respectively, and || mare: specifies a random intercept at the mare level (see [ME] mixed). Also, for consistency with example 13, we assume an AR(1) within-group error correlation structure:

```
. mixed follicles sin1 cos1 || mare:, residuals(ar 1, t(time)) nolog
```

```
Mixed-effects ML regression
Number of obs = 308
Group variable: mare
Number of groups = 11
Obs per group:
  min = 25
  avg = 28.0
  max = 31
Wald chi2(2) = 39.00
Log likelihood = -776.51731  Prob > chi2 = 0.0000

Follicles | Coef. Std. Err. z  P>|z|  [95% Conf. Interval]
-------------- | -------------- | -------- | ------ | ------------------
     sin1 | -2.958619  .4935054  -6.00 0.000  -3.925872 -1.991366
     cos1 | -0.8798847  .5031764  -1.75 0.080  -1.866092  .1063228
    _cons |  12.18963  .9017435  13.52 0.000  10.42224 13.95701

Random-effects Parameters | Estimate  Std. Err.  [95% Conf. Interval]
------------------------------ | ---------- | ----------- | ---------------------
 mare: Identity var(_cons) | 7.095503  3.764874  2.508047  20.07385
 Residual: AR(1) rho | 0.5974664  0.0547217  0.4795551  0.6941854
                    var(e) | 13.08097  1.765326  10.04078  17.0417
```

LR test vs. linear model: chi2(2) = 242.63  Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.

We will now use the estimates of the fixed effects shown in the output table as initial values for menl by specifying the initial() option. We use 1 as the initial value for $/phi_3$. There are three ways to specify initial values in the initial() option; see Specifying initial values. Here we will use the specification where we repeatedly list a parameter name followed by its initial value; also see Examples of specifying initial values.
local xb phi1:_cons 12.2 /phi2 -3.0 /phi3 1 /phi4 -.88

menl follicles = {phi1: U1[mare], xb} + {phi2}*sin(2*_pi*stime*{phi3}) +
> {phi4}*cos(2*_pi*stime*{phi3}), rescorrelation(ar 1, t(time)) init('xb')

Alternating PNLS/LME algorithm:
Iteration 1:  linearization log likelihood = -775.627845
Iteration 2:  linearization log likelihood = -775.624331
Iteration 3:  linearization log likelihood = -775.624331

Computing standard errors:
Mixed-effects ML nonlinear regression  Number of obs = 308
Group variable: mare  Number of groups = 11
Obs per group:
  min = 25
  avg = 28.0
  max = 31

Linearization log likelihood = -775.62433

   phil:  U1[mare], xb

|       | Coef.   | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|-------|---------|-----------|-------|------|---------------------|
| phil  |         |           |       |      |                     |
| _cons | 12.18125| .9055129  | 13.45 | 0.000| 10.40647             |
|       |         |           |       |      | 13.95602             |
| /phi2 | -2.874414| .5389564  | -5.33 | 0.000| -3.930749             |
|       |         |           |       |      | -1.818079             |
| /phi3 | .9191143| .0512334  | 17.94 | 0.000| .8186987              |
|       |         |           |       |      | 1.01953               |
| /phi4 | -1.675311| .6766108  | -2.48 | 0.013| -3.001444             |
|       |         |           |       |      | -.3491784             |

Random-effects Parameters Estimate Std. Err. [95% Conf. Interval]
mare: Identity
  var(U1)  7.207072  3.755604  2.595362  20.01335
Residual: AR(1),
  time time
  var(e)  12.63377  1.646897  9.785277  16.31146
  corr   .5823733  .0544508  .4656903  .679153

In the above, we initialized only fixed-effects parameters and used naive initial estimates of 1 for random-intercept and error variances and 0 for the correlation. We could have specified initial()'s fixed suboption to use the EM algorithm to compute initial estimates for the random-effects parameters; see Examples of specifying initial values for details.

With the linearization approach, we can also use estimates of the random-effects parameters from the mixed command to initialize the corresponding parameters of menl. This is an advantage of the linearization approach over the other two approaches we discuss in subsequent sections. One complication with the initialization of random-effects parameters is that the initial values must be supplied in the estimation metric, the metric used during estimation, instead of the parameter original metric. For example, instead of variances, we must supply estimates of log-standard deviations, and instead of covariances or correlations, we must supply inverse hyperbolic tangents of correlation parameters. Luckily for us, mixed stores results using the same metric as menl and provides the estmetric option to display parameters in that metric.

In our example, the random-effects parameters are the random-intercept variance, the within-group error variance, and the correlation between error terms. We refit the earlier mixed command but now with the estmetric option to obtain the estimates of the random-effects parameters as they are stored in e(b).
. mixed follicles sin1 cos1 || mare:, residuals(ar 1, t(time)) nolog estmetric

Mixed-effects ML regression
Number of obs = 308
Group variable: mare
Number of groups = 11
Obs per group:
  min = 25
  avg = 28.0
  max = 31

Log likelihood = -776.51731

Wald chi2(2) = 39.00
Prob > chi2 = 0.0000

| Coef.  | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|--------|-----------|-----|-----|----------------------|
| follicles |          |     |     |                      |
| sin1    | -2.958619 | .4935054 | -6.00 | 0.000 | -3.925872 -1.991366 |
| cos1    | -.8798847 | .5031764 | -1.75 | 0.080 | -1.866092 .1063228 |
| _cons   | 12.18963  | .9017435 | 13.52 | 0.000 | 10.42224 13.95701 |

| Coef.  | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|--------|-----------|-----|-----|----------------------|
| lns1_1_1 |          |     |     |                      |
| _cons  | .9797306  | .2653 | .5762499 | 1.665722 |

| Coef.  | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|--------|-----------|-----|-----|----------------------|
| lnsig_e |          |     |     |                      |
| _cons  | 1.285579  | .0674768 | 19.05 | 0.000 | 1.153327 1.417832 |

| Coef.  | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|--------|-----------|-----|-----|----------------------|
| r_atr1 |          |     |     |                      |
| _cons  | .6891977  | .0850992 | 8.10 | 0.000 | .5224063 .8559892 |

menl uses the same ordering of the parameters as mixed does, so we can simply list all the estimates directly in the initial() option. When we list the values without parameter names, we must specify initial()’s copy suboption and specify the values for all parameters. In our example, we specify four fixed-effects coefficients and three random-effects parameters.

. menl follicles = {phi1: U1[mare], xb} + {phi2}*sin(2*_pi*stime*{phi3}) +
>   {phi4}*cos(2*_pi*stime*{phi3}), rescorrelation(ar 1, t(time))
>   initial(12.2 -3.0 1 -.88 .98 1.29 .69, copy)

Alternating PNLS/LME algorithm:
Iteration 1:  linearization log likelihood = -775.624332
Iteration 2:  linearization log likelihood = -775.624331

Computing standard errors:

Mixed-effects ML nonlinear regression
Number of obs = 308
Group variable: mare
Number of groups = 11
Obs per group:
  min = 25
  avg = 28.0
  max = 31

Linearization log likelihood = -775.62433

| Coef.  | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|--------|-----------|-----|-----|----------------------|
| phi1   |          |     |     |                      |
| _cons  | 12.18125  | .9055135 | 13.45 | 0.000 | 10.40647 13.95602 |

| Coef.  | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|--------|-----------|-----|-----|----------------------|
| /phi2  | -2.874434 | .5389241 | -5.33 | 0.000 | -3.930706 -1.818162 |
| /phi3  | .919119   | .0512356 | 17.94 | 0.000 | .818699 1.019539 |
| /phi4  | -1.675261 | .6766409 | -2.48 | 0.013 | -3.001452 -.3490689 |
The results are different from those in example 14. The value of the linearization log likelihood in this example, $-775.62$, is larger than that from example 14, $-789.43$. So it appears that we have converged to a local maximum of the linearization log likelihood in example 14.

