9 Panel-data models

A panel dataset has multiple observations on the same economic units. For instance, we may have multiple observations on the same households or firms over time. In panel data, each element has two subscripts, the group identifier i and a within-group index denoted by t in econometrics, because it usually identifies time.

Given panel data, we can define several models that arise from the most general linear representation:

$$y_{it} = \sum_{k=1}^{k} x_{kit}\beta_{kit} + \epsilon_{it}, \;\; i = 1, \ldots, N, \;\; t = 1, \ldots, T \tag{9.1}$$

where N is the number of individuals and T is the number of periods.

In sections 9.1–9.3, I present methods designed for "large N, small T" panels in which there are many individuals and a few periods. These methods use the large number of individuals to construct the large-sample approximations. The small T puts limits on what can be estimated.

Assume a *balanced* panel in which there are T observations for each of the N individuals. Since this model contains $k \times N \times T$ regression coefficients, it cannot be estimated from $N \times T$ observations. We could ignore the nature of the panel data and apply pooled ordinary least squares, which would assume that $\beta = \beta_j \; \forall \; j, i, t$, but that model might be overly restrictive and can have a 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 allows for heterogeneity across panel units (and possibly across time) but confines that heterogeneity to the intercept terms of the relationship. I discuss these techniques, the *fixed-effects* (FE) and *random-effects* (RE) models, in the next section. They impose restrictions on the above model of $\beta_{jit} = \beta \, \forall i, t, \; j > 1$, thereby allowing only the constant to differ over i.

These estimation techniques can be extended to deal with endogenous regressors. The following section discusses several IV estimators that accommodate endogenous regressors. I then present the dynamic panel data (DPD) estimator, which is appropriate when lagged dependent variables are included in the set of regressors. The DPD estimator is applied to "large N, small T" panels, such as a few years of annual data on each of several hundred firms.

Section 9.4 discusses applying seemingly unrelated regression (SUR) estimators to "small N, large T" panels, in which there are a few individuals and many periods—for instance, financial variables of the 10 largest U.S. manufacturing firms, observed over the last 40 calendar quarters.

The last section of the chapter revisits the notion of moving-window estimation, demonstrating how to compute a moving-window regression for each unit of a panel.

9.1 FE and RE models

The structure represented in (9.1) 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

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta}_k + \mathbf{z}_i\boldsymbol{\delta} + u_i + \epsilon_{it} \tag{9.2}$$

where $\mathbf{x}_{it}$ is a $1 \times k$ vector of variables that vary over individual and time, $\boldsymbol{\beta}$ is the $k \times 1$ vector of coefficients on $\mathbf{x}$, $\mathbf{z}_i$ is a $1 \times p$ vector of time-invariant variables that vary only over individuals, $\boldsymbol{\delta}$ is the $p \times 1$ vector of coefficients on $\mathbf{z}$, u_i is the individual-level effect, and ϵ_{it} is the disturbance term.

The u_i are either correlated or uncorrelated with the regressors in $\mathbf{x}_{it}$ and $\mathbf{z}_i$. (The u_i are always assumed to be uncorrelated with ϵ_{it}.)

If the u_i are uncorrelated with the regressors, they are known as RE, but if the u_i are correlated with the regressors, they are known as FE. The origin of the term RE is clear: when u_i are uncorrelated with everything else in the model, the individual-level effects are simply parameterized as additional random disturbances. The sum $u_i + \epsilon_{it}$ is sometimes referred to as the composite-error term and the model is sometimes known as an error-components model. The origin of the term FE is more elusive. When the u_i are correlated with some of the regressors in the model, one estimation strategy is to treat them like parameters or FE. But simply including a parameter for every individual is not feasible, because it would imply an infinite number of parameters in our large-N, large-sample approximations. The solution is to remove the u_i from the estimation problem by a transformation that still identifies some of the coefficients of interest.

RE estimators use the assumptions that the u_i are uncorrelated with the regressors to identify the $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ coefficients. In the process of removing the u_i, FE estimators lose the ability to identify the $\boldsymbol{\delta}$ coefficients. An additional cost of using the FE estimator is that all inference is conditional on the u_i in the sample. In contrast, inference using RE estimators pertains to the population from which the RE were drawn.

We could treat a time-varying intercept term similarly, as either an FE (giving rise to an additional coefficient) or as a component of a composite-error term. We concentrate here on the one-way FE and RE models in which only the individual intercept is considered in the "large N, small T" context most commonly found in microeconomic research.[1]

9.1.1 One-way FE

The FE model modestly relaxes the assumption that the regression function is constant over time and space. A one-way FE model permits each cross-sectional unit to have its own constant term while the slope estimates ($\boldsymbol{\beta}$) are constrained across units, as is the σ_ϵ^2. This estimator is often termed the least-squares dummy variable (LSDV) model, since it is equivalent to including $N-1$ dummy variables in the OLS regression of y on $\mathbf{x}$ (including a units vector). However, the name LSDV is fraught with problems because it implies an infinite number of parameters in our estimator. A better way to understand the FE estimator is to see that removing panel-level averages from each side of (9.2) removes the FE from the model. Let $\overline{y}_i = (1/T)\sum_{t=1}^T y_{it}$, $\overline{\mathbf{x}}_i = (1/T)\sum_{t=1}^T \mathbf{x}_{it}$, and $\overline{\epsilon}_i = (1/T)\sum_{t=1}^T \epsilon_{it}$. Also note that $\mathbf{z}_i$ and u_i are panel-level averages. Then simple algebra on (9.2) implies

$$y_{it} - \overline{y}_i = (\mathbf{x}_{it} - \overline{\mathbf{x}}_i)\boldsymbol{\beta} + (\mathbf{z}_i - \mathbf{z}_i)\boldsymbol{\delta} + u_i - u_i + \epsilon_{it} - \overline{\epsilon}_i$$

which implies that

$$\widetilde{y}_{it} = (\widetilde{\mathbf{x}}_{it})\,\boldsymbol{\beta} + \widetilde{\epsilon}_{it} \tag{9.3}$$

Equation (9.3) implies that OLS on the within-transformed data will produce consistent estimates of $\boldsymbol{\beta}$. We call this estimator $\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}$. Equation (9.3) also shows that sweeping out the u_i also removes the $\boldsymbol{\delta}$. The large-sample estimator of the VCE of $\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}$ is just the standard OLS estimator of the VCE that has been adjusted for the degrees of freedom used up by the within transform

$$s^2\left(\sum_{i=1}^N\sum_{t=1}^T \widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}_{it}'\right)^{-1}$$

where $s^2 = \{1/(NT-N-k-1)\}\sum_{i=1}^N\sum_{t=1}^T \widehat{\widetilde{\epsilon}}_{it}^{\,2}$ and $\widehat{\widetilde{\epsilon}}_{it}$ are the residuals from the OLS regression of $\widetilde{y}_{it}$ on $\widetilde{\mathbf{x}}_{it}$.

This model will have explanatory power *only if* the individual's y above or below the individual's mean is significantly correlated with the individual's $\mathbf{x}$ values above or below the individual's vector of mean $\mathbf{x}$ 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 $\mathbf{x}$ values because it is only the within variation that will show up as explanatory power.[2] This outcome clearly implies that any characteristic that does not vary over time for each unit cannot be included

1. Stata's set of **xt** commands extends these panel-data models in a variety of ways. For more information, see [XT] **xt**.
2. 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.

