Models for Longitudinal / Panel Data and Mixed Models

Christopher F Baum

Boston College and DIW Berlin

University of Adelaide, June 2010

Forms of panel data

To define the problems of data management, consider a dataset in which each variable contains information on N panel units, each with T time-series observations. The second dimension of panel data need not be calendar time, but many estimation techniques assume that it can be treated as such, so that operations such as first differencing make sense.

These data may be commonly stored in either the *long form* or the *wide form*, in Stata parlance. In the long form, each observation has both an i and t subscript.

Long form data:

```
. list, noobs sepby(state)
```

state	year	pop
CT	1990	3291967
CT	1995	3324144
CT	2000	3411750
MA	1990	6022639
MA	1995	6141445
MA	2000	6362076
RI	1990	1005995
RI	1995	1017002
RI	2000	1050664

However, you often encounter data in the wide form, in which different variables (or columns of the data matrix) refer to different time periods.

Wide form data:

```
. list, noobs

   state    pop1990    pop1995    pop2000

      CT    3291967    3324144    3411750
      MA    6022639    6141445    6362076
      RI    1005995    1017002    1050664
```

In a variant on this theme, the wide form data could also index the observations by the time period, and have the same measurement for different units stored in different variables.

The former kind of wide-form data, where time periods are arrayed across the columns, is often found in spreadsheets or on-line data sources.

These examples illustrate a *balanced panel*, where each unit is represented in each time period. That is often not available, as different units may enter and leave the sample in different periods (companies may start operations or liquidate, household members may die, etc.) In those cases, we must deal with *unbalanced panels*. Stata's data transformation commands are uniquely handy in that context.

The reshape command

The solution to this problem is Stata's `reshape` command, an immensely powerful tool for reformulating a dataset in memory without recourse to external files. In statistical packages lacking a data-reshape feature, common practice entails writing the data to one or more external text files and reading it back in. With the proper use of `reshape`, this is not necessary in Stata. But `reshape` requires, first of all, that the data to be reshaped are labelled in such a way that they can be handled by the mechanical rules that the command applies.

In situations beyond the simple application of `reshape`, it may require some experimentation to construct the appropriate command syntax. This is all the more reason for enshrining that code in a do-file as some day you are likely to come upon a similar application for `reshape`.

The `reshape` command works with the notion of $x_{i,j}$ data. Its syntax lists the variables to be stacked up, and specifies the *i* and *j* variables, where the *i* variable indexes the rows and the *j* variable indexes the columns in the existing form of the data. If we have a dataset in the wide form, with time periods incorporated in the variable names, we could use

```
. reshape long expp revpp avgsal math4score math7score, i(distid) j(year)
(note: j = 1992 1994 1996 1998)
Data                                    wide   ->   long

Number of obs.                           550   ->    2200
Number of variables                       21   ->       7
j variable (4 values)                          ->   year
xij variables:
        expp1992 expp1994 ... expp1998   ->   expp
     revpp1992 revpp1994 ... revpp1998   ->   revpp
   avgsal1992 avgsal1994 ... avgsal1998  ->   avgsal
math4score1992 math4score1994 ... math4score1998->math4score
math7score1992 math7score1994 ... math7score1998->math7score
```

You use `reshape long` because the data are in the wide form and we want to place them in the long form. You provide the variable names to be stacked *without* their common suffixes: in this case, the `year` embedded in their wide-form variable name. The `i` variable is `distid` and the `j` variable is `year`: together, those variables uniquely identify each measurement.

Stata's description of `reshape` speaks of *i* defining a unique observation and *j* defining a subobservation logically related to that observation. Any additional variables that do not vary over *j* are not specified in the `reshape` statement, as they will be automatically replicated for each *j*.

What if you wanted to reverse the process, and translate the data from the long to the wide form?

```
. reshape wide expp revpp avgsal math4score math7score, i(distid) j(year)
(note: j = 1992 1994 1996 1998)
Data                                  long   ->   wide

Number of obs.                        2200   ->     550
Number of variables                      7   ->      21
j variable (4 values)                 year   ->   (dropped)
xij variables:
                                      expp   ->   expp1992 expp1994 ... expp1998
                                     revpp   ->   revpp1992 revpp1994 ... revpp199
> 8
                                    avgsal   ->   avgsal1992 avgsal1994 ... avgsal
> 1998
                                 math4score  ->   math4score1992 math4score1994 ..
> . math4score1998
                                 math7score  ->   math7score1992 math7score1994 ..
> . math7score1998
```

This example highlights the importance of having appropriate variable names for `reshape`. If our wide-form dataset contained the variables `expp1992, Expen94, xpend_96` and `expstu1998` there would be no way to specify the common stub labeling the choices. However, one common case can be handled without the renaming of variables. Say that we have the variables `exp92pp, exp94pp, exp96pp, exp98pp`. The command

```
reshape long exp@pp, i(distid) j(year)
```

will deal with that case, with the `@` as a placeholder for the location of the *j* component of the variable name.

Estimation for panel data

We first consider estimation of models that satisfy the zero conditional mean assumption for OLS regression: that is, the conditional mean of the error process, conditioned on the regressors, is zero. This does not rule out non-*i.i.d.* errors, but it does rule out endogeneity of the regressors and, generally, the presence of lagged dependent variables. We will deal with these exceptions later.

The most commonly employed model for panel data, the *fixed effects* estimator, addresses the issue that no matter how many individual-specific factors you may include in the regressor list, there may be *unobserved heterogeneity* in a pooled OLS model. This will generally cause OLS estimates to be biased and inconsistent.

Given longitudinal data $\{y\ X\}$, each element of which has two subscripts: the unit identifier i and the time identifier t, we may define a number of models that arise from the most general linear representation:

$$y_{it} = \sum_{k=1}^{K} X_{kit}\beta_{kit} + \epsilon_{it}, \ i = 1, N, \ t = 1, T \tag{1}$$

Assume a balanced panel of $N \times T$ observations. Since this model contains $K \times N \times T$ regression coefficients, it cannot be estimated from the data. We could ignore the nature of the panel data and apply pooled ordinary least squares, which would assume that $\beta_{kit} = \beta_k \ \forall \ k, i, t$, but that model might be viewed as overly restrictive and is likely to have a very complicated error process (e.g., heteroskedasticity across panel units, serial correlation within panel units, and so forth). Thus the pooled OLS solution is not often considered to be practical.

