Panel data management, estimation and forecasting

Christopher F Baum

Boston College and DIW Berlin

IMF Institute, Spring 2011

Forms of panel data

To define the problems of panel 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 reshape 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.

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:

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

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^2_ϵ. 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 $s^2[X'M_D X]^{-1}$, 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, or the nature of a country as landlocked. 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

Results of the one-way fixed effects model:

```
. 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).

Note the empirical correlation labeled `corr(u_i, Xb)` of -0.8804. This correlation indicates that the unobserved heterogeneity term, proxied by the estimated fixed effect, is strongly correlated with a linear combination of the included regressors. That is not a problem for the fixed effects model, but as we shall see it is an important magnitude.

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.

In Stata 11, that is very easy to do using factor variables (e.g., `i.year`). Previously, it could be achieved with the `xi` command. The joint significance of those variables may be assessed with `testparm`.

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. Rather than using factor variables, these 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 in cross-country studies 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.

We illustrate with the traffic fatality dataset:

```
. 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

Note that cross-sectional (interstate) variation in `beertax` and `unrate` has no explanatory power in this specification, whereas they are highly significant when the within estimator is employed.

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:

$$y_{it} = X^*_{it}\beta^* + [u_i + \epsilon_{it}] \tag{7}$$

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 can 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 i^{th} 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:

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

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

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

where

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

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 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.

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 labels `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 official 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.

The `corr(u_i, Xb)` in this context is assumed to be zero: a necessary condition for the random effects estimator to yield consistent estimates. Recall that when the fixed effect estimator was used, this correlation was reported as -0.8804.

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.

The command `hausman` *setconsist seteff* will then 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 sizable estimated correlation reported in the fixed effects estimator also supports this rejection.

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 Mark Schaffer's `xtivreg2` from SSC provides this option as the `fd` model.

We illustrate the first difference estimator with the traffic data set.

```
. xtivreg2 fatal beertax spircons unrate perincK, fd nocons small
FIRST DIFFERENCES ESTIMATION
----------------------------

Number of groups =        48                    Obs per group: min =          6
                                                               avg =        6.0
                                                               max =          6

OLS estimation
--------------

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

                                                      Number of obs =       288
                                                      F(  4,   284) =      6.29
                                                      Prob > F      =    0.0001
Total (centered) SS     =  11.21286023                Centered R2   =    0.0812
Total (uncentered) SS   =  11.21590589                Uncentered R2 =    0.0814
Residual SS             =  10.30276586                Root MSE      =     .1905

------------------------------------------------------------------------------
     D.fatal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |
         D1. |   .1187701   .2728036     0.44   0.664    -.4182035    .6557438
    spircons |
         D1. |    .523584   .1408249     3.72   0.000     .2463911     .800777
      unrate |
         D1. |    .003399   .0117009     0.29   0.772    -.0196325    .0264304
     perincK |
         D1. |   .1417981   .0372814     3.80   0.000     .0684152     .215181
------------------------------------------------------------------------------

Included instruments: D.beertax D.spircons D.unrate D.perincK
```

We may note that, as in the between estimation results, the `beertax` and `unrate` variables have lost their significance. The larger Root MSE for the `fd` equation, compared to that for `fe`, illustrates the relative inefficiency of the first difference estimator when there are more than two time periods.

The seemingly unrelated regression estimator

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, or annual data on the G7 countries for the last 30 years.

The SURE technique (implemented in Stata as `sureg`) requires that the number of time periods exceeds the number of cross-sectional units.

The concept of 'seemingly unrelated' regressions is that we have several panel units, for which we could separately estimate proper OLS equations: that is, there is no simultaneity linking the units' equations. The units might be firms operating in the same industry, or industries in a particular economy, or countries in the same region.

We might be interested in estimating these equations jointly in order to take account of the likely correlation, across equations, of their error terms. These correlations represent common shocks. Incorporating those correlations in the estimation can provide gains in efficiency.