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 $\mathbf{x}$ 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.

We can fit the one-way individual FE model with the Stata command `xtreg` by using the `fe` (FE) option. The command has a syntax similar to that of `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 fits a model in which the u_i of (9.2) are taken as deviations from one constant term, displayed as `_cons`. The empirical correlation between u_i and the fitted values is also displayed as `corr(u_i, Xb)`. The FE estimator does not require a balanced panel as long as there are at least 2 observations per unit.[3]

We wish to test whether the individual-specific heterogeneity of u_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. A rejection of this null hypothesis indicates that pooled OLS would produce inconsistent estimates. The one-way FE model also assumes that the errors are not contemporaneously correlated across units of the panel. De Hoyos and Sarafidis (2006) describe some new tests for contemporaneous correlation, and their command `xtscd` is available from SSC. Likewise, a departure from the assumed homoskedasticity of ϵ_{it} across units of the panel—that is, a form of groupwise heteroskedasticity as discussed in section 6.2.2—may be tested by an LM statistic (Greene 2003, 328), available as the author's `xttest3` routine from `ssc` (Baum 2001). `xttest3` will operate on unbalanced panels.

The example below uses 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, in thousands (`perincK`). Descriptive statistics for these variables of the `traffic.dta` dataset are given below.

3. An alternative command for this model is `areg`, which used to provide options unavailable from `xtreg`. With Stata version 9 or better, there is no advantage to using `areg`.

```
. use http://www.stata-press.com/data/imeus/traffic, clear
. xtsum fatal beertax spircons unrate perincK state year
```

Variable		Mean	Std. Dev.	Min	Max	Observations	
fatal	overall	2.040444	.5701938	.82121	4.21784	N =	336
	between		.5461407	1.110077	3.653197	n =	48
	within		.1794253	1.45556	2.962664	T =	7
beertax	overall	.513256	.4778442	.0433109	2.720764	N =	336
	between		.4789513	.0481679	2.440507	n =	48
	within		.0552203	.1415352	.7935126	T =	7
spircons	overall	1.75369	.6835745	.79	4.9	N =	336
	between		.6734649	.8614286	4.388572	n =	48
	within		.147792	1.255119	2.265119	T =	7
unrate	overall	7.346726	2.533405	2.4	18	N =	336
	between		1.953377	4.1	13.2	n =	48
	within		1.634257	4.046726	12.14673	T =	7
perincK	overall	13.88018	2.253046	9.513762	22.19345	N =	336
	between		2.122712	9.95087	19.51582	n =	48
	within		.8068546	11.43261	16.55782	T =	7
state	overall	30.1875	15.30985	1	56	N =	336
	between		15.44883	1	56	n =	48
	within		0	30.1875	30.1875	T =	7
year	overall	1985	2.002983	1982	1988	N =	336
	between		0	1985	1985	n =	48
	within		2.002983	1982	1988	T =	7

The results for the panel identifier, `state`, and time variable, `year`, illustrate the importance of the additional information provided by `xtsum`. By construction, the panel identifier `state` does not vary within the panels; i.e., it is time invariant. `xtsum` informs us of this fact by reporting that the within standard deviation is zero. Any variable with a within standard deviation of zero will be dropped from the FE model. The coefficients on variables with small within standard deviations are not well identified. The above output indicates that the coefficient on `beertax` may not be as well identified as the others. Similarly, the between standard deviation of `year` is zero by construction.

(*Continued on next page*)

The results of the one-way FE 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 having income a positive effect (as expected because driving is a normal good). The estimate of `rho` suggests that almost all the variation in `fatal` is related to interstate differences in fatality rates. The F test following the regression indicates that there are significant individual (state level) effects, implying that pooled OLS would be inappropriate.

9.1.2 Time effects and two-way FE

Stata lacks a command to automatically fit two-way FE models. If the number of periods is reasonably small, we can fit a two-way FE model by creating a set of time indicator variables and including all but one in the regression.[4] The joint test that all the coefficients on those indicator variables are zero will be a test of the significance of time FE. Just as the individual FE model requires regressors' variation over time within each unit, a time FE (implemented with a time indicator variable) requires regressors' variation over units within each period. Estimating an equation from individual or firm microdata implies that we cannot include a macrofactor such as the rate of GDP growth or price inflation in a model with a time FE because those factors do not vary across individuals. `xtsum` can be used to check that the between standard deviation is greater than zero.

4. In the context of a balanced panel, Hsiao (1986) proposes an algebraic solution involving "double demeaning", which allows estimation of a two-way FE model with no i or t indicator variables.

We consider the two-way FE 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 (as discussed in section 7.1.1) by subtracting the indicator for the excluded class from each of the other indicator variables. This transformation expresses the time effects as variations from the conditional mean of the sample rather than deviations from the excluded class (1988).

```
. quietly tabulate year, generate(yr)
. local j 0
. forvalues i=82/87 {
  2.         local ++j
  3.         rename yr`j' yr`i'
  4.         quietly 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 FE model retain their sign and significance in the two-way FE 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 much variation explained by the individual FE.

9.1.3 The between estimator

Another estimator for a panel dataset is the *between estimator*, in which the group means of y are regressed on the group means of $\mathbf{x}$ in a regression of N observations. This estimator ignores all the individual-specific variation in y that is considered by the within estimator, replacing each observation for an individual with his or her mean behavior. The between estimator is the OLS estimator of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ from the model

$$\overline{y}_i = \overline{\mathbf{x}}_i \boldsymbol{\beta} + \overline{\mathbf{z}}_i \boldsymbol{\delta} + u_i + \overline{\epsilon}_i \tag{9.4}$$

Equation (9.4) shows that if the u_i are correlated with any of the regressors in the model, the zero-conditional-mean assumption does not hold and the between estimator will produce inconsistent results.

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 1 observation per unit. We could do so explicitly with Stata's `collapse` command, which would generate a new dataset of that nature (see section 3.3). However, you need not form that dataset to use the between estimator because the command `xtreg` with the `be` (between) option will invoke it. Using the between estimator requires that $N > k$. Any macro factor that is constant over individuals cannot be included in the between estimator because 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. With panel data, we can identify whether the interesting sources of variation are 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.

To show why we account for all the information present in the panel, we refit the first model above with the between estimator (the second model, containing year FE, is not appropriate, since the time dimension is suppressed by the between estimator). Interestingly, two of the factors that played an important role in the one- and two-way FE model, `beertax` and `unrate`, play no significant role in this regression on group (state) means.