One set of panel data estimators allow for heterogeneity across panel units (and possibly across time), but confine that heterogeneity to the intercept terms of the relationship. These techniques, the *fixed effects* and *random effects* models, we consider below. They impose restrictions on the model above of $\beta_{kit} = \beta_k \ \forall i, t, \ k > 1$, assuming that β_1 refers to the constant term in the relationship.

An alternative technique which may be applied to "small *N*, large *T*" panels is the method of *seemingly unrelated regressions* or SURE. The "small *N*, large *T*" setting refers to the notion that we have a relatively small number of panel units, each with a lengthy time series: for instance, financial variables of the ten largest U.S. manufacturing firms, observed over the last 40 calendar quarters. The SURE technique (implemented in Stata as `sureg`) requires that the number of time periods exceeds the number of cross-sectional units.

The fixed effects estimator

The general structure above may be restricted to allow for heterogeneity across units without the full generality (and infeasibility) that this equation implies. In particular, we might restrict the slope coefficients to be constant over both units and time, and allow for an intercept coefficient that varies by unit or by time. For a given observation, an intercept varying over units results in the structure:

$$y_{it} = \sum_{k=2}^{K} X_{kit}\beta_k + u_i + \epsilon_{it} \qquad (2)$$

There are two interpretations of u_i in this context: as a parameter to be estimated in the model (a so-called *fixed effect*) or alternatively, as a component of the disturbance process, giving rise to a composite error term $[u_i + \epsilon_{it}]$: a so-called *random effect*. Under either interpretation, u_i is taken as a random variable.

If we treat it as a fixed effect, we assume that the u_i may be correlated with some of the regressors in the model. The fixed-effects estimator removes the fixed-effects parameters from the estimator to cope with this incidental parameter problem, which implies that all inference is conditional on the fixed effects in the sample. Use of the random effects model implies additional orthogonality conditions—that the u_i are not correlated with the regressors—and yields inference about the underlying population that is not conditional on the fixed effects in our sample.

We could treat a time-varying intercept term similarly: as either a fixed effect (giving rise to an additional coefficient) or as a component of a composite error term. We concentrate here on so-called *one-way fixed (random) effects* models in which only the individual effect is considered in the "large N, small T" context most commonly found in economic and financial research.

Stata's set of `xt` commands include those which extend these panel data models in a variety of ways. For more information, see `help xt`.

One-way fixed effects: the within estimator

Rewrite the equation to express the individual effect u_i as

$$y_{it} = X_{it}^* \beta^* + Z_i \alpha + \epsilon_{it} \tag{3}$$

In this context, the X^* matrix does not contain a units vector. The heterogeneity or individual effect is captured by Z, which contains a constant term and possibly a number of other individual-specific factors. Likewise, β^* contains $\beta_2 \ldots \beta_K$ from the equation above, constrained to be equal over i and t. If Z contains only a units vector, then pooled OLS is a consistent and efficient estimator of $[\beta^* \ \alpha]$. However, it will often be the case that there are additional factors specific to the individual unit that must be taken into account, and omitting those variables from Z will cause the equation to be misspecified.

The *fixed effects* model deals with this problem by relaxing the assumption that the regression function is constant over time and space in a very modest way. A one-way fixed effects model permits each cross-sectional unit to have its own constant term while the slope estimates (β^*) are constrained across units, as is the σ_ϵ^2. This estimator is often termed the *LSDV* (least-squares dummy variable) model, since it is equivalent to including ($N - 1$) dummy variables in the OLS regression of y on X (including a units vector). The *LSDV* model may be written in matrix form as:

$$y = X\beta + D\alpha + \epsilon \tag{4}$$

where D is a $NT \times N$ matrix of dummy variables d_i (assuming a balanced panel of $N \times T$ observations).

The model has $(K - 1) + N$ parameters (recalling that the β^* coefficients are all slopes) and when this number is too large to permit estimation, we rewrite the least squares solution as

$$b = (X'M_D X)^{-1}(X'M_D y) \tag{5}$$

where

$$M_D = I - D(D'D)^{-1}D' \tag{6}$$

is an idempotent matrix which is block–diagonal in $M_0 = I_T - T^{-1}\iota\iota'$ (ι a T–element units vector).

Premultiplying any data vector by M_0 performs the demeaning transformation: if we have a T–vector Z_i, $M_0 Z_i = Z_i - \bar{Z}_i\iota$. The regression above estimates the slopes by the projection of demeaned y on demeaned X without a constant term.

The estimates a_i may be recovered from $a_i = \bar{y}_i - b'\bar{X}_i$, since for each unit, the regression surface passes through that *unit's* multivariate point of means. This is a generalization of the OLS result that in a model with a constant term the regression surface passes through the *entire sample's* multivariate point of means.

The large-sample VCE of b is $s^2[X'M_DX]^{-1}$, with s^2 based on the least squares residuals, but taking the proper degrees of freedom into account: $NT - N - (K - 1)$.

This model will have explanatory power *if and only if* the variation of the individual's *y* above or below the individual's mean is significantly correlated with the variation of the individual's *X* values above or below the individual's vector of mean *X* values. For that reason, it is termed the *within estimator*, since it depends on the variation *within* the unit.

It does not matter if some individuals have, e.g., very high *y* values and very high *X* values, since it is only the within variation that will show up as explanatory power. This is the panel analogue to the notion that OLS on a cross-section does not seek to "explain" the mean of *y*, but only the variation around that mean.

This has the clear implication that any characteristic which does not vary over time for each *unit* cannot be included in the model: for instance, an individual's gender, or a firm's three-digit SIC (industry) code. The unit-specific intercept term absorbs all heterogeneity in y and X that is a function of the identity of the unit, and any variable constant over time for each unit will be perfectly collinear with the unit's indicator variable.

The one-way individual fixed effects model may be estimated by the Stata command `xtreg` using the `fe` (fixed effects) option. The command has a syntax similar to `regress`:

> `xtreg` *depvar indepvars,* `fe` [*options*]

As with standard regression, options include `robust` and `cluster()`. The command output displays estimates of σ_u^2 (labeled `sigma_u`), σ_ϵ^2 (labeled `sigma_e`), and what Stata terms `rho`: the fraction of variance due to u_i. Stata estimates a model in which the u_i of Equation (2) are taken as deviations from a single constant term, displayed as `_cons`; therefore testing that all u_i are zero is equivalent in our notation to testing that all α_i are identical. The empirical correlation between u_i and the regressors in X^* is also displayed as `corr(u_i, Xb)`.