The SURE model is considerably more flexible than the fixed-effect model for panel data, as it allows for coefficients that may differ across units (but may be tested, or constrained to be identical) as well as separate estimates of the error variance for each equation. In fact, the regressor list for each equation may differ: for a particular country, for example, the price of an important export commodity might appear, but only in that country's equation. To use `sureg`, your data must be stored in the 'wide' format: the same variable for different units must be named for that unit.

Its limitation, as mentioned above, is that it cannot be applied to models in which $N > T$, as that will imply that the residual covariance matrix is singular. SURE is a generalized least squares (GLS) technique which makes use of the inverse of that covariance matrix.

A limitation of official Stata's `sureg` command is that it can only deal with balanced panels. This may be problematic in the case of firm-level or country-level data where firms are formed, or merged, or liquidated during the sample period, or when new countries emerge, as in Eastern Europe.

I wrote an extended version of `sureg`, named `suregub`, which will handle SURE in the case of unbalanced panels as long as the degree of imbalance is not too severe: that is, there must be some time periods in common across panel units. A copy of `suregub` has been provided in your materials.

One special case of note: if the equations contain exactly the same regressors (that is, numerically identical), SURE results will exactly reproduce equation-by-equation OLS results. This situation is likely to arise when you are working with a set of demand equations (for goods or factors) or a set of portfolio shares, wherein the explanatory variables should be the same for each equation.

Although SURE will provide no efficiency gain in this setting, you may still want to employ the technique on such a set of equations, as by estimating them as a system you gain the ability to perform hypothesis tests across equations, or estimate them subject to a set of linear constraints. The `sureg` command supports linear constraints, defined in the same manner as single-equation `cnsreg`.

We illustrate `sureg` with a macro example using the Penn World Tables (v6.3) dataset, `pwt6_3`. For simplicity, we choose three countries from that dataset: Spain, Italy, and Greece for 1960–2007. Our 'model' considers the consumption share of real GDP per capita (`kc`) as a function of openness (`openc`) and the lagged ratio of GNP to GDP (`cgnp`).

```
. // keep three countries for 1960-, reshape to wide for sureg
. use pwt6_3, clear
(Penn World Tables 6.3, August 2009)
. keep if inlist(isocode, "ITA", "ESP", "GRC")
(10846 observations deleted)
. keep isocode year kc openc cgnp
. keep if year >= 1960
(30 observations deleted)
. levelsof isocode, local(ctylist)
`"ESP"´ `"GRC"´ `"ITA"´
. reshape wide kc openc cgnp, i(year) j(isocode) string
(note: j = ESP GRC ITA)
Data                               long   ->   wide
-----------------------------------------------------------------------------
Number of obs.                      144   ->      48
Number of variables                   5   ->      10
j variable (3 values)           isocode   ->   (dropped)
xij variables:
                                     kc   ->   kcESP kcGRC kcITA
                                  openc   ->   opencESP opencGRC opencITA
                                   cgnp   ->   cgnpESP cgnpGRC cgnpITA
-----------------------------------------------------------------------------
. tsset year, yearly
        time variable:  year, 1960 to 2007
                delta:  1 year
```

We build up a list of equations for `sureg` using the list of country codes created by `levelsof`:

```
 . // build up list of equations for sureg
. loc eqns
. foreach c of local ctylist {
  2.         loc eqns "`eqns´ (kc`c´ openc`c´ L.cgnp`c´)"
  3. }
. display "`eqns´"
 (kcESP opencESP L.cgnpESP) (kcGRC opencGRC L.cgnpGRC) (kcITA opencITA L.cgnpIT
> A)
```

```
. sureg "`eqns´", corr
Seemingly unrelated regression
```

Equation	Obs	Parms	RMSE	"R-sq"	chi2	P
kcESP	47	2	.9379665	0.6934	104.50	0.0000
kcGRC	47	2	4.910707	0.3676	40.29	0.0000
kcITA	47	2	1.521322	0.4051	45.56	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
kcESP						
opencESP	-.1205816	.012307	-9.80	0.000	-.1447028	-.0964603
cgnpESP L1.	-.97201	.373548	-2.60	0.009	-1.704151	-.2398694
_cons	157.6905	37.225	4.24	0.000	84.73086	230.6502
kcGRC						
opencGRC	.4215421	.0670958	6.28	0.000	.2900367	.5530476
cgnpGRC L1.	.5918787	.5900844	1.00	0.316	-.5646655	1.748423
_cons	-16.48375	60.74346	-0.27	0.786	-135.5387	102.5712