Our initial values based on mixed turned out to be better than those computed by default by menl. This is not surprising. In general, menl’s EM algorithm should produce reasonable initial values for many nonlinear models, but the initial values may not necessarily be optimal for all of those models. In this example, our initial values were tailored to the ovary data and the model.

In general, sensitivity to initial values is one of the key issues in NLME models, especially for models that involve periodic functions. Therefore, it is important to try different sets of initial values to verify global convergence before reporting your final results. Sometimes, you may even have to rely on your knowledge of the science behind the problem to decide which set of results is more reasonable.

### Graphical approach to finding initial values

If your model has parameters that have natural physical interpretations, you may be able to obtain starting values from a graph of the data.

Draper and Smith (1998) presented a dataset in which the trunk circumference $\text{circumf}$ (in mm) of five different orange trees was measured over seven different time points, stored in $\text{age}$. Pinheiro and Bates (2000) suggested the following model for these data:

$$
\text{circumf}_{ij} = \frac{\phi_{1j}}{1 + \exp\left\{-\left(\text{age}_{ij} - \phi_{2j}\right)/\phi_{3j}\right\}} + \epsilon_{ij}
$$

In this model, $\phi_{1j}$ is the asymptotic trunk circumference for the $j$th tree as $\text{age}_{ij} \to \infty$, $\phi_{2j}$ is the age at which the $j$th tree attains half of its asymptotic trunk circumference $\phi_{1j}$, and $\phi_{3j}$ is a scale parameter; see the graph below.

The stage 2 specification of this model is

$$
\phi_j = \begin{bmatrix}
\phi_{1j} \\
\phi_{2j} \\
\phi_{3j}
\end{bmatrix} = \begin{bmatrix}
\beta_1 + u_{1j} \\
\beta_2 \\
\beta_3
\end{bmatrix}
$$

where

$$
u_{1j} \sim N\left(0, \sigma_{u_1}^2\right), \quad \epsilon_{ij} \sim N\left(0, \sigma_{\epsilon}^2\right)
$$
Because the model parameters have graphical interpretations, we can plot our data and obtain initial values from the graph.

```
use https://www.stata-press.com/data/r16/orange
(Growth of orange trees (Draper and Smith, 1998))
twoway connected circumf age, connect(L) yline(175) xline(1582)
>   yline(87.5, lpattern(dash)) xline(700, lpattern(dash))
>   yline(131.25, lpattern("-...")) xline(1000, lpattern("-..."))
>   xlabel(0 118 484 700 1000 1372 1582) ylabel(#5 87.5 131.25 175)
```

From the above graph, the mean asymptotic trunk circumference can be estimated as 175 mm, which is roughly the mean of the circumference values at age 1,582 (in days). The trees attain half of their asymptotic trunk circumference, \(175/2 = 87.5\), at about age 700 (in days). Therefore, we use the initial estimates \(\beta_1 = 175\) for the asymptotic trunk circumference and \(\beta_2 = 700\) for the location of the inflection point. To obtain an initial estimate for \(\beta_3\), we note that when \(\text{age} = \beta_2 + \beta_3\) in (24), \(E(\text{circumf}_{ij}) = \beta_1/(1 + \exp(-1)) = 0.73\beta_1\), which we will approximate as \(0.75\beta_1\) for the purpose of the graph. That is, the logistic curve reaches approximately 3/4 of its asymptotic value, \(0.75 \times 175 = 131.25\), at \(\text{age} = \beta_2 + \beta_3\). The above graph suggests that the trees attain 3/4 of their final trunk circumference at about 1,000 days (= \(\beta_2 + \beta_3\)), giving an initial estimate of \(\beta_3 = 1000 - 700 = 300\). We can now supply these values to `menl` in the `initial()` option.

Unfortunately, the graph does not provide us with the estimates for variance components. In this case, we can use `initial()`’s `fixed` suboption to specify that the EM algorithm still be used to initialize variance components, while the supplied values be used to initialize fixed effects. If we do not specify `fixed`, `menl` will use naïve initial estimates for variance components such as ones for variances and zeros for covariances.
We now fit the model using our own initial estimates for fixed effects:

```
.menl circumf = {phi1: U1[tree], xb}/(1+exp(-(age-{phi2})/{phi3})),
> initial(phi1:_cons 175 /phi2 700 /phi3 300, fixed)
```

Obtaining starting values by EM:

Alternating PNLS/LME algorithm:

- Iteration 1: linearization log likelihood = -131.584945
- Iteration 2: linearization log likelihood = -131.584579
- Iteration 3: linearization log likelihood = -131.584579

Computing standard errors:

<table>
<thead>
<tr>
<th>Mixed-effects ML nonlinear regression</th>
<th>Number of obs = 35</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group variable: tree</td>
<td>Number of groups = 5</td>
</tr>
<tr>
<td>Obs per group:</td>
<td></td>
</tr>
<tr>
<td>min = 7</td>
<td>avg = 7.0</td>
</tr>
<tr>
<td>max = 7</td>
<td></td>
</tr>
</tbody>
</table>

Linearization log likelihood = -131.58458

<table>
<thead>
<tr>
<th>phi1: U1[tree], xb</th>
</tr>
</thead>
<tbody>
<tr>
<td>circumf</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>_cons</td>
</tr>
<tr>
<td>/phi2</td>
</tr>
<tr>
<td>/phi3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random-effects Parameters</th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>tree: Identity var(U1)</td>
<td>991.1514</td>
<td>639.4636</td>
<td>279.8776 - 3510.038</td>
</tr>
<tr>
<td>var(Residual)</td>
<td>61.56371</td>
<td>15.89568</td>
<td>37.11466 - 102.1184</td>
</tr>
</tbody>
</table>

For comparison, we fit the same model but now using the default initial values for fixed effects:

```bash
.menl circumf = {phi1: U1[tree], xb}/(1+exp(-(-age-{phi2})/{phi3}))
```

Obtaining starting values by EM:

Alternating PNLS/LME algorithm:

Iteration 1: linearization log likelihood = -131.584579

Computing standard errors:

Mixed-effects ML nonlinear regression

| Number of obs | 35 |
| Number of groups | 5 |

| Obs per group: | min | avg | max |
| | 7 | 7.0 | 7 |

Linearization log likelihood = -131.58458

```
phi:  U1[tree], xb
```

| circumf | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] |
| --- | --- | --- | --- | --- | --- |
| phi1 | _cons | 191.049 | 16.15403 | 11.83 | 0.000 | 159.3877 | 222.7103 |
| /phi2 | 722.556 | 35.15082 | 20.56 | 0.000 | 653.6616 | 791.4503 |
| /phi3 | 344.1624 | 27.14739 | 12.68 | 0.000 | 290.9545 | 397.3703 |

<table>
<thead>
<tr>
<th>Random-effects Parameters</th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>tree: Identity</td>
<td>var(U1)</td>
<td>991.1514</td>
<td>639.4636</td>
</tr>
<tr>
<td>var(Residual)</td>
<td>61.56371</td>
<td>15.89568</td>
<td>37.11466</td>
</tr>
</tbody>
</table>

The results are identical except for the iteration log.