The fixed effects estimator does not require a balanced panel. As long as there are at least two observations per unit, it may be applied. However, since the individual fixed effect is in essence estimated from the observations of each unit, the precision of that effect (and the resulting slope estimates) will depend on N_i.

We wish to test whether the individual-specific heterogeneity of α_i is necessary: are there distinguishable intercept terms across units? `xtreg,fe` provides an *F*-test of the null hypothesis that the constant terms are equal across units. If this null is rejected, pooled OLS would represent a misspecified model. The one-way fixed effects model also assumes that the errors are not contemporaneously correlated across units of the panel. This hypothesis can be tested (provided $T > N$) by the Lagrange multiplier test of Breusch and Pagan, available as the author's `xttest2` routine (`findit xttest2`).

In this example, we have 1982–1988 state-level data for 48 U.S. states on traffic fatality rates (deaths per 100,000). We model the highway fatality rates as a function of several common factors: `beertax`, the tax on a case of beer, `spircons`, a measure of spirits consumption and two economic factors: the state unemployment rate (`unrate`) and state per capita personal income, $000 (`perincK`). We present descriptive statistics for these variables of the `traffic.dta` dataset.

```
. use traffic, clear

. summarize fatal beertax spircons unrate perincK
    Variable │        Obs        Mean    Std. Dev.        Min        Max
─────────────┼──────────────────────────────────────────────────────────
       fatal │        336    2.040444    .5701938     .82121    4.21784
     beertax │        336     .513256    .4778442   .0433109   2.720764
     spircons │       336     1.75369    .6835745        .79        4.9
      unrate │        336    7.346726    2.533405        2.4         18
      perincK │       336    13.88018    2.253046   9.513762   22.19345
```

The results of the one-way fixed effects model are:

```
. xtreg fatal beertax spircons unrate perincK, fe
```

Fixed-effects (within) regression	Number of obs	=	336
Group variable (i): state	Number of groups	=	48

R-sq:	within = 0.3526	Obs per group: min =	7
	between = 0.1146	avg =	7.0
	overall = 0.0863	max =	7
		F(4,284) =	38.68
corr(u_i, Xb) = -0.8804		Prob > F =	0.0000

fatal	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]
beertax	-.4840728	.1625106	-2.98	0.003	-.8039508 -.1641948
spircons	.8169652	.0792118	10.31	0.000	.6610484 .9728819
unrate	-.0290499	.0090274	-3.22	0.001	-.0468191 -.0112808
perincK	.1047103	.0205986	5.08	0.000	.064165 .1452555
_cons	-.383783	.4201781	-0.91	0.362	-1.210841 .4432754

sigma_u	1.1181913	
sigma_e	.15678965	
rho	.98071823	(fraction of variance due to u_i)

F test that all u_i=0:   F(47, 284) =   59.77   Prob > F = 0.0000

All explanatory factors are highly significant, with the unemployment rate having a negative effect on the fatality rate (perhaps since those who are unemployed are income-constrained and drive fewer miles), and income a positive effect (as expected because driving is a normal good).

We have considered one-way fixed effects models, where the effect is attached to the individual. We may also define a two-way fixed effect model, where effects are attached to each unit and time period. Stata lacks a command to estimate two-way fixed effects models. If the number of time periods is reasonably small, you may estimate a two-way FE model by creating a set of time indicator variables and including all but one in the regression.

The joint test that all of the coefficients on those indicator variables are zero will be a test of the significance of time fixed effects. Just as the individual fixed effects (LSDV) model requires regressors' variation over time within each *unit*, a time fixed effect (implemented with a time indicator variable) requires regressors' variation over units within each *time period*. If we are estimating an equation from individual or firm microdata, this implies that we cannot include a "macro factor" such as the rate of GDP growth or price inflation in a model with time fixed effects, since those factors do not vary across individuals.

We consider the two-way fixed effects model by adding time effects to the model of the previous example. The time effects are generated by `tabulate`'s `generate` option, and then transformed into "centered indicators" by subtracting the indicator for the excluded class from each of the other indicator variables. This expresses the time effects as variations from the conditional mean of the sample rather than deviations from the excluded class (the year 1988).

```
. qui tabulate year, generate(yr)
. local j 0
. forvalues i=82/87 {
  2.          local ++j
  3.          rename yr`j´ yr`i´
  4.          qui replace yr`i´ = yr`i´ - yr7
  5.          }
. drop yr7
```

```
. xtreg fatal beertax spircons unrate perincK yr*, fe
```

```
Fixed-effects (within) regression              Number of obs      =        336
Group variable (i): state                      Number of groups   =         48

R-sq:  within  = 0.4528                         Obs per group: min =          7
       between = 0.1090                                        avg =        7.0
       overall = 0.0770                                        max =          7

                                                F(10,278)          =      23.00
corr(u_i, Xb)  = -0.8728                         Prob > F           =     0.0000
```

fatal	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
beertax	-.4347195	.1539564	-2.82	0.005	-.7377878	-.1316511
spircons	.805857	.1126425	7.15	0.000	.5841163	1.027598
unrate	-.0549084	.0103418	-5.31	0.000	-.0752666	-.0345502
perincK	.0882636	.0199988	4.41	0.000	.0488953	.1276319
yr82	.1004321	.0355629	2.82	0.005	.0304253	.170439
yr83	.0470609	.0321574	1.46	0.144	-.0162421	.1103638
yr84	-.0645507	.0224667	-2.87	0.004	-.1087771	-.0203243
yr85	-.0993055	.0198667	-5.00	0.000	-.1384139	-.0601971
yr86	.0496288	.0232525	2.13	0.034	.0038554	.0954021
yr87	.0003593	.0289315	0.01	0.990	-.0565933	.0573119
_cons	.0286246	.4183346	0.07	0.945	-.7948812	.8521305

sigma_u	1.0987683					
sigma_e	.14570531					
rho	.98271904	(fraction of variance due to u_i)				

```
F test that all u_i=0:       F(47, 278) =      64.52                Prob > F = 0.0000
```

```
. test yr82 yr83 yr84 yr85 yr86 yr87
( 1)  yr82 = 0
( 2)  yr83 = 0
( 3)  yr84 = 0
( 4)  yr85 = 0
( 5)  yr86 = 0
( 6)  yr87 = 0

      F(  6,   278) =    8.48
         Prob > F =    0.0000
```

The four quantitative factors included in the one-way fixed effects model retain their sign and significance in the two-way fixed effects model. The time effects are jointly significant, suggesting that they should be included in a properly specified model. Otherwise, the model is qualitatively similar to the earlier model, with a sizable amount of variation explained by the individual (state) fixed effect.

