Title

---

xtmixed — Multilevel mixed-effects linear regression

Syntax

xtmixed *depvar fe_equation* $\left[\text{II } re\_equation \right]$ $\left[\text{II } re\_equation \dots \right]$ $\left[, options \right]$

where the syntax of *fe_equation* is

$\left[indepvars \right]$ $\left[if \right]$ $\left[in \right]$ $\left[, fe\_options \right]$

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

for random coefficients and intercepts

levelvar: $\left[varlist \right]$ $\left[, re\_options \right]$

for random effects among the values of a factor variable

levelvar: R.*varname* $\left[, re\_options \right]$

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

fe_options	description
Model	
<u>no</u>constant	suppress constant term from the fixed-effects equation

re_options	description
Model	
<u>cov</u>ariance(*vartype*)	variance–covariance structure of the random effects
<u>no</u>constant	suppress constant term from the random-effects equation
<u>col</u>linear	keep collinear variables

vartype	description
<u>ind</u>ependent	one unique variance parameter per random effect, all covariances zero; the default unless a factor variable is specified
<u>exc</u>hangeable	equal variances for random effects, and one common pairwise covariance
<u>id</u>entity	equal variances for random effects, all covariances zero
<u>un</u>structured	all variances and covariances distinctly estimated

options	description
Model	
<u>resid</u>uals(*rspec*)	structure of residual errors
Estimation	
reml	fit model via maximum restricted likelihood; the default
mle	fit model via maximum likelihood
Reporting	
level(*#*)	set confidence level; default is level(95)
variance	show random-effects parameter estimates as variances and covariances
noretable	suppress random-effects table
nofetable	suppress fixed-effects table
estmetric	show parameter estimates in the estimation metric
noheader	suppress output header
nogroup	suppress table summarizing groups
nostderr	do not estimate standard errors of random-effects parameters
nolrtest	do not perform LR test comparing to linear regression
display_options	control spacing and display of omitted variables and base and empty cells
EM options	
emiterate(*#*)	number of EM iterations; default is 20
emtolerance(*#*)	EM convergence tolerance; default is 1e-10
emonly	fit model exclusively using EM
emlog	show EM iteration log
emdots	show EM iterations as dots
Maximization	
maximize_options	control the maximization process; seldom used
matsqrt	parameterize variance components using matrix square roots; the default
matlog	parameterize variance components using matrix logarithms
[†]<u>coefl</u>egend	display coefficients' legend instead of coefficient table

[†]coeflegend does not appear in the dialog box.

indepvars may contain factor variables; see [U] **11.4.3 Factor variables**.
depvar, *indepvars*, and *varlist* may contain time-series operators; see [U] **11.4.4 Time-series varlists**.
bootstrap, by, jackknife, mi estimate, rolling, and statsby are allowed; see [U] **11.1.10 Prefix commands**.
See [U] **20 Estimation and postestimation commands** for more capabilities of estimation commands.

Menu

Statistics > Longitudinal/panel data > Multilevel mixed-effects models > Mixed-effects linear regression

Description

xtmixed fits linear mixed models. Mixed models are characterized as containing both *fixed effects* and *random effects*. The fixed effects are analogous to standard regression coefficients and are estimated directly. The random effects are not directly estimated but are summarized according to their estimated variances and covariances. Although random effects are not directly estimated, you can form best linear unbiased predictions (BLUPs) of them (and standard errors) by using predict after xtmixed; see [XT] **xtmixed postestimation**. Random effects may take the form of either random intercepts or random coefficients, and the grouping structure of the data may consist of multiple levels of nested groups. The overall error distribution of the linear mixed model is assumed to be Gaussian, but heteroskedasticity and correlations within lowest-level groups also may be modeled.

Options

  Model

noconstant suppresses the constant (intercept) term and may be specified for the fixed-effects equation and for any or all of the random-effects equations.

covariance(*vartype*), where *vartype* is

$$\texttt{independent} \mid \texttt{exchangeable} \mid \texttt{identity} \mid \texttt{unstructured}$$

specifies the structure of the covariance matrix for the random effects and may be specified for each random-effects equation. An independent covariance structure allows for a distinct variance for each random effect within a random-effects equation and assumes that all covariances are zero. exchangeable structure 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 zero. unstructured allows for all variances and covariances to be distinct. If an equation consists of p random-effects terms, the unstructured covariance matrix will have $p(p+1)/2$ unique parameters.

covariance(independent) is the default, except when the random-effects equation is a factor-variable specification R.*varname*, in which case covariance(identity) is the default, and only covariance(identity) and covariance(exchangeable) are allowed.

collinear specifies that xtmixed not omit collinear variables from the random-effects equation. Usually there is no reason to leave collinear variables in place, and in fact doing so usually causes the estimation to fail because of the matrix singularity caused by the collinearity. However, with certain models (for example, a random-effects model with a full set of contrasts), the variables may be collinear, yet the model is fully identified because of restrictions on the random-effects covariance structure. In such cases, using the collinear option allows the estimation to take place with the random-effects equation intact.

residuals(*rspec*), where *rspec* is

$$restype \; \big[\, , \; residual\_options \, \big]$$

specifies the structure of the residual errors within the lowest-level groups of the linear mixed model. For example, if you are modeling random effects for classes nested within schools, then residuals() refers to the residual variance–covariance structure of the observations within classes, the lowest-level groups.

restype is

$$\underline{\text{ind}}\text{ependent} \mid \underline{\text{ex}}\text{changeable} \mid \text{ar } \# \mid \text{ma } \# \mid \underline{\text{uns}}\text{tructured} \mid$$

$$\underline{\text{ban}}\text{ded } \# \mid \underline{\text{t}}\text{oeplitz } \# \mid \underline{\text{ex}}\text{ponential}$$

By default, *restype* is independent, which means that all residuals are i.i.d. Gaussian with one common variance. When combined with by(*varname*), independence is still assumed, but you estimate a distinct variance for each level of *varname*. Unlike with the structures described below, *varname* does not need to be constant within groups.

restype exchangeable estimates two parameters, one common within-group variance and one common pairwise covariance. When combined with by(*varname*), these two parameters are distinctly estimated for each level of *varname*. Because you are modeling a within-group covariance, *varname* must be constant within lowest-level groups.

restype ar # assumes that within-group errors have an autoregressive (AR) structure of order #; ar 1 is the default. The t(*varname*) option is required, where *varname* is an integer-valued time variable used to order the observations within groups and to determine the lags between successive observations. Any nonconsecutive time values will be treated as gaps. For this structure, # + 1 parameters are estimated (# AR coefficients and one overall error variance). *restype* ar may be combined with by(*varname*), but *varname* must be constant within groups.

restype ma # assumes that within-group errors have a moving average (MA) structure of order #; ma 1 is the default. The t(*varname*) option is required, where *varname* is an integer-valued time variable used to order the observations within groups and to determine the lags between successive observations. Any nonconsecutive time values will be treated as gaps. For this structure, # + 1 parameters are estimated (# MA coefficients and one overall error variance). *restype* ma may be combined with by(*varname*), but *varname* must be constant within groups.

restype unstructured is the most general structure; it estimates distinct variances for each within-group error and distinct covariances for each within-group error pair. The t(*varname*) option is required, where *varname* is a nonnegative-integer–valued variable that identifies the observations within each group. The groups may be unbalanced in that not all levels of t() need to be observed within every group, but you may not have repeated t() values within any particular group. When you have p levels of t(), then $p(p+1)/2$ parameters are estimated. *restype* unstructured may be combined with by(*varname*), but *varname* must be constant within groups.

restype banded # 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 zero. The t(*varname*) option is required, where *varname* is a nonnegative-integer–valued variable that identifies the observations within each group. # is an integer between zero and $p-1$, where p is the number of levels of t(). By default, # is $p-1$; that is, all elements of the covariance matrix are estimated. When # is zero, only the diagonal elements of the covariance matrix are estimated. *restype* banded may be combined with by(*varname*), but *varname* must be constant within groups.

restype toeplitz # assumes that within-group errors have Toeplitz structure of order #, for which correlations are constant with respect to time lags less than or equal to # and are zero for lags greater than #. The t(*varname*) option is required, where *varname* is an integer-valued time variable used to order the observations within groups and to determine the lags between successive observations. # is an integer between one and the maximum observed lag (the default). Any nonconsecutive time values will be treated as

gaps. For this structure, $\# + 1$ parameters are estimated (# correlations and one overall error variance). *restype* toeplitz may be combined with by(*varname*), but *varname* must be constant within groups.

restype exponential is a generalization of the autoregressive (AR) covariance model that allows for unequally spaced and noninteger time values. The t(*varname*) option is required, where *varname* is real-valued. For the exponential covariance model, the correlation between two errors is the parameter ρ, raised to a power equal to the absolute value of the difference between the t() values for those errors. For this structure, two parameters are estimated (the correlation parameter ρ and one overall error variance). *restype* exponential may be combined with by(*varname*), but *varname* must be constant within groups.

residual_options are by(*varname*) and t(*varname*).

by(*varname*) is for use within the residuals() option and specifies that a set of distinct residual-error parameters be estimated for each level of *varname*. In other words, you use by() to model heteroskedasticity.

t(*varname*) is for use within the residuals() option to specify a time variable for the ar, ma, toeplitz, and exponential structures, or to ID the observations when *restype* is unstructured or banded.

Estimation

reml and mle specify the statistical method for fitting the model.

reml, the default, specifies that the model be fit using restricted maximum likelihood (REML), also known as residual maximum likelihood.

mle specifies that the model be fit using maximum likelihood (ML).

Reporting

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

variance displays the random-effects and residual-error parameter estimates as variances and co-variances. The default is to display them 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 the estimation metric. Fixed-effects estimates are unchanged from those normally displayed, but random-effects parameter estimates are displayed as log-standard deviations and hyperbolic arctangents of correlations, with equation names that organize them by model level. Residual-variance parameter estimates are also displayed in their original estimation metric.

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 xtmixed 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. nostderr is available only when residuals are modeled as independent with constant variance.

nolrtest prevents xtmixed from fitting a reference linear regression model and using this model to calculate a likelihood-ratio test comparing the mixed model to ordinary regression. This option may also be specified on replay to suppress this test from the output.

display_options: <u>noomitted</u>, vsquish, <u>noempty</u>cells, <u>base</u>levels, <u>allbase</u>levels; see [R] **estimation options**.

⌐ EM options ¬

These options control the EM (expectation-maximization) iterations that take place before estimation switches to a gradient-based method. When residuals are modeled as independent with constant variance, EM will either converge to the solution or bring parameter estimates close to the solution. For other residual structures, EM is used to obtain starting values.

emiterate(#) specifies the number of EM iterations to perform. The default is emiterate(20).

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 a gradient-based method, unless emonly is specified, in which case maximization stops.

emonly specifies that the likelihood be maximized exclusively using EM. The advantage of specifying emonly is that EM iterations are typically much faster than those for gradient-based methods. The disadvantages are that EM iterations can be slow to converge (if at all) and that EM provides no facility for estimating standard errors for the random-effects parameters. emonly is available only when residuals are modeled as independent with constant variance.

emlog specifies that the EM iteration log be shown. The EM iteration log is, by default, not displayed unless the emonly option is specified.

emdots specifies that the EM iterations be shown as dots. This option can be convenient because the EM algorithm may require many iterations to converge.

⌐ Maximization ¬

maximize_options: <u>difficult</u>, <u>tech</u>nique(*algorithm_spec*), <u>iter</u>ate(#), [no]<u>log</u>, <u>trace</u>, <u>grad</u>ient, <u>showstep</u>, <u>hess</u>ian, <u>showtol</u>erance, <u>tol</u>erance(#), <u>ltol</u>erance(#), <u>nrtol</u>erance(#), <u>nonrtol</u>erance; see [R] **maximize**. Those that require special mention for xtmixed are listed below.

For the technique() option, the default is technique(nr). The bhhh algorithm may not be specified.

matsqrt (the default), during optimization, parameterizes variance components by using the matrix square roots of the variance–covariance matrices formed by these components at each model level.

matlog, during optimization, parameterizes variance components by using the matrix logarithms of the variance–covariance matrices formed by these components at each model level.

Both the matsqrt and matlog parameterizations ensure that variance–covariance matrices are positive semidefinite. For most problems, the matrix square root is more stable near the boundary of the parameter space. However, if convergence is problematic, one option may be to try the alternate matlog parameterization. When convergence is not an issue, both parameterizations yield equivalent results.

The following option is available with xtmixed but is not shown in the dialog box:

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

Remarks

Remarks are presented under the following headings:

> *Introduction*
> *One-level models*
> *Covariance structures*
> *Likelihood versus restricted likelihood*
> *Two-level models*
> *Blocked-diagonal covariance structures*
> *Heteroskedastic random effects*
> *Heteroskedastic residual errors*
> *Other residual-error structures*
> *Random-effects factor notation and crossed-effects models*
> *Diagnosing convergence problems*
> *Distribution theory for likelihood-ratio tests*

Introduction

Linear mixed models are models containing both fixed effects and random effects. They are a generalization of linear regression allowing for the inclusion of random deviations (effects) other than those associated with the overall error term. In matrix notation,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon} \tag{1}$$

where $\mathbf{y}$ is the $n \times 1$ vector of responses, $\mathbf{X}$ is an $n \times p$ design/covariate matrix for the fixed effects $\boldsymbol{\beta}$, and $\mathbf{Z}$ is the $n \times q$ design/covariate matrix for the random effects $\mathbf{u}$. The $n \times 1$ vector of errors, $\boldsymbol{\epsilon}$, is assumed to be multivariate normal with mean zero and variance matrix $\sigma_\epsilon^2 \mathbf{R}$.