(continued)

kcITA						
opencITA	.0684288	.0269877	2.54	0.011	.0155339	.1213237
cgnpITA						
L1.	-1.594811	.3426602	-4.65	0.000	-2.266412	-.923209
_cons	211.6658	34.58681	6.12	0.000	143.8769	279.4547

Correlation matrix of residuals:

	kcESP	kcGRC	kcITA
kcESP	1.0000		
kcGRC	-0.2367	1.0000	
kcITA	-0.0786	-0.2618	1.0000

Breusch-Pagan test of independence: chi2(3) = 6.145, Pr = 0.1048

Note from the displayed correlation matrix of residuals and the Breusch–Pagan test of independence that there is weak evidence of cross-equation correlation of the residuals.

Given our systems estimates, we may test hypotheses on coefficients in different equations: for instance, that the coefficients on `openc` are equal across equations. Note that in the `test` command we must specify in which equation each coefficient appears.

```
. // test cross-equation hypothesis of coefficient equality
. test [kcESP]opencESP = [kcGRC]opencGRC = [kcITA]opencITA
 ( 1)  [kcESP]opencESP - [kcGRC]opencGRC = 0
 ( 2)  [kcESP]opencESP - [kcITA]opencITA = 0
           chi2(  2) =  100.55
         Prob > chi2 =    0.0000
```

We can produce *ex post* or *ex ante* forecasts from `sureg` with `predict`, specifying a different variable name for each equation's predictions:

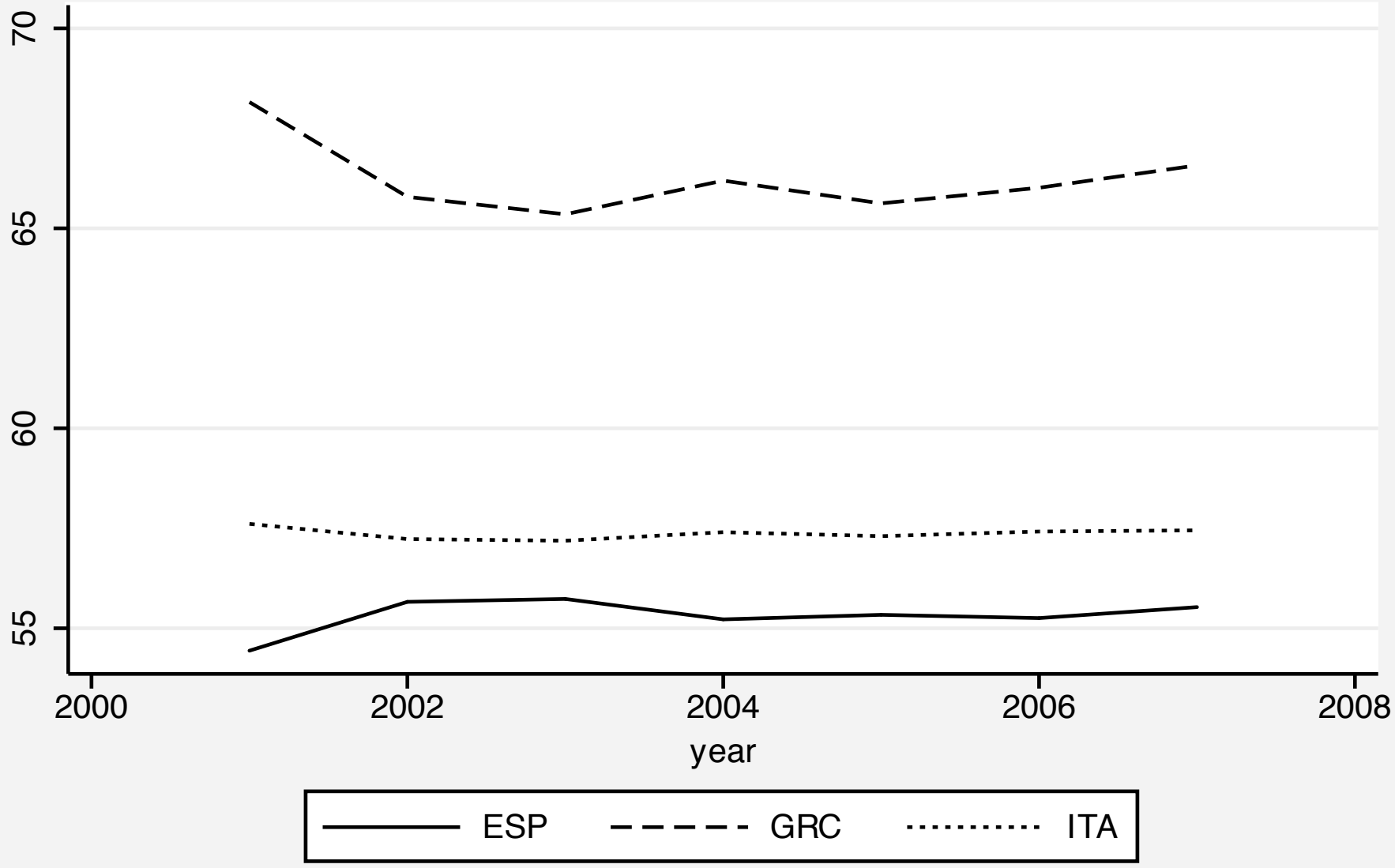
```
. sureg "`eqns´" if year <= 2000, notable
Seemingly unrelated regression
```