The between estimator

Another estimator that may be defined for a panel data set is the *between estimator*, in which the group means of *y* are regressed on the group means of *X* in a regression of *N* observations. This estimator *ignores* all of the individual-specific variation in *y* and *X* that is considered by the within estimator, replacing each observation for an individual with their mean behavior.

This estimator is not widely used, but has sometimes been applied where the time series data for each individual are thought to be somewhat inaccurate, or when they are assumed to contain random deviations from long-run means. If you assume that the inaccuracy has mean zero over time, a solution to this measurement error problem can be found by averaging the data over time and retaining only one observation per unit.

This could be done explicitly with Stata's `collapse` command. However, you need not form that data set to employ the between estimator, since the command `xtreg` with the `be` (between) option will invoke it. Use of the between estimator requires that $N > K$. Any macro factor that is constant over *individuals* cannot be included in the between estimator, since its average will not differ by individual.

We can show that the pooled OLS estimator is a matrix weighted average of the within and between estimators, with the weights defined by the relative precision of the two estimators. We might ask, in the context of panel data: where are the interesting sources of variation? In individuals' variation around their means, or in those means themselves? The within estimator takes account of only the former, whereas the between estimator considers only the latter.

The random effects estimator

As an alternative to considering the individual-specific intercept as a "fixed effect" of that unit, we might consider that the individual effect may be viewed as a random draw from a distribution:

$$y_{it} = X_{it}^* \beta^* + [u_i + \epsilon_{it}] \qquad (7)$$

where the bracketed expression is a composite error term, with the u_i being a single draw per unit. This model could be consistently estimated by OLS or by the between estimator, but that would be inefficient in not taking the nature of the composite disturbance process into account.

A crucial assumption of this model is that u_i is independent of X^*: individual i receives a random draw that gives her a higher wage. That u_i must be independent of individual i's measurable characteristics included among the regressors X^*. If this assumption is not sustained, the random effects estimator will yield inconsistent estimates since the regressors will be correlated with the composite disturbance term.

If the individual effects can be considered to be strictly independent of the regressors, then we might model the individual-specific constant terms (reflecting the unmodeled heterogeneity across units) as draws from an independent distribution. This greatly reduces the number of parameters to be estimated, and conditional on that independence, allows for inference to be made to the population from which the survey was constructed.

In a large survey, with thousands of individuals, a random effects model will estimate K parameters, whereas a fixed effects model will estimate $(K - 1) + N$ parameters, with the sizable loss of $(N - 1)$ degrees of freedom. In contrast to fixed effects, the random effects estimator can identify the parameters on time-invariant regressors such as race or gender at the individual level.

Therefore, where its use can be warranted, the random effects model is more efficient and allows a broader range of statistical inference. The assumption of the individual effects' independence is testable and should always be tested.

To implement the one-way random effects formulation of Equation (7), we assume that both u and ϵ are meanzero processes, distributed independent of X^*; that they are each homoskedastic; that they are distributed independently of each other; and that each process represents independent realizations from its respective distribution, without correlation over individuals (nor time, for ϵ). For the T observations belonging to the ith unit of the panel, we have the composite error process

$$\eta_{it} = u_i + \epsilon_{it} \tag{8}$$

This is known as the *error components* model with conditional variance

$$E[\eta_{it}^2 | X^*] = \sigma_u^2 + \sigma_\epsilon^2 \tag{9}$$

and conditional covariance within a unit of

$$E[\eta_{it}\eta_{is} | X^*] = \sigma_u^2, \quad t \neq s. \tag{10}$$

The covariance matrix of these *T* errors may then be written as

$$\Sigma = \sigma_\epsilon^2 I_T + \sigma_u^2 \iota_T \iota_T'. \tag{11}$$

Since observations *i* and *j* are independent, the full covariance matrix of η across the sample is block-diagonal in Σ: $\Omega = I_n \otimes \Sigma$ where $\otimes$ is the Kronecker product of the matrices.

Generalized least squares (GLS) is the estimator for the slope parameters of this model:

$$
\begin{aligned}
b_{RE} &= (X^{*'}\Omega^{-1}X^*)^{-1}(X^{*'}\Omega^{-1}y) \\
&= \left(\sum_i X_i^{*'}\Sigma^{-1}X_i^*\right)^{-1}\left(\sum_i X_i^{*'}\Sigma^{-1}y_i\right) \qquad (12)
\end{aligned}
$$

To compute this estimator, we require $\Omega^{-1/2} = [I_n \otimes \Sigma]^{-1/2}$, which involves

$$
\Sigma^{-1/2} = \sigma_\epsilon^{-1}[I - T^{-1}\theta\iota_T\iota_T'] \qquad (13)
$$

where

$$
\theta = 1 - \frac{\sigma_\epsilon^2}{\sqrt{\sigma_\epsilon^2 + T\sigma_u^2}} \qquad (14)
$$

The *quasi-demeaning* transformation defined by $\Sigma^{-1/2}$ is then $\sigma_\epsilon^{-1}(y_{it} - \theta \bar{y}_i)$: that is, rather than subtracting the entire individual mean of y from each value, we should subtract some fraction of that mean, as defined by θ. Compare this to the LSDV model in which we define the within estimator by setting $\theta = 1$. Like pooled OLS, the GLS random effects estimator is a matrix weighted average of the within and between estimators, but in this case applying optimal weights, as based on

$$\lambda = \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + T\sigma_u^2} = (1 - \theta)^2 \tag{15}$$

where λ is the weight attached to the covariance matrix of the between estimator. To the extent that λ differs from unity, pooled OLS will be inefficient, as it will attach too much weight on the between-units variation, attributing it all to the variation in X rather than apportioning some of the variation to the differences in ϵ_i across units.

The setting $\lambda = 1$ ($\theta = 0$) is appropriate if $\sigma_u^2 = 0$, that is, if there are no random effects; then a pooled OLS model is optimal. If $\theta = 1$, $\lambda = 0$ and the appropriate estimator is the LSDV model of individual fixed effects. To the extent that λ differs from zero, the within (LSDV) estimator will be inefficient, in that it applies zero weight to the between estimator.

The GLS random effects estimator applies the optimal λ in the unit interval to the between estimator, whereas the fixed effects estimator arbitrarily imposes $\lambda = 0$. This would only be appropriate if the variation in ϵ was trivial in comparison with the variation in u, since then the indicator variables that identify each unit would, taken together, explain almost all of the variation in the composite error term.