```
. xtreg fatal beertax spircons unrate perincK, be
Between regression (regression on group means)  Number of obs      =       336
Group variable (i): state                       Number of groups   =        48
R-sq:  within  = 0.0479                         Obs per group: min =         7
       between = 0.4565                                        avg =       7.0
       overall = 0.2583                                        max =         7
                                                F(4,43)            =      9.03
sd(u_i + avg(e_i.))=  .4209489                  Prob > F           =    0.0000

       fatal       Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]

     beertax    .0740362   .1456333     0.51   0.614    -.2196614    .3677338
    spircons    .2997517   .1128135     2.66   0.011     .0722417    .5272618
      unrate    .0322333    .038005     0.85   0.401    -.0444111    .1088776
     perincK   -.1841747   .0422241    -4.36   0.000    -.2693277   -.0990218
       _cons    3.796343   .7502025     5.06   0.000     2.283415    5.309271
```

9.1.4 One-way RE

Rather than considering the individual-specific intercept as an FE of that unit, the RE model specifies the individual effect as a random draw that is uncorrelated with the regressors and the overall disturbance term

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{z}_i\boldsymbol{\delta} + (u_i + \epsilon_{it}) \tag{9.5}$$

where $(u_i + \epsilon_{it})$ is a composite error term and the u_i are the individual effects. A crucial assumption of this model is that the u_i are uncorrelated with the regressors $\mathbf{x}_{it}$ and $\mathbf{z}_i$. This orthogonality assumption implies that the parameters can be consistently estimated by OLS and the between estimator, but neither of these estimators is efficient. The RE estimator uses the assumption that the u_i are uncorrelated with regressors to construct a more efficient estimator. If the regressors are correlated with the u_i, they are correlated with the composite error term and the RE estimator is inconsistent.

The RE model uses the orthogonality between the u_i and the regressors to greatly reduce the number of estimated parameters. In a large survey, with thousands of individuals, an RE model has $k+p$ coefficients and two variance parameters, whereas an FE model has $k-1+N$ coefficients and one variance parameter. The coefficients on time-invariant variables are identified in the RE model. Because the RE model identifies the population parameter that describes the individual-level heterogeneity, inference from the RE model pertains to the underlying population of individuals. In contrast, because the FE model cannot estimate the parameters that describe the individual-level heterogeneity, inference from the FE model is conditional on the FE in the sample. Therefore, the RE model is more efficient and allows a broader range of statistical inference. The key assumption that the u_i are uncorrelated with the regressors can and should be tested.

To implement the one-way RE formulation of (9.5), we assume that both u and ϵ are mean-zero processes, uncorrelated with the regressors; that they are each homoskedastic; that they are uncorrelated with each other; and that there is no correlation over individuals or time. For the T observations belonging to the ith unit of the panel, the composite error process

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

gives rise to the *error-components* model with conditional variance

$$E[\eta_{it}^2|\mathbf{x}^*] = \sigma_u^2 + \sigma_\epsilon^2$$

and conditional covariance within a unit of

$$E[\eta_{it}\eta_{is}|\mathbf{x}^*] = \sigma_u^2, \quad t \neq s$$

The covariance matrix of these T errors can then be written as

$$\Sigma = \sigma_\epsilon^2 I_T + \sigma_u^2 \iota_T \iota_T'$$

Since observations i and j are uncorrelated, the full covariance matrix of η across the sample is block diagonal in $\mathbf{\Sigma}$: $\mathbf{\Omega} = \mathbf{I}_n \otimes \mathbf{\Sigma}$.[5,6]

The GLS estimator for the slope parameters of this model is

$$\begin{aligned}\widehat{\beta}_{\text{RE}} &= (\mathbf{X}^{*'}\mathbf{\Omega}^{-1}\mathbf{X}^*)^{-1}(\mathbf{X}^{*'}\mathbf{\Omega}^{-1}\mathbf{y}) \\ &= \left(\sum_i \mathbf{X}_i^{*'}\mathbf{\Sigma}^{-1}\mathbf{X}_i^*\right)^{-1}\left(\sum_i \mathbf{X}_i^{*'}\mathbf{\Sigma}^{-1}\mathbf{y}_i\right)\end{aligned}$$

To compute this estimator, we require $\mathbf{\Omega}^{-1/2} = (\mathbf{I}_n \otimes \mathbf{\Sigma})^{-1/2}$, which involves

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

where

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

and the *quasidemeaning* transformation defined by $\mathbf{\Sigma}^{-1/2}$ is then $\sigma_\epsilon^{-1}(y_{it} - \theta\overline{y}_i)$; 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. The quasidemeaning transformation reduces to the within transformation when $\theta = 1$. Like pooled OLS, the GLS RE estimator is a matrix-weighted average of the within and between estimators, but we apply optimal weights, as based on

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

5. The operator $\otimes$ denotes the Kronecker product of the two matrices. For any matrices $\mathbf{A}_{K\times L}, \mathbf{B}_{M\times N}$, $\mathbf{A}\otimes\mathbf{B} = \mathbf{C}_{KM\times LN}$. To form the product matrix, each element of $\mathbf{A}$ scalar multiplies the entire matrix $\mathbf{B}$. See Greene (2003, 824–825).

6. I give the expressions for a balanced panel. Unbalanced panels merely complicate the algebra.

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 $\mathbf{x}$ 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 RE, then a pooled OLS model is optimal. If $\theta = 1$, $\lambda = 0$ and the FE estimator is appropriate. To the extent that λ differs from zero, the FE estimator will be inefficient, in that it applies zero weight to the between estimator. The GLS RE estimator applies the optimal λ in the unit interval to the between estimator, whereas the FE estimator arbitrarily imposes $\lambda = 0$. This imposition would be appropriate only if the variation in ϵ was trivial in comparison with the variation in u.

To implement the FGLS estimator of the model, all we need are consistent estimates of σ_ϵ^2 and σ_u^2. Because the FE 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 estimate θ and transform the data for the GLS model.[7] Because the GLS model uses quasidemeaning, it can include time-invariant variables (such as gender or race).

The FGLS estimator may be executed in Stata by using the command `xtreg` with the `re` (RE) option. The command will display estimates of σ_u^2, σ_ϵ^2, and what Stata calls `rho`: the fraction of the total variance due to ϵ_i. Breusch and Pagan (1980) have developed a Lagrange multiplier test for $\sigma_u^2 = 0$, which may be computed following an RE estimation via the command `xttest0` (see [XT] **xtreg** for details).

We can also estimate the parameters of the RE model with full maximum likelihood. Typing `xtreg, mle` requests that estimator. The application of maximum likelihood estimation continues to assume that the regressors and u are uncorrelated, 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 RE estimator and implement a test of the orthogonality assumption under which RE is appropriate and preferred, we estimate the parameters of the RE model that corresponds to the FE model above.

7. A possible complication: as generally defined, the two estimators above are not guaranteed to generate a positive estimate of σ_ϵ^2 in finite samples. Then the variance estimates without degrees-of-freedom corrections, which will still be consistent, may be used.