The fixed portion of (1), $\mathbf{X}\boldsymbol{\beta}$, is analogous to the linear predictor from a standard OLS regression model with $\boldsymbol{\beta}$ being the regression coefficients to be estimated. For the random portion of (1), $\mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$, we assume that $\mathbf{u}$ has variance–covariance matrix $\mathbf{G}$ and that $\mathbf{u}$ is orthogonal to $\boldsymbol{\epsilon}$ so that

$$\text{Var}\begin{bmatrix} \mathbf{u} \\ \boldsymbol{\epsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \sigma_\epsilon^2 \mathbf{R} \end{bmatrix}$$

The random effects $\mathbf{u}$ are not directly estimated (although they may be predicted), but instead are characterized by the elements of $\mathbf{G}$, known as *variance components*, that are estimated along with the overall residual variance σ_ϵ^2 and the residual-variance parameters that are contained within $\mathbf{R}$.

The general forms of the design matrices $\mathbf{X}$ and $\mathbf{Z}$ allow estimation for a broad class of linear models: blocked designs, split-plot designs, growth curves, multilevel or hierarchical designs, etc. They also allow a flexible method of modeling within-panel correlation. Subjects within the same panel can be correlated as a result of a shared random intercept, or through a shared random slope on (say) age, or both. The general specification of $\mathbf{G}$ also provides additional flexibility—the random intercept and random slope could themselves be modeled as independent, or correlated, or independent with equal variances, and so forth. The general structure of $\mathbf{R}$ also allows for residual errors to be heteroskedastic and correlated, and allows flexibility in exactly how these characteristics can be modeled.

Comprehensive treatments of mixed models are provided by, among others, Searle, Casella, and McCulloch (1992); McCulloch, Searle, and Neuhaus (2008); Verbeke and Molenberghs (2000); Raudenbush and Bryk (2002); Demidenko (2004); and Pinheiro and Bates (2000). In particular, chapter 2 of Searle, Casella, and McCulloch (1992) provides an excellent history.

The key to fitting mixed models lies in estimating the variance components, and for that there exist many methods. Most of the early literature in mixed models dealt with estimating variance components in ANOVA models. For simple models with balanced data, estimating variance components amounts to solving a system of equations obtained by setting expected mean-squares expressions equal to their observed counterparts. Much of the work in extending the "ANOVA method" to unbalanced data for general ANOVA designs is due to Henderson (1953).

The ANOVA method, however, has its shortcomings. Among these is a lack of uniqueness in that alternative, unbiased estimates of variance components could be derived using other quadratic forms of the data in place of observed mean squares (Searle, Casella, and McCulloch 1992, 38–39). As a result, ANOVA methods gave way to more modern methods, such as minimum norm quadratic unbiased estimation (MINQUE) and minimum variance quadratic unbiased estimation (MIVQUE); see Rao (1973) for MINQUE and LaMotte (1973) for MIVQUE. Both methods involve finding optimal quadratic forms of the data that are unbiased for the variance components.

The most popular methods, however, are maximum likelihood (ML) and restricted maximum-likelihood (REML), and these are the two methods that are supported by `xtmixed`. The ML estimates are based on the usual application of likelihood theory, given the distributional assumptions of the model. The basic idea behind REML (Thompson Jr. 1962) is that you can form a set of linear contrasts of the response that do not depend on the fixed effects, β, but instead depend only on the variance components to be estimated. You then apply ML methods by using the distribution of the linear contrasts to form the likelihood.

Returning to (1): in panel-data situations, it is convenient not to consider all n observations at once but instead to organize the mixed model as a series of M independent panels

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\epsilon}_i \tag{2}$$

for $i = 1, \ldots, M$, with panel i consisting of n_i observations. The response, $\mathbf{y}_i$, comprises the rows of $\mathbf{y}$ corresponding to the ith panel, with $\mathbf{X}_i$ and $\boldsymbol{\epsilon}_i$ defined analogously. The random effects, $\mathbf{u}_i$, can now be thought of as M realizations of a $q \times 1$ vector that is normally distributed with mean $\mathbf{0}$ and $q \times q$ variance matrix $\boldsymbol{\Sigma}$. The matrix $\mathbf{Z}_i$ is the $n_i \times q$ design matrix for the ith panel random effects. Relating this to (1), note that

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Z}_M \end{bmatrix}; \quad \mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_M \end{bmatrix}; \quad \mathbf{G} = \mathbf{I}_M \otimes \boldsymbol{\Sigma}; \quad \mathbf{R} = \mathbf{I}_M \otimes \boldsymbol{\Lambda} \tag{3}$$

The mixed-model formulation (2) is from Laird and Ware (1982) and offers two key advantages. First, it makes specifications of random-effects terms easier. If the panels are schools, you can simply specify a random effect "at the school level", as opposed to thinking of what a school-level random effect would mean when all the data are considered as a whole (if it helps, think Kronecker products). Second, representing a mixed-model with (2) generalizes easily to more than one level of random variation. For example, if classes are nested within schools, then (2) can be generalized to allow random effects at both the school and the class-within-school levels. This we demonstrate later.

Finally, using formulation (2) and its multilevel extensions requires one important convention of terminology. Model (2) is what we call a *one-level* model, with extensions to two, three, or any number of levels. In our hypothetical two-level model with classes nested within schools, the schools are considered the first level and classes, the second level of the model. This is consistent with terminology used elsewhere, for example, Pinheiro and Bates (2000), but differs from that of the literature on hierarchical models, for example, Skrondal and Rabe-Hesketh (2004). In that literature,

our schools and classes model would be considered a three-level model, with the students forming the first level, classes the second, and schools the third. Not only is there one more level (students) but the order is reversed.

In the sections that follow, we assume that residuals are independent with constant variance; that is, in (3) we treat Λ equal to the identity matrix and limit ourselves to estimating one overall residual variance, σ_ϵ^2. Beginning in *Heteroskedastic residual errors*, we relax this assumption.

One-level models

We begin with a simple application of (2).

▷ Example 1

Consider a longitudinal dataset used by both Ruppert, Wand, and Carroll (2003) and Diggle et al. (2002), consisting of weight measurements of 48 pigs on 9 successive weeks. Pigs are identified by variable id. Below is a plot of the growth curves for the first 10 pigs.

```
. use http://www.stata-press.com/data/r11/pig
(Longitudinal analysis of pig weights)

. twoway connected weight week if id<=10, connect(L)
```
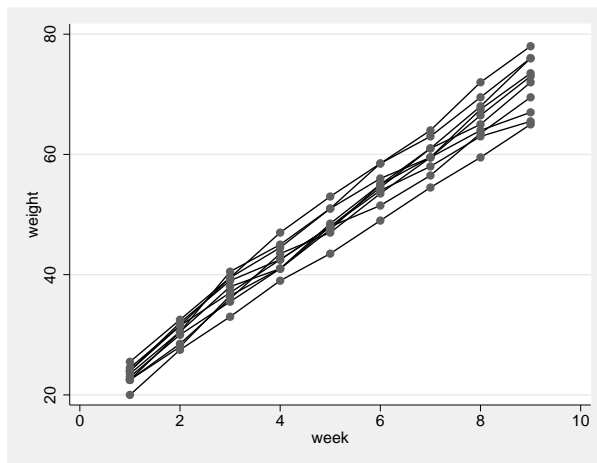

It seems clear that each pig experiences a linear trend in growth and that overall weight measurements vary from pig to pig. Because we are not really interested in these particular 48 pigs per se, we instead treat them as a random sample from a larger population and model the between-pig variability as a random effect or, in the terminology of (2), as a random-intercept term at the pig level. We thus wish to fit the model

$$\texttt{weight}_{ij} = \beta_0 + \beta_1 \texttt{week}_{ij} + u_i + \epsilon_{ij} \tag{4}$$

for $i = 1, \ldots, 48$ pigs and $j = 1, \ldots, 9$ weeks. The fixed portion of the model, $\beta_0 + \beta_1 \texttt{week}_{ij}$, simply states that we want one overall regression line representing the population average. The random effect, u_i, serves to shift this regression line up or down according to each pig. Because the random effects occur at the pig level (id), we fit the model by typing

```
. xtmixed weight week || id:

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:    log restricted-likelihood = -1016.8984
Iteration 1:    log restricted-likelihood = -1016.8984

Computing standard errors:

Mixed-effects REML regression                   Number of obs      =        432
Group variable: id                              Number of groups   =         48

                                                Obs per group: min =          9
                                                               avg =        9.0
                                                               max =          9

                                                Wald chi2(1)       =   25271.50
Log restricted-likelihood = -1016.8984          Prob > chi2        =     0.0000
```

weight	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
week	6.209896	.0390633	158.97	0.000	6.133333	6.286458
_cons	19.35561	.6031391	32.09	0.000	18.17348	20.53774

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
id: Identity				
sd(_cons)	3.891253	.41432	3.158334	4.794253
sd(Residual)	2.096356	.0757444	1.953034	2.250195

```
LR test vs. linear regression: chibar2(01) =    473.15 Prob >= chibar2 = 0.0000
```

At this point, a guided tour of the model specification and output is in order:

1. By typing "`weight week`", we specified the response, `weight`, and the fixed portion of the model in the same way that we would if we were using `regress` or any other estimation command. Our fixed effects are a coefficient on `week` and a constant term.

2. When we added "`|| id:`", we specified random effects at the level identified by group variable `id`, that is, the pig level. Because we wanted only a random intercept, that is all we had to type.

3. The estimation log consists of three parts:

   a. A set of expectation-maximization (EM) iterations used to refine starting values. By default, the iterations themselves are not displayed, but you can display them with the `emlog` option.

   b. A set of "gradient-based" iterations. By default, these are Newton–Raphson iterations, but other methods are available by specifying the appropriate `maximize` options; see [R] **maximize**.

   c. The message "Computing standard errors:". This is just to inform you that `xtmixed` has finished its iterative maximization and is now reparameterizing from a matrix-based parameterization (see *Methods and formulas*) to the natural metric of variance components and their estimated standard errors.

4. The output title, "Mixed-effects REML regression", informs us that our model was fit using REML, the default. For ML estimates, use the `mle` option.

   Because this model is a simple random-intercept model, specifying the `mle` option would be equivalent to using `xtreg` with its `mle` option.

5. The first estimation table reports the fixed effects. We estimate $\beta_0 = 19.36$ and $\beta_1 = 6.21$.

6. The second estimation table shows the estimated variance components. The first section of the table is labeled "id: Identity", meaning that these are random effects at the id (pig) level and that their variance–covariance matrix is a multiple of the identity matrix; that is, $\Sigma = \sigma_u^2 \mathbf{I}$. Because we have only one random effect at this level, xtmixed knew that Identity is the only possible covariance structure. In any case, σ_u is estimated as 3.89 with standard error 0.414.

If you prefer variance estimates, $\hat{\sigma}_u^2$, to standard deviation estimates, $\hat{\sigma}_u$, specify the variance option either at estimation or on replay.

7. The row labeled "sd(Residual)" displays the estimated standard deviation of the overall error term; that is, $\hat{\sigma}_\epsilon = 2.10$.

8. Finally, a likelihood-ratio test comparing the model with ordinary linear regression, model (4) without u_i, is provided and is highly significant for these data.

We now store our estimates for later use:

```
. estimates store randint
```

◁

▷ Example 2

Extending (4) to allow for a random slope on week yields the model

$$\text{weight}_{ij} = \beta_0 + \beta_1 \text{week}_{ij} + u_{0i} + u_{1i}\text{week}_{ij} + \epsilon_{ij} \tag{5}$$

fit using xtmixed:

```
. xtmixed weight week || id: week
Performing EM optimization:
Performing gradient-based optimization:
Iteration 0:   log restricted-likelihood = -870.51473
Iteration 1:   log restricted-likelihood = -870.51473
Computing standard errors:
Mixed-effects REML regression                   Number of obs      =        432
Group variable: id                              Number of groups   =         48
                                                Obs per group: min =          9
                                                               avg =        9.0
                                                               max =          9
                                                Wald chi2(1)       =    4592.10
Log restricted-likelihood = -870.51473          Prob > chi2        =     0.0000
```

weight	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
week	6.209896	.0916387	67.77	0.000	6.030287	6.389504
_cons	19.35561	.4021144	48.13	0.000	18.56748	20.14374

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
id: Independent				
sd(week)	.6135475	.0673971	.4947037	.7609413
sd(_cons)	2.630134	.3028832	2.09872	3.296107
sd(Residual)	1.26443	.0487971	1.172317	1.363781

```
LR test vs. linear regression:       chi2(2) =    765.92   Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
```

```
. estimates store randslope
```

Because we did not specify a covariance structure for the random effects $(u_{0i}, u_{1i})'$, xtmixed used the default Independent structure; that is,

$$\boldsymbol{\Sigma} = \text{Var}\begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix} = \begin{bmatrix} \sigma_{u0}^2 & 0 \\ 0 & \sigma_{u1}^2 \end{bmatrix} \tag{6}$$

with $\widehat{\sigma}_{u0} = 2.63$ and $\widehat{\sigma}_{u1} = 0.61$. Our point estimates of the fixed effects are essentially identical to those from model (4), but note that this does not hold generally. Given the 95% confidence interval for $\widehat{\sigma}_{u1}$, it would seem that the random slope is significant, and we can use lrtest and our two saved estimation results to verify this fact:

```
. lrtest randslope randint
Likelihood-ratio test                              LR chi2(1)  =     292.77
(Assumption: randint nested in randslope)          Prob > chi2 =     0.0000
Note: The reported degrees of freedom assumes the null hypothesis is not on
      the boundary of the parameter space.  If this is not true, then the
      reported test is conservative.
Note: LR tests based on REML are valid only when the fixed-effects
      specification is identical for both models.
```

The near-zero significance level favors the model that allows for a random pig-specific regression line over the model that allows only for a pig-specific shift.

◁

❏ Technical note

At the bottom of the previous xtmixed output, there is a note stating that the likelihood ratio (LR) test comparing our model to standard linear regression is conservative. Also, our lrtest output warns us that our test comparing the random-slope model with the random-intercept model may be conservative if the null hypothesis is on the boundary. For the former, the null hypothesis is $H_0 : \sigma_{u0}^2 = \sigma_{u1}^2 = 0$. For the latter, the null hypothesis is $H_0 : \sigma_{u1}^2 = 0$. Because variances are constrained to be positive, both null hypotheses are on the boundaries of their respective parameter spaces. xtmixed is capable of detecting this automatically because it compares with linear regression. lrtest, on the other hand, can be used to compare a wide variety of nested mixed models, making automatic detection of boundary conditions impractical. With lrtest, the onus is on the user to verify testing on the boundary.