**Smart regressions approach to finding initial values**

Consider the following NLME model,

\[ y_{ij} = \phi_1 + (\phi_2 - \phi_1) \exp \{- \exp(\phi_3) x_{ij}\} + \epsilon_{ij} \]

where

\[ \phi_j = \begin{bmatrix} \phi_{1j} \\ \phi_{2j} \\ \phi_{3j} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 + u_{1j} \\ \beta_3 \end{bmatrix} \]

Here \( \phi_{1j} \) is the asymptote as \( x_{ij} \to \infty \) and \( \phi_{2j} \) is the value of \( y_{ij} \) at \( x_{ij} = 0 \). Thus initial estimates, \( \beta_1^{(0)} \) and \( \beta_2^{(0)} \), may be obtained by using the graphical approach as described in **Graphical approach to finding initial values**. To obtain an initial estimate for \( \beta_3 \), notice that, ignoring the error term \( \epsilon_{ij} \) and setting \( u_{1j} = 0 \),

\[
\log (|y_{ij} - \beta_1|) = \log (\beta_2 - \beta_1) + \{- \exp(\beta_3)\} x_{ij}
\]

Therefore, we can regress \( \log(|y - \beta_1^{(0)}|) \) on \( x \) and use the estimated slope, \( \widehat{\beta}_x = -\exp(\beta_3^{(0)}) \), to obtain the initial value for \( \beta_3^{(0)} = \log(-\widehat{\beta}_x) \).
Examples of specifying initial values

When you want to assign initial values for a subset of the model parameters, for example, fixed effects or random-effects covariance parameters, you will often need to know their estimation names or, in other words, how `menl` labels them in `e(b)`. To learn the names, you can fit the model with the `iterate(0)` and `coeflegend` options first.

```
. menl ... , ... iterate(0) coeflegend
```

The `iterate(0)` option specifies to bypass maximization and only report the initial values and the likelihood evaluated at those values. The `coeflegend` option specifies that the legend of the parameters and how to specify them in an expression be displayed rather than displaying the statistics for the parameters.

Keep in mind, however, that `menl` does not perform estimation in the original parameter metric. For computational stability, the estimation is performed, loosely speaking, in a metric that transforms all parameters to be defined on a real line. For example, a log transformation is used for standard deviations, and an inverse hyperbolic tangent transformation is used for correlations. When you specify initial values, you must specify them for parameters in the estimation metric and not the original metric.

`coeflegend` displays parameter names as they are stored in `e(b)`, which, for `menl`, are the names of estimation parameters. If you also want to see parameters in the original metric, you can specify `coeflegend` on replay.

```
. menl ... , ... iterate(0)
. menl, coeflegend
```

For example, recall the NLME model for the soybean data from example 9. Suppose that we want to supply our own initial values.
We fit the model with iterate(0) and coeflegend:

```
. menl weight = {phi1:}/(1+exp(-(time-{phi2:})/{phi3:})),
> define(phi1: U1[plot], xb)
> define(phi2: U2[plot], xb)
> define(phi3: U3[plot], xb)
> covariance(U*, unstructured) iterate(0) coeflegend
```

Obtaining starting values by EM:

```
Computing standard errors:
```

Mixed-effects ML nonlinear regression
Number of obs = 412
Group variable: plot
Number of groups = 48

<table>
<thead>
<tr>
<th>Obs per group:</th>
</tr>
</thead>
<tbody>
<tr>
<td>min = 8</td>
</tr>
<tr>
<td>avg = 8.6</td>
</tr>
<tr>
<td>max = 10</td>
</tr>
</tbody>
</table>

Linearization log likelihood = -740.06177

```
phi1: U1[plot], xb
phi2: U2[plot], xb
phi3: U3[plot], xb
```

<table>
<thead>
<tr>
<th>weight</th>
<th>Coef. Legend</th>
</tr>
</thead>
<tbody>
<tr>
<td>phi1 _cons</td>
<td>19.26527 _b[phi1:_cons]</td>
</tr>
<tr>
<td>phi2 _cons</td>
<td>55.05299 _b[phi2:_cons]</td>
</tr>
<tr>
<td>phi3 _cons</td>
<td>8.385531 _b[phi3:_cons]</td>
</tr>
<tr>
<td>/plot</td>
<td></td>
</tr>
<tr>
<td>lnsd(U1)</td>
<td>1.650846 _b[/plot:lnsd(U1)]</td>
</tr>
<tr>
<td>lnsd(U2)</td>
<td>1.436634 _b[/plot:lnsd(U2)]</td>
</tr>
<tr>
<td>lnsd(U3)</td>
<td>.4081525 _b[/plot:lnsd(U3)]</td>
</tr>
<tr>
<td>athcorr(U2, U1)</td>
<td>.9055785 _b[/plot:athcorr(U2,U1)]</td>
</tr>
<tr>
<td>athcorr(U3, U1)</td>
<td>.8482105 _b[/plot:athcorr(U3,U1)]</td>
</tr>
<tr>
<td>athcorr(U3, U2)</td>
<td>1.537798 _b[/plot:athcorr(U3,U2)]</td>
</tr>
<tr>
<td>/Residual</td>
<td></td>
</tr>
<tr>
<td>lnsigma</td>
<td>.1069986 _b[/Residual:lnsigma]</td>
</tr>
</tbody>
</table>

Warning: convergence not achieved

Parameter names are listed within the _b[] specifier.

In what follows, we will outline only the syntax of the specifications. If you actually want to run all the examples to see the initialization in action, we suggest that you specify iterate(0) for speed.

Let’s first specify initial values for fixed effects only. The fixed-effects parameters are phi1:_cons, phi2:_cons, and phi3:_cons. Suppose that we want to initialize them with 19, 55, and 8.

We can type

```
. menl ..., ... initial(phi1:_cons 19 phi2:_cons 55 phi3:_cons 8)
```

Or, more compactly, we can type

```
. local fe phi1:_cons 19 phi2:_cons 55 phi3:_cons 8
. menl ..., ... initial('fe')
```
When you specify the `initial()` option, `menl` does not perform the EM algorithm to initialize the parameters but instead uses the values you supplied. If you specify values for only a subset of parameters, the remaining parameters will be initialized with naive initial values such as zeros for fixed effects and correlations and ones for variances. Often, you may have good initial values for fixed effects but not for variance components. In this situation, `menl` provides `initial()`’s `fixed` suboption. This option specifies that the supplied values be used for fixed effects but that the EM algorithm still be used to obtain initial values for variance components. If you specify only a subset of values for fixed effects, the remaining fixed effects will still be initialized with zeros even if `fixed` is specified. We recommend that you specify `fixed` when you intend to supply initial values only for the fixed effects.

```
local fe phi1:_cons 19 phi2:_cons 55 phi3:_cons 8
menl ..., ... initial('fe', fixed)
```

Now suppose that we also want to assign initial values for random-effects parameters. As we mentioned earlier, remember that we assign initial values for standard deviations in the log metric and for correlation in the inverse hyperbolic tangent or atanh metric. For example, if you want to assign an initial value of 2 to $\sigma_\varepsilon$, then you should supply $\log(2)$ to the `initial()` option. Similarly, if you want to assign a value of 0.7 to the correlation of two random effects, then you should provide `atanh(0.7)` to the `initial()` option.

Continuing with example 9, suppose that we want to specify the following initial values for the random-effects covariance parameters:

\[
\begin{pmatrix}
U1[plot] & U2[plot] & U3[plot] \\
\sigma_1 & 5 \\
\rho_{21} & 0.72 & \sigma_2 & 4 \\
\rho_{31} & 0.71 & \rho_{32} & 0.94 & \sigma_3 & 1.4
\end{pmatrix}
\]

The names of the parameters in the estimation metric that correspond to $\sigma_1$, $\sigma_2$, and $\sigma_3$ are `/plot:lnsd(U1)`, `/plot:lnsd(U2)`, and `/plot:lnsd(U3)` and that correspond to $\rho_{21}$, $\rho_{31}$, and $\rho_{32}$ are `/plot:athcorr(U2, U1)`, `/plot:athcorr(U3, U1)`, and `/plot:athcorr(U3, U2)`.