Equation	Obs	Parms	RMSE	"R-sq"	chi2	P
kcESP	40	2	.985171	0.5472	48.72	0.0000
kcGRC	40	2	5.274077	0.3076	27.49	0.0000
kcITA	40	2	1.590656	0.4364	42.14	0.0000

```
. foreach c of local ctylist {
  2.          predict double `c´hat if year > 2000, xb equation(kc`c´)
  3.          label var `c´hat "`c´"
  4. }
(41 missing values generated)
(41 missing values generated)
(41 missing values generated)
. su *hat if year > 2000
```

Variable	Obs	Mean	Std. Dev.	Min	Max
ESPhat	7	55.31007	.4318259	54.43892	55.7324
GRChat	7	66.24322	.932017	65.35107	68.15631
ITAhat	7	57.37146	.1436187	57.18819	57.60937

```
. tsline *hat if year>2000, scheme(s2mono) legend(rows(1)) ///
> ti("Predicted consumption share, real GDP per capita") t2("ex ante prediction
> s")
```


Predicted consumption share, real GDP per capita
ex ante predictions
70
65
60
55
2000
2002
2004
2006
2008
year
ESP
GRC
ITA

Instrumental variables estimators for panel data

Linear instrumental variables (IV) models for panel data may be estimated with Stata's `xtivreg`, a panel-capable analog to `ivregress`. This command only fits standard two-stage least squares models, and does not support IV-GMM nor LIML. By specifying options, you may choose among the random effects (`re`), fixed effects (`fe`), between effects (`be`) and first-differenced (`fd`) estimators.

If you want to use IV-GMM or LIML in a panel setting, you may use Mark Schaffer's `xtivreg2` routine, which is a 'wrapper' for Baum–Schaffer–Stillman's `ivreg2`, providing all of its capabilities in a panel setting. However, `xtivreg2` only implements the fixed-effects and first-difference estimators.

We spoke in an earlier lecture about the use of *cluster-robust standard errors*: a specification of the error term's VCE in which we allow for arbitrary correlation within *M* clusters of observations. Most Stata commands, including `regress, ivregress` and `xtreg`, support the option of `vce(cluster` *varname)* to produce the cluster-robust VCE.