To implement the feasible GLS estimator of the model all we need are consistent estimates of σ_ϵ^2 and σ_u^2. Because the fixed effects model is consistent its residuals can be used to estimate σ_ϵ^2. Likewise, the residuals from the pooled OLS model can be used to generate a consistent estimate of $(\sigma_\epsilon^2 + \sigma_u^2)$. These two estimators may be used to define θ and transform the data for the GLS model.

Because the GLS model uses quasi-demeaning, it is capable of including variables that do not vary at the individual level (such as gender or race). Since such variables cannot be included in the LSDV model, an alternative estimator must be defined based on the between estimator's consistent estimate of $(\sigma_u^2 + T^{-1}\sigma_\epsilon^2)$.

The feasible GLS estimator may be executed in Stata using the command `xtreg` with the `re` (random effects) option. The command will display estimates of σ_u^2, σ_ϵ^2 and what Stata calls `rho`: the fraction of variance due to ϵ_i. Breusch and Pagan have developed a Lagrange multiplier test for $\sigma_u^2 = 0$ which may be computed following a random-effects estimation via the command `xttest0`.

You can also estimate the parameters of the random effects model with full maximum likelihood. The `mle` option on the `xtreg, re` command requests that estimator. The application of MLE continues to assume that X^* and u are independently distributed, adding the assumption that the distributions of u and ϵ are Normal. This estimator will produce a likelihood ratio test of $\sigma_u^2 = 0$ corresponding to the Breusch–Pagan test available for the GLS estimator.

To illustrate the one-way random effects estimator and implement a test of the assumption of independence under which random effects would be appropriate and preferred, we estimate the same model in a random effects context.

```
. xtreg fatal beertax spircons unrate perincK, re
```

| Random-effects GLS regression | Number of obs | = | 336 |
| Group variable (i): state | Number of groups | = | 48 |

R-sq: within = 0.2263   Obs per group: min =   7
       between = 0.0123   avg =   7.0
       overall = 0.0042   max =   7

Random effects u_i ~ Gaussian   Wald chi2(4)   =   49.90
corr(u_i, X)   = 0 (assumed)   Prob > chi2   =   0.0000

fatal	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]
beertax	.0442768	.1204613	0.37	0.713	−.191823	.2803765
spircons	.3024711	.0642954	4.70	0.000	.1764546	.4284877
unrate	−.0491381	.0098197	−5.00	0.000	−.0683843	−.0298919
perincK	−.0110727	.0194746	−0.57	0.570	−.0492423	.0270968
_cons	2.001973	.3811247	5.25	0.000	1.254983	2.748964
sigma_u	.41675665					
sigma_e	.15678965					
rho	.87601197	(fraction of variance due to u_i)				

In comparison to the fixed effects model, where all four regressors were significant, we see that the `beertax` and `perincK` variables do not have significant effects on the fatality rate. The latter variable's coefficient switched sign.

A Hausman test may be used to test the null hypothesis that the extra orthogonality conditions imposed by the random effects estimator are valid. The fixed effects estimator, which does not impose those conditions, is consistent regardless of the independence of the individual effects. The fixed effects estimates are inefficient if that assumption of independence is warranted. The random effects estimator is efficient under the assumption of independence, but inconsistent otherwise.

Therefore, we may consider these two alternatives in the Hausman test framework, estimating both models and comparing their common coefficient estimates in a probabilistic sense. If both fixed and random effects models generate consistent point estimates of the slope parameters, they will not differ meaningfully. If the assumption of independence is violated, the inconsistent random effects estimates will differ from their fixed effects counterparts.

To implement the Hausman test, you estimate each form of the model, using the commands `estimates store` *set* after each estimation, with *set* defining that set of estimates: for instance, set might be `fix` for the fixed effects model. Then the command `hausman` *setconsist* *seteff* will invoke the Hausman test, where *setconsist* refers to the name of the fixed effects estimates (which are consistent under the null and alternative) and *seteff* referring to the name of the random effects estimates, which are only efficient under the null hypothesis of independence. This test is based on the difference of the two estimated covariance matrices (which is not guaranteed to be positive definite) and the difference between the fixed effects and random effects vectors of slope coefficients.

We illustrate the Hausman test with the two forms of the motor vehicle fatality equation:

```
. qui xtreg fatal beertax spircons unrate perincK, fe

. estimates store fix

. qui xtreg fatal beertax spircons unrate perincK, re

. hausman fix .
```

| | ──── Coefficients ──── | | | |
	(b) fix	(B) .	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
beertax	-.4840728	.0442768	-.5283495	.1090815
spircons	.8169652	.3024711	.514494	.0462668
unrate	-.0290499	-.0491381	.0200882	.
perincK	.1047103	-.0110727	.115783	.0067112

```
                       b = consistent under Ho and Ha; obtained from xtreg
          B = inconsistent under Ha, efficient under Ho; obtained from xtreg
   Test:  Ho:  difference in coefficients not systematic
            chi2(4) = (b-B)´[(V_b-V_B)^(-1)](b-B)
                    =        130.93
          Prob>chi2 =      0.0000
          (V_b-V_B is not positive definite)
```

As we might expect from the quite different point estimates generated by the random effects estimator, the Hausman test's null—that the random effects estimator is consistent—is soundly rejected. The state-level individual effects cannot be considered independent of the regressors: hardly surprising, given the wide variation in some of the regressors over states.

The first difference estimator

The within transformation used by fixed effects models removes unobserved heterogeneity at the unit level. The same can be achieved by first differencing the original equation (which removes the constant term). In fact, if $T = 2$, the fixed effects and first difference estimates are identical. For $T > 2$, the effects will not be identical, but they are both consistent estimators of the original model. Stata's `xtreg` does not provide the first difference estimator, but `xtivreg2` provides this option as the `fd` model.

The ability of first differencing to remove unobserved heterogeneity also underlies the family of estimators that have been developed for dynamic panel data (DPD) models. These models contain one or more lagged dependent variables, allowing for the modeling of a partial adjustment mechanism.

A serious difficulty arises with the one-way fixed effects model in the context of a *dynamic panel data* (DPD) model particularly in the "small T, large N" context. As Nickell (1981) shows, this arises because the demeaning process which subtracts the individual's mean value of y and each X from the respective variable creates a correlation between regressor and error.

The mean of the lagged dependent variable contains observations 0 through $(T-1)$ on y, and the mean error—which is being conceptually subtracted from each ϵ_{it}—contains contemporaneous values of ϵ for $t = 1 \ldots T$. The resulting correlation creates a bias in the estimate of the coefficient of the lagged dependent variable which is not mitigated by increasing N, the number of individual units.