When specifying initial values for free parameters such as random-effects covariance parameters, you can omit the forward slash (`/`) at the beginning of their names. Keeping in mind that initial values for covariance parameters are supplied in the log and atanh metrics, we can type

```
local re_cov plot:lnsd(U1) log(5) // log(5)
local re_cov 're_cov' plot:lnsd(U2) 1.4 // log(4)
local re_cov 're_cov' plot:lnsd(U3) 0.34 // log(1.4)
local re_cov 're_cov' plot:athcorr(U2, U1) atanh(0.72) // atanh(0.72)
local re_cov 're_cov' plot:athcorr(U3, U1) 0.89 // atanh(0.71)
local re_cov 're_cov' plot:athcorr(U3, U2) 1.7 // atanh(0.94)
menl ... , ... initial('fe', 're_cov', Residual:lnsigma 0.5)
```

In the above, we also specified an initial value of 0.5 for the log of the error standard deviation. For parameters `/plot:lnsd(U1)` and `/plot:athcorr(U2, U1)`, instead of specifying the values, we specified the corresponding expression. This is allowed, as long as your expression is simple and does not contain spaces.

Instead of using parameter names, we can specify a list of values directly in the `initial()` option, in which case we must also specify `initial()`’s `copy` suboption.

```
menl ... , ... initial(19 55 8 1.6 1.4 0.34 0.9 0.89 1.7 0.5, copy)
```
Or we can provide these values as a matrix:

```
. matrix initvals = (19, 55, 8, 1.6, 1.4, 0.34, 0.9, 0.89, 1.7, 0.5)
. matrix list initvals
initvals[1,10]
c1   c2   c3   c4   c5   c6   c7   c8   c9   c10
r1   19   55   8   1.6  1.4  0.34  0.9  0.89  1.7  0.5
```

If we label the columns of the `initvals` matrix properly, we do not need to specify `copy`:

```
. local fullcolnames : colfullnames e(b)
. matrix colnames initvals = 'fullcolnames'
. matrix list initvals
initvals[1,10]
  phi1: phi2: phi3: /plot: /plot: /plot:
    _cons _cons _cons  lnsd(U1)  lnsd(U2)
  r1   19   55   8   1.6  1.4
  /plot: /plot: /plot: /plot: /Residual:  
    athcorr(U2, athcorr(U3, athcorr(U3, lnsd(U3) U1) U1) U2) lnsigma
  r1   .34   .9  .89  1.7  .5
```

Using a properly labeled initial-value matrix, we can also specify initial values for a subset of parameters. For example, we can specify initial values for fixed effects only as follows:

```
. matrix initvals = initvals[1,1..3]
. matrix list initvals
initvals[1,3]
  phi1: phi2: phi3: _cons _cons _cons
  r1   19   55   8
```

```
Stored results
```

`menl` stores the following in `e()`:

Scalars

- `e(N)` number of observations
- `e(N_nonmiss)` number of nonmissing `depvar` observations, if `tmsmissing` is specified
- `e(N_miss)` number of missing `depvar` observations, if `tmsmissing` is specified
- `e(N_ic)` number of nonmissing `depvar` observations to be used for BIC computation when `tmsmissing` is specified
- `e(k)` number of parameters
- `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(k_res)` number of within-group error parameters
- `e(k_eq)` number of equations
- `e(k_feq)` number of fixed-effects equations
- `e(kreq)` number of random-effects equations
- `e(k_reseq)` number of within-group error equations
- `e(df_m)` model degrees of freedom
- `e(df_c)` degrees of freedom for comparison test
- `e(ll)` linearization log (restricted) likelihood
<table>
<thead>
<tr>
<th>Macro</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>e(ll_c)</td>
<td>log likelihood, comparison model</td>
</tr>
<tr>
<td>e(chi2)</td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>e(chi2_c)</td>
<td>$\chi^2$ for comparison test</td>
</tr>
<tr>
<td>e(p)</td>
<td>$p$-value for model test</td>
</tr>
<tr>
<td>e(p_c)</td>
<td>$p$-value for comparison test</td>
</tr>
<tr>
<td>e(rank)</td>
<td>rank of $e(V)$</td>
</tr>
<tr>
<td>e(rc)</td>
<td>return code</td>
</tr>
<tr>
<td>e(converged)</td>
<td>1 if converged, 0 otherwise</td>
</tr>
</tbody>
</table>

**Macros**

- `e(cmd)`
  - menl

- `e(cmdline)`
  - command as typed

- `e(depvar)`
  - name of dependent variable

- `e(ivars)`
  - grouping variables

- `e(title)`
  - title in estimation output

- `e(varlist)`
  - variables used in the specified equation

- `e(key_Nic)`
  - nonmissing obs, if `tsmissing` is specified

- `e(tsmissing)`
  - `tsmissing`, if specified

- `e(eq depvar)`
  - user-specified equation

- `e(eq depvar)`
  - `tsinit()` specification for L.$\{depvar\}$

- `e(expressions)`
  - names of defined expressions, $expr_1, expr_2, \ldots, expr_k$

- `e(expr expr_i)`
  - defined expression $expr_i, i=1, \ldots, k$

- `e(tsinit_expr)`
  - `tsinit()` specification for L.$\{expr\}$

- `e(hierarchy)`
  - random-effects hierarchy structure, ($path$:$covtype$:REs) (...)

- `e(revars)`
  - names of random effects

- `e(restrctlab)`
  - within-group error covariance output label

- `e(timevar)`
  - within-group error covariance $t()$ variable, if specified

- `e(indexvar)`
  - within-group error covariance $index()$ variable, if specified

- `e(covbyvar)`
  - within-group error covariance by() variable, if specified

- `e(stratavar)`
  - within-group error variance $strata()$ variable, if specified

- `e(corrbyvar)`
  - within-group error correlation by() variable, if specified

- `e(rescovopt)`
  - within-group error covariance option, if $rescovariance()$ specified

- `e(resvaropt)`
  - within-group error variance option, if $resvariance()$ specified

- `e(rescorropt)`
  - within-group error correlation option, if $rescorrelation()$ specified

- `e(groupvar)`
  - lowest-level group() variable, if specified

- `e(chi2type)`
  - Wald: type of model $\chi^2$ test

- `e(vce)`
  - conventional

- `e(method)`
  - MLE or REML

- `e(opt)`
  - type of optimization, lbates

- `e(crittype)`
  - optimization criterion

- `e(properties)`
  - b V

- `e(estat_cmd)`
  - program used to implement `estat`

- `e(predict)`
  - program used to implement `predict`

- `e(marginsok)`
  - predictions allowed by `margins`

- `e(marginsnotok)`
  - predictions disallowed by `margins`

- `e(marginsdefault)`
  - default `predict()` specification for `margins`

- `e(asbalanced)`
  - factor variables `fvset` as `asbalanced`

- `e(asobserved)`
  - factor variables `fvset` as `asobserved`

**Matrices**

- `e(b)`
  - coefficient vector

- `e(Cns)`
  - factor-variable constraint matrix

- `e(V)`
  - variance–covariance matrix of the estimators

- `e(V_modelbased)`
  - model-based variance

- `e(b_sd)`
  - random-effects and within-group error estimates in the standard deviation metric

- `e(V_sd)`
  - VCE for parameters in the standard deviation metric

- `e(b_var)`
  - random-effects and within-group error estimates in the variance metric