In fact, if you use `xtreg, fe` with the `robust` option, the VCE estimates are generated as cluster-robust, as Stock and Watson demonstrated (*Econometrica*, 2008) that it is necessary to allow for clustering to generate a consistent robust VCE when $T > 2$.

However, Stata's `xtivreg` does not implement the `cluster` option, although the construction of a cluster-robust VCE in an IV setting is appropriate analytically.

To circumvent this limitation, you may use `xtivreg2` to estimate fixed-effects or first-difference IV models with cluster-robust standard errors. In a panel context, you may also want to consider *two-way clustering*: the notion that dependence between observations' errors may not only appear within the time series observations of a given panel unit, but could also appear across units at each point in time.

The extension of cluster-robust VCE estimates to two- and multi-way clustering is an area of active econometric research. Please see the Baum–Nichols–Schaffer slides (UKSUG10) in your materials for an overview.

Computation of the two-way cluster-robust VCE is straightforward, as Thompson (SSRN WP, 2006) illustrates. The VCE may be calculated from

$$VCE(\hat{\beta}) = VCE_1(\hat{\beta}) + VCE_2(\hat{\beta}) - VCE_{12}(\hat{\beta})$$

where the three VCE estimates are derived from one-way clustering on the first dimension, the second dimension and their intersection, respectively. As these one-way cluster-robust VCE estimates are available from most Stata estimation commands, computing the two-way cluster-robust VCE involves only a few matrix manipulations.

One concern that arises with two-way (and multi-way) clustering is the number of clusters in each dimension. With one-way clustering, we should be concerned if the number of clusters G is too small to produce unbiased estimates. The theory underlying two-way clustering relies on asymptotics in the smaller number of clusters: that is, the dimension containing fewer clusters. The two-way clustering approach is thus most sensible if there are a sizable number of clusters in each dimension.

We illustrate with a fixed-effect IV model of `kc` from the Penn World Tables data set, in which regressors are again specified as `openc` and `cgnp`, each instrumented with two lags. The model is estimated for an unbalanced panel of 99 countries for 38–46 years per country. We fit the model with classical standard errors (IID), cluster-robust by country (clCty) and cluster-robust by country and year (clCtyYr).

Table: Panel IV estimates of kc, 1960-2007

	(1) IID	(2) clCty	(3) clCtyYr
openc	-0.036***	-0.036*	-0.036*
	(0.007)	(0.018)	(0.018)
cgnp	0.800***	0.800***	0.800***
	(0.033)	(0.146)	(0.146)
N	4508	4508	4508

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The two-way cluster-robust standard errors are very similar to those produced by the one-way cluster-robust VCE. Both sets are considerably larger than those produced by the *i.i.d.* error assumption, suggesting that classical standard errors are severely biased in this setting.

Dynamic panel data estimators

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 (*Econometrica*, 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 $N \rightarrow \infty$ 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 $N \rightarrow \infty$ 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.

One 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, included in your materials. 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 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 using the `traffic` data, 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.

We also illustrate DPD estimation using the Penn World Table cross-country panel. We specify a model for `kc` depending on its own lag, `cgnp`, and a set of time fixed effects, which we compute with the `xi` command, as `xtabond2` does not support factor variables. We first estimate the two-step 'difference GMM' form of the model with (cluster-)robust VCE, using data for 1991–2007. We could use `testparm _I*` after estimation to evaluate the joint significance of time effects (listing of which has been suppressed).

```
. xi i.year
i.year            _Iyear_1991-2007    (naturally coded; _Iyear_1991 omitted)
. xtabond2 kc L.kc cgnp _I*, gmm(L.kc openc cgnp, lag(2 9)) iv(_I*) ///
> twostep robust noleveleq nodiffsargan
Favoring speed over space. To switch, type or click on mata: mata set matafavor
>  space, perm.
Dynamic panel-data estimation, two-step difference GMM
------------------------------------------------------------------------------

Group variable: iso                             Number of obs      =      1485
Time variable : year                            Number of groups   =        99
Number of instruments = 283                     Obs per group: min =        15
Wald chi2(17) =     94.96                                      avg =     15.00
Prob > chi2   =     0.000                                      max =        15
------------------------------------------------------------------------------
             |              Corrected
          kc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          kc |
         L1. |   .6478636   .1041122     6.22   0.000     .4438075    .8519197
             |
        cgnp |    .233404   .1080771     2.16   0.031     .0215768    .4452312
...
-------------+----------------------------------------------------------------
```