```
. 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)
------------------------------------------------------------------------------
```

Compared with the FE 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.

9.1.5 Testing the appropriateness of RE

We can use a Hausman test (presented in section 8.11) to test the null hypothesis that the extra orthogonality conditions imposed by the RE estimator are valid. If the regressors are correlated with the u_i, the FE estimator is consistent but the RE estimator is not consistent. If the regressors are uncorrelated with the u_i, the FE estimator is still consistent, albeit inefficient, whereas the RE estimator is consistent and efficient. Therefore, we may consider these two alternatives in the Hausman test framework, fitting both models and comparing their common coefficient estimates in a probabilistic sense. If both FE and RE models generate consistent point estimates of the slope parameters, they will not differ meaningfully. If the orthogonality assumption is violated, the inconsistent RE estimates will significantly differ from their FE counterparts.

To implement the Hausman test, we fit each model and store its results by typing `estimates store` *set* after each estimation (*set* defines that set of estimates: for instance, *set* might be `fix` for the FE model). Then typing `hausman` *setconsist seteff* will invoke the Hausman test, where *setconsist* refers to the name of the FE estimates (which are consistent under the null and alternative) and *seteff* refers to the name of the RE estimates, which are consistent and efficient only under the null hypothesis. This test uses the difference of the two estimated covariance matrices (which is not guaranteed to be positive definite) to weight the difference between the FE and RE vectors of slope coefficients.

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

```
. quietly xtreg fatal beertax spircons unrate perincK, fe
. estimates store fix
. quietly xtreg fatal beertax spircons unrate perincK, re
. estimates store ran
. hausman fix ran
                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |      fix          ran         Difference          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 different point estimates generated by the RE estimator, the Hausman test's null hypothesis—that the RE estimator is consistent—is soundly rejected. The state-level individual effects do appear to be correlated with the regressors.[8]

9.1.6 Prediction from one-way FE and RE

Following `xtreg`, the `predict` command may be used to generate a variety of series. The default result is `xb`, the linear prediction of the model. Stata normalizes the unit-specific effects (whether fixed or random) as deviations from the intercept term `_cons`; therefore, the `xb` prediction ignores the individual effect. We can generate predictions that include the RE or FE by specifying the `xbu` option; the individual effect itself may be predicted with option `u`;[9] and the ϵ_{it} error component (or "true" residual) may be predicted with option `e`. The three last predictions are available only in sample for either the FE or RE model, whereas the linear prediction `xb` and the "combined residual" (option `ue`) by default will be computed out of sample as well, just as with predictions from `regress`.

8. Here Stata signals that the difference of the estimated VCEs is not positive definite.
9. Estimates of u_i are not consistent with $N \rightarrow \infty$ and fixed T.

9.2 IV models for panel data

If the Hausman test indicates that the RE u_i cannot be considered orthogonal to the individual-level error, an IV estimator may be used to generate consistent estimates of the coefficients on the time-invariant variables. The Hausman–Taylor estimator (Hausman and Taylor 1981) assumes that some of the regressors in $\mathbf{x}_{it}$ and $\mathbf{z}_i$ are correlated with u but that none are correlated with ϵ. This estimator is available in Stata as `xthtaylor`. This approach begins by writing (9.2) as

$$y_{it} = \mathbf{x}_{1,it}\boldsymbol{\beta}_1 + \mathbf{x}_{2,it}\boldsymbol{\beta}_2 + \mathbf{z}_{1,i}\boldsymbol{\delta}_1 + \mathbf{z}_{2,i}\boldsymbol{\delta}_2 + u_i + \epsilon_{it}$$

where the $\mathbf{x}$ variables are time varying, the $\mathbf{z}$ variables are time invariant, the variables subscripted with a "1" are exogenous, and the variables subscripted with a "2" are correlated with the u_i. Identifying the parameters requires that k_1 (the number of $\mathbf{x}_{1,it}$ variables) be at least as large as ℓ_2 (the number of $\mathbf{z}_{2,i}$ variables). Applying the Hausman–Taylor estimator circumvents the problem that the $\mathbf{x}_{2,it}$ and $\mathbf{z}_{2,i}$ variables are correlated with u_i, but it requires that we find variables that are not correlated with the individual-level effect.

Stata also provides an IV estimator for the FE and RE models in which some of the $\mathbf{x}_{it}$ and $\mathbf{z}_i$ variables are correlated with the disturbance term ϵ_{it}. These are different assumptions about the nature of any suspected correlation between the regressor and the composite error term from those underlying the Hausman–Taylor estimator. The `xtivreg` command offers FE, RE, between-effects, and first-differenced IV estimators in a panel-data context.

9.3 Dynamic panel-data models

A serious difficulty arises with the one-way FE model in the context of a dynamic panel-data (DPD) model, one containing a lagged dependent variable (and possibly other regressors), particularly in the "small T, large N" context. As Nickell (1981) shows, this problem arises because the within-transform N, the lagged dependent variable, is correlated with the error term. As Nickell (1981) shows, the resulting correlation creates a large-sample bias in the estimate of the coefficient of the lagged dependent variable, which is not mitigated by increasing N, the number of individual units. In the simplest setup of a pure AR(1) model without additional regressors:

$$\begin{aligned} y_{it} &= \beta + \rho y_{i,t-1} + u_i + \epsilon_{it} \\ y_{it} - \overline{y}_{ix} &= \rho(y_{i,t-1} - \overline{L.y_i}) + (\epsilon_{it} - \epsilon_{i\cdot}) \end{aligned}$$

$\overline{L.y_i}$ is correlated with $(\epsilon_{it}-\epsilon_i)$ by definition. Nickell demonstrates that the inconsistency of $\widehat{\rho}$ as $N \to \infty$ is of order $1/T$, which may be sizable in a "small T" context. If $\rho > 0$, the bias is invariably negative, so the persistence of y will be underestimated. For reasonably large values of T, the limit of $(\widehat{\rho} - \rho)$ as $N \to \infty$ will be approximately $-(1+\rho)/(T-1)$, which is a sizable value. With $T = 10$ and $\rho = 0.5$, the bias will be -0.167, or about 1/3 of the true value. Including more regressors does not remove this bias. If the regressors are correlated with the lagged dependent variable to some degree, their coefficients may be seriously biased as well. This bias is not caused by an autocorrelation in the error process ϵ and 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 RE 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 regressor $\mathbf{x}$:

$$y_{it} = \beta_1 + \rho y_{i,t-1} + \mathbf{x}_{it}\boldsymbol{\beta}_2 + u_i + \epsilon_{it}$$

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

$$\Delta y_{it} = \rho \Delta y_{i,t-1} + \Delta \mathbf{x}_{it}\boldsymbol{\beta}_2 + \Delta\epsilon_{it}$$

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 FE swept out, a straightforward IV 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.[10] Even if we believed 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 time series for each unit is long enough to do so).

The DPD approach of Arellano and Bond (1991) is based on the notion that the IV approach noted above does not exploit all the information available in the sample. By doing so in a GMM context, we can construct more efficient estimates of the DPD model. The Arellano–Bond estimator can be thought of as an extension to the Anderson–Hsiao estimator implemented by `xtivreg, fd`. Arellano and Bond argue that the Anderson–Hsiao estimator, although consistent, fails to take all the potential orthogonality conditions into account. Consider the equations

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}\boldsymbol{\beta}_1 + \mathbf{w}_{it}\boldsymbol{\beta}_2 + v_{it} \\ v_{it} &= u_i + \epsilon_{it} \end{aligned}$$

where $\mathbf{x}_{it}$ includes strictly exogenous regressors and $\mathbf{w}_{it}$ 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 begins by specifying the model as a system of equations, one per period, and allows the instruments

10. The degree to which these instruments are not weak depends on the true value of ρ. See Arellano and Bover (1995) and Blundell and Bond (1998).

applicable to each equation to differ (for instance, in later periods, more 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 a great many instruments, since by period τ all lags prior to, say, $(\tau - 2)$ might be individually considered as instruments. If T is nontrivial, we may need to use the option that 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` (see [XT] **xtabond**).