The demeaning operation creates a regressor which *cannot* be distributed independently of the error term. Nickell demonstrates that the inconsistency of $\hat{\rho}$ as $N \to \infty$ is of order $1/T$, which may be quite sizable in a "small T" context. If $\rho > 0$, the bias is invariably negative, so that the persistence of y will be underestimated.

For reasonably large values of T, the limit of $(\hat{\rho} - \rho)$ as $N \to \infty$ will be approximately $-(1 + \rho)/(T - 1)$: a sizable value, even if $T = 10$. With $\rho = 0.5$, the bias will be -0.167, or about 1/3 of the true value. The inclusion of additional regressors does not remove this bias. Indeed, if the regressors are correlated with the lagged dependent variable to some degree, their coefficients may be seriously biased as well.

Note also that this bias is not caused by an autocorrelated error process ϵ. The bias arises even if the error process is *i.i.d*. If the error process is autocorrelated, the problem is even more severe given the difficulty of deriving a consistent estimate of the *AR* parameters in that context.

The same problem affects the one-way random effects model. The u_i error component enters every value of y_{it} by assumption, so that the lagged dependent variable *cannot* be independent of the composite error process.

A solution to this problem involves taking first differences of the original model. Consider a model containing a lagged dependent variable and a single regressor X:

$$y_{it} = \beta_1 + \rho y_{i,t-1} + X_{it}\beta_2 + u_i + \epsilon_{it} \tag{16}$$

The first difference transformation removes both the constant term and the individual effect:

$$\Delta y_{it} = \rho \Delta y_{i,t-1} + \Delta X_{it}\beta_2 + \Delta\epsilon_{it} \tag{17}$$

There is still correlation between the differenced lagged dependent variable and the disturbance process (which is now a first-order moving average process, or $MA(1)$): the former contains $y_{i,t-1}$ and the latter contains $\epsilon_{i,t-1}$.

But with the individual fixed effects swept out, a straightforward instrumental variables estimator is available. We may construct instruments for the lagged dependent variable from the second and third lags of y, either in the form of differences or lagged levels. If ϵ is *i.i.d.*, those lags of y will be highly correlated with the lagged dependent variable (and its difference) but uncorrelated with the composite error process.

Even if we had reason to believe that ϵ might be following an $AR(1)$ process, we could still follow this strategy, "backing off" one period and using the third and fourth lags of y (presuming that the timeseries for each unit is long enough to do so).

Dynamic panel data estimators

The *DPD* (Dynamic Panel Data) approach of Arellano and Bond (1991) is based on the notion that the instrumental variables approach noted above does not exploit all of the information available in the sample. By doing so in a Generalized Method of Moments (GMM) context, we may construct more efficient estimates of the dynamic panel data model. The Arellano–Bond estimator can be thought of as an extension of the Anderson–Hsiao estimator implemented by `xtivreg, fd`.

Arellano and Bond argue that the Anderson–Hsiao estimator, while consistent, fails to take all of the potential orthogonality conditions into account. Consider the equations

$$
\begin{aligned}
y_{it} &= X_{it}\beta_1 + W_{it}\beta_2 + v_{it} \\
v_{it} &= u_i + \epsilon_{it}
\end{aligned}
\tag{18}
$$

where X_{it} includes strictly exogenous regressors, W_{it} are predetermined regressors (which may include lags of y) and endogenous regressors, all of which may be correlated with u_i, the unobserved individual effect. First-differencing the equation removes the u_i and its associated omitted-variable bias. The Arellano–Bond estimator sets up a generalized method of moments (*GMM*) problem in which the model is specified as a system of equations, one per time period, where the instruments applicable to each equation differ (for instance, in later time periods, additional lagged values of the instruments are available).

The instruments include suitable lags of the levels of the endogenous variables (which enter the equation in differenced form) as well as the strictly exogenous regressors and any others that may be specified. This estimator can easily generate an immense number of instruments, since by period τ all lags prior to, say, $(\tau - 2)$ might be individually considered as instruments. If T is nontrivial, it is often necessary to employ the option which limits the maximum lag of an instrument to prevent the number of instruments from becoming too large. This estimator is available in Stata as `xtabond`. A more general version, allowing for autocorrelated errors, is available as `xtdpd`.

A potential weakness in the Arellano–Bond *DPD* estimator was revealed in later work by Arellano and Bover (1995) and Blundell and Bond (1998). The lagged levels are often rather poor instruments for first differenced variables, especially if the variables are close to a random walk. Their modification of the estimator includes lagged levels as well as lagged differences.

The original estimator is often entitled *difference GMM*, while the expanded estimator is commonly termed *System GMM*. The cost of the System GMM estimator involves a set of additional restrictions on the initial conditions of the process generating *y*. This estimator is available in Stata as `xtdpdsys`.

An excellent alternative to Stata's built-in commands is David Roodman's `xtabond2`, available from SSC (`findit xtabond2`). It is very well documented in his paper, referenced above. The `xtabond2` routine handles both the difference and system GMM estimators and provides several additional features—such as the orthogonal deviations transformation—not available in official Stata's commands.

As any of the DPD estimators are instrumental variables methods, it is particularly important to evaluate the Sargan–Hansen test results when they are applied. Roodman's `xtabond2` provides *C* tests (as discussed in re `ivreg2`) for groups of instruments. In his routine, instruments can be either "GMM-style" or "IV-style". The former are constructed per the Arellano–Bond logic, making use of multiple lags; the latter are included as is in the instrument matrix. For the system GMM estimator (the default in `xtabond2` instruments may be specified as applying to the differenced equations, the level equations or both.

Another important diagnostic in DPD estimation is the *AR* test for autocorrelation of the residuals. By construction, the residuals of the differenced equation should possess serial correlation, but if the assumption of serial independence in the original errors is warranted, the differenced residuals should not exhibit significant *AR*(2) behavior. These statistics are produced in the `xtabond` and `xtabond2` output. If a significant *AR*(2) statistic is encountered, the second lags of endogenous variables will not be appropriate instruments for their current values.

A useful feature of `xtabond2` is the ability to specify, for GMM-style instruments, the limits on how many lags are to be included. If *T* is fairly large (more than 7–8) an unrestricted set of lags will introduce a huge number of instruments, with a possible loss of efficiency. By using the lag limits options, you may specify, for instance, that only lags 2–5 are to be used in constructing the GMM instruments.