(continued)

```
Instruments for first differences equation
  Standard
    D.(_Iyear_1992 _Iyear_1993 _Iyear_1994 _Iyear_1995 _Iyear_1996 _Iyear_1997
    _Iyear_1998 _Iyear_1999 _Iyear_2000 _Iyear_2001 _Iyear_2002 _Iyear_2003
    _Iyear_2004 _Iyear_2005 _Iyear_2006 _Iyear_2007)
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    L(2/9).(L.kc openc cgnp)
------------------------------------------------------------------------------
Arellano-Bond test for AR(1) in first differences: z =  -2.94  Pr > z =  0.003
Arellano-Bond test for AR(2) in first differences: z =   0.23  Pr > z =  0.815
------------------------------------------------------------------------------
Sargan test of overid. restrictions: chi2(266)  = 465.53  Prob > chi2 =  0.000
  (Not robust, but not weakened by many instruments.)
Hansen test of overid. restrictions: chi2(266)  =  87.81  Prob > chi2 =  1.000
  (Robust, but can be weakened by many instruments.)
```

Given the relatively large number of time periods available, I have specified that the GMM instruments only be constructed for lags 2–9 to keep the number of instruments manageable. I am treating `openc` as a GMM-style instrument. The autoregressive coefficient is 0.648, and the `cgnp` coefficient is positive and significant. Although not shown, the test for joint significance of the time effects has p-value 0.0270.

We could also fit this model with the 'system GMM' estimator, which will be able to utilize one more observation per country in the level equation, and estimate a constant term in the relationship. I am treating lagged `openc` as a IV-style instrument in this specification.

```
. xtabond2 kc L.kc cgnp _I*, gmm(L.kc cgnp, lag(2 8)) iv(_I* L.openc) ///
> twostep robust nodiffsargan
Dynamic panel-data estimation, two-step system GMM
```

Group variable: iso		Number of obs	= 1584
Time variable : year		Number of groups	= 99
Number of instruments = 207		Obs per group: min	= 16
Wald chi2(17) = 8193.54		avg	= 16.00
Prob > chi2 = 0.000		max	= 16

kc	Coef.	Corrected Std. Err.	z	P>|z|	[95% Conf.	Interval]
kc						
L1.	.9452696	.0191167	49.45	0.000	.9078014	.9827377
cgnp	.097109	.0436338	2.23	0.026	.0115882	.1826297
...						
_cons	-6.091674	3.45096	-1.77	0.078	-12.85543	.672083

(continued)

```
------------------------------------------------------------------------------
Instruments for first differences equation
  Standard
    D.(_Iyear_1992 _Iyear_1993 _Iyear_1994 _Iyear_1995 _Iyear_1996 _Iyear_1997
    _Iyear_1998 _Iyear_1999 _Iyear_2000 _Iyear_2001 _Iyear_2002 _Iyear_2003
    _Iyear_2004 _Iyear_2005 _Iyear_2006 _Iyear_2007 L.openc)
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    L(2/8).(L.kc cgnp)
Instruments for levels equation
  Standard
    _cons
    _Iyear_1992 _Iyear_1993 _Iyear_1994 _Iyear_1995 _Iyear_1996 _Iyear_1997
    _Iyear_1998 _Iyear_1999 _Iyear_2000 _Iyear_2001 _Iyear_2002 _Iyear_2003
    _Iyear_2004 _Iyear_2005 _Iyear_2006 _Iyear_2007 L.openc
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    DL.(L.kc cgnp)
------------------------------------------------------------------------------
Arellano-Bond test for AR(1) in first differences: z =  -3.29  Pr > z =  0.001
Arellano-Bond test for AR(2) in first differences: z =   0.42  Pr > z =  0.677
------------------------------------------------------------------------------
Sargan test of overid. restrictions: chi2(189)  = 353.99  Prob > chi2 =  0.000
  (Not robust, but not weakened by many instruments.)
Hansen test of overid. restrictions: chi2(189)  =  88.59  Prob > chi2 =  1.000
  (Robust, but can be weakened by many instruments.)
```

Note that the autoregressive coefficient is much larger: 0.945 in this context. The `cgnp` coefficient is again positive and significant, but has a much smaller magnitude when the system GMM estimator is used.