By "conservative", we mean that when boundary conditions exist, the reported significance level is an upper bound on the actual significance; see *Distribution theory for likelihood-ratio tests* later in this entry for further details.

❏

❏ Technical note

LR tests with REML require identical fixed-effects specifications for both models. As stated in Ruppert, Wand, and Carroll (2003), "The reason for this is that restricted likelihood is the likelihood of the residuals after fitting the fixed effects and so is not appropriate when there is more than one fixed effects model under consideration." To compare models with different fixed-effects specifications, use a Wald test or fit the models by ML (the mle option).

In our example, the fixed-effects specifications for both models are identical ($\beta_0 + \beta_1$week), so our REML-based test is valid.

◻

Covariance structures

In example 2, we fit a model with the default Independent covariance given in (6). Within any random-effects level specification, we can override this default by specifying an alternative covariance structure via the covariance() option.

▷ Example 3

We generalize (6) to allow u_{0i} and u_{1i} to be correlated; that is,

$$\boldsymbol{\Sigma} = \text{Var}\begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix} = \begin{bmatrix} \sigma_{u0}^2 & \sigma_{01} \\ \sigma_{01} & \sigma_{u1}^2 \end{bmatrix}$$

```
. xtmixed weight week || id: week, covariance(unstructured) variance
  (output omitted)
```

```
Mixed-effects REML regression                    Number of obs      =        432
Group variable: id                               Number of groups   =         48

                                                 Obs per group: min =          9
                                                                avg =        9.0
                                                                max =          9

                                                 Wald chi2(1)       =    4552.31
Log restricted-likelihood = -870.43562           Prob > chi2        =     0.0000
```

weight	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
week	6.209896	.0920383	67.47	0.000	6.029504	6.390288
_cons	19.35561	.4038678	47.93	0.000	18.56405	20.14718

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
id: Unstructured				
var(week)	.3799962	.0839024	.2465106	.5857642
var(_cons)	6.986472	1.616359	4.439436	10.99482
cov(week,_cons)	-.1033635	.2627315	-.6183078	.4115808
var(Residual)	1.596829	.123198	1.372735	1.857505

```
LR test vs. linear regression:       chi2(3) =    766.07   Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
```

But we do not find the correlation to be at all significant.

```
. lrtest . randslope
Likelihood-ratio test                            LR chi2(1)  =       0.16
(Assumption: randslope nested in .)              Prob > chi2 =     0.6908
Note: LR tests based on REML are valid only when the fixed-effects
      specification is identical for both models.
```

In addition to specifying an alternate covariance structure, we specified the `variance` option to display variance components in the variance–covariance metric, rather than the default, which displays them as standard deviations and correlations.

◁

Instead, we could have also specified `covariance(identity)`, restricting u_{0i} and u_{1i} to not only be independent but also to have common variance, or we could have specified `covariance(exchangeable)`, which imposes a common variance but allows for a nonzero correlation.

Likelihood versus restricted likelihood

Thus far, all our examples have used restricted maximum likelihood (REML) to estimate variance components. We could have just as easily asked for ML estimates. Refitting the model in example 2 by ML, we get

```
. xtmixed weight week || id: week, ml
 (output omitted )
Mixed-effects ML regression                     Number of obs      =        432
Group variable: id                              Number of groups   =         48

                                                Obs per group: min =          9
                                                               avg =        9.0
                                                               max =          9

                                                Wald chi2(1)       =    4689.51
Log likelihood = -869.03825                     Prob > chi2        =     0.0000
```

weight	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
week	6.209896	.0906819	68.48	0.000	6.032163	6.387629
_cons	19.35561	.3979159	48.64	0.000	18.57571	20.13551

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
id: Independent				
sd(week)	.6066851	.0660294	.4901417	.7509396
sd(_cons)	2.599301	.2969073	2.077913	3.251515
sd(Residual)	1.264441	.0487958	1.17233	1.363789

```
LR test vs. linear regression:       chi2(2) =   764.42   Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
```

Although ML estimators are based on the usual likelihood theory, the idea behind REML is to transform the response into a set of linear contrasts whose distribution is free of the fixed effects β. The restricted likelihood is then formed by considering the distribution of the linear contrasts. Not only does this make the maximization problem free of β, it also incorporates the degrees of freedom used to estimate β into the estimation of the variance components. This follows because, by necessity, the rank of the linear contrasts must be less than the number of observations.

As a simple example, consider a constant-only regression where $y_i \sim N(\mu, \sigma^2)$ for $i = 1, \dots, n$. The ML estimate of σ^2 can be derived theoretically as the n-divided sample variance. The REML estimate can be derived by considering the first $n-1$ error contrasts, $y_i - \overline{y}$, whose joint distribution is free of μ. Applying maximum likelihood to this distribution results in an estimate of σ^2, that is, the $(n-1)$ divided sample variance, which is unbiased for σ^2.

The unbiasedness property of REML extends to all mixed models when the data are balanced, and thus REML would seem the clear choice in balanced-data problems, although in large samples the difference between ML and REML is negligible. One disadvantage of REML is that LR tests based on REML are inappropriate for comparing models with different fixed-effects specifications. ML is appropriate for such LR tests and has the advantage of being easy to explain and being the method of choice for other estimators. The question of which method to use thus remains a matter of personal taste.

Examining the ML output, we find that the estimates of the variance components are slightly smaller than the REML estimates. This is typical, because ML estimates, which do not incorporate the degrees of freedom used to estimate the fixed effects, tend to be biased downward.

Two-level models

The panel-data representation of the mixed model given in (2) can be extended to two nested levels. Formally,

$$\mathbf{y}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}^{(1)}\mathbf{u}_i^{(1)} + \mathbf{Z}_{ij}^{(2)}\mathbf{u}_{ij}^{(2)} + \boldsymbol{\epsilon}_{ij} \tag{7}$$

for $i = 1, \ldots, M$ first-level groups and $j = 1, \ldots, M_i$ second-level groups that are nested within group i. Group i, j consists of n_{ij} observations, so $\mathbf{y}_{ij}$, $\mathbf{X}_{ij}$, and $\boldsymbol{\epsilon}_{ij}$ each have row dimension n_{ij}. $\mathbf{Z}_{ij}^{(1)}$ is the $n_{ij} \times q_1$ design matrix for the first-level random effects $\mathbf{u}_i^{(1)}$, and $\mathbf{Z}_{ij}^{(2)}$ is the $n_{ij} \times q_2$ design matrix for the second-level random effects $\mathbf{u}_{ij}^{(2)}$. Furthermore, assume that

$$\mathbf{u}_i^{(1)} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_1); \quad \mathbf{u}_{ij}^{(2)} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_2); \quad \boldsymbol{\epsilon}_{ij} \sim N(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I})$$

and that $\mathbf{u}_i^{(1)}$, $\mathbf{u}_{ij}^{(2)}$, and $\boldsymbol{\epsilon}_{ij}$ are independent.

Fitting a two-level model requires you to specify two random-effects "equations", one for each level. The variable list for the first equation represents $\mathbf{Z}_{ij}^{(1)}$, and the second equation, $\mathbf{Z}_{ij}^{(2)}$.

▷ Example 4

Baltagi, Song, and Jung (2001) estimate a Cobb–Douglas production function examining the productivity of public capital in each state's private output. Originally provided by Munnell (1990), the data were recorded over 1970–1986 for 48 states grouped into nine regions.

```
. use http://www.stata-press.com/data/r11/productivity
(Public Capital Productivity)
. describe

Contains data from http://www.stata-press.com/data/r11/productivity.dta
  obs:           816                          Public Capital Productivity
  vars:           11                          29 Mar 2009 10:57
  size:       32,640 (99.7% of memory free)   (_dta has notes)
```

variable name	storage type	display format	value label	variable label
state	byte	%9.0g		states 1-48
region	byte	%9.0g		regions 1-9
year	int	%9.0g		years 1970-1986
public	float	%9.0g		public capital stock
hwy	float	%9.0g		log(highway component of public)
water	float	%9.0g		log(water component of public)
other	float	%9.0g		log(bldg/other component of public)
private	float	%9.0g		log(private capital stock)
gsp	float	%9.0g		log(gross state product)
emp	float	%9.0g		log(non-agriculture payrolls)
unemp	float	%9.0g		state unemployment rate

```
Sorted by:
```

Because the states are nested within regions, we fit a two-level mixed model with random intercepts at both the region and the state-within-region levels. That is, we use (7) with both $\mathbf{Z}_{ij}^{(1)}$ and $\mathbf{Z}_{ij}^{(2)}$ set to the $n_{ij} \times 1$ column of ones, and $\boldsymbol{\Sigma}_1 = \sigma_1^2$ and $\boldsymbol{\Sigma}_2 = \sigma_2^2$ are both scalars.

```
. xtmixed gsp private emp hwy water other unemp || region: || state:
  (output omitted)
Mixed-effects REML regression                   Number of obs      =       816
```

Group Variable	No. of Groups	Observations per Group		
		Minimum	Average	Maximum
region	9	51	90.7	136
state	48	17	17.0	17

```
                                                Wald chi2(6)       =  18382.38
Log restricted-likelihood =  1404.7101          Prob > chi2        =    0.0000
```

gsp	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
private	.2660308	.0215471	12.35	0.000	.2237993	.3082623
emp	.7555059	.0264556	28.56	0.000	.7036539	.807358
hwy	.0718857	.0233478	3.08	0.002	.0261249	.1176465
water	.0761552	.0139952	5.44	0.000	.0487251	.1035853
other	-.1005396	.0170173	-5.91	0.000	-.1338929	-.0671863
unemp	-.0058815	.0009093	-6.47	0.000	-.0076636	-.0040994
_cons	2.126995	.1574865	13.51	0.000	1.818327	2.435663

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
region: Identity				
sd(_cons)	.0435474	.0186293	.0188289	.1007164
state: Identity				
sd(_cons)	.0802738	.0095512	.0635762	.1013567
sd(Residual)	.0368008	.0009442	.034996	.0386986

LR test vs. linear regression: chi2(2) = 1162.40   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

Some items of note:

1. Our model now has two random-effects equations, separated by ||. The first is a random intercept (constant only) at the region level, and the second is a random intercept at the state level. The order in which these are specified (from left to right) is significant—xtmixed assumes that state is nested within region.

2. The information on groups is now displayed as a table, with one row for each model level. You can suppress this table with the nogroup or the noheader option, which will suppress the rest of the header, as well.

3. The variance-component estimates are now organized and labeled according to level.

After adjusting for the nested-level error structure, we find that the highway and water components of public capital had significant positive effects on private output, whereas the other public buildings component had a negative effect.

◁

❑ Technical note

In the previous example, the states are coded 1–48 and are nested within nine regions. xtmixed treated the states as nested within regions, regardless of whether the codes for each state are unique between regions. That is, even if codes for states were duplicated between regions, xtmixed would have enforced the nesting and produced the same results.

The group information at the top of xtmixed output and that produced by the postestimation command estat group (see [XT] **xtmixed postestimation**) take the nesting into account. The statistics are thus not necessarily what you would get if you instead tabulated each group variable individually.

❑

Model (7) extends in a straightforward manner to more than two nested levels of random effects, as does the specification of such models in xtmixed.

Blocked-diagonal covariance structures

Covariance matrices of random effects within an equation can be modeled either as a multiple of the identity matrix, diagonal (that is, Independent), exchangeable, or as general symmetric (Unstructured). These may also be combined to produce more complex block-diagonal covariance structures, effectively placing constraints on the variance components.

▷ Example 5

Returning to our productivity data, we now add random coefficients on `hwy` and `unemp` at the `region` level. This only slightly changes the estimates of the fixed effects, so we focus our attention on the variance components:

```
. xtmixed gsp private emp hwy water other unemp || region: hwy unemp || state:,
> nolog nogroup nofetable
Mixed-effects REML regression                   Number of obs     =        816
                                                Wald chi2(6)      =   16803.58
Log restricted-likelihood =  1423.3455          Prob > chi2       =     0.0000
```

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
region: Independent				
sd(hwy)	.0052747	.0108958	.000092	.302341
sd(unemp)	.0052895	.001545	.002984	.0093766
sd(_cons)	.0595987	.0759006	.0049114	.7232107
state: Identity				
sd(_cons)	.0807544	.0098873	.0635255	.1026559
sd(Residual)	.0353932	.000914	.0336464	.0372307

```
LR test vs. linear regression:       chi2(4) =  1199.67   Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
. estimates store prodrc
```

This model is the same as that fit in example 4, except that $\mathbf{Z}_{ij}^{(1)}$ is now the $n_{ij} \times 3$ matrix with columns determined by the values of `hwy`, `unemp`, and an intercept term (one), in that order, and (because we used the default Independent structure) Σ_1 is

$$
\begin{array}{ccc}
\text{hwy} & \text{unemp} & \_\text{cons}
\end{array}
$$
$$
\Sigma_1 = \begin{pmatrix} \sigma_a^2 & 0 & 0 \\ 0 & \sigma_b^2 & 0 \\ 0 & 0 & \sigma_c^2 \end{pmatrix}
$$

The random-effects specification at the state level remains unchanged; that is, Σ_2 is still treated as the scalar variance of the random intercepts at the state level.

An LR test comparing this model with that from example 4 favors the inclusion of the two random coefficients, a fact we leave to the interested reader to verify.

Examining the estimated variance components reveals that the variances of the random coefficients on `hwy` and `unemp` could be treated as equal. That is,

$$
\begin{array}{ccc}
\text{hwy} & \text{unemp} & \_\text{cons}
\end{array}
$$
$$
\Sigma_1 = \begin{pmatrix} \sigma_a^2 & 0 & 0 \\ 0 & \sigma_a^2 & 0 \\ 0 & 0 & \sigma_c^2 \end{pmatrix}
$$

looks plausible. We can impose this equality constraint by treating Σ_1 as block diagonal: the first block is a 2×2 multiple of the identity matrix, that is, $\sigma_a^2 \mathbf{I}_2$; the second is a scalar, equivalently, a 1×1 multiple of the identity.