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, whereas 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.

Both the difference GMM and system GMM estimators have one-step and two-step variants. The two-step estimates of the difference GMM standard errors have been shown to have a severe downward bias. To evaluate the precision of the two-step estimators for hypothesis tests, we should apply the "Windmeijer finite-sample correction" (see Windmeijer 2005) to these standard errors. Bond (2002) provides an excellent guide to the DPD estimators.

All the features described above are available in David Roodman's improved version of official Stata's estimator. His version, `xtabond2`, offers a much more flexible syntax than official Stata's `xtabond`, which does not allow the same specification of instrument sets, nor does it provide the system GMM approach or the Windmeijer correction to the standard errors of the two-step estimates. On the other hand, Stata's `xtabond` has a simpler syntax and is faster, so you may prefer to use it.

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 be used with the Windmeijer correction. The `noleveleq` option specifies the original Arellano–Bond estimator in differences:[11]

```
. use http://www.stata-press.com/data/imeus/traffic, clear
. tsset
       panel variable:  state, 1 to 56
        time variable:  year, 1982 to 1988
```

11. The estimated parameters of the difference GMM model do not include a constant term because it is differenced out.

```
. xtabond2 fatal L.fatal spircons year,
> gmmstyle(beertax spircons unrate perincK)
> ivstyle(year) twostep robust noleveleq
Favoring space over speed. To switch, type or click on mata: mata set matafavor
>  speed.
Warning: Number of instruments may be large relative to number of observations.
Suggested rule of thumb: keep number of instruments <= number of groups.
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
------------------------------------------------------------------------------
             |              Corrected
             |      Coef.   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(82) =  47.26    Prob > chi2 =  0.999
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 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.

To contrast the difference GMM and system GMM approaches, we use the latter estimator by dropping the `noleveleq` option:

(*Continued on next page*)

```
. xtabond2 fatal L.fatal spircons year,
> gmmstyle(beertax spircons unrate perincK) ivstyle(year) twostep robust
Favoring space over speed. To switch, type or click on mata: mata set matafavor
> speed.
Warning: Number of instruments may be large relative to number of observations.
Suggested rule of thumb: keep number of instruments <= number of groups.
Arellano-Bond dynamic panel-data estimation, two-step system GMM results

Group variable: state                          Number of obs      =       288
Time variable : year                           Number of groups   =        48
Number of instruments = 48                     Obs per group: min =         6
Wald chi2(3)  =   1336.50                                     avg =      6.00
Prob > chi2   =     0.000                                     max =         6

                           Corrected
                  Coef.    Std. Err.      z    P>|z|     [95% Conf. Interval]

fatal
         L1.   .8670531   .0272624    31.80   0.000     .8136198    .9204865
    spircons  -.0333786   .0166285    -2.01   0.045    -.0659697   -.0007874
        year   .0135718   .0051791     2.62   0.009     .0034209    .0237226
       _cons  -26.62532   10.27954    -2.59   0.010    -46.77285   -6.477799

Hansen test of overid. restrictions: chi2(110) =  44.26   Prob > chi2 =  1.000
Arellano-Bond test for AR(1) in first differences: z =  -3.71  Pr > z =  0.000
Arellano-Bond test for AR(2) in first differences: z =   1.77  Pr > z =  0.077
```

Although the other summary measures from this estimator are acceptable, the marginally significant negative coefficient on `spircons` casts doubt on this specification.

9.4 Seemingly unrelated regression models

Often we want to estimate a similar specification for several different units, a production function or cost function for each industry. If the equation to be estimated for a given unit meets the zero-conditional-mean assumption of (4.2), we can estimate each equation independently. However, we may want to estimate the equations jointly: first, to allow cross-equation restrictions to be imposed or tested, and second, to gain efficiency, since we might expect the error terms across equations to be *contemporaneously correlated*. Such equations are often called *seemingly unrelated regressions* (SURs), and Zellner (1962) proposed an estimator for this problem: the SUR estimator. Unlike the FE and RE estimators, whose large-sample justification is based on "small T, large N" datasets in which $N \to \infty$, the SUR estimator is based on the large-sample properties of "large T, small N" datasets in which $T \to \infty$, so it may be considered a multiple time-series estimator.

Equation i of the SUR model is

$$y_i = \mathbf{x}_i \boldsymbol{\beta}_i + \epsilon_i, \;\; i = 1, \ldots, N$$

where y_i is the ith equation's dependent variable and $\mathbf{X}_i$ is the $T \times k_i$ matrix of observations on the regressors for the ith equation. The disturbance process $\boldsymbol{\epsilon} = (\epsilon'_1, \epsilon'_2, \ldots, \epsilon'_N)'$ is assumed to have an expectation of zero and an $NT \times NT$ covariance matrix of $\boldsymbol{\Omega}$. We will consider only the case where we have T observations per equation, although we could fit the model with an unbalanced panel. Each equation may have a differing set of regressors, and apart from the constant term, there may be no variables in common across the $\mathbf{x}_i$. Applying SUR requires that the T observations per unit exceed N, the number of units, to render $\boldsymbol{\Omega}$ of full rank and invertible. If this constraint is not satisfied, we cannot use SUR. In practice, T should be much larger than N for the large-sample approximations to work well.

We assume that $E[\epsilon_{it}\epsilon_{js}] = \sigma_{ij},\ t = s$, and otherwise zero, which implies that we are allowing for the error terms in different equations to be *contemporaneously correlated*, but assuming that they are not correlated at other points (including within a unit: they are assumed independent). Thus for any two error vectors,

$$\begin{aligned} E[\epsilon_i \epsilon'_j] &= \sigma_{ij}\mathbf{I}_T \\ \boldsymbol{\Omega} &= \boldsymbol{\Sigma} \otimes \mathbf{I}_T \end{aligned}$$

where $\boldsymbol{\Sigma}$ is the $N \times N$ covariance matrix of the N error vectors and $\otimes$ is the Kronecker matrix product.

The efficient estimator for this problem is GLS, in which we can write $\mathbf{y}$ as the stacked set of $\mathbf{y}_i$ vectors and $\mathbf{X}$ as the block-diagonal matrix of $\mathbf{X}_i$. Since the GLS estimator is

$$\widehat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y})$$

and

$$\boldsymbol{\Omega}^{-1} = \boldsymbol{\Sigma}^{-1} \otimes \mathbf{I} \quad .$$

We can write the (infeasible) GLS estimator as

$$\widehat{\boldsymbol{\beta}}_{\text{GLS}} = \{\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{X}\}^{-1}\{\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{y}\}$$

which if expanded demonstrates that each block of the $\mathbf{X}'_i\mathbf{X}_j$ matrix is weighted by the scalar σ_{ij}^{-1}. The large-sample VCE of $\widehat{\boldsymbol{\beta}}_{\text{GLS}}$ is the first term of this expression.

When will this estimator provide a gain in efficiency over equation-by-equation OLS? First, if the $\sigma_{ij},\ i \neq j$ are actually zero, there is no gain. Second, if the $\mathbf{X}_i$ matrices are identical across equations—not merely having the same variable names, but containing the same numerical values—GLS is identical to equation-by-equation OLS, and there is no gain. Beyond these cases, the gain in efficiency depends on the magnitude of the cross-equation contemporaneous correlations of the residuals. The higher those correlations are, the greater the gain will be. Furthermore, if the $\mathbf{X}_i$ matrices' columns are highly correlated across equations, the gains will be smaller.

The feasible SUR estimator requires a consistent estimate of $\boldsymbol{\Sigma}$, the $N \times N$ contemporaneous covariance matrix of the equations' disturbance processes. We can estimate the representative element σ_{ij}, the contemporaneous correlation between ϵ_i, ϵ_j, from equation-by-equation OLS residuals as

$$s_{ij} = \frac{e_i' e_j}{T}$$

assuming that each unit's equation is estimated from T observations.[12] We use these estimates to perform the "Zellner step", where the algebra of partitioned matrices will show that the Kronecker products may be rewritten as products of the blocks in the expression for $\widehat{\beta}_{\text{GLS}}$. The estimator may be iterated. The GLS estimates will produce a new set of residuals, which may be used in a second Zellner step, and so on. Iteration will make the GLS estimates equivalent to maximum likelihood estimates of the system.

The SUR estimator is available in Stata via the `sureg` command; see [R] **sureg**. SUR can be applied to panel-data models in the wide format.[13] SUR is a more attractive estimator than pooled OLS, or even FE, in that SUR allows each unit to have its own coefficient vector.[14] Not only does the constant term differ from unit to unit, but each of the slope parameters and σ_ϵ^2 differ across units. In contrast, the slope and variance parameters are constrained to be equal across units in pooled OLS, FE, or RE estimators. We can use standard F tests to compare the unrestricted SUR results with those that may be generated in the presence of linear constraints, such as cross-equation restrictions (see [R] **constraint**). Cross-equation constraints correspond to the restriction that a particular regressor's effect is the same for each panel unit. We can use the `isure` option to iterate the estimates, as described above.

We can test whether applying SUR has yielded a significant gain in efficiency by using a test for the diagonality of $\mathbf{\Sigma}$ proposed by Breusch and Pagan (1980).[15] Their LM statistic sums the squared correlations between residual vectors $\mathbf{i}$ and $\mathbf{j}$, with a null hypothesis of diagonality (zero contemporaneous covariance between the errors of different equations). This test is produced by `sureg` when the `corr` option is specified.

We apply SUR to detrended annual output and factor input prices of five U.S. industries (SIC codes 32–35) for 1958–1996, stored in the wide format.[16] The descriptive statistics of the price series are given below.