We can also estimate the model using the forward orthogonal deviations (FOD) transformation of Arellano and Bover, as described in Roodman's paper. The first-difference transformation applied in DPD estimators has the unfortunate feature of magnifying any gaps in the data, as one period of missing data is replaced with two missing differences. FOD transforms each observation by subtracting the average of all *future* observations, which will be defined (regardless of gaps) for all but the last observation in each panel. To illustrate:

```
. xtabond2 kc L.kc cgnp _I*, gmm(L.kc cgnp, lag(2 8)) iv(_I* L.openc) ///
> twostep robust nodiffsargan orthog
Dynamic panel-data estimation, two-step system GMM
```

Group variable: iso		Number of obs	= 1584
Time variable : year		Number of groups	= 99
Number of instruments = 207		Obs per group: min	= 16
Wald chi2(17) = 8904.24		avg	= 16.00
Prob > chi2 = 0.000		max	= 16

kc	Coef.	Corrected Std. Err.	z	P>|z|	[95% Conf.	Interval]
kc						
L1.	.9550247	.0142928	66.82	0.000	.9270114	.983038
cgnp	.0723786	.0339312	2.13	0.033	.0058746	.1388825
...						
_cons	-4.329945	2.947738	-1.47	0.142	-10.10741	1.447515

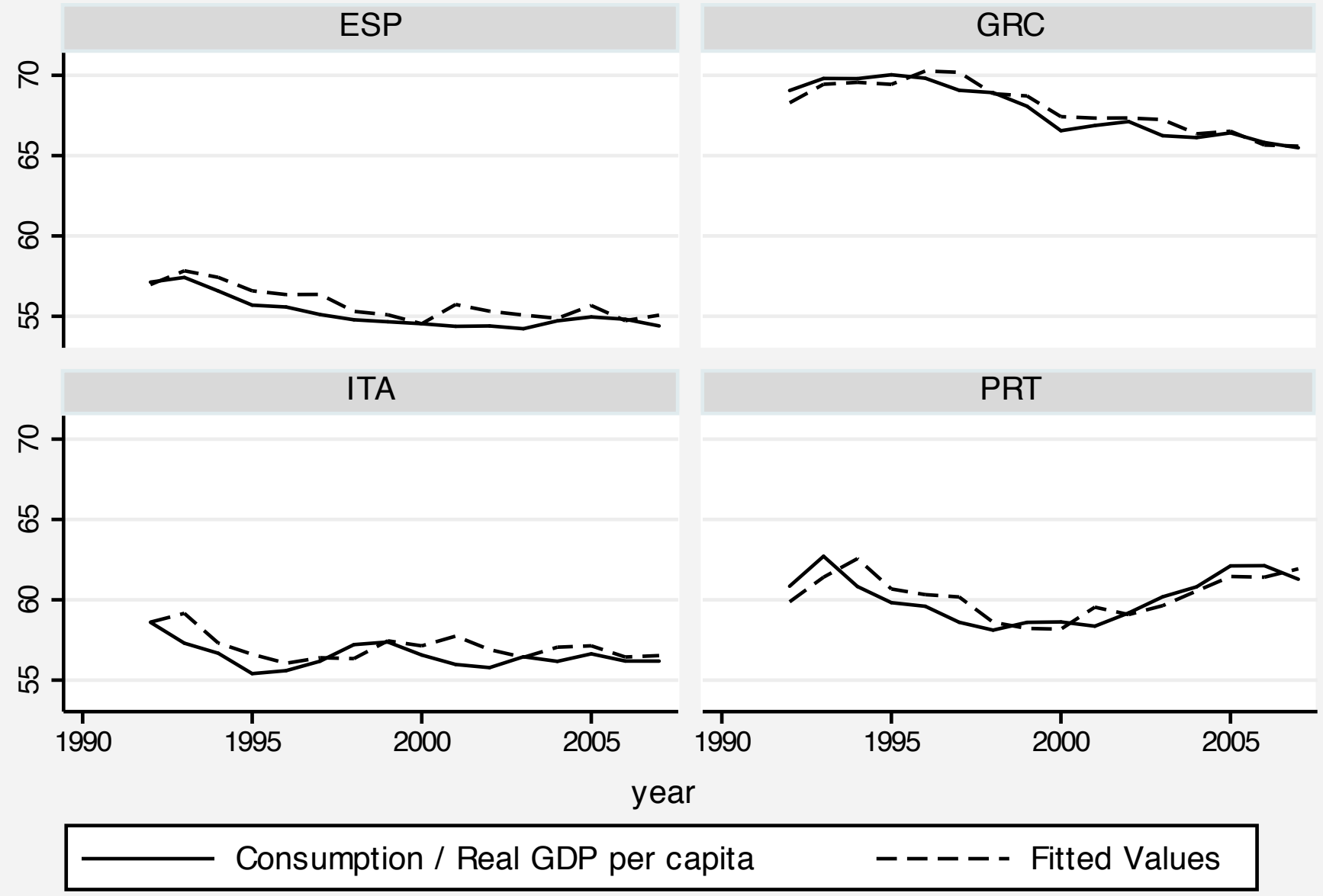
(continued)

```
------------------------------------------------------------------------------
Instruments for orthogonal deviations equation
  Standard
    FOD.(_Iyear_1992 _Iyear_1993 _Iyear_1994 _Iyear_1995 _Iyear_1996
    _Iyear_1997 _Iyear_1998 _Iyear_1999 _Iyear_2000 _Iyear_2001 _Iyear_2002
    _Iyear_2003 _Iyear_2004 _Iyear_2005 _Iyear_2006 _Iyear_2007 L.openc)
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    L(2/8).(L.kc cgnp)
Instruments for levels equation
  Standard
    _cons
    _Iyear_1992 _Iyear_1993 _Iyear_1994 _Iyear_1995 _Iyear_1996 _Iyear_1997