We construct block-diagonal covariances by repeating level specifications:

```
. xtmixed gsp private emp hwy water other unemp || region: hwy unemp,
> cov(identity) || region: || state:, nolog nogroup nofetable
Mixed-effects REML regression                   Number of obs    =      816
                                                Wald chi2(6)     = 16803.40
Log restricted-likelihood = 1423.3455           Prob > chi2      =   0.0000
```

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
region: Identity				
sd(hwy unemp)	.0052896	.0015446	.0029844	.0093752
region: Identity				
sd(_cons)	.0595037	.0318237	.0208595	.1697396
state: Identity				
sd(_cons)	.0807521	.0097453	.0637425	.1023007
sd(Residual)	.0353932	.0009139	.0336465	.0372306

```
LR test vs. linear regression:      chi2(3) =  1199.67   Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
```

We specified two equations for the region level: the first for the random coefficients on hwy and unemp with covariance set to Identity and the second for the random intercept _cons, whose covariance defaults to Identity because it is of dimension one. xtmixed labeled the estimate of σ_a as "sd(hwy unemp)" to designate that it is common to the random coefficients on both hwy and unemp.

An LR test shows that the constrained model fits equally well.

```
. lrtest . prodrc
Likelihood-ratio test                           LR chi2(1) =        0.00
(Assumption: . nested in prodrc)                Prob > chi2 =      0.9989
Note: The reported degrees of freedom assumes the null hypothesis is not on
      the boundary of the parameter space.  If this is not true, then the
      reported test is conservative.
Note: LR tests based on REML are valid only when the fixed-effects
      specification is identical for both models.
```
◁

Because the null hypothesis for this test is one of equality ($H_0 : \sigma_a^2 = \sigma_b^2$), it is not on the boundary of the parameter space. As such, we can take the reported significance as precise rather than a conservative estimate.

You can repeat level specifications as often as you like, defining successive blocks of a block-diagonal covariance matrix. However, repeated-level equations must be listed consecutively; otherwise, xtmixed will give an error.

❑ Technical note

In the previous estimation output, there was no constant term included in the first region equation, even though we did not use the noconstant option. When you specify repeated-level equations, xtmixed knows not to put constant terms in each equation because such a model would be unidentified. By default, it places the constant in the last repeated-level equation, but you can use noconstant creatively to override this.

❑

Heteroskedastic random effects

Blocked-diagonal covariance structures and repeated-level specifications of random effects can also be used to model heteroskedasticity among random effects at a given level.

▷ Example 6

Following Rabe-Hesketh and Skrondal (2008, sec. 5.10), we analyze data from Asian children in a British community who were weighed up to four times, roughly between the ages of 6 weeks and 27 months. The dataset is a random sample of data previously analyzed by Goldstein (1986) and Prosser, Rasbash, and Goldstein (1991).

```
. use http://www.stata-press.com/data/r11/childweight
(Weight data on Asian children)
. describe

Contains data from http://www.stata-press.com/data/r11/childweight.dta
  obs:           198                          Weight data on Asian children
  vars:            5                          23 May 2009 15:12
  size:         3,960 (99.9% of memory free)  (_dta has notes)

              storage   display    value
variable name   type    format     label      variable label

id              int     %8.0g                 child identifier
age             float   %8.0g                 age in years
weight          float   %8.0g                 weight in Kg
brthwt          int     %8.0g                 Birth weight in g
girl            float   %9.0g       bg        gender

Sorted by:  id  age
. graph twoway (line weight age, connect(ascending)), by(girl)
> xtitle(Age in years) ytitle(Weight in kg)
```
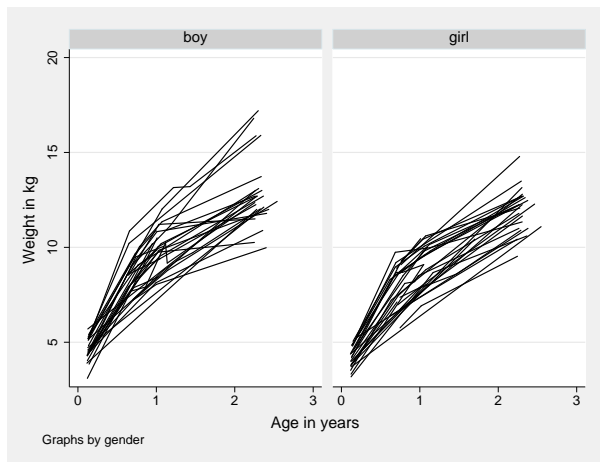

Ignoring gender effects for the moment, we begin with the following model for the jth measurement on the ith child:

$$\text{weight}_{ij} = \beta_0 + \beta_1\text{age}_{ij} + \beta_2\text{age}_{ij}^2 + u_{i0} + u_{i1}\text{age}_{ij} + \epsilon_{ij}$$

The above models overall mean growth as quadratic in age and allows for two child-specific random effects: a random intercept, u_{i0}, that represents each child's vertical shift from the overall mean (β_0), and a random age slope, u_{i1}, that represents each child's deviation in linear growth rate from the overall mean linear growth rate (β_1). For reasons of simplicity, we do not consider child-specific changes in the quadratic component of growth.

```
. xtmixed weight age c.age#c.age || id: age, mle nolog
Mixed-effects ML regression                     Number of obs      =        198
Group variable: id                              Number of groups   =         68

                                                Obs per group: min =          1
                                                               avg =        2.9
                                                               max =          5

                                                Wald chi2(2)       =    1863.46
Log likelihood = -258.51915                     Prob > chi2        =     0.0000
```

weight	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
age	7.693701	.2381076	32.31	0.000	7.227019	8.160384
c.age#c.age	-1.654542	.0874987	-18.91	0.000	-1.826037	-1.483048
_cons	3.497628	.1416914	24.68	0.000	3.219918	3.775338

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
id: Independent				
sd(age)	.5465535	.075708	.4166057	.7170347
sd(_cons)	.7087917	.0996506	.5380794	.9336647
sd(Residual)	.5561382	.0426951	.4784488	.6464426

```
LR test vs. linear regression:       chi2(2) =    114.70   Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
```

◁

Because there is no reason to believe that the random effects are uncorrelated, it is always a good idea to first fit a model with the covariance(unstructured) option. We do not include the output for such a model because for these data the correlation between random effects is not significant, but we did check this before reverting to xtmixed's default Independent structure.

Next we introduce gender effects into the fixed portion of the model by including a main gender effect and gender/age interaction for overall mean growth:

```
. xtmixed weight i.girl i.girl#c.age c.age#c.age || id: age, mle nolog
```

```
Mixed-effects ML regression                    Number of obs     =        198
Group variable: id                             Number of groups  =         68

                                               Obs per group: min =          1
                                                              avg =        2.9
                                                              max =          5

                                               Wald chi2(4)      =    1942.30
Log likelihood =   -253.182                    Prob > chi2       =     0.0000
```

weight	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
1.girl	-.5104676	.2145529	-2.38	0.017	-.9309835	-.0899516
girl#c.age						
0	7.806765	.2524583	30.92	0.000	7.311956	8.301574
1	7.577296	.2531318	29.93	0.000	7.081166	8.073425
c.age#c.age	-1.654323	.0871752	-18.98	0.000	-1.825183	-1.483463
_cons	3.754275	.1726404	21.75	0.000	3.415906	4.092644

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
id: Independent				
sd(age)	.5265782	.0730408	.4012307	.6910851
sd(_cons)	.6385054	.0969921	.4740922	.8599364
sd(Residual)	.5596163	.0426042	.4820449	.6496707

```
LR test vs. linear regression:        chi2(2) =    104.39   Prob > chi2 = 0.0000
```
Note: LR test is conservative and provided only for reference.
```
. estimates store homoskedastic
```

The main gender effect is significant at the 5% level, but the gender/age interaction is not:

```
. test 0.girl#c.age = 1.girl#c.age
 ( 1)  [weight]0b.girl#c.age - [weight]1.girl#c.age = 0
           chi2( 1) =    1.66
         Prob > chi2 =    0.1978
```

On average, boys are heavier than girls but their average linear growth rates are not significantly different.

In the above model, we introduced a gender effect on average growth, but we still assumed that the variability in child-specific deviations from this average was the same for boys and girls. To check this assumption, we introduce gender into the random component of the model. Because support for factor-variable notation is limited in specifications of random effects (see *Random-effects factor notation and crossed-effects models* below), we need to generate the interactions ourselves.

```
. gen boy = !girl

. gen boyXage = boy*age

. gen girlXage = girl*age

. xtmixed weight i.girl i.girl#c.age c.age#c.age || id: boy boyXage, noconstant
> || id: girl girlXage, noconstant mle nolog nofetable
```

```
Mixed-effects ML regression                     Number of obs     =         198
Group variable: id                              Number of groups  =          68

                                                Obs per group: min =           1
                                                               avg =         2.9
                                                               max =           5

                                                Wald chi2(4)      =     2358.11
Log likelihood = -248.94752                     Prob > chi2       =      0.0000
```

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
id: Independent				
sd(boy)	.5622358	.138546	.3468691	.9113211
sd(boyXage)	.6880757	.1144225	.4966919	.9532031
id: Independent				
sd(girl)	.7614904	.1286769	.5467994	1.060476
sd(girlXage)	.257805	.1073047	.1140251	.582884
sd(Residual)	.5548717	.0418872	.4785591	.6433534

```
LR test vs. linear regression:        chi2(4) =    112.86   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

. estimates store heteroskedastic
```

In the above, we suppress displaying the fixed portion of the model (the `nofetable` option) because it does not differ much from that of the previous model.

Our previous model had the random effects specification

```
|| id: age
```

which we have replaced with the dual repeated-level specification

```
|| id: boy boyXage, noconstant || id: girl girlXage, noconstant
```

The former models a random intercept and random slope on age, and does so treating all children as a random sample from one population. The latter also specifies a random intercept and random slope on age, but allows for the variability of the random intercepts and slopes to differ between boys and girls. In other words, it allows for heteroskedasticity in random effects due to gender. We use the `noconstant` option so that we can separate the overall random intercept (automatically provided by the former syntax) into one specific to boys and one specific to girls.

There seems to be a large gender effect in the variability of linear growth rates. We can compare both models with a likelihood-ratio test, recalling that we saved the previous estimation results under the name `homoskedastic`:

```
. lrtest homoskedastic heteroskedastic

Likelihood-ratio test                           LR chi2(2)  =        8.47
(Assumption: homoskedastic nested in heteroskedas~c)  Prob > chi2 =      0.0145

Note: The reported degrees of freedom assumes the null hypothesis is not on
      the boundary of the parameter space.  If this is not true, then the
      reported test is conservative.
```

Because the null hypothesis here is one of equality of variances and not that variances are zero, the above does not test on the boundary, and thus we can treat the significance level as precise and not conservative. Either way, the results favor the new model with heteroskedastic random effects.

Heteroskedastic residual errors

Up to this point, we have assumed that the residual errors—the ϵ's in the stated models—have been i.i.d. Gaussian with variance σ_ϵ^2. This is demonstrated in xtmixed output in the random-effects table, where up until now we have estimated a single residual-error standard deviation or variance, labeled as sd(Residual) or var(Residual), respectively.

To relax the assumptions of homoskedasticity or independence of residual errors, use the resid-uals() option.

▷ Example 7

West, Welch, and Galecki (2007, chap. 7) analyze data studying the effect of ceramic dental veneer placement on gingival (gum) health. Data on 55 teeth located in the maxillary arches of 12 patients were considered.

```
. use http://www.stata-press.com/data/r11/veneer, clear
(Dental veneer data)
. describe
Contains data from http://www.stata-press.com/data/r11/veneer.dta
  obs:          110                          Dental veneer data
 vars:            7                          24 May 2009 06:20
 size:        1,540 (99.9% of memory free)   (_dta has notes)
```

variable name	storage type	display format	value label	variable label
patient	byte	%8.0g		Patient ID
tooth	byte	%8.0g		Tooth number with patient
gcf	byte	%8.0g		Gingival crevicular fluid (GCF)
age	byte	%8.0g		Patient age
base_gcf	byte	%8.0g		Baseline GCF
cda	float	%9.0g		Average contour difference after veneer placement
followup	byte	%9.0g	t	Follow-up time: 3 or 6 months

```
Sorted by:
```

Veneers were placed to match the original contour of the tooth as closely as possible, and researchers were interested in how contour differences (variable cda) impacted gingival health. Gingival health was measured as the amount of gingival crevical fluid (GCF) at each tooth, measured at baseline (variable base_gcf) and at two posttreatment follow-ups at 3 and 6 months. Variable gcf records GCF at follow-up, and variable followup records the follow-up time.