- `e(V_var)`
  - VCE for parameters in the variance metric

- `e(cov_#)`
  - random-effects covariance structure at the hierarchical level $k-#+1$ in a $k$-level model

- `e(hierstats)`
  - group-size statistics for each hierarchy

**Functions**

- `e(sample)`
  - marks estimation sample
Methods and formulas

Methods and formulas are presented under the following headings:

Introduction
Variance-components parameters
Inference based on linearization
Initial values

Introduction

Recall (1), a two-level NLME model, from the Introduction,

\[ y_{ij} = \mu \left( x_{ij}', \beta, u_j \right) + \epsilon_{ij} \quad i = 1, \ldots, n_j; \ j = 1, \ldots, M \]

where \( M \) is the number of clusters and, for each cluster \( j \), \( n_j \) is the number of observations in that cluster; \( y_j = (y_{ij1}, y_{ij2}, \ldots, y_{ijn_j})' \) is the \( n_j \times 1 \) response vector; \( X_j = (x_{1j}, x_{2j}, \ldots, x_{njj})' \) is the \( n_j \times l \) matrix of covariates, including within-subject and between-subjects covariates; \( \beta \) is the \( p \times 1 \) vector of unknown parameters; \( u_j \) is the \( q \times 1 \) vector of random effects; and \( \epsilon_j = (\epsilon_{1j}, \epsilon_{2j}, \ldots, \epsilon_{njj})' \) is the \( n_j \times 1 \) vector of within-group or within-cluster errors. \( u_j \)'s follow a multivariate normal distribution with mean 0 and \( q \times q \) variance–covariance matrix \( \Sigma \), and \( \epsilon_j \)'s follow a multivariate normal distribution with mean 0 and \( n_j \times n_j \) variance–covariance matrix \( \sigma^2 \Lambda_j \); \( u_j \)'s are assumed to be independent of \( \epsilon_j \)'s. Depending on the form of \( \Lambda_j \), \( \sigma^2 \) is either a within-group error variance \( \sigma^2_e \) or a squared scale parameter \( \sigma^2 \). For example, when errors are i.i.d., that is, when \( \Lambda_j \) is the identity matrix, \( \sigma^2 = \sigma^2_e \) is the within-group error variance. When \( \Lambda_j \) corresponds to the heteroskedastic power structure, \( \sigma^2 \) is a multiplier or a scale parameter.

Positive-definite matrices \( \Sigma/\sigma^2 \) and \( \Lambda_j \) are expressed as functions of unconstrained parameter vectors \( \alpha_u \) and \( \alpha_w \), respectively, to recast a constrained optimization problem into an unconstrained one. Thus \( \alpha_u \) contains unconstrained random-effects covariance parameters and \( \alpha_w \) contains unconstrained within-group error covariance parameters. \( \Lambda_j \) may also depend on the random effects \( u_j \) and the fixed effects \( \beta \). For more details about \( \Sigma \) and \( \Lambda_j \) and about functional forms of parameter vectors \( \alpha_u \) and \( \alpha_w \) given different covariance structures, see Variance-components parameters.

Based on (1), the marginal, with respect to \( u_j \)'s, log likelihood for \( (\beta, \alpha, \sigma^2) \) is

\[
L(\beta, \alpha, \sigma^2) = \log \left\{ \prod_{j=1}^{M} \int f(y_j | X_j, u_j; \beta, \alpha_w, \sigma^2) f(u_j; \alpha_u) \, du_j \right\} \tag{25}
\]

where \( \alpha = (\alpha_u', \alpha_w')' \), \( f(y_j | X_j, u_j; \beta, \alpha_w, \sigma^2) \) is the conditional density of \( y_j \) given \( X_j \) and \( u_j \), and \( f(u_j; \alpha_u) \) is the density of \( u_j \).

In general, there are no closed-form expressions for (25) or the marginal moments of an NLME model. This is because the random effects \( u_j \) enter the model nonlinearly, making the \( q \)-dimensional integral in (25) analytically intractable in all but simpler cases. Several estimation techniques have been proposed for estimating parameters \( \beta, \alpha, \) and \( \sigma^2 \), including numerical integration of the integral in (25) by using an adaptive Gaussian quadrature and a linearization of the mean function in (1) by using a Taylor-series expansion.

\texttt{menl} implements the linearization method of Lindstrom and Bates (1990), with extensions from Pinheiro and Bates (1995), which is described in Inference based on linearization.
Variance-components parameters

For numerical stability, maximization of (25) is performed with respect to the unique elements of the matrix $G = \Sigma / \sigma^2$ expressed as logarithms of standard deviations for the diagonal elements and hyperbolic arctangents of the correlations for off-diagonal elements. Let $\alpha_u$ be the vector containing these elements. For example, if we assume that the elements of the random-effects vector $u_j$ are independent, then $\Sigma$ is diagonal and $\alpha_u$ will contain $q$ distinct parameters—$q$ logarithms of standard deviations. Table 1 lists the vectors of parameters $\alpha_u$ for all random-effects covariance structures supported by `menl` in the `covariance(vartype)` option.

<table>
<thead>
<tr>
<th>vartype</th>
<th>$\alpha_u$</th>
</tr>
</thead>
<tbody>
<tr>
<td>independent</td>
<td>$(g_1, g_2, \ldots, g_q)$</td>
</tr>
<tr>
<td>exchangeable</td>
<td>$(g_1, g_{12})$</td>
</tr>
<tr>
<td>identity</td>
<td>$g_1$</td>
</tr>
<tr>
<td>unstructured</td>
<td>$(g_1, g_2, \ldots, g_q, g_{12}, g_{13}, \ldots, g_{q-1})$</td>
</tr>
</tbody>
</table>

Notes: $g_u = \text{log} (\sqrt{|G|}_{uu})$, $g_{uv} = \text{atanh}(|G|_{uv})$.
unstructured has $q(q+1)/2$ parameters.

The within-group error covariance matrix is parameterized as follows,

$$\text{Var}(e_j | u_j) = \sigma^2 A_j (X_j, \beta, u_j, \alpha_w) = \sigma^2 S_j (\delta, v_j) C_j (\rho) S_j (\delta, v_j)$$

where $\alpha_w = (\delta^*, \rho^*)'$ and $\delta^*$ and $\rho^*$ are unconstrained versions of $\delta$ and $\rho$ defined in table 2 and table 3, respectively. For example, for a positive $\delta_1$, $\delta_1^* = \log (\delta_1)$. $S_j = S_j (\delta, v_j)$ is an $n_j \times n_j$ diagonal matrix with nonnegative diagonal elements $g (\delta, v_{1j}), g (\delta, v_{2j}), \ldots, g (\delta, v_{n_j})$ such that $\text{Var}(e_{ij}) = \sigma^2 [S_j]_{ii} = \sigma^2 g^2 (\delta, v_{ij})$, where $v_{ij}$'s are the values of a variance covariate or the values of a mean function $\mu (x_{ij}', \beta, u_j)$, in which case $A_j$ will depend on $X_j$, $\beta$, and $u_j$. $C_j = C_j (\rho)$ is a correlation matrix such that $\text{corr}(e_{ij}, e_{kj}) = [C_j]_{ik} = h (|t_{ij} - t_{kj}|, \rho)$, where $t_{ij}$ is a value of a time variable for time-dependent correlation structures such as AR, MA, and Toeplitz structures or an index variable for banded and unstructured correlation structures. A list of the supported $g(\cdot)$ and $h(\cdot)$ functions is given in table 2 and table 3, respectively.