    _Iyear_1998 _Iyear_1999 _Iyear_2000 _Iyear_2001 _Iyear_2002 _Iyear_2003
    _Iyear_2004 _Iyear_2005 _Iyear_2006 _Iyear_2007 L.openc
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    DL.(L.kc cgnp)
------------------------------------------------------------------------------
Arellano-Bond test for AR(1) in first differences: z =  -3.31  Pr > z =  0.001
Arellano-Bond test for AR(2) in first differences: z =   0.42  Pr > z =  0.674
------------------------------------------------------------------------------
Sargan test of overid. restrictions: chi2(189)  = 384.95  Prob > chi2 =  0.000
  (Not robust, but not weakened by many instruments.)
Hansen test of overid. restrictions: chi2(189)  =  83.69  Prob > chi2 =  1.000
  (Robust, but can be weakened by many instruments.)
```

Using the FOD transformation, the autoregressive coefficient is a bit larger, and the `cgnp` coefficient a bit smaller, although its significance is retained.

After any DPD estimation command, we may save predicted values or residuals and graph them against the actual values:

```
. predict double kchat if inlist(country, "Italy", "Spain", "Greece", "Portugal
> ")
(option xb assumed; fitted values)
(1619 missing values generated)
. label var kc "Consumption / Real GDP per capita"
. xtline kc kchat if !mi(kchat), scheme(s2mono)
```


ESP
GRC
ITA
PRT
55
60
65
70
1990
1995
2000
2005
year
Consumption / Real GDP per capita
Fitted Values
Graphs by ISO country code

Although the DPD estimators are linear estimators, they are highly sensitive to the particular specification of the model and its instruments: more so in my experience than any other regression-based estimation approach. 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.