Because two measurements were taken for each tooth and there exist multiple teeth per patient, we fit a model with two levels of random effects: a random intercept and random slope on follow-up time at the patient level, and a random intercept at the tooth level. For the kth measurement of the jth tooth from the ith patient, we have

$$\text{gcf}_{ijk} = \beta_0 + \beta_1 \text{followup}_{ijk} + \beta_2 \text{base\_gcf}_{ijk} + \beta_3 \text{cda}_{ijk} + \beta_4 \text{age}_{ijk} +$$
$$u_{0i} + u_{1i} \text{followup}_{ijk} + v_{0ij} + \epsilon_{ijk}$$

which we can fit using xtmixed as

```
. xtmixed gcf followup base_gcf cda age || patient: followup, cov(un) || tooth:,
> nolog
```

Mixed-effects REML regression   Number of obs   =   110

Group Variable	No. of Groups	Observations per Group		
		Minimum	Average	Maximum
patient	12	2	9.2	12
tooth	55	2	2.0	2

Log restricted-likelihood = -420.92761

Wald chi2(4)   =   7.48
Prob > chi2   =   0.1128

gcf	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
followup	.3009815	1.936863	0.16	0.877	-3.4952	4.097163
base_gcf	-.0183127	.1433094	-0.13	0.898	-.299194	.2625685
cda	-.329303	.5292525	-0.62	0.534	-1.366619	.7080128
age	-.5773932	.2139656	-2.70	0.007	-.9967582	-.1580283
_cons	45.73862	12.55497	3.64	0.000	21.13133	70.34591

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
patient: Unstructured				
sd(followup)	6.472072	1.452392	4.168943	10.04756
sd(_cons)	22.91255	5.521438	14.28736	36.74472
corr(followup,_cons)	-.9469371	.0394744	-.9878843	-.7827271
tooth: Identity				
sd(_cons)	6.888932	1.207033	4.886635	9.711668
sd(Residual)	6.990496	.7513934	5.662578	8.629822

LR test vs. linear regression: chi2(4) =   91.12   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

Among the other features of the model fit, we note that the residual standard deviation, σ_ϵ, was estimated as 6.99 and that our model assumed that the residuals were independent with constant variance (homoskedastic). Because it may be the case that the precision of gcf measurements could change over time, we modify the above to estimate two distinct error standard deviations: one for the 3-month follow-up and one for the 6-month follow-up.

To fit this model, we add the residuals(independent, by(followup)) option, which maintains independence of residual errors but allows for heteroskedasticity with respect to follow-up time.

```
. xtmixed gcf followup base_gcf cda age || patient: followup, cov(un) || tooth:,
> residuals(independent, by(followup)) nolog
```

Mixed-effects REML regression   Number of obs   =   110

Group Variable	No. of Groups	Observations per Group		
		Minimum	Average	Maximum
patient	12	2	9.2	12
tooth	55	2	2.0	2

Log restricted-likelihood = -420.4576

Wald chi2(4)   =   7.51
Prob > chi2   =   0.1113

gcf	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
followup	.2703944	1.933096	0.14	0.889	-3.518405	4.059193
base_gcf	.0062144	.1419121	0.04	0.965	-.2719283	.284357
cda	-.2947235	.5245126	-0.56	0.574	-1.322749	.7333023
age	-.5743755	.2142249	-2.68	0.007	-.9942487	-.1545024
_cons	45.15089	12.51452	3.61	0.000	20.62288	69.6789

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
patient: Unstructured				
sd(followup)	6.461555	1.449333	4.163051	10.02911
sd(_cons)	22.69806	5.55039	14.0554	36.65509
corr(followup,_cons)	-.9480776	.0395764	-.9885662	-.7800707
tooth: Identity				
sd(_cons)	6.881798	1.198038	4.892355	9.680234
Residual: Independent, by followup				
3 months: sd(e)	7.833764	1.17371	5.840331	10.5076
6 months: sd(e)	6.035612	1.240554	4.034281	9.029765

LR test vs. linear regression: chi2(5) =   92.06   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

Comparison of both models via a likelihood-ratio test reveals the difference in residual standard deviations as not significant, something we leave to you to verify as an exercise.

◁

The default residual-variance structure is independent, and when not specified with by() is equivalent to the default behavior of xtmixed: estimating one overall residual standard deviation/variance for the entire model.

Other residual-error structures

Besides the default independent residual-error structure, xtmixed supports four other structures that allow for correlation between residual errors within the lowest-level (smallest) groups. For purposes of notation, in what follows we assume a model with one level of grouping, with the obvious extension to models with multiple nested levels of random effects.

The exchangeable structure assumes one overall variance and one common pairwise covariance; that is,

$$\text{Var}(\boldsymbol{\epsilon}_i) = \text{Var} \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{in_i} \end{bmatrix} = \begin{bmatrix} \sigma_\epsilon^2 & \sigma_1 & \cdots & \sigma_1 \\ \sigma_1 & \sigma_\epsilon^2 & \cdots & \sigma_1 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma_\epsilon^2 \end{bmatrix}$$

By default, xtmixed will report estimates of the two parameters as estimates of the common standard deviation, σ_ϵ, and of pairwise correlation. If the variance option is specified, you obtain estimates of σ_ϵ^2 and the covariance σ_1. When the by(*varname*) option is also specified, these two parameters are estimated for each level *varname*.

The ar p structure assumes that the errors have an autoregressive structure of order p. That is,

$$\epsilon_{ij} = \phi_1 \epsilon_{i,j-1} + \cdots + \phi_p \epsilon_{i,j-p} + u_{ij}$$

where u_{ij} are i.i.d. Gaussian with mean zero and variance σ_u^2. xtmixed reports estimates of $\phi_1, \ldots, \phi_p$ and the overall error standard deviation σ_ϵ (or variance if the variance option is specified), which can be derived from the above expression. The t(*varname*) option is required, where *varname* is a time variable used to order the observations within lowest-level groups and to determine any gaps between observations. When the by(*varname*) option is also specified, the set of $p + 1$ parameters is estimated for each level of *varname*. If $p = 1$, then the estimate of ϕ_1 is reported as "rho", because in this case it represents the correlation between successive error terms.

The ma q structure assumes that the errors are a moving average process of order q. That is,

$$\epsilon_{ij} = u_{ij} + \theta_1 u_{i,j-1} + \cdots + \theta_q u_{i,j-q}$$

where u_{ij} are i.i.d. Gaussian with mean zero and variance σ_u^2. xtmixed reports estimates of $\theta_1, \ldots, \theta_q$ and the overall error standard deviation σ_ϵ (or variance if the variance option is specified), which can be derived from the above expression. The t(*varname*) option is required, where *varname* is a time variable used to order the observations within lowest level groups and to determine any gaps between observations. When the by(*varname* option) is also specified, the set of $q + 1$ parameters is estimated for each level of *varname*.

The unstructured structure is the most general and estimates unique variances and unique pairwise covariances for all residuals within the lowest level grouping. Because the data may be unbalanced and the ordering of the observations is arbitrary, the t(*varname*) option is required, where *varname* is an ID variable that matches error terms in different groups. If *varname* has n distinct levels, then $n(n+1)/2$ parameters are estimated. Not all n levels need to be observed within each group, but duplicated levels of *varname* within a given group are not allowed because they would cause a singularity in the estimated error variance matrix for that group. When the by(*varname*) option is also specified, the set of $n(n+1)/2$ parameters is estimated for each level of *varname*.

The banded q structure is a special case of unstructured that confines estimation to within the first q off-diagonal elements of the residual variance–covariance matrix and sets the covariances outside this band to zero. As is the case with unstructured, the t(*varname*) is required, where *varname* is an ID variable that matches error terms in different groups. However, with banded variance structures, the ordering of the values in *varname* is significant because it determines which covariances are to be estimated and which are to be set to zero. For example, if *varname* has $n = 5$ distinct values $t = 1, 2, 3, 4, 5$, then a banded variance–covariance structure of order $q = 2$ would estimate the following:

$$\text{Var}(\boldsymbol{\epsilon}) = \text{Var} \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & 0 & 0 \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} & 0 \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{34} & \sigma_{35} \\ 0 & \sigma_{24} & \sigma_{34} & \sigma_4^2 & \sigma_{45} \\ 0 & 0 & \sigma_{35} & \sigma_{45} & \sigma_5^2 \end{bmatrix}$$

In other words, you would have an unstructured variance matrix that constrains $\sigma_{14} = \sigma_{15} = \sigma_{25} = 0$. If *varname* has n distinct levels, then $(q+1)(2n-q)/2$ parameters are estimated. Not all n levels need to be observed within each group, but duplicated levels of *varname* within a given group are not allowed because they would cause a singularity in the estimated error variance matrix for that group. When the by(*varname*) option is also specified, the set of parameters is estimated for each level of *varname*. If q is left unspecified, then banded is equivalent to unstructured; that is, all variances and covariances are estimated. When $q = 0$, $\text{Var}(\boldsymbol{\epsilon}_i)$ is treated as diagonal and can thus be used to model uncorrelated, yet heteroskedastic residual errors.

The toeplitz q structure assumes that the residual errors are homoskedastic and that the correlation between two errors is determined by the time lag between the two. That is, $\text{Var}(\epsilon_{ij}) = \sigma_{\epsilon}^2$ and

$$\text{Corr}(\epsilon_{ij}, \epsilon_{i,j+k}) = \rho_k$$

If the lag k is less than or equal to q, then the pairwise correlation ρ_k is estimated; if the lag is greater than q, then ρ_k is assumed to be zero. If q is left unspecified, then ρ_k is estimated for each observed lag k. The t(*varname*) option is required, where *varname* is a time variable t used to determine the lags between pairs of residual errors. As such, t() must be integer-valued. $q + 1$ parameters are estimated, one overall variance σ_{ϵ}^2 and q correlations. When the by(*varname*) option is also specified, the set of $q + 1$ parameters is estimated for each level of *varname*.

The exponential structure is a generalization of the AR structure that allows for noninteger and irregularly spaced time lags. That is, $\text{Var}(\epsilon_{ij}) = \sigma_{\epsilon}^2$ and

$$\text{Corr}(\epsilon_{ij}, \epsilon_{ik}) = \rho^{|j-k|}$$

for $0 \leq \rho \leq 1$, with j and k not required to be integers. The t(*varname*) option is required, where *varname* is a time variable used to determine j and k for each residual-error pair. t() is real-valued. xtmixed reports estimates of σ_{ϵ}^2 and ρ. When the by(*varname*) option is also specified, these two parameters are estimated for each level of *varname*.

▷ Example 8

Pinheiro and Bates (2000, chap. 5) analyze data from a study of the estrus cycles of mares. Originally analyzed in Pierson and Ginther (1987), the data record the number of ovarian follicles larger than 10mm, daily over a period ranging from three days before ovulation to three days after the subsequent ovulation.

```
. use http://www.stata-press.com/data/r11/ovary
(Ovarian follicles in mares)

. describe

Contains data from http://www.stata-press.com/data/r11/ovary.dta
  obs:           308                          Ovarian follicles in mares
  vars:            6                          20 May 2009 13:49
  size:        6,776 (99.9% of memory free)   (_dta has notes)
```

variable name	storage type	display format	value label	variable label
mare	byte	%9.0g		mare ID
stime	float	%9.0g		Scaled time
follicles	byte	%9.0g		Number of ovarian follicles > 10 mm in diameter
sin1	float	%9.0g		sine(2*pi*stime)
cos1	float	%9.0g		cosine(2*pi*stime)
time	float	%9.0g		time order within mare

```
Sorted by:  mare  stime
```

The stime variable is time that has been scaled so that ovulation occurs at scaled times 0 and 1, and the time variable records the time ordering within mares. Because graphical evidence suggests a periodic behavior, the analysis includes the sin1 and cos1 variables, which are sine and cosine transformations of scaled time, respectively.

We consider the following model for the jth measurement on the ith mare:

$$\texttt{follicles}_{ij} = \beta_0 + \beta_1 \texttt{sin1}_{ij} + \beta_2 \texttt{cos1}_{ij} + u_i + \epsilon_{ij}$$

The above model incorporates the cyclical nature of the data as affecting the overall average number of follicles and includes mare-specific random effects u_i. Because we believe successive measurements within each mare are probably correlated (even after controlling for the periodicity in the average), we also model the within-mare errors as being autoregressive of order 2.

```
. xtmixed follicles sin1 cos1 || mare:, residuals(ar 2, t(time)) nolog
Mixed-effects REML 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)       =      34.72
Log restricted-likelihood = -772.59855     Prob > chi2        =     0.0000
```

follicles	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]
sin1	-2.899227	.5110786	-5.67	0.000	-3.900923 -1.897532
cos1	-.8652936	.5432925	-1.59	0.111	-1.930127 .1995402
_cons	12.14455	.9473712	12.82	0.000	10.28774 14.00136

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
mare: Identity				
sd(_cons)	2.663195	.8264567	1.449618	4.892742
Residual: AR(2)				
phi1	.5386104	.0624897	.4161329	.661088
phi2	.1446712	.0632039	.0207939	.2685486
sd(e)	3.775055	.3225427	3.19298	4.463242

LR test vs. linear regression: chi2(3) =   251.67   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

We picked an order of 2 as a guess, but we could have used likelihood-ratio tests of competing AR models to determine the optimal order, because models of smaller order are nested within those of larger order.

◁

▷ Example 9

Fitzmaurice, Laird, and Ware (2004, chap. 7) analyzed data on 37 subjects who participated in an exercise therapy trial.

```
. use http://www.stata-press.com/data/r11/exercise
(Exercise Therapy Trial)
. describe
Contains data from http://www.stata-press.com/data/r11/exercise.dta
  obs:          259                          Exercise Therapy Trial
 vars:            4                          24 Jun 2010 18:35
 size:        2,072 (99.9% of memory free)   (_dta has notes)
```

variable name	storage type	display format	value label	variable label
id	byte	%9.0g		Person ID
day	byte	%9.0g		Day of measurement
program	byte	%9.0g		1 = reps increase; 2 = weights increase
strength	byte	%9.0g		Strength measurement

```
Sorted by:  id  day
```

Subjects (variable id) were placed on either an increased-repetition regimen (program==1) or a program that kept the repetitions constant but increased weight (program==2). Muscle-strength measurements (variable strength) were taken at baseline (day==0) and then at every two days over the next twelve days.