Carroll and Ruppert (1988) introduced various variance functions $g (\delta, v_{ij})$ to model heteroskedasticity, which were further studied in the context of NLM models by Davidian and Giltinan (1995). Table 2 lists variance functions supported by the `resvariance(resvarfunc ...)` option.

<table>
<thead>
<tr>
<th>resvarfunc</th>
<th>$g (\delta, v_{ij})$</th>
<th>$\delta'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>identity</td>
<td>$1$</td>
<td>$-$</td>
</tr>
<tr>
<td>linear</td>
<td>$\sqrt{v_{ij}}$</td>
<td>$-$</td>
</tr>
<tr>
<td>power</td>
<td>$c +</td>
<td>v_{ij}</td>
</tr>
<tr>
<td>power, noconstant</td>
<td>$</td>
<td>v_{ij}</td>
</tr>
<tr>
<td>exponential</td>
<td>$\exp (\delta v_{ij})$</td>
<td>$\delta$</td>
</tr>
<tr>
<td>distinct</td>
<td>$\sum_{i=1}^{L} \delta_l I (v_{ij} = l)$</td>
<td>$(\delta_1 = 1, \delta_2, \ldots, \delta_L)$</td>
</tr>
</tbody>
</table>
In table 2, the variance function distinct models a distinct parameter $\delta_l$ for each level $l$ ($l = 1, 2, \ldots, L$) of the index variable $v_{ij} \in \{1, 2, \ldots, L\}$ such that for $v_{ij} = l$, $\text{Var}(\epsilon_{ij}) = \sigma^{2}_{l} = \sigma^{2}\delta^{2}_{l}$, where $\delta_{1} = 1$ for identifiability purposes and $\delta_{l} = \sigma_{l}/\sigma$. menl estimates and stores results as $\delta$'s but displays results as variances $\sigma^{2}_{l}$, $l = 1, \ldots, L$.

The variance function $g(\cdot)$ and thus the within-group error covariance may depend on $\beta$ and $u_j$ through $\mu(\cdot)$, when $v_{ij} = \mu_{ij} = \mu(x'_{ij}, u_j, \beta)$ in table 2. This is particularly appealing in PK applications, where there is often considerable intraindividual heterogeneity that is modeled, for example, as a power function of the mean.

The within-group error correlation structure is governed by the $h(\cdot)$ function. Table 3 lists correlation structures that are supported by the rescorrelation(rescorr ... ) option and also have a closed-form expression. In addition, the AR and MA correlation structures are defined below.

The ar $p$ structure assumes that the errors have an AR structure of order $p$. That is,

$$\epsilon_{ij} = \phi_{1}\epsilon_{i-1,j} + \cdots + \phi_{p}\epsilon_{i-p,j} + z_{ij}$$

where $z_{ij}$ are i.i.d. Gaussian with mean 0 and variance $\sigma^{2}_{z}$. menl reports estimates of $\phi_{1}, \ldots, \phi_{p}$ and the overall error variance $\sigma^{2}_{\epsilon}$, which can be derived from the above expression. This structure has a closed-form expression only for $p = 1$, in which case $\phi_{1} = \rho$ is the correlation between error terms.

The ma $q$ structure assumes that the errors are an MA process of order $q$. That is,

$$\epsilon_{ij} = Z_{i} + \theta_{1}Z_{i-1} + \cdots + \theta_{q}Z_{i-q}$$

where $Z_{i}$ are i.i.d. Gaussian with mean 0 and variance $\sigma^{2}_{Z}$. menl reports estimates of $\theta_{1}, \ldots, \theta_{q}$ and the overall error variance $\sigma^{2}_{\epsilon}$, which can be derived from the above expression.

Table 3. Within-group error correlation functions $h(\cdot)$

| rescorr        | $h(|t_{ij} - t_{ij}|, \rho)$ | Expression | $\rho$     |
|----------------|-------------------------------|------------|------------|
| identity       | $h(k)$                        | $I(k = 0)$ | –          |
| exchangeable   | $h(k, \rho)$                  | $\rho, \ k = 1, 2, \ldots$ | $\rho, \ |\rho| < 1$ |
| ar 1           | $h(k, \rho)$                  | $\rho^{k}, \ k = 0, 1, \ldots$ | $\rho, \ |\rho| < 1$ |
| ar $p, p > 1$  | $h(k, \phi)$                  | no closed form | $(\phi_{1}, \phi_{2}, \ldots, \phi_{p})$ |
| ctar1          | $h(s, \rho)$                  | $\rho^{s}, \ s \geq 0$ | $\rho, \ |\rho| < 1$ |
| ma $q$         | $h(k, \theta)$                | $\left\{ \begin{array}{ll} \sum_{j=0}^{q-|k|} \theta_{j}\theta_{j+|k|} & \text{if } k \leq q \\ \sum_{j=0}^{q} \theta_{j}^{2} & \text{if } k > q \end{array} \right.$ | $(\theta_{0} = 1, \theta_{1}, \ldots, \theta_{q})$ |
| toeplitz       | $h(k, \rho)$                  | $\rho_{k}I(k \leq q), \ k = 1, 2, \ldots, q$ | $(\rho_{1}, \rho_{2}, \ldots, \rho_{q})$ |
| banded         | $h(|i - l|, \rho)$            | $\rho_{il}I(|i - l| \leq q), \ 1 \leq i < l \leq n_{j}$ | $\{\rho_{il}: 0 < l - i \leq q\}$ |
| unstructured   | $h(|i - l|, \rho)$            | $\rho_{il}, \ 1 \leq i < l \leq n_{j}$ | $(\rho_{12}, \ldots, \rho_{(n_{j}-1)n_{j}})$ |

You can build many flexible within-group error covariance structures by combining different functions $g(\cdot)$ and $h(\cdot)$, that is, by combining the resvariance() and rescorrelation() options. For example, you can combine an AR(1) correlation structure with a heteroskedastic structure that is expressed as a power function of the mean by specifying rescorrelation(ar 1, t(timevar)) and resvariance(power _yhat).
Inference based on linearization

Let’s write (1), equivalently, in matrix form as

\[ y_j = \mu (X_j, \beta, \mathbf{u}_j) + \Lambda_j^{\frac{1}{2}} (X_j, \beta, \mathbf{u}_j, \alpha_w) e_j \]

Here \( \mu (X_j, \beta, \mathbf{u}_j) \) depends on \( \beta \) and \( \mathbf{u}_j \) through the function \( d(\cdot) \) in (2), and \( e_j \)'s \( \sim N (0, \sigma^2 I_{n_j}) \), where \( I_{n_j} \) is the identity matrix of dimension \( n_j \). In what follows, for brevity, we suppress the dependence of \( \mu \) and \( \Lambda_j \) on \( X_j \).

Following Lindstrom and Bates (1990), we will initially assume that \( \Lambda_j \) does not depend on \( X_j, \beta, \) and \( \mathbf{u}_j \) or, equivalently, on \( \phi_j \) but rather on \( j \) only through its dimension; that is, \( \Lambda_j = \Lambda_j (\alpha_w) \). Therefore, heteroskedastic structures that depend on the mean are not yet allowed in this context. Toward the end of this section, we will present a modified version of the algorithm that accounts for the dependence of \( \Lambda_j \) on \( \phi_j \).