```
. use http://www.stata-press.com/data/imeus/4klem_wide_defl, clear
(35KLEM: Jorgensen industry sector data)
. tsset
          time variable:  year, 1958 to 1996
```

12. A degrees-of-freedom correction could be used in the denominator, but relying on large-sample properties, it is not warranted.
13. If the data are set up in the long format more commonly used with panel data, the `reshape` command (see [D] **reshape**) may be used to place them in the wide format; see section 3.8.
14. See [XT] **xtgls** for a SUR estimator that imposes a common coefficient vector on a panel-data model.
15. This test should not be confused with these authors' test for heteroskedasticity described in section 6.2.1.
16. The price series have been detrended with a cubic polynomial time trend.

```
. summarize *d year, sep(5)
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       pi32d |        39     .611359      .02581    .566742   .6751782
       pk32d |        39    .7335128    .0587348   .5981754    .840534
       pl32d |        39    .5444872    .0198763   .4976022   .5784216
       pe32d |        39    .5592308    .0786871   .4531953   .7390293
       pm32d |        39    .5499744    .0166443   .5171617   .5823871
-------------+--------------------------------------------------------
       pi33d |        39    .4948205    .0149315   .4624915   .5163859
       pk33d |        39    .5190769     .035114   .4277323   .5760419
       pl33d |        39    .5200256    .0424153   .4325826   .6127931
       pe33d |        39    .5706154     .093766   .4387668   .8175654
       pm33d |        39    .5192564    .0151137   .4870717   .5421571
-------------+--------------------------------------------------------
       pi34d |        39    .5013333    .0178689   .4659021   .5258276
       pk34d |        39    .5157692    .0558735    .377311   .6376742
       pl34d |        39    .5073077    .0169301    .468933   .5492905
       pe34d |        39    .5774359    .0974223   .4349643   .8020797
       pm34d |        39    .5440256    .0180344   .5070866   .5773573
-------------+--------------------------------------------------------
       pi35d |        39    .5159487    .0168748   .4821945   .5484785
       pk35d |        39    .7182051    .1315394    .423117   1.061852
       pl35d |        39    .4984872    .0216141   .4493805   .5516838
       pe35d |        39    .5629231    .0865252   .4476493   .7584586
       pm35d |        39    .5684615    .0234541   .5317762   .6334837
-------------+--------------------------------------------------------
        year |        39        1977    11.40175       1958       1996
```

We regress each industry's output price on its lagged value and four factor input prices: those for capital (k), labor (l), energy (e), and materials (m). The sureg command requires the specification of each equation in parentheses. We build up the equations' specification by using a forvalues loop over the industry codes.

```
. forvalues i=32/35 {
  2.         local eqn "`eqn' (pi`i'd L.pi`i'd pk`i'd pl`i'd pe`i'd pm`i'd) "
  3. }
. sureg `eqn', corr
Seemingly unrelated regression
----------------------------------------------------------------------
Equation          Obs  Parms        RMSE    "R-sq"       chi2        P
----------------------------------------------------------------------
pi32d              38      5    .0098142    0.8492     219.14   0.0000
pi33d              38      5    .0027985    0.9615    1043.58   0.0000
pi34d              38      5    .0030355    0.9677    1182.37   0.0000
pi35d              38      5    .0092102    0.6751      78.10   0.0000
----------------------------------------------------------------------
```

		Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
pi32d							
	pi32d						
	L1.	-.0053176	.1623386	-0.03	0.974	-.3234953	.3128602
	pk32d	-.0188711	.0344315	-0.55	0.584	-.0863556	.0486133
	pl32d	-.5575705	.1166238	-4.78	0.000	-.786149	-.328992
	pe32d	.0402698	.0592351	0.68	0.497	-.0758289	.1563684
	pm32d	1.587711	.3252302	4.88	0.000	.9502717	2.225151
	_cons	.0362004	.1104716	0.33	0.743	-.1803199	.2527208
pi33d							
	pi33d						
	L1.	.1627936	.0495681	3.28	0.001	.065642	.2599453
	pk33d	-.0199381	.0250173	-0.80	0.425	-.0689712	.0290949
	pl33d	-.0655277	.0225466	-2.91	0.004	-.1097181	-.0213372
	pe33d	-.0657604	.008287	-7.94	0.000	-.0820027	-.0495181
	pm33d	1.133285	.084572	13.40	0.000	.9675273	1.299043
	_cons	-.0923547	.0185494	-4.98	0.000	-.1287109	-.0559985
pi34d							
	pi34d						
	L1.	.3146301	.0462574	6.80	0.000	.2239673	.405293
	pk34d	.0137423	.009935	1.38	0.167	-.0057298	.0332145
	pl34d	.0513415	.0373337	1.38	0.169	-.0218312	.1245142
	pe34d	-.0483202	.0115829	-4.17	0.000	-.0710222	-.0256182
	pm34d	.8680835	.0783476	11.08	0.000	.7145251	1.021642
	_cons	-.1338766	.0241593	-5.54	0.000	-.1812279	-.0865252
pi35d							
	pi35d						
	L1.	.2084134	.1231019	1.69	0.090	-.0328619	.4496887
	pk35d	-.0499452	.0125305	-3.99	0.000	-.0745046	-.0253858
	pl35d	.0129142	.0847428	0.15	0.879	-.1531786	.179007
	pe35d	.1071003	.0641549	1.67	0.095	-.018641	.2328415
	pm35d	.0619171	.2051799	0.30	0.763	-.3402282	.4640624
	_cons	.3427017	.1482904	2.31	0.021	.0520579	.6333454