Following Fitzmaurice, Laird, and Ware (2004, chap. 7), and to demonstrate fitting residual-error structures to data collected at uneven time points, we confine our analysis to those data collected at baseline (day 0) and at days 4, 6, 8, and 12. We fit a full two-way factorial model of strength on program and day, with an unstructured residual-error covariance matrix over those repeated measurements taken on the same subject:

```
. keep if inlist(day, 0, 4, 6, 8, 12)
(74 observations deleted)

. xtmixed strength i.program##i.day || id:, noconstant
> residuals(unstructured, t(day)) nolog
```

Mixed-effects REML regression				Number of obs	=	173
Group variable: id				Number of groups	=	37
				Obs per group: min =		3
				avg =		4.7
				max =		5
				Wald chi2(9)	=	43.31
Log restricted-likelihood = -298.66995				Prob > chi2	=	0.0000

strength	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
2.program	1.360119	1.031824	1.32	0.187	-.6622186	3.382457
day						
4	1.125	.3416829	3.29	0.001	.4553137	1.794686
6	1.360171	.3873741	3.51	0.000	.6009323	2.119411
8	1.584576	.5048458	3.14	0.002	.5950966	2.574056
12	1.623562	.5537523	2.93	0.003	.5382271	2.708896
program#day						
2 4	-.169159	.4548986	-0.37	0.710	-1.060744	.7224258
2 6	.2112571	.5123692	0.41	0.680	-.792968	1.215482
2 8	-.1309187	.6714468	-0.19	0.845	-1.44693	1.185093
2 12	.3205179	.7532375	0.43	0.670	-1.1558	1.796836
_cons	79.6875	.777346	102.51	0.000	78.16393	81.21107

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
id: (empty)				
Residual: Unstructured				
sd(e0)	3.109384	.37163	2.460026	3.930149
sd(e4)	3.542604	.4236122	2.802456	4.478229
sd(e6)	3.262155	.3913399	2.578644	4.126842
sd(e8)	3.740389	.4528256	2.950306	4.742053
sd(e12)	3.734223	.4619605	2.930206	4.758853
corr(e0,e4)	.92373	.024856	.8569737	.9599995
corr(e0,e6)	.884732	.0370518	.7867596	.9392225
corr(e0,e8)	.8437011	.0495939	.7146234	.9172149
corr(e0,e12)	.8101738	.0610426	.6523359	.9006621
corr(e4,e6)	.9597348	.0135117	.9227054	.9792163
corr(e4,e8)	.9493895	.0173757	.9015375	.9743001
corr(e4,e12)	.9016953	.034642	.8068802	.9512163
corr(e6,e8)	.9577075	.0162689	.9108242	.9801978
corr(e6,e12)	.9112644	.0304673	.8283424	.955113
corr(e8,e12)	.9394158	.0223506	.8764459	.9707924

```
LR test vs. linear regression:      chi2(14) =    296.13   Prob > chi2 = 0.0000
```

Note: The reported degrees of freedom assumes the null hypothesis is not on
      the boundary of the parameter space. If this is not true, then the
      reported test is conservative.

   Because we are using variable id only to group the repeated measurements and not to introduce
random effects at the subject level, we use the noconstant option to omit any subject-level effects.
The unstructured covariance matrix is the most general and contains many parameters. In this example,

we estimate a distinct residual standard error for each day and a distinct correlation for each pair of days.

That there is very high positive correlation between all pairs of measurements is evident, but what is not as evident is whether the pairwise correlation may be more parsimoniously represented. One option would be to explore whether the correlation diminishes as the time gap between strength measurements increases and whether it diminishes systematically. Given the irregularity of the time intervals, an exponential structure would be more appropriate than, say, an autoregressive or moving-average structure.

```
. estimates store unstructured

. xtmixed strength i.program##i.day || id:, noconstant
> residuals(exponential, t(day)) nolog nofetable
```

```
Mixed-effects REML regression                    Number of obs      =        173
Group variable: id                               Number of groups   =         37

                                                 Obs per group: min =          3
                                                                avg =        4.7
                                                                max =          5

                                                 Wald chi2(9)       =      34.63
Log restricted-likelihood = -309.27293           Prob > chi2        =     0.0001
```

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
id: (empty)				
Residual: Exponential				
rho	.9785563	.0052913	.9653241	.9868081
sd(e)	3.445999	.3689702	2.793671	4.250647

```
LR test vs. linear regression:        chi2(1) =    274.93   Prob > chi2 = 0.0000
```

Note: The reported degrees of freedom assumes the null hypothesis is not on
      the boundary of the parameter space. If this is not true, then the
      reported test is conservative.

In the above example, we suppressed displaying the main regression parameters because they did not differ much from those of the previous model. While the unstructured model estimated fifteen variance–covariance parameters, the exponential model claims to get the job done with just two, a fact that is not disputed by an LR test comparing the two nested models (at least not at the 0.05 level).

```
. lrtest unstructured .
```

```
Likelihood-ratio test                            LR chi2(13) =      21.21
(Assumption: . nested in unstructured)           Prob > chi2 =     0.0690
```

Note: The reported degrees of freedom assumes the null hypothesis is not on
      the boundary of the parameter space. If this is not true, then the
      reported test is conservative.
Note: LR tests based on REML are valid only when the fixed-effects
      specification is identical for both models.

◁

Random-effects factor notation and crossed-effects models

Not all mixed models contain nested levels of random effects.

▷ Example 10

Returning to our longitudinal analysis of pig weights, suppose that instead of (5) we wish to fit

$$\texttt{weight}_{ij} = \beta_0 + \beta_1 \texttt{week}_{ij} + u_i + v_j + \epsilon_{ij} \tag{8}$$

for the $i = 1, \ldots, 48$ pigs and $j = 1, \ldots, 9$ weeks and

$$u_i \sim N(0, \sigma_u^2); \quad v_j \sim N(0, \sigma_v^2); \quad \epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$$

all independently. Both (5) and (8) assume an overall population-average growth curve $\beta_0 + \beta_1 \texttt{week}$ and a random pig-specific shift.

The models differ in how week enters into the random part of the model. In (5), we assume that the effect due to week is linear and pig specific (a random slope); in (8), we assume that the effect due to week, v_j, is systematic to that week and common to all pigs. The rationale behind (8) could be that, assuming that the pigs were measured contemporaneously, we might be concerned that week-specific random factors such as weather and feeding patterns had significant systematic effects on all pigs.

Model (8) is an example of a two-way *crossed-effects* model, with the pig effects, u_i, being crossed with the week effects, v_j. One way to fit such models is to consider all the data as one big panel and treat the u_i and v_j as a series of $48 + 9 = 57$ random coefficients on indicator variables for pig and week. In the notation of (2),

$$\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_{48} \\ v_1 \\ \vdots \\ v_9 \end{bmatrix} \sim N(\mathbf{0}, \mathbf{G}); \quad \mathbf{G} = \begin{bmatrix} \sigma_u^2 \mathbf{I}_{48} & \mathbf{0} \\ \mathbf{0} & \sigma_v^2 \mathbf{I}_9 \end{bmatrix}$$

Because $\mathbf{G}$ is block diagonal, it can be represented in xtmixed as repeated-level equations. All we need is an ID variable to identify all the observations as one big group and a way to tell xtmixed to treat pig and week as factor variables (or equivalently, as two sets of overparameterized indicator variables identifying pigs and weeks, respectively). xtmixed supports the special group designation _all for the former and the factor notation R.*varname* for the latter.

```
. use http://www.stata-press.com/data/r11/pig
(Longitudinal analysis of pig weights)
. xtmixed weight week || _all: R.id || _all: R.week
Performing EM optimization:
Performing gradient-based optimization:
Iteration 0:   log restricted-likelihood = -1015.4214
Iteration 1:   log restricted-likelihood = -1015.4214
Computing standard errors:
Mixed-effects REML regression              Number of obs      =        432
Group variable: _all                       Number of groups   =          1
                                           Obs per group: min =        432
                                                          avg =      432.0
                                                          max =        432
                                           Wald chi2(1)       =   11515.87
Log restricted-likelihood = -1015.4214     Prob > chi2        =     0.0000
```

weight	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
week	6.209896	.0578677	107.31	0.000	6.096477	6.323314
_cons	19.35561	.649402	29.81	0.000	18.08281	20.62842

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
_all: Identity				
sd(R.id)	3.89265	.4141712	3.159942	4.795256
_all: Identity				
sd(R.week)	.3337658	.1611831	.129532	.8600158
sd(Residual)	2.072916	.0755914	1.92993	2.226495

```
LR test vs. linear regression:       chi2(2) =    476.10   Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
. estimates store crossed
```

Thus we estimate $\hat{\sigma}_u = 3.89$ and $\hat{\sigma}_v = 0.33$. Both (5) and (8) estimate a total of five parameters, two fixed effects and three variance components. The models, however, are not nested within each other, which precludes the use of an LR test to compare both models. Refitting model (5) and looking at the AIC values by using `estimates stats`,

```
. quietly xtmixed weight week || id:week
. estimates stats crossed .
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
crossed	432	.	-1015.421	5	2040.843	2061.185
.	432	.	-870.5147	5	1751.029	1771.372

Note: N=Obs used in calculating BIC; see **[R] BIC note**

definitely favors model (5). This finding is not surprising, given that our rationale behind (8) was somewhat fictitious. In our `estimates stats` output, the values of `ll(null)` are missing. `xtmixed` does not fit a constant-only model as part of its usual estimation of the full model, but you can use `xtmixed` to fit a constant-only model directly, if you wish.

◁

The R. *varname* notation is equivalent to giving a list of overparameterized (none dropped) indicator variables for use in a random-effects specification. When you use R. *varname*, xtmixed handles the calculations internally rather than creating the indicators in the data. Because the set of indicators is overparameterized, R. *varname* implies noconstant. You can include factor variables in the fixed-effects specification by using standard methods; see [U] **11.4.3 Factor variables**. However, random-effects equations support only the R. *varname* factor specification. For more complex factor specifications (such as interactions) in random-effects equations, use generate to form the variables manually, as we demonstrated in example 6.

❑ Technical note

Although we were able to fit the crossed-effects model (8), it came at the expense of increasing the column dimension of our random-effects design from two in model (5) to 57 in model (8). Computation time and memory requirements grow (roughly) quadratically with the dimension of the random effects. As a result, fitting such crossed-effects models is feasible only when the total column dimension is small to moderate.

Reexamining model (8), we note that if we drop v_j, we end up with a model equivalent to (4), meaning that we could have fit (4) by typing

```
. xtmixed weight week || _all: R.id
```

instead of

```
. xtmixed weight week || id:
```

as we did when we originally fit the model. The results of both estimations are identical, but the latter specification, organized at the panel (pig) level with random-effects dimension one (a random intercept) is much more computationally efficient. Whereas with the first form we are limited in how many pigs we can analyze, there is no such limitation with the second form.

Furthermore, we fit model (8) by using

```
. xtmixed weight week || _all: R.id || _all: R.week
```

as a direct way to demonstrate factor notation. However, we can technically treat pigs as nested within the "_all" group, yielding the equivalent and more efficient (total column dimension 10) way to fit (8):

```
. xtmixed weight week || _all: R.week || id:
```

We leave it to you to verify that both produce identical results. See Rabe-Hesketh and Skrondal (2008, chap. 11) for more techniques for making calculations more efficient in more complex models.

❑

▷ Example 11

As another example of how the same model may be fit in different ways by using xtmixed (and as a way to demonstrate covariance(exchangeable)), consider the model used in example 4:

$$\mathbf{y}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + u_i^{(1)} + u_{ij}^{(2)} + \boldsymbol{\epsilon}_{ij}$$

where $\mathbf{y}_{ij}$ represents the logarithms of gross state products for the $n_{ij} = 17$ observations from state j in region i, $\mathbf{X}_{ij}$ is a set of regressors, $u_i^{(1)}$ is a random intercept at the region level, and $u_{ij}^{(2)}$ is a random intercept at the state (nested within region) level. We assume that $u_i^{(1)} \sim N(0, \sigma_1^2)$ and $u_{ij}^{(2)} \sim N(0, \sigma_2^2)$ independently. Define

$$
\mathbf{v}_i = \begin{bmatrix} u_i^{(1)} + u_{i1}^{(2)} \\ u_i^{(1)} + u_{i2}^{(2)} \\ \vdots \\ u_i^{(1)} + u_{iM_i}^{(2)} \end{bmatrix}
$$

where M_i is the number of states in region i. Making this substitution, we can stack the observations for all the states within region i to get

$$
\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{v}_i + \boldsymbol{\epsilon}_i
$$

where $\mathbf{Z}_i$ is a set of indicators identifying the states within each region; that is,

$$
\mathbf{Z}_i = \mathbf{I}_{M_i} \otimes \mathbf{J}_{17}
$$

for a k-column vector of ones $\mathbf{J}_k$, and

$$
\boldsymbol{\Sigma} = \mathrm{Var}(\mathbf{v}_i) = \begin{bmatrix} \sigma_1^2 + \sigma_2^2 & \sigma_1^2 & \cdots & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 + \sigma_2^2 & \cdots & \sigma_1^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 + \sigma_2^2 \end{bmatrix}_{M_i \times M_i}
$$

Because $\boldsymbol{\Sigma}$ is an exchangeable matrix, we can fit this alternative form of the model by specifying the `exchangeable` covariance structure.