Lindstrom and Bates discuss a natural extension of the methods for the LME models to NLME models. For a known \( \alpha \) (and thus known \( \Sigma \) and \( \Lambda_j \)) and \( \sigma^2 \), the estimates of \( \beta \) and \( \mathbf{u}_j \) jointly minimize

\[
\sum_{j=1}^{M} \left[ \log|\Sigma(\alpha_u)| + \mathbf{u}_j' \{\Sigma(\alpha_u)\}^{-1} \mathbf{u}_j + \log \sigma^2 \Lambda_j (\alpha_w) \right] \\
+ \sigma^{-2} \{y_j - \mu(\beta, \mathbf{u}_j)\}' \Lambda_j^{-1} (\alpha_w) \{y_j - \mu(\beta, \mathbf{u}_j)\}
\]

which is twice the negative log likelihood for \( \beta \) when \( \mathbf{u}_j \) is fixed or twice the negative log of the posterior density of \( \mathbf{u}_j \) when \( \beta \) is fixed. Consequently, one strategy for estimating \( \beta \) and (predicting) \( \mathbf{u}_j \) is to minimize the above objective function with respect to \( \beta \) and \( \mathbf{u}_j \) given suitable estimates of \( \alpha \) and \( \sigma^2 \). Estimation of \( \alpha \) and \( \sigma^2 \) can be accomplished by using MLE with respect to the marginal density of \( y_j \), in which \( \mathbf{u}_j \)'s are integrated out. But because no closed-form expression for this density is available, we approximate the conditional distribution of \( y_j \) given \( \mathbf{u}_j \) by a multivariate normal distribution with an expectation that is linear in \( \mathbf{u}_j \) and \( \beta \). This is illustrated in step 2 of the algorithm below.

Lindstrom and Bates (1990) propose the following two-step estimation method or alternating algorithm.

**Step 1 (PNLS step).** Given current estimates \( \hat{\alpha} \) (and thus \( \hat{\alpha}_u \) and \( \hat{\alpha}_w \)) of \( \alpha \) and \( \hat{\sigma}^2 \) of \( \sigma^2 \), minimize with respect to \( \beta \) and \( \mathbf{u}_j \)

\[
\sum_{j=1}^{M} \left[ \log|\Sigma(\hat{\alpha}_u)| + \mathbf{u}_j' \{\Sigma(\hat{\alpha}_u)\}^{-1} \mathbf{u}_j + \log \hat{\sigma}^2 \Lambda_j (\hat{\alpha}_w) \right] \\
+ \hat{\sigma}^{-2} \{y_j - \mu(\beta, \mathbf{u}_j)\}' \Lambda_j^{-1} (\hat{\alpha}_w) \{y_j - \mu(\beta, \mathbf{u}_j)\}
\]  

(26)

Define \( \Delta \) such that \( \sigma^2 \Sigma^{-1} = \Delta' \Delta \). Note that \( \Delta = \Delta (\alpha_u) \), but for notational convenience, this dependency is suppressed throughout the rest of this section. Equation (26) is equivalent to minimizing the penalized least-squares objective function
PNLS step: \[
\sum_{j=1}^{M} \left[ \left\{ \Lambda_j' (\alpha_w) \right\}^{-1/2} \left\{ y_j - \mu (\beta, u_j) \right\}^2 + \| \Delta u_j \|^2 \right] \]

with respect to \( \beta \) and \( u_j \) while holding the current estimates of \( \alpha \) (and, consequently, of \( \Delta \) and of \( \Lambda_j \)) fixed. \text{pnlsopts}(iterate(#) \) iterations are performed at this step, unless the convergence criterion (CC) is met. The CC for PNLS optimization is controlled by \text{pnlsopts}(ntolerance(#) \) and one of \text{pnlsopts}(ltolerance(#) \) or \text{pnlsopts}(tolerance(#)); see \text{menlmaxopts} for details.

Denote the resulting estimates as \( \hat{u}_j \) and \( \hat{\beta} \).

In the absence of random effects in the model (see example 19), the previous formulas no longer include the random effects and related components. In particular, \( u_j \) and \( \Delta \) are set to 0, and \( \alpha = \alpha_w \). In this case, the PNLS step reduces to what we call a GNLS estimation step. Furthermore, if no within-group error covariance structure is specified, that is, when all observations are assumed i.i.d., \( \Lambda_j (\alpha_w) \) is set to the identity matrix \( I \), and the PNLS step reduces to the classical NLS estimation.

**Step 2 (LME step).** Perform a first-order Taylor-series expansion of the model mean function around the current estimates of \( \beta \) and of the conditional modes of the random effects \( u_j \), yielding

\[
y_j = \mu \left( \hat{\beta}, \hat{u}_j \right) + \hat{X}_j (\beta - \hat{\beta}) + \hat{Z}_j (u_j - \hat{u}_j) + \Lambda_j^{\frac{1}{2}} (\alpha_w) e_j \tag{27}
\]

where

\[
\hat{X}_j = \frac{\partial \mu (\beta, u_j)}{\partial \beta} \bigg|_{\beta=\hat{\beta}, u_j=\hat{u}_j}, \quad \hat{Z}_j = \frac{\partial \mu (\beta, u_j)}{\partial u_j} \bigg|_{\beta=\hat{\beta}, u_j=\hat{u}_j}
\]

Model (27) is essentially an LME model, and we use notations \( \hat{X}_j \) and \( \hat{Z}_j \) for the derivatives to emphasize this. That is, \( \hat{X}_j \) and \( \hat{Z}_j \) represent the corresponding fixed-effects and random-effects design matrices of an LME model.

Thus the approximate conditional distribution of \( y_j \) is

\[
y_j | u_j \sim N \left\{ \mu \left( \beta, \hat{u}_j \right) + \hat{X}_j (\beta - \hat{\beta}) + \hat{Z}_j (u_j - \hat{u}_j), \sigma^2 \Lambda_j \right\}
\]

Because the expectation is now linear in random effects \( u_j \), the approximate conditional distribution of \( y_j \), along with distribution of \( u_j \), allows us to approximate the marginal distribution of \( y_j \) as

\[
y_j \sim N \left\{ \mu \left( \hat{\beta}, \hat{u}_j \right) + \hat{X}_j (\beta - \hat{\beta}) - \hat{Z}_j \hat{u}_j, \sigma^2 V_j(\alpha) \right\} \tag{28}
\]

where \( V_j(\alpha) = \hat{Z}_j \Delta^{-1} (\Delta^{-1})' \hat{Z}_j' + \Lambda_j (\alpha_w) \).

Let \( \hat{w}_j = y_j - \mu \left( \hat{\beta}, \hat{u}_j \right) + \hat{X}_j \hat{\beta} + \hat{Z}_j \hat{u}_j \). Estimation of \( \alpha \) and \( \sigma^2 \) can now be accomplished by maximizing the log likelihood corresponding to the approximate marginal distribution in (28),
\[
l_{\text{LB}}(\alpha, \beta, \sigma^2) = \frac{-n}{2} \log (2\pi\sigma^2) - \frac{1}{2} \sum_{j=1}^{M} \left\{ \log |V_j(\alpha)| \right. \\
+ \sigma^{-2} \left( \hat{w}_j - \hat{X}_j \beta \right)' V_j^{-1}(\alpha) \left( \hat{w}_j - \hat{X}_j \beta \right) \right\} \tag{29}
\]

where \( n = \sum_{j=1}^{M} n_j \).

Alternatively, when the reml option is specified, we take an REML approach and maximize

\[
l_{\text{LB},R}(\alpha, \sigma^2) = l_{\text{LB}}(\alpha, \hat{\beta}(\alpha), \sigma^2) - \frac{1}{2} \sum_{j=1}^{M} \log \left| \sigma^{-2} \hat{X}_j' V_j^{-1}(\alpha) \hat{X}_j \right| \tag{30}
\]

The LME step (step 2) of the alternating algorithm consists of optimizing an LME log likelihood, in which the response vector is given by \( \hat{w}_j \) and the fixed- and random-effects design matrices are given by \( \hat{X}_j \) and \( \hat{Z}_j \), respectively. \( \text{lmeopts}(\text{iterate}(\#)) \) iterations are performed at this step, unless the CC is met. The CC for LME optimization is controlled by \( \text{lmeopts}(\text{ntolerance}(\#)) \) and one of \( \text{lmeopts}(\text{ltolerance}(\#)) \) or \( \text{lmeopts}(\text{tolerance}(\#)) \); see \text{lmenmaxopts} for details.