```
Correlation matrix of residuals:

         pi32d     pi33d     pi34d     pi35d
pi32d   1.0000
pi33d  -0.3909    1.0000
pi34d  -0.2311    0.2225    1.0000
pi35d  -0.1614   -0.1419    0.1238    1.0000

Breusch-Pagan test of independence: chi2(6) =    12.057, Pr = 0.0607
```

The summary output indicates that each equation explains almost all the variation in the industry's output price. The `corr` option displays the estimated VCE of residuals and tests for independence of the residual vectors. Sizable correlations—both positive and negative—appear in the correlation matrix, and the Breusch–Pagan test rejects its null of independence of these residual series at the 10% level.

We can test cross-equation constraints in the sureg framework with test, combining multiple hypotheses as expressions in parentheses. We consider the null hypothesis that each industry's coefficient on the energy price index is the same.

```
. test ([pi32d]pe32d = [pi33d]pe33d) ([pi32d]pe32d = [pi34d]pe34d)
>      ([pi32d]pe32d = [pi35d]pe35d)
 ( 1)  [pi32d]pe32d - [pi33d]pe33d = 0
 ( 2)  [pi32d]pe32d - [pi34d]pe34d = 0
 ( 3)  [pi32d]pe32d - [pi35d]pe35d = 0
            chi2(  3) =   11.38
          Prob > chi2 =    0.0098
```

The joint test decisively rejects these equality constraints. To illustrate using constrained estimation with sureg, we impose the restriction that the coefficient on the energy price index should be identical over industries. This test involves the definition of three constraints on the coefficient vector. Imposing constraints cannot improve the fit of each equation but may be warranted if the data accept the restriction.

```
. constraint define 1 [pi32d]pe32d = [pi33d]pe33d
. constraint define 2 [pi32d]pe32d = [pi34d]pe34d
. constraint define 3 [pi32d]pe32d = [pi35d]pe35d
. sureg 'eqn', notable c(1 2 3)
Seemingly unrelated regression
Constraints:
 ( 1)  [pi32d]pe32d - [pi33d]pe33d = 0
 ( 2)  [pi32d]pe32d - [pi34d]pe34d = 0
 ( 3)  [pi32d]pe32d - [pi35d]pe35d = 0
```

Equation	Obs	Parms	RMSE	"R-sq"	chi2	P
pi32d	38	5	.0098793	0.8472	236.78	0.0000
pi33d	38	5	.0029664	0.9567	719.32	0.0000
pi34d	38	5	.0030594	0.9672	1212.12	0.0000
pi35d	38	5	.0101484	0.6055	110.37	0.0000

These constraints considerably increase the RMSE (or Root MSE) values for each equation, as we would expect from the results of the test command.

9.4.1 SUR with identical regressors

The second case discussed above, in which SUR will generate the same point and interval estimates—the case of numerically identical regressors—arises often in economic theory and financial theory. For instance, the demand for each good should depend on the set of prices and income, or the portfolio share of assets held in a given class should depend on the returns to each asset and on total wealth. Here there is no reason to use anything other than OLS for efficiency. However, SUR estimation is often used in this case because it allows us to test cross-equation constraints or to estimate with those constraints in place.

If we try to apply SUR to a system with adding-up constraints, such as a *complete set* of cost share or portfolio share equations, the SUR estimator will fail because the error covariance matrix is singular. This assertion holds not only for the unobservable errors but also for the least-squares residuals. A bit of algebra will show that if there are adding-up constraints across equations—for instance, if the set of y_i variables is a complete set of portfolio shares or demand shares—the OLS residuals will sum to zero across equations, and their empirical covariance matrix will be singular *by construction*.

We may still want to use systems estimation to impose the cross-equation constraints arising from economic theory. Here we drop one of the equations and estimate the system of $N-1$ equations with SUR. The parameters of the Nth equation, in point and interval form, can be algebraically derived from those estimates. The FGLS estimates will be sensitive to which equation is dropped, but iterated SUR will restore the invariance property of the maximum likelihood estimator of the problem. For more details, see Greene (2003, 362–369). Poi (2002) shows how to fit singular systems of nonlinear equations.

9.5 Moving-window regression estimates

As with `mvsumm` and `mvcorr` (discussed in section 3.5.3), we may want to compute moving-window regression estimates in a panel context. As with `mvsumm`, we can compute regression estimates for nonoverlapping subsamples with Stata's `statsby` command. However, that command cannot deal with overlapping subsamples, as that would correspond to the same observation's being a member of several by-groups. The functionality to compute moving-window regression estimates is available from the author's `rollreg` routine, available from `ssc`.

With a moving-window regression routine, how should we design the window? One obvious scheme would mimic `mvsumm` and allow for a window of fixed width that is to be passed through the sample, one period at a time: the `move(#)` option.[17] In other applications, we may want an "expanding window": that is, starting with the first τ periods, we compute a set of estimates that consider observations $1 \ldots (\tau + 1)$, $1 \ldots (\tau + 2)$, and so on. This sort of window corresponds to the notion of the information set available to an economic agent at a point in time (and to the scheme used to generate instruments in a DPD model; see [XT] **xtabond**). Thus `rollreg` also offers that functionality via its `add(`τ`)` option. For completeness, the routine also offers the `drop(`τ`)` option, which implements a window that initially takes into account the last τ periods and then expands the window back toward the beginning of the sample. This sort of moving-window estimate can help us determine the usefulness of past information in generating an ex ante forecast, using more or less of that information in the computation. We must use one of these three options when executing `rollreg`.

17. One could imagine something like a 12-month window that is to be advanced to end-of-quarter months, but that could be achieved by merely discarding the intermediate window estimates from `rollreg`.

A moving-window regression will generate sequences of results corresponding to each estimation period. A Stata routine could store those sequences in the columns of a matrix (which perhaps makes them easier to present in tabular format) or as additional variables in the current dataset (which perhaps makes them easier to include in computations or in graphical presentations using `tsline`). The latter, on balance, seems handier and is implemented in `rollreg` via the mandatory `stub(`*string*`)` option, which specifies that new variables should be created with names beginning with *string*.

All the features of `rollreg` (including built-in graphics with the `graph()` option) are accessible with panel data when applied to one time series within the panel by using an `if` *exp* or `in` *range* qualifier. However, rolling regressions certainly have their uses with a full panel. For instance, a finance researcher may want to calculate a "CAPM beta" for each firm in a panel using a moving window of observations, simulating the information set used by the investor at each point in time. Therefore, `rollreg` has been designed to operate with panels where the same sequence of rolling regressions is computed for each time series within the panel.[18] In this context, the routine's graphical output is not available. Although `rollreg` does not produce graphics when multiple time series are included from a panel, it is easy to generate graphics using the results left behind by the routine. For example,