```
. use http://www.stata-press.com/data/r11/productivity
(Public Capital Productivity)
```

```
. xtmixed gsp private emp hwy water other unemp || region: R.state,
> cov(exchangeable) variance
```
  (*output omitted*)

| Mixed-effects REML regression | | | | Number of obs | = | 816 |
| Group variable: region | | | | Number of groups | = | 9 |

				Obs per group: min =	51
				avg =	90.7
				max =	136

| | | | | Wald chi2(6) | = | 18382.38 |
| Log restricted-likelihood = | 1404.7101 | | | Prob > chi2 | = | 0.0000 |

gsp	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
private	.2660308	.0215471	12.35	0.000	.2237993	.3082623
emp	.7555059	.0264556	28.56	0.000	.7036539	.8073579
hwy	.0718857	.0233478	3.08	0.002	.0261249	.1176464
water	.0761552	.0139952	5.44	0.000	.0487251	.1035853
other	-.1005396	.0170173	-5.91	0.000	-.1338929	-.0671862
unemp	-.0058815	.0009093	-6.47	0.000	-.0076636	-.0040994
_cons	2.126995	.1574864	13.51	0.000	1.818327	2.435663

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
region: Exchangeable				
var(R.state)	.0083402	.0020718	.0051254	.0135715
cov(R.state)	.0018963	.0016225	-.0012836	.0050763
var(Residual)	.0013543	.0000695	.0012247	.0014976

LR test vs. linear regression: chi2(2) = 1162.40   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

The estimates of the fixed effects and their standard errors are equivalent to those from example 4, and remapping the variance components from $(\sigma_1^2 + \sigma_2^2, \sigma_1^2, \sigma_\epsilon^2)$, as displayed here, to $(\sigma_1, \sigma_2, \sigma_\epsilon)$, as displayed in example 4, will show that they are equivalent as well.

Of course, given the discussion in the previous technical note, it is more efficient to fit this model as we did originally, as a two-level model.

◁

Diagnosing convergence problems

Given the flexibility of the class of linear mixed models, you will find that some models "fail to converge" when used with your data. The default gradient-based method used by xtmixed is the Newton–Raphson algorithm, requiring the calculation of a gradient vector and Hessian (second-derivative) matrix; see [R] **ml**.

A failure to converge can take any one of three forms:

1. repeated "nonconcave" or "backed-up" iterations without convergence;

2. a Hessian (second-derivative) calculation that has become asymmetric, unstable, or has missing values; or

3. the message "standard-error calculation has failed" when computing standard errors.

All three situations essentially amount to the same thing: the Hessian calculation has become unstable, most likely because of a ridge in the likelihood function, a subsurface of the likelihood in which all points give the same value of the likelihood and for which there is no unique solution.

Such behavior is usually the result of either

A. a model that is not identified given the data, for example, fitting the two-level nested random intercept model

$$y_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta} + u_i^{(1)} + u_{ij}^{(2)} + \epsilon_{ij}$$

without any replicated measurements at the (i, j) level. This model is unidentified for such data because the random intercepts $u_{ij}^{(2)}$ are confounded with the overall errors ϵ_{ij}; or

B. a model that contains a variance component whose estimate is really close to zero. When this occurs, a ridge is formed by an interval of values near zero, which produce the same likelihood and look equally good to the optimizer.

In models with independent and homoskedastic residuals, one useful way to diagnose problems of nonconvergence is to rely on the expectation-maximization (EM) algorithm (Dempster, Laird, and Rubin 1977), normally used by xtmixed only as a means of refining starting values. The advantages of EM are that it does not require a Hessian calculation, each successive EM iteration will result in a larger likelihood, iterations can be calculated quickly, and iterations will quickly bring parameter estimates into a neighborhood of the solution. The disadvantages of EM are that, once in a neighborhood of the solution, it can be slow to converge, if at all, and EM provides no facility for estimating standard errors of the estimated variance components. One useful property of EM is that it is always willing to provide a solution if you allow it to iterate enough times, if you are satisfied with being in a neighborhood of the optimum rather than right on the optimum, and if standard errors of variance components are not crucial to your analysis. If you encounter a nonconvergent model, try using the emonly option to bypass gradient-based optimization. Use emiterate(#) to specify the maximum number of EM iterations, which you will usually want to set much higher than the default of 20. If your EM solution shows an estimated variance component that is near zero, this provides evidence that reason B is the cause of the nonconvergence of the gradient-based method, in which case the solution would be to drop the offending variance component from the model. If no estimated variance components are near zero, reason A could be the culprit.

If your data and model are nearly unidentified, as opposed to fully unidentified, you may be able to obtain convergence with standard errors by changing some of the settings of the gradient-based optimization. Adding the difficult option can be particularly helpful if you are seeing many "nonconcave" messages; you may also consider changing the technique() or using the nonrtolerance option; see [R] **maximize**.

Distribution theory for likelihood-ratio tests

When determining the asymptotic distribution of a likelihood-ratio (LR) test comparing two nested models fit by xtmixed, issues concerning boundary problems imposed by estimating strictly positive quantities (that is, variances) can complicate the situation. When performing LR tests involving mixed models (whether comparing with linear regression within xtmixed or comparing two separate mixed models with lrtest), you may thus sometimes see a test labeled as "chibar" rather than the usual "chi2", or you may see a chi2 test with a note attached stating that the test is conservative or possibly conservative depending on the hypothesis being tested.

At the heart of the issue is the number of variances being restricted to zero in the reduced model. If there are none, the usual asymptotic theory holds, and the distribution of the test statistic is χ^2 with degrees of freedom equal to the difference in the number of estimated parameters between both models.

When there is only one variance being set to zero in the reduced model, the asymptotic distribution of the LR test statistic is a 50:50 mixture of a χ^2_k and a χ^2_{k+1} distribution, where k is the number of other restricted parameters in the reduced model that are unaffected by boundary conditions. Stata labels such test statistics as chibar and adjusts the significance levels accordingly. See Self and Liang (1987) for the appropriate theory or Gutierrez, Carter, and Drukker (2001) for a Stata-specific discussion.

When more than one variance parameter is being set to zero in the reduced model, however, the situation becomes more complicated. For example, consider a comparison test versus linear regression for a mixed model with two random coefficients and unstructured covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}$$

Because the random component of the mixed model comprises three parameters $(\sigma_0^2, \sigma_{01}, \sigma_1^2)$, on the surface it would seem that the LR comparison test would be distributed as χ^2_3. However, two complications need to be considered. First, the variances σ_0^2 and σ_1^2 are restricted to be positive, and second, constraints such as $\sigma_1^2 = 0$ implicitly restrict the covariance σ_{01} to be zero as well. From a technical standpoint, it is unclear how many parameters must be restricted to reduce the model to linear regression.

Because of these complications, appropriate and sufficiently general distribution theory for the more-than-one-variance case has yet to be developed. Theory (for example, Stram and Lee [1994]) and empirical studies (for example, McLachlan and Basford [1988]) have demonstrated that, whatever the distribution of the LR test statistic, its tail probabilities are bounded above by those of the χ^2 distribution with degrees of freedom equal to the full number of restricted parameters (three in the above example).

xtmixed uses this reference distribution, the χ^2 with full degrees of freedom, to produce a conservative test and places a note in the output labeling the test as such. Because the displayed significance level is an upper bound, rejection of the null hypothesis based on the reported level would imply rejection on the basis of the actual level.

❑ Technical note

It may seem that xtmixed does not follow Stata's standard syntax for multiple-equation models, but it does. In example 2, we typed

        . xtmixed weight week || id:

but we could have used the standard multiequation syntax:

        . xtmixed (weight week) (id:)

xtmixed will understand either and produce the same results. We prefer the syntax using || because it better emphasizes the nested structure of the levels.

❑

Saved results

xtmixed saves the following in e():

Scalars
e(N)	number of observations
e(k)	number of parameters
e(k_f)	number of FE parameters
e(k_r)	number of RE parameters
e(k_rs)	number of standard deviations
e(k_rc)	number of correlations
e(k_res)	number of residual-error parameters
e(nrgroups)	number of residual-error by() groups
e(ar_p)	AR order of residual errors, if specified
e(ma_q)	MA order of residual errors, if specified
e(res_order)	order of residual-error structure, if appropriate
e(df_m)	model degrees of freedom
e(ll)	log (restricted) likelihood
e(chi2)	χ^2 statistic
e(p)	p-value for χ^2
e(ll_c)	log likelihood, comparison model
e(chi2_c)	χ^2, comparison model
e(df_c)	degrees of freedom, comparison model
e(p_c)	p-value, comparison model
e(rank)	rank of e(V)
e(rc)	return code
e(converged)	1 if converged, 0 otherwise

Macros
e(cmd)	xtmixed
e(cmdline)	command as typed
e(depvar)	name of dependent variable
e(ivars)	grouping variables
e(title)	title in estimation output
e(redim)	random-effects dimensions
e(vartypes)	variance-structure types
e(revars)	random-effects covariates
e(resopt)	residuals() specification, as typed
e(rstructure)	residual-error structure
e(rstructlab)	residual-error structure output label
e(rbyvar)	residual-error by() variable, if specified
e(rglabels)	residual-error by() groups labels
e(timevar)	residual-error t() variable, if specified
e(chi2type)	Wald; type of model χ^2 test
e(vce)	bootstrap or jackknife, if defined
e(vcetype)	title used to label Std. Err.
e(method)	ML or REML
e(opt)	type of optimization
e(optmetric)	matsqrt or matlog; random-effects matrix parameterization
e(ml_method)	type of ml method
e(technique)	maximization technique
e(crittype)	optimization criterion
e(properties)	b V
e(estat_cmd)	program used to implement estat
e(predict)	program used to implement predict
e(asbalanced)	factor variables fvset as asbalanced
e(asobserved)	factor variables fvset as asobserved

Matrices
    e(b)   coefficient vector
    e(N_g)   group counts
    e(g_min)   group-size minimums
    e(g_avg)   group-size averages
    e(g_max)   group-size maximums
    e(tmap)   ID mapping for unstructured residual errors
    e(V)   variance–covariance matrix of the estimator
Functions
    e(sample)   marks estimation sample

Methods and formulas

xtmixed is implemented as an ado-file that uses Mata.

As given by (1), we have the linear mixed model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$$

where $\mathbf{y}$ is the $n \times 1$ vector of responses, $\mathbf{X}$ is an $n \times p$ design/covariate matrix for the fixed effects $\boldsymbol{\beta}$, and $\mathbf{Z}$ is the $n \times q$ design/covariate matrix for the random effects $\mathbf{u}$. The $n \times 1$ vector of errors, $\boldsymbol{\epsilon}$, is for now assumed to be multivariate normal with mean zero and variance matrix $\sigma_\epsilon^2 \mathbf{I}_n$. We also assume that $\mathbf{u}$ has variance–covariance matrix $\mathbf{G}$ and that $\mathbf{u}$ is orthogonal to $\boldsymbol{\epsilon}$ so that

$$\text{Var}\begin{bmatrix} \mathbf{u} \\ \boldsymbol{\epsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \sigma_\epsilon^2 \mathbf{I}_n \end{bmatrix}$$

Considering the combined error term $\mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$, we see that $\mathbf{y}$ is multivariate normal with mean $\mathbf{X}\boldsymbol{\beta}$ and $n \times n$ variance–covariance matrix

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \sigma_\epsilon^2 \mathbf{I}_n$$

Defining $\boldsymbol{\theta}$ as the vector of unique elements of $\mathbf{G}$ results in the log likelihood

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_\epsilon^2) = -\frac{1}{2}\left\{ n\log(2\pi) + \log|\mathbf{V}| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \tag{9}$$

which is maximized as a function of $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, and σ_ϵ^2. As explained in chapter 6 of Searle, Casella, and McCulloch (1992), considering instead the likelihood of a set of linear contrasts, $\mathbf{K}\mathbf{y}$, that do not depend on $\boldsymbol{\beta}$ results in the restricted log likelihood

$$L_R(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_\epsilon^2) = L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_\epsilon^2) - \frac{1}{2}\log\left|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right| \tag{10}$$

Given the high dimension of $\mathbf{V}$, however, the log-likelihood and restricted log-likelihood criteria are not usually computed by brute-force application of the above expressions. Instead, you can simplify the problem by subdividing the data into independent panels (and subpanels if possible) and using matrix decomposition methods on the smaller matrices that result from treating each panel one at a time.

Consider the one-level model described previously in (2)

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\epsilon}_i$$

for $i = 1, \ldots, M$ panels with panel i containing n_i observations, with $\text{Var}(\mathbf{u}_i) = \boldsymbol{\Sigma}$, a $q \times q$ matrix.

