Panel data estimation and forecasting

Christopher F Baum

Boston College and DIW Berlin

NCER, Queensland University of Technology, March 2014
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)
```

<table>
<thead>
<tr>
<th>state</th>
<th>year</th>
<th>pop</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT</td>
<td>1990</td>
<td>3291967</td>
</tr>
<tr>
<td>CT</td>
<td>1995</td>
<td>3324144</td>
</tr>
<tr>
<td>CT</td>
<td>2000</td>
<td>3411750</td>
</tr>
<tr>
<td>MA</td>
<td>1990</td>
<td>6022639</td>
</tr>
<tr>
<td>MA</td>
<td>1995</td>
<td>6141445</td>
</tr>
<tr>
<td>MA</td>
<td>2000</td>
<td>6362076</td>
</tr>
<tr>
<td>RI</td>
<td>1990</td>
<td>1005995</td>
</tr>
<tr>
<td>RI</td>
<td>1995</td>
<td>1017002</td>
</tr>
<tr>
<td>RI</td>
<td>2000</td>
<td>1050664</td>
</tr>
</tbody>
</table>
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:

\[
\text{. list, noobs}
\]

<table>
<thead>
<tr>
<th>state</th>
<th>pop1990</th>
<th>pop1995</th>
<th>pop2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT</td>
<td>3291967</td>
<td>3324144</td>
<td>3411750</td>
</tr>
<tr>
<td>MA</td>
<td>6022639</td>
<td>6141445</td>
<td>6362076</td>
</tr>
<tr>
<td>RI</td>
<td>1005995</td>
<td>1017002</td>
<td>1050664</td>
</tr>
</tbody>
</table>

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.
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-
\textit{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 \textit{fixed effects} estimator, addresses the issue that no matter how many individual-specific factors you may include in the regressor list, there may be \textit{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:

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

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 $\beta_1$ 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:

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

(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

\[
y_{it} = X_{it}^{*} \beta^{*} + Z_{i} \alpha + \epsilon_{it} \quad (3)
\]

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, \( \beta^{*} \) 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 (\(\beta^*\)) are constrained across units, as is the \(\sigma^2_{\epsilon}\).

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:

\[
y = X\beta + D\alpha + \epsilon
\]

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 \(\beta^*\) coefficients are all slopes) and when this number is too large to permit estimation, we rewrite the least squares solution as

\[
b = (X' M_D X)^{-1} (X' M_D y)
\]  

(5)

where

\[
M_D = I - D (D' D)^{-1} D'
\]

(6)

is an idempotent matrix which is block–diagonal in \(M_0 = I_T - T^{-1} \iota \iota'\) (\(\iota\) 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 \textit{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 \textit{within estimator}, since it depends on the variation \textit{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`:

```
xtrg  depvar indepvars, fe [options]
```

As with standard regression, options include `robust` and `cluster()`. The command output displays estimates of $\sigma_u^2$ (labeled `sigma_u`), $\sigma_e^2$ (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 $\alpha_j$ 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 $\alpha_i$ 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, using the traffic dataset, 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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>fatal</td>
<td>336</td>
<td>2.040444</td>
<td>0.5701938</td>
<td>0.82121</td>
<td>4.21784</td>
</tr>
<tr>
<td>beertax</td>
<td>336</td>
<td>0.513256</td>
<td>0.4778442</td>
<td>0.0433109</td>
<td>2.720764</td>
</tr>
<tr>
<td>spircons</td>
<td>336</td>
<td>1.75369</td>
<td>0.6835745</td>
<td>0.79</td>
<td>4.9</td>
</tr>
<tr>
<td>unrate</td>
<td>336</td>
<td>7.346726</td>
<td>2.533405</td>
<td>2.4</td>
<td>18</td>
</tr>
<tr>
<td>perincK</td>
<td>336</td>
<td>13.88018</td>
<td>2.253046</td>
<td>9.513762</td>
<td>22.19345</td>
</tr>
</tbody>
</table>
The one-way fixed effects model

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

Fixed-effects (within) regression

|                  | Coef.  | Std. Err. | t     | P>|t| | 95% Conf. Interval |
|------------------|--------|-----------|-------|------|-------------------|
| fatal            | -.4840728 | .1625106 | -2.98 | 0.003 | -.8039508, -.1641948 |
| beertax          | .8169652 | .0792118 | 10.31 | 0.000 | .6610484, .9728819 |
| spircons         | -.0290499 | .0090274 | -3.22 | 0.001 | -.0468191, -.0112808 |
| unrate           | .1047103 | .0205986 | 5.08  | 0.000 | .064165, .1452555 |
| perincK          | -.383783 | .4201781 | -0.91 | 0.362 | -.1210841, .4432754 |
| _cons            | 1.1181913 |          |       |      |                   |
| sigma_u          | .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 $\text{corr}(u_i, Xb)$ 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 onward, that is very easy to do using factor variables by specifying `i.year` in the regressor list. The joint significance of those variables may be assessed with `testparm`, as we illustrate below.
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.
### Estimation for panel data

The fixed effects estimator

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

Fixed-effects (within) regression

<table>
<thead>
<tr>
<th>Group variable: state</th>
<th>Number of obs = 336</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-sq: within = 0.4528</td>
<td>Obs per group: min = 7</td>
</tr>
<tr>
<td>between = 0.1090</td>
<td>avg = 7.0</td>
</tr>
<tr>
<td>overall = 0.0770</td>
<td>max = 7</td>
</tr>
<tr>
<td>F(10,278) = 23.00</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>corr(u_i, Xb) = -0.8728</td>
<td></td>
</tr>
</tbody>
</table>

| 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 |
| year 1983 | -.0533713 | .030209 | -1.77 | 0.078 | -.1128387 | .0060962 |
| 1984 | -.1649828 | .037482 | -4.40 | 0.000 | -.2387674 | -.0911983 |
| 1985 | -.1997376 | .0415808 | -4.80 | 0.000 | -.2815908 | -.1178845 |
| 1986 | -.0508034 | .0515808 | -0.99 | 0.325 | -.1522647 | .050658 |
| 1987 | -.1000728 | .05906 | -1.69 | 0.091 | -.2163345 | .0161889 |
| 1988 | -.134057 | .0677696 | -1.98 | 0.049 | -.2674638 | -.0006503 |
| _cons | .1290568 | .4310663 | 0.30 | 0.765 | -.7195118 | .9776253 |

sigma_u | 1.0987683 |
sigma_e | .14570531 |

Christopher F Baum  (BC / DIW)
Panel data models
NCER/QUT, 2014
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.
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
sd(u_i + avg(e_i.))= .4209489
F(4,43) = 9.03  Prob > F  =  0.0000
```

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

$$y_{it} = X_{it}^* \beta^* + [u_i + \epsilon_{it}]$$  \hspace{1cm} (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.
In actual empirical work, it is extremely unusual to find that the key assumption underlying the random effects model is satisfied. Beyond textbook examples, it is difficult to find instances where the unobserved random effect can plausibly be uncorrelated with all observable attributes of the unit.

For instance, if you applied the estimator to country-level data on GDP growth, you would attribute the country-specific random component of the error term to a draw from nature that is uncorrelated with all observable characteristics of the country’s performance. Thus, we will not discuss this estimator in any greater detail.
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 \textit{beertax} and \textit{unrate} variables have lost their significance. The larger Root MSE for the \textit{fd} equation, compared to that for \textit{fe}, illustrates the relative inefficiency of the first difference estimator when there are more than two time periods.
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.
Seemingly unrelated regression estimators

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.
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 (`k_c`) as a function of openness (`open_c`) and the lagged ratio of GNP to GDP (`c_gnp`).
Seemingly unrelated regression estimators

```
. // 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)
```