To illustrate the use of the *DPD* estimators, we first specify a model of `fatal` as depending on the prior year's value (`L.fatal`), the state's `spircons` and a time trend (`year`). We provide a set of instruments for that model with the `gmm` option, and list `year` as an `iv` instrument. We specify that the two-step Arellano–Bond estimator is to be employed with the Windmeijer correction. The `noleveleq` option specifies the original Arellano–Bond estimator in differences.

```
. xtabond2 fatal L.fatal spircons year, ///
>   gmmstyle(beertax spircons unrate perincK) ///
>   ivstyle(year) twostep robust noleveleq
Favoring space over speed. See help matafavor.
Warning: Number of instruments may be large relative to number of observations.
Arellano-Bond dynamic panel-data estimation, two-step difference GMM results
```

```
Group variable: state                         Number of obs      =        240
Time variable : year                          Number of groups   =         48
Number of instruments = 48                    Obs per group: min =          5
Wald chi2(3)   =      51.90                                  avg =       5.00
Prob > chi2    =      0.000                                  max =          5
```

	Coef.	Corrected Std. Err.	z	P>|z|	[95% Conf. Interval]	
fatal						
L1.	.3205569	.071963	4.45	0.000	.1795121	.4616018
spircons	.2924675	.1655214	1.77	0.077	-.0319485	.6168834
year	.0340283	.0118935	2.86	0.004	.0107175	.0573391

```
Hansen test of overid. restrictions: chi2(45) =  47.26    Prob > chi2 =   0.381
Arellano-Bond test for AR(1) in first differences: z =   -3.17  Pr > z =   0.002
Arellano-Bond test for AR(2) in first differences: z =    1.24  Pr > z =   0.216
```

This model is moderately successful in terms of relating `spircons` to the dynamics of the fatality rate. The Hansen test of overidentifying restrictions is satisfactory, as is the test for AR(2) errors. We expect to reject the test for AR(1) errors in the Arellano–Bond model.

Although the DPD estimators are linear estimators, they are highly sensitive to the particular specification of the model and its instruments. There is no substitute for experimentation with the various parameters of the specification to ensure that your results are reasonably robust to variations in the instrument set and lags used. If you are going to work with DPD models, you should study David Roodman's "How to do `xtabond2`" paper (available from `http://ideas.repec.org`) so that you fully understand the nuances of this estimation strategy.

Introduction to mixed models

Stata supports the estimation of several types of *multilevel mixed models*, also known as hierarchical models, random-coefficient models, and in the context of panel data, repeated-measures or growth-curve models. These models share the notion that individual observations are grouped in some way by the design of the data.

Mixed models are characterized as containing both fixed 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 in terms of their estimated variances and covariances. Random effects may take the form of random intercepts or random coefficients.

For instance, in hierarchical models, individual students may be associated with schools, and schools with school districts. There may be coefficients or random effects at each level of the hierarchy. Unlike traditional panel data, these data may not have a time dimension.

In repeated-measures or growth-curve models, we consider multiple observations associated with the same subject: for instance, repeated blood-pressure readings for the same patient. This may also involve a hierarchical component, as patients may be grouped by their primary care physician (PCP), and physicians may be grouped by hospitals or practices.

Linear mixed models

The simplest sort of model of this type is the *linear mixed model*, a regression model with one or more random effects. A special case of this model is the random effects panel data model implemented by `xtreg, re` which we have already discussed. If the only random coefficient is a random intercept, that command should be used to estimate the model.

For more complex models, the command `xtmixed` may be used to estimate a multilevel mixed-effects regression. Consider a dataset in which students are grouped within schools (from Rabe-Hesketh and Skrondal, 2008). We are interested in evaluating the relationship between a student's age-16 score on the GCSE exam and their age-11 score on the LRT instrument.

As the authors illustrate, we could estimate a separate regression equation for each school in which there are at least five students in the dataset:

```
. use gcse, clear
. egen nstu = count(gcse + lrt), by(school)
. statsby alpha = _b[_cons] beta = _b[lrt], by(school) ///
> saving(indivols, replace) nodots: regress gcse lrt if nstu > 4
      command:  regress gcse lrt if nstu > 4
        alpha:  _b[_cons]
         beta:  _b[lrt]
           by:  school
```

That approach gives us a set of 64 αs (intercepts) and βs (slopes) for the relationship between `gcse` and `lrt`. We can consider these estimates as data and compute their covariance matrix:

```
. use indivols, clear
(statsby: regress)

. summarize alpha beta

    Variable |        Obs        Mean    Std. Dev.        Min        Max
-------------+--------------------------------------------------------------
       alpha |         64   -.1805974    3.291357   -8.519253   6.838716
        beta |         64    .5390514    .1766135    .0380965   1.076979

. correlate alpha beta, covariance
(obs=64)

             |    alpha       beta
-------------+----------------------
       alpha |   10.833
        beta |  .208622   .031192
```

To estimate a single model, we could consider a fixed-effects approach (`xtreg, fe`), but the introduction of random intercepts and slopes for each school would lead to a regression with 130 coefficients.

Furthermore, if we consider the schools as a random sample of schools, we are not interested in the individual coefficients for each school's regression line, but rather in the mean intercept, mean slope, and the covariation in the intercepts and slopes in the population of schools.

A more sensible approach is to specify a model with a school-specific random intercept and school-specific random slope for the ith student in the jth school:

$$y_{i,j} = (\beta_1 + \delta_{1,j}) + (\beta_2 + \delta_{2,j})x_{i,j} + \epsilon_{i,j}$$

We assume that the covariate x and the idiosyncratic error ϵ are both independent of δ_1, δ_2.

The random intercept and random slope are assumed to follow a bivariate Normal distribution with covariance matrix:

$$\Psi = \begin{pmatrix} \psi_{11} & \psi_{21} \\ \psi_{21} & \psi_{22} \end{pmatrix}$$

Implying that the correlation between the random intercept and slope is

$$\rho_{12} = \frac{\psi_{21}}{\sqrt{\psi_{11}\psi_{22}}}$$

We could estimate a special case of this model, in which only the intercept contains a random component, with either `xtreg, re` or `xtmixed, mle`.