Efficient methods for computing (9) and (10) are given in chapter 2 of Pinheiro and Bates (2000). Namely, for the one-level model, define $\boldsymbol{\Delta}$ to be the Cholesky factor of $\sigma_\epsilon^2 \boldsymbol{\Sigma}^{-1}$, such that $\sigma_\epsilon^2 \boldsymbol{\Sigma}^{-1} = \boldsymbol{\Delta}' \boldsymbol{\Delta}$. For $i = 1, \ldots, M$, decompose

$$\begin{bmatrix} \mathbf{Z}_i \\ \boldsymbol{\Delta} \end{bmatrix} = \mathbf{Q}_i \begin{bmatrix} \mathbf{R}_{11i} \\ \mathbf{0} \end{bmatrix}$$

using an orthogonal-triangular (QR) decomposition, with $\mathbf{Q}_i$ a $(n_i + q)$-square matrix and $\mathbf{R}_{11i}$ a q-square matrix. We then apply $\mathbf{Q}_i$ as follows:

$$\begin{bmatrix} \mathbf{R}_{10i} \\ \mathbf{R}_{00i} \end{bmatrix} = \mathbf{Q}_i' \begin{bmatrix} \mathbf{X}_i \\ \mathbf{0} \end{bmatrix}; \qquad \begin{bmatrix} \mathbf{c}_{1i} \\ \mathbf{c}_{0i} \end{bmatrix} = \mathbf{Q}_i' \begin{bmatrix} \mathbf{y}_i \\ \mathbf{0} \end{bmatrix}$$

Stack the $\mathbf{R}_{00i}$ and $\mathbf{c}_{0i}$ matrices, and perform the additional QR decomposition

$$\begin{bmatrix} \mathbf{R}_{001} & \mathbf{c}_{01} \\ \vdots & \vdots \\ \mathbf{R}_{00M} & \mathbf{c}_{0M} \end{bmatrix} = \mathbf{Q}_0 \begin{bmatrix} \mathbf{R}_{00} & \mathbf{c}_0 \\ \mathbf{0} & \mathbf{c}_1 \end{bmatrix}$$

Pinheiro and Bates (2000) show that ML estimates of $\boldsymbol{\beta}$, σ_ϵ^2, and $\boldsymbol{\Delta}$ (the unique elements of $\boldsymbol{\Delta}$, that is) are obtained by maximizing the profile log likelihood (profiled in $\boldsymbol{\Delta}$)

$$L(\boldsymbol{\Delta}) = \frac{n}{2} \left\{ \log n - \log(2\pi) - 1 \right\} - n \log ||\mathbf{c}_1|| + \sum_{i=1}^{M} \log \left| \frac{\det(\boldsymbol{\Delta})}{\det(\mathbf{R}_{11i})} \right| \tag{11}$$

where $|| \cdot ||$ denotes the 2-norm, and following this maximization with

$$\widehat{\boldsymbol{\beta}} = \mathbf{R}_{00}^{-1} \mathbf{c}_0; \quad \widehat{\sigma}_\epsilon^2 = n^{-1} ||\mathbf{c}_1||^2 \tag{12}$$

REML estimates are obtained by maximizing

$$\begin{aligned} L_R(\boldsymbol{\Delta}) = \frac{n-p}{2} &\left\{ \log(n-p) - \log(2\pi) - 1 \right\} - (n-p) \log ||\mathbf{c}_1|| \\ &- \log |\det(\mathbf{R}_{00})| + \sum_{i=1}^{M} \log \left| \frac{\det(\boldsymbol{\Delta})}{\det(\mathbf{R}_{11i})} \right| \end{aligned} \tag{13}$$

followed by

$$\widehat{\boldsymbol{\beta}} = \mathbf{R}_{00}^{-1} \mathbf{c}_0; \quad \widehat{\sigma}_\epsilon^2 = (n-p)^{-1} ||\mathbf{c}_1||^2$$

For numerical stability, maximization of (11) and (13) is not performed with respect to the unique elements of $\boldsymbol{\Delta}$ but instead with respect to the unique elements of the matrix square root (or matrix logarithm if the matlog option is specified) of $\boldsymbol{\Sigma}/\sigma_\epsilon^2$; define $\boldsymbol{\gamma}$ to be the vector containing these elements.

Once maximization with respect to $\boldsymbol{\gamma}$ is completed, $(\boldsymbol{\gamma}, \sigma_\epsilon^2)$ is reparameterized to $\{\boldsymbol{\alpha}, \log(\sigma_\epsilon)\}$, where $\boldsymbol{\alpha}$ is a vector containing the unique elements of $\boldsymbol{\Sigma}$, expressed as logarithms of standard deviations for the diagonal elements and hyperbolic arctangents of the correlations for off-diagonal elements. This last step is necessary to (a) obtain a joint variance–covariance estimate of the elements of $\boldsymbol{\Sigma}$ and σ_ϵ^2; (b) obtain a parameterization under which parameter estimates can be interpreted individually, rather than as elements of a matrix square root (or logarithm); and (c) parameterize these elements such that their ranges each encompass the entire real line.

Obtaining a joint variance–covariance matrix for the estimated $\{\boldsymbol{\alpha}, \log(\sigma_\epsilon)\}$ requires the evaluation of the log likelihood (or log-restricted likelihood) with only $\boldsymbol{\beta}$ profiled out. For ML, we have

$$L^*\{\boldsymbol{\alpha}, \log(\sigma_\epsilon)\} = L\{\boldsymbol{\Delta}(\boldsymbol{\alpha}, \sigma_\epsilon^2), \sigma_\epsilon^2\}$$

$$= -\frac{n}{2}\log(2\pi\sigma_\epsilon^2) - \frac{||\mathbf{c}_1||^2}{2\sigma_\epsilon^2} + \sum_{i=1}^{M} \log\left|\frac{\det(\boldsymbol{\Delta})}{\det(\mathbf{R}_{11i})}\right|$$

with the analogous expression for REML.

The variance–covariance matrix of $\widehat{\boldsymbol{\beta}}$ is estimated as

$$\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\beta}}) = \widehat{\sigma}_\epsilon^2 \mathbf{R}_{00}^{-1}\left(\mathbf{R}_{00}^{-1}\right)'$$

but this does not mean that $\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\beta}})$ is identical under both ML and REML because $\mathbf{R}_{00}$ depends on $\boldsymbol{\Delta}$. Because $\widehat{\boldsymbol{\beta}}$ is asymptotically uncorrelated with $\{\widehat{\boldsymbol{\alpha}}, \log(\widehat{\sigma}_\epsilon)\}$, the covariance of $\widehat{\boldsymbol{\beta}}$ with the other estimated parameters is treated as zero.

Parameter estimates are stored in e(b) as $\{\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}}, \log(\widehat{\sigma}_\epsilon)\}$, with the corresponding (block-diagonal) variance–covariance matrix stored in e(V). Parameter estimates can be displayed in this metric by specifying the estmetric option. However, in xtmixed output, variance components are most often displayed either as variances and covariances or as standard deviations and correlations.

EM iterations are derived by considering the $\mathbf{u}_i$ in (2) as missing data. Here we describe the procedure for maximizing the log likelihood via EM; the procedure for maximizing the restricted log likelihood is similar. The log likelihood for the full data $(\mathbf{y}, \mathbf{u})$ is

$$L_F(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \sigma_\epsilon^2) = \sum_{i=1}^{M}\left\{\log f_1(\mathbf{y}_i|\mathbf{u}_i, \boldsymbol{\beta}, \sigma_\epsilon^2) + \log f_2(\mathbf{u}_i|\boldsymbol{\Sigma})\right\}$$

where $f_1()$ is the density function for multivariate normal with mean $\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i$ and variance $\sigma_\epsilon^2 \mathbf{I}_{n_i}$, and $f_2()$ is the density for multivariate normal with mean $\mathbf{0}$ and $q \times q$ covariance matrix $\boldsymbol{\Sigma}$. As before, we can profile $\boldsymbol{\beta}$ and σ_ϵ^2 out of the optimization, yielding the following EM iterative procedure:

1. For the current iterated value of $\boldsymbol{\Sigma}^{(t)}$, fix $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\boldsymbol{\Sigma}^{(t)})$ and $\widehat{\sigma}_\epsilon^2 = \widehat{\sigma}_\epsilon^2(\boldsymbol{\Sigma}^{(t)})$ according to (12).

2. Expectation step: Calculate

$$D(\boldsymbol{\Sigma}) \equiv E\left\{L_F(\widehat{\boldsymbol{\beta}}, \boldsymbol{\Sigma}, \widehat{\sigma}_\epsilon^2)|\mathbf{y}\right\}$$

$$= C - \frac{M}{2}\log\det(\boldsymbol{\Sigma}) - \frac{1}{2}\sum_{i=1}^{M} E\left(\mathbf{u}_i'\boldsymbol{\Sigma}^{-1}\mathbf{u}_i|\mathbf{y}\right)$$

where C is a constant that does not depend on $\boldsymbol{\Sigma}$, and the expected value of the quadratic form $\mathbf{u}_i'\boldsymbol{\Sigma}^{-1}\mathbf{u}_i$ is taken with respect to the conditional density $f(\mathbf{u}_i|\mathbf{y}, \widehat{\boldsymbol{\beta}}, \boldsymbol{\Sigma}^{(t)}, \widehat{\sigma}_\epsilon^2)$.

3. Maximization-step: Maximize $D(\boldsymbol{\Sigma})$ to produce $\boldsymbol{\Sigma}^{(t+1)}$.

For general, symmetric $\boldsymbol{\Sigma}$, the maximizer of $D(\boldsymbol{\Sigma})$ can be derived explicitly, making EM iterations quite fast.

For general residual-error structures,

$$\mathrm{Var}(\boldsymbol{\epsilon}_i) = \sigma_\epsilon^2 \boldsymbol{\Lambda}_i$$

where the subscript i merely represents that ϵ_i and Λ_i vary in dimension in unbalanced data, the data are first transformed according to

$$\mathbf{y}_i^* = \widehat{\mathbf{\Lambda}}_i^{-1/2}\mathbf{y}; \qquad \mathbf{X}_i^* = \widehat{\mathbf{\Lambda}}_i^{-1/2}\mathbf{X}; \qquad \mathbf{Z}_i^* = \widehat{\mathbf{\Lambda}}_i^{-1/2}\mathbf{Z};$$

and the likelihood-evaluation techniques described above are applied to $\mathbf{y}_i^*$, $\mathbf{X}_i^*$, and $\mathbf{Z}_i^*$ instead. The unique elements of $\mathbf{\Lambda}$, $\boldsymbol{\rho}$, are estimated along with the fixed effects and variance components. Because σ_ϵ^2 is always estimated and multiplies the entire $\mathbf{\Lambda}_i$ matrix, $\widehat{\boldsymbol{\rho}}$ is parameterized to take this into account.

EM iterations always assume an independent and homoskedastic error structure. As such, when error structures are more complex, EM is used only to obtain starting values.

For extensions to two or more nested levels of random effects, see Bates and Pinheiro (1998).

Charles Roy Henderson (1911–1989) was born in Iowa and grew up on the family farm. His education in animal husbandry, animal nutrition, and statistics at Iowa State was interspersed with jobs in the Iowa Extension Service, Ohio University, and the U.S. Army. After completing his PhD, Henderson joined the Animal Science faculty at Cornell. He developed and applied statistical methods in the improvement of farm livestock productivity through genetic selection, with particular focus on dairy cattle. His methods are general and have been used worldwide in livestock breeding and beyond agriculture. Henderson's work on variance components and best linear unbiased predictions has proved to be one of the main roots of current mixed-model methods.

Acknowledgments

References

Andrews, M., T. Schank, and R. Upward. 2006. Practical fixed-effects estimation methods for the three-way error-components model. *Stata Journal* 6: 461–481.

Baltagi, B. H., S. H. Song, and B. C. Jung. 2001. The unbalanced nested error component regression model. *Journal of Econometrics* 101: 357–381.

Bates, D. M., and J. C. Pinheiro. 1998. Computational methods for multilevel modelling. In *Technical Memorandum BL0112140-980226-01TM*. Murray Hill, NJ: Bell Labs, Lucent Technologies. http://stat.bell-labs.com/NLME/CompMulti.pdf.

Cameron, A. C., and P. K. Trivedi. 2009. *Microeconometrics Using Stata*. College Station, TX: Stata Press.

Demidenko, E. 2004. *Mixed Models: Theory and Applications*. Hoboken, NJ: Wiley.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* 39: 1–38.

Diggle, P. J., P. J. Heagerty, K.-Y. Liang, and S. L. Zeger. 2002. *Analysis of Longitudinal Data*. 2nd ed. Oxford: Oxford University Press.

Fitzmaurice, G. M., N. M. Laird, and J. H. Ware. 2004. *Applied Longitudinal Analysis*. Hoboken, NJ: Wiley.

Goldstein, H. 1986. Efficient statistical modelling of longitudinal data. *Annals of Human Biology* 13: 129–141.

Gutierrez, R. G., S. Carter, and D. M. Drukker. 2001. sg160: On boundary-value likelihood-ratio tests. *Stata Technical Bulletin* 60: 15–18. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 269–273. College Station, TX: Stata Press.

Harville, D. A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 72: 320–338.

Henderson, C. R. 1953. Estimation of variance and covariance components. *Biometrics* 9: 226–252.

Hocking, R. R. 1985. *The Analysis of Linear Models.* Monterey, CA: Brooks/Cole.

Laird, N. M., and J. H. Ware. 1982. Random-effects models for longitudinal data. *Biometrics* 38: 963–974.

LaMotte, L. R. 1973. Quadratic estimation of variance components. *Biometrics* 29: 311–330.

Marchenko, Y. 2006. Estimating variance components in Stata. *Stata Journal* 6: 1–21.

McCulloch, C. E., S. R. Searle, and J. M. Neuhaus. 2008. *Generalized, Linear, and Mixed Models.* 2nd ed. Hoboken, NJ: Wiley.

McLachlan, G. J., and K. E. Basford. 1988. *Mixture Models.* New York: Dekker.

Munnell, A. 1990. Why has productivity growth declined? Productivity and public investment. *New England Economic Review* Jan./Feb.: 3–22.

Nichols, A. 2007. Causal inference with observational data. *Stata Journal* 7: 507–541.

Pierson, R. A., and O. J. Ginther. 1987. Follicular population dynamics during the estrous cycle of the mare. *Animal Reproduction Science* 14: 219–231.

Pinheiro, J. C., and D. M. Bates. 2000. *Mixed-Effects Models in S and S-PLUS.* New York: Springer.

Prosser, R., J. Rasbash, and H. Goldstein. 1991. *ML3 Software for 3-Level Analysis: User's Guide for V. 2.* London: Institute of Education, University of London.

Rabe-Hesketh, S., and A. Skrondal. 2008. *Multilevel and Longitudinal Modeling Using Stata.* 2nd ed. College Station, TX: Stata Press.

Rao, C. R. 1973. *Linear Statistical Inference and Its Applications.* 2nd ed. New York: Wiley.

Raudenbush, S. W., and A. S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods.* 2nd ed. Thousand Oaks, CA: Sage.

Ruppert, D., M. P. Wand, and R. J. Carroll. 2003. *Semiparametric Regression.* Cambridge: Cambridge University Press.

Searle, S. R. 1989. Obituary: Charles Roy Henderson 1911–1989. *Biometrics* 45: 1333–1335.

Searle, S. R., G. Casella, and C. E. McCulloch. 1992. *Variance Components.* New York: Wiley.

Self, S. G., and K.-Y. Liang. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* 82: 605–610.

Skrondal, A., and S. Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models.* Boca Raton, FL: Chapman & Hall/CRC.

Stram, D. O., and J. W. Lee. 1994. Variance components testing in the longitudinal mixed effects model. *Biometrics* 50: 1171–1177.

Thompson Jr., W. A. 1962. The problem of negative estimates of variance components. *Annals of Mathematical Statistics* 33: 273–289.

Verbeke, G., and G. Molenberghs. 2000. *Linear Mixed Models for Longitudinal Data.* New York: Springer.

West, B. T., K. B. Welch, and A. T. Galecki. 2007. *Linear Mixed Models: A Practical Guide Using Statistical Software.* Boca Raton, FL: Chapman & Hall/CRC.

Also see

[XT] **xtmixed postestimation** — Postestimation tools for xtmixed

[XT] **xtmelogit** — Multilevel mixed-effects logistic regression

[XT] **xtmepoisson** — Multilevel mixed-effects Poisson regression

[XT] **xtreg** — Fixed-, between-, and random-effects, and population-averaged linear models

[XT] **xtrc** — Random-coefficients model

[XT] **xtgee** — Fit population-averaged panel-data models by using GEE

[U] **20 Estimation and postestimation commands**