```
. use http://www.stata-press.com/data/imeus/invest2, clear
. keep if company<5
(20 observations deleted)
. tsset company time
       panel variable:  company, 1 to 4
        time variable:  time, 1 to 20
. rollreg market L(0/1).invest time, move(8) stub(mktM)
. local dv `r(depvar)'
. local rl `r(reglist)'
. local stub `r(stub)'
. local wantcoef invest
. local m "`r(rolloption)'(`r(rollobs)')"
. generate fullsample = .
(80 missing values generated)
. forvalues i = 1/4 {
  2.        qui regress `dv' `rl' if company==`i'
  3.        qui replace fullsample = _b[`wantcoef'] if company==`i' & time > 8
  4. }
. label var `stub'_`wantcoef' "moving beta"
. xtline `stub'_`wantcoef', saving("`wantcoef'.gph",replace)
> byopts(title(Moving coefficient of market on invest)
> subtitle("Full-sample coefficient displayed") yrescale legend(off))
> addplot(line fullsample time if fullsample < .)
(file invest.gph saved)
```

Here an 8-year moving window is used to generate the regression estimates of a model where the firm's market value is regressed on current and once-lagged investment ex-

18. I thank Todd Prono for suggesting that this feature be added to the routine.

penditures and a time trend. The trajectory of the resulting coefficient for current investment expenditures is graphed in figure 9.1 for each firm.

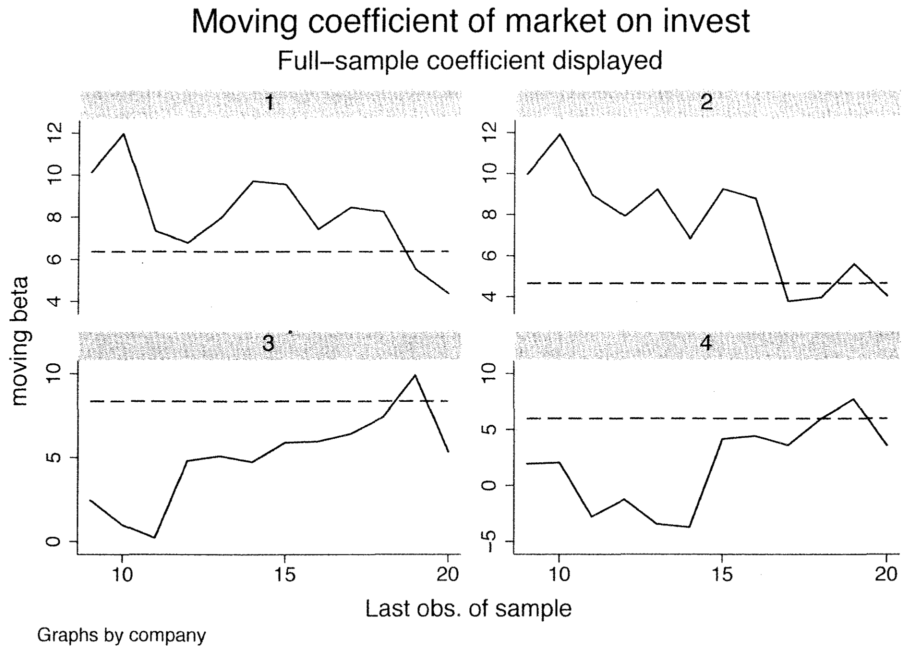

Figure 9.1: Moving-window regression estimates

Companies 1 and 2 display broadly similar trajectories, as do companies 3 and 4; the second pair is different from the first pair. A clear understanding of the temporal stability of the coefficient estimates is perhaps more readily obtained graphically. Although they are not displayed on this graph, `rollreg` also creates series of coefficients' standard errors, from which we can compute confidence intervals, as well as the `Root MSE` of the equation and its R^2.

Or we could use Stata's `rolling` prefix to specify that the moving-window regression be run over each firm.[19] Below we save the estimated coefficients (`_b`) in a new dataset, which we may then merge with the original dataset for further analysis or producing graphics.

```
. use http://www.stata-press.com/data/imeus/invest2, clear
. keep if company<5
(20 observations deleted)
```

19. The `add` and `drop` options of `rollreg` are available using the `rolling` prefix as options `recursive` and `rrecursive`, respectively.

```
. tsset company time
       panel variable:  company, 1 to 4
        time variable:  time, 1 to 20

. rolling _b, window(8) saving(roll_invest, replace) nodots:
> regress market L(0/1).invest time
file roll_invest.dta saved

. use http://www.stata-press.com/data/imeus/roll_invest, clear
(rolling: regress)

. tsset company start
       panel variable:  company, 1 to 4
        time variable:  start, 1 to 13

. describe

Contains data from roll_invest.dta
  obs:            52                          rolling: regress
 vars:             7                          9 Jun 2006 14:08
 size:         1,664 (99.8% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
company         float  %9.0g
start           float  %9.0g
end             float  %9.0g
_b_invest       float  %9.0g                  _b[invest]
_stat_2         float  %9.0g                  _b[L.invest]
_b_time         float  %9.0g                  _b[time]
_b_cons         float  %9.0g                  _b[_cons]
-------------------------------------------------------------------------------
Sorted by:  company  start
```

We could produce a graph of each firm's moving coefficient estimate for `invest` with the commands

```
. label var _b_invest "moving beta"
. xtline _b_invest,  byopts(title(Moving coefficient of market on invest))
```

using the `roll_invest` dataset produced by `rolling`.

Exercises

1. The `cigconsump` dataset contains 48 states' annual data for 1985–1995. Fit an FE model of demand for cigarettes, `packpc`, as a function of price (`avgprs`) and per capita income (`incpc`). What are the expected signs? Are they borne out by the estimates? If not, how might you explain the estimated coefficients? Can you reject the pooled OLS model of demand?
2. Store the estimates from the FE model, and refit the model with RE. How do these estimates compare? Does a Hausman test accept RE as the more appropriate estimator?
3. Refit the FE model in constant-elasticity form by using `lpackpc`, `lavgprs`, and `lincpc`. How do the results compare to those on the levels variables? Is this form of the model more in line with economic theory?

4. Refit the constant-elasticity form of the model as a dynamic model, including `L.packpc` as a regressor. Use the two-step robust DPD estimator of `xtabond2` with `lpackpc` as a GMM instrument and `year, L.avgprs` as IV instruments. Do the results support the dynamic formulation of the model? Is the model more in line with economic theory than the static form? Is it adequate for the test of overidentifying restrictions and second-order serial correlation?

5. The `cigconsumpNE` dataset contains the log demand, price, and per capita income variables for the six New England states in wide format. Use that dataset to fit the constant-elasticity form of the model for the six states as a seemingly unrelated regression model with `sureg`. Are there meaningful correlations across the equations' residuals? How do the results differ state by state?