The syntax of the `xtmixed` command is

`xtmixed` *depvar fe_eqn* [`||` *re_eqn*] [`||` *re_eqn*] [, *options*]

The `fe_eqn` specifies the fixed-effects part of the model, while the `re_eqn` components optionally specify the random-effects part(s), separated by the double vertical bars (`||`). If a `re_eqn` includes only the level of a variable, it is listed followed by a colon (`:`). It may also specify a linear model including an additional *varlist*.

```
. xtmixed gcse lrt || school: if nstu > 4, mle nolog
Mixed-effects ML regression                    Number of obs      =      4057
Group variable: school                         Number of groups   =        64

                                               Obs per group: min =         8
                                                              avg =      63.4
                                                              max =       198

                                               Wald chi2(1)       =   2041.42
Log likelihood = -14018.571                    Prob > chi2        =    0.0000
```

gcse	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
lrt	.5633325	.0124681	45.18	0.000	.5388955	.5877695
_cons	.0315991	.4018891	0.08	0.937	-.7560891	.8192873

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
school: Identity				
sd(_cons)	3.042017	.3068659	2.496296	3.70704
sd(Residual)	7.52272	.0842097	7.35947	7.689592

```
LR test vs. linear regression: chibar2(01) =    403.32 Prob >= chibar2 = 0.0000
```

By specifying `gcse lrt || school:`, we indicate that the fixed-effects part of the model should include a constant term and slope coefficient for |tt lrt. The only random effect is that for `school`, which is specified as a random intercept term which varies the school's intercept around the estimated (mean) constant term.

These results display the `sd(_cons)` as the standard deviation of the random intercept term. The likelihood ratio (LR) test shown at the foot of the output indicates that the linear regression model in which a single intercept is estimated is strongly rejected by the data.

This specification restricts the school-specific regression lines to be parallel in `lrt-gcse` space. To relax that assumption, and allow each school's regression line to have its own slope, we add `lrt` to the random-effects specification. We also add the `cov(unstructured)` option, as the default is to set the covariance (ψ_{21}) and the corresponding correlation to zero.

```
. xtmixed gcse lrt || school: lrt if nstu > 4, mle nolog ///
> covariance(unstructured)
```

Mixed-effects ML regression   Number of obs   =   4057
Group variable: school   Number of groups   =   64

                                               Obs per group: min =   8
                                                              avg =   63.4
                                                              max =   198

                                               Wald chi2(1)   =   779.93
Log likelihood = -13998.423   Prob > chi2   =   0.0000

gcse	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lrt	.5567955	.0199374	27.93	0.000	.5177189	.5958721
_cons	-.1078456	.3993155	-0.27	0.787	-.8904895	.6747984

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
school: Unstructured				
sd(lrt)	.1205424	.0189867	.0885252	.1641394
sd(_cons)	3.013474	.305867	2.469851	3.676752
corr(lrt,_cons)	.497302	.1490473	.1563124	.7323728
sd(Residual)	7.442053	.0839829	7.279257	7.608491

LR test vs. linear regression: chi2(3) =   443.62   Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.

These estimates present the standard deviation of the random slope parameters (`sd(lrt))`) as well as the estimated correlation between the two random parameters (`corr(lrt,_cons)`). We can obtain the corresponding covariance matrix with `estat`:

```
. estat recovariance
Random-effects covariance matrix for level school
                         lrt        _cons
─────────────┼─────────────────────────────
        lrt  │   .0145305
      _cons  │   .1806457    9.081027
```

These estimates may be compared with those generated by school-specific regressions. As before, the likelihood ratio (LR) test of the model against the linear regression in which these three parameters are set to zero soundly rejects the linear regression model.

The dataset also contains a school-level variable, `schgend`, which is equal to 1 for schools of mixed gender, 2 for boys-only schools, and 3 for girls-only schools. We interact this qualitative factor with the continuous `lrt` model to allow both intercept and slope to differ by the type of school:

```
. xtmixed gcse c.lrt##i.schgend || school: lrt if nstu > 4, mle nolog ///
> covariance(unstructured)
```

```
Mixed-effects ML regression                    Number of obs      =       4057
Group variable: school                         Number of groups   =         64

                                               Obs per group: min =          8
                                                              avg =       63.4
                                                              max =        198

                                               Wald chi2(5)       =     804.34
Log likelihood = -13992.533                    Prob > chi2        =     0.0000
```

gcse	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
lrt	.5712245	.0271235	21.06	0.000	.5180634	.6243855
schgend						
2	.8546836	1.08629	0.79	0.431	-1.274405	2.983772
3	2.47453	.8473229	2.92	0.003	.8138071	4.135252
schgend#c.lrt						
2	-.0230016	.057385	-0.40	0.689	-.1354742	.0894709
3	-.0289542	.0447088	-0.65	0.517	-.1165818	.0586734
_cons	-.9975795	.5074132	-1.97	0.049	-1.992091	-.0030679

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
school: Unstructured				
sd(lrt)	.1198846	.0189169	.0879934	.163334
sd(_cons)	2.801682	.2895906	2.287895	3.43085
corr(lrt,_cons)	.5966466	.1383159	.2608112	.8036622
sd(Residual)	7.442949	.0839984	7.280122	7.609417

LR test vs. linear regression: chi2(3) =   381.44   Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.

The coefficients on `schgend` levels 2 and 3 indicate that girls-only schools have a significantly higher intercept than the other school types. However, the slopes for all three school types are statistically indistinguishable. Allowing for this variation in the intercept term has reduced the estimated variability of the random intercept (`sd(_cons)`).

Just as `xtmixed` can estimate multilevel mixed-effects linear regression models, `xtmelogit` can be used to estimate logistic regression models incorporating mixed effects, and `xtmepoisson` can be used for Poisson regression (count data) models with mixed effects. More complex models, such as ordinal logit models with mixed effects, can be estimated with the user-written software `gllamm` by Rabe-Hesketh and Skrondal (`ssc describe gllamm` for details).

Case study: Panel data estimation

- Access the `grunfeld` data file of firm-level characteristics:

  `use grunfeld, clear`

- describe the data:

  `describe`

- Describe the time-series calendar for the dataset:

  `tsset`

- Describe the panel dataset:

  `xtdes`

  Is this a balanced panel?

- Fit a fixed-effects model of investment and store the estimates:

  `xtreg invest mvalue kstock, fe`

  `est store fe`

  Are there significant fixed effects?

Case study: Panel data estimation

- Estimate the between estimator:

  `xtreg invest mvalue kstock, be`
  Why do the point and interval estimates differ so greatly from the fixed effects results?

- Estimate a random-effects model of investment and store the estimates:

  `xtreg invest mvalue kstock, re`

  `est store re`

- Perform the Hausman test:

  `hausman fe re`
  What do you conclude about the appropriateness of random effects?