The LME step produces estimates \( \hat{\alpha} \) and \( \hat{\sigma}^2 \). (The estimates \( \hat{\beta} \) can also be obtained at this step, but it is generally more computationally efficient to compute them at the PNLS step.) These estimates will now be used in step 1, the PNLS step.

In the absence of random effects in the model (see example 19), \( u_j, \hat{u}_j, \Delta, \) and \( \hat{Z}_j \) are all set to 0, and \( \alpha = \alpha_w \). In this case, the LME step is referred to as the ML step or, if the reml option is specified, the REML step in the \text{menl} output. Furthermore, if all observations are assumed i.i.d., then step 2 of the alternating algorithm is not needed, and only step 1 (NLS) is performed.

**Stopping rules.** One PNLS step and one LME step correspond to one iteration of the alternating algorithm. The log likelihood reported by \text{menl} at each iteration is the log likelihood (29) or, if the reml option is specified, (30) from the last iteration of the LME step. \text{menl} refers to this log likelihood as “linearization log likelihood” because it corresponds to the log likelihood of the LME model, which was the result of the linearization of the NLME model. The algorithm stops when the linearization likelihoods from successive iterations satisfy \text{ltolerance}(\#), when the parameter estimates from successive iterations satisfy \text{tolerance}(\#), or if the model does not converge, when the maximum number of iterations in \text{iterate()} is reached; see \text{menlmaxopts} for details about maximization options. Because the alternating algorithm does not provide a joint Hessian matrix for all parameters, there is no check for the tolerance of the scaled gradient; thus the convergence cannot be established in its strict sense. The convergence is declared based on the stopping rules described above.

When \( \Lambda_j = \Lambda_j(\beta, u_j, \alpha_w) \) depends on \( u_j \) and \( \beta \), which is the case, for example, with \text{resvariance(power_yhat)} and \text{resvariance(exponential_yhat)}, an intermediate step between the PNLS and the LME step is performed to replace the fixed effects and random effects in \( \Lambda_j \), or more precisely in the variance function \( g(\cdot) \), by their current estimates from the PNLS step. After that, \( \Lambda_j(\alpha_w; \hat{\beta}, \hat{u}_j) = \Lambda_j(\alpha_w) \) depends only on \( \alpha_w \) because both \( u_j \) and \( \beta \) are held fixed at their current estimates throughout the LME step.

Efficient methods for computing (29) or (30) are given in chapters 2 and 5 of Pinheiro and Bates (2000). Namely, to simplify the optimization problem, one can express the optimal values of \( \beta \) and \( \sigma^2 \) as functions of \( \alpha \) (and thus of \( \Delta \) and \( \alpha_w \)) and work with the profiled log likelihood of \( \alpha \).
For the PNLS step, the objective function to be minimized is the penalized sum of squares

\[
\sum_{j=1}^{M} \left[ ||(\Lambda_j')^{-1/2} \{y_j - \mu(\beta, u_j)\} ||^2 + ||\Delta u_j||^2 \right]
\]

By adding “pseudo”-observations to the data, the PNLS problem can be reexpressed as a standard nonlinear least-squares problem. Thus step 1 of the alternating algorithm is sometimes called the “pseudodata step”. Define pseudoobservations \( \tilde{y}_j \) as follows:

\[
\tilde{y}_j = \begin{bmatrix} (\Lambda_j')^{-1/2} y_j \\ 0 \end{bmatrix} \quad \tilde{\mu}(\beta, u_j) = \begin{bmatrix} (\Lambda_j')^{-1/2} \mu(\beta, u_j) \\ \Delta u_j \end{bmatrix}
\]

Then, the PNLS step can be rewritten as

\[
\sum_{j=1}^{M} ||\tilde{y}_j - \tilde{\mu}(\beta, u_j)||^2
\]

Hence, for values of \( \alpha \) and \( \sigma^2 \) fixed at the current estimates, the estimation of \( \beta \) and \( u_j \) in the PNLS step can be regarded as a standard nonlinear least-squares problem. A popular iterative estimation technique for standard nonlinear least-squares is the Gauss–Newton method (see Pinheiro and Bates [2000, chap. 7] for more details).

After the completion of the alternating algorithm, an extra LME iteration is performed, with fixed effects profiled-out of the likelihood, to reparameterize \( [\alpha, \log(\sigma)] \) to their natural metric and to compute their standard errors with the delta method. This step is labeled Computing standard errors: in the output of menl. If you are interested only in standard errors for fixed effects, you can skip this step by specifying the nostderr option, in which case standard errors for the random-effects and within-group error parameters will not be computed and will be shown as missing in the output table. The standard errors for the fixed effects are obtained from the PNLS step, and the standard errors for random-effects parameters are obtained from the LME step.

Inference on the parameters of the NLME model is based on the approximating LME model with log likelihood and restricted log likelihood functions defined in (29) and (30). Therefore, all the inferential machinery available within the context of LME models can be used. For example, under the LME approximation, the distribution of the (restricted) MLE \( \hat{\beta} \) of the fixed effects is

\[
\hat{\beta} \sim N \left( \beta, \sigma^2 \left( \sum_{j=1}^{M} \hat{X}_j' \hat{V}_j^{-1}(\alpha) \hat{X}_j \right)^{-1} \right)
\]

and for random-effects and within-group error parameters is

\[
\begin{bmatrix} \hat{\alpha} \\ \log\hat{\sigma} \end{bmatrix} \sim N \left( \begin{bmatrix} \alpha \\ \log\sigma \end{bmatrix}, I^{-1}(\alpha, \sigma) \right)
\]

where

\[
I(\alpha, \sigma) = - \begin{bmatrix} \frac{\partial^2 l_{LB_p}}{\partial \alpha \partial \alpha'} & \frac{\partial^2 l_{LB_p}}{\partial \log\sigma \partial \alpha'} \\ \frac{\partial^2 l_{LB_p}}{\partial \alpha \partial \log\sigma} & \frac{\partial^2 l_{LB_p}}{\partial^2 \log\sigma} \end{bmatrix}
\]
and \( t_{LB_p} = t_{LB_p}(\alpha, \sigma) \) is the approximated log likelihood from the LME step with fixed effects profiled out. Because inference is based on the LME approximation of the original NLME model, asymptotic results are technically “approximately asymptotic” and are thus less accurate than the asymptotic inferential results for LME models as described in [ME] mixed.

**Initial values**

The PNLS step requires starting values for \( \beta \) and \( u_j \). These are obtained from the EM algorithm; see, for example, Bates and Pinheiro (1998) for details. You can control optimization within the EM algorithm by specifying the `emtolerance()` and `emiterate()` options. You can also supply your own initial values; see Examples of specifying initial values. NLME models are often sensitive to initial values, so it is good practice to try different sets of initial values to verify that your results are robust to them.

**References**


Also see

[ME] `menl postestimation` — Postestimation tools for `menl`

[ME] `meglm` — Multilevel mixed-effects generalized linear model

[ME] `mixed` — Multilevel mixed-effects linear regression

[ME] `me` — Introduction to multilevel mixed-effects models

[R] `nl` — Nonlinear least-squares estimation

[U] 20 *Estimation and postestimation commands*