<table>
<thead>
<tr>
<th>Data</th>
<th>long</th>
<th>-&gt;</th>
<th>wide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of obs.</td>
<td>144</td>
<td>-&gt;</td>
<td>48</td>
</tr>
<tr>
<td>Number of variables</td>
<td>5</td>
<td>-&gt;</td>
<td>10</td>
</tr>
<tr>
<td>j variable (3 values)</td>
<td>isocode</td>
<td>-&gt;</td>
<td>(dropped)</td>
</tr>
<tr>
<td>xij variables:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kc</td>
<td>kcESP kcGRC kcITA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>openc</td>
<td>opencESP opencGRC opencITA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cgnp</td>
<td>cgnpESP cgnpGRC cgnpITA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```
. 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 {
    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)
```

Christopher F Baum (BC / DIW) Panel data models NCER/QUT, 2014 47 / 126
. sureg "eqns", corr

Seemingly unrelated regression

<table>
<thead>
<tr>
<th>Equation</th>
<th>Obs</th>
<th>Parms</th>
<th>RMSE</th>
<th>&quot;R-sq&quot;</th>
<th>chi2</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>kcESP</td>
<td>47</td>
<td>2</td>
<td>.9379665</td>
<td>0.6934</td>
<td>104.50</td>
<td>0.0000</td>
</tr>
<tr>
<td>kcGRC</td>
<td>47</td>
<td>2</td>
<td>4.910707</td>
<td>0.3676</td>
<td>40.29</td>
<td>0.0000</td>
</tr>
<tr>
<td>kcITA</td>
<td>47</td>
<td>2</td>
<td>1.521322</td>
<td>0.4051</td>
<td>45.56</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

| Equation | Coef.  | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|----------|--------|-----------|-------|-------|----------------------|
| kcESP    |        |           |       |       |                      |
| opencESP | -.1205816 | .012307  | -9.80 | 0.000 | -.1447028            | -.0964603             |
| cgnpESP  | -.97201 | .373548   | -2.60 | 0.009 | -1.704151            | -.2398694             |
| L1.      |        |           |       |       |                      |
| _cons    | 157.6905 | 37.225    | 4.24  | 0.000 | 84.73086             | 230.6502              |
| kcGRC    |        |           |       |       |                      |
| opencGRC | .4215421 | .0670958  | 6.28  | 0.000 | .2900367             | .5530476              |
| cgnpGRC  | .5918787 | .5900844  | 1.00  | 0.316 | -1.5646655           | 1.748423              |
| L1.      |        |           |       |       |                      |
| _cons    | -16.48375 | 60.74346  | -0.27 | 0.786 | -135.5387            | 102.5712              |
(continued)

<table>
<thead>
<tr>
<th></th>
<th>kcITA</th>
<th>opencITA</th>
<th>.0684288</th>
<th>.0269877</th>
<th>2.54</th>
<th>0.011</th>
<th>.0155339</th>
<th>.1213237</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cgnpITA</td>
<td>L1.</td>
<td>-1.594811</td>
<td>.3426602</td>
<td>-4.65</td>
<td>0.000</td>
<td>-2.266412</td>
<td>-.923209</td>
</tr>
<tr>
<td></td>
<td>_cons</td>
<td></td>
<td>211.6658</td>
<td>34.58681</td>
<td>6.12</td>
<td>0.000</td>
<td>143.8769</td>
<td>279.4547</td>
</tr>
</tbody>
</table>

Correlation matrix of residuals:

<table>
<thead>
<tr>
<th></th>
<th>kcESP</th>
<th>kcGRC</th>
<th>kcITA</th>
<th>kcESP</th>
<th>kcESP</th>
<th>kcGRC</th>
<th>kcITA</th>
<th>kcESP</th>
<th>kcESP</th>
<th>kcGRC</th>
<th>kcITA</th>
</tr>
</thead>
<tbody>
<tr>
<td>kcESP</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kcGRC</td>
<td>-0.2367</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kcITA</td>
<td>-0.0786</td>
<td>-0.2618</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 $\text{openc}$ are equal across equations. Note that in the `test` command we must specify in which equation each coefficient appears.

```plaintext
. // test cross-equation hypothesis of coefficient equality
  ( 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

<table>
<thead>
<tr>
<th>Equation</th>
<th>Obs</th>
<th>Parms</th>
<th>RMSE</th>
<th>&quot;R-sq&quot;</th>
<th>chi2</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>kcESP</td>
<td>40</td>
<td>2</td>
<td>.985171</td>
<td>0.5472</td>
<td>48.72</td>
<td>0.0000</td>
</tr>
<tr>
<td>kcGRC</td>
<td>40</td>
<td>2</td>
<td>5.274077</td>
<td>0.3076</td>
<td>27.49</td>
<td>0.0000</td>
</tr>
<tr>
<td>kcITA</td>
<td>40</td>
<td>2</td>
<td>1.590656</td>
<td>0.4364</td>
<td>42.14</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

```
. 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)

```
. su * hat if year>2000
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESPhat</td>
<td>7</td>
<td>55.31007</td>
<td>.4318259</td>
<td>54.43892</td>
<td>55.7324</td>
</tr>
<tr>
<td>GRChat</td>
<td>7</td>
<td>66.24322</td>
<td>.932017</td>
<td>65.35107</td>
<td>68.15631</td>
</tr>
<tr>
<td>ITAhat</td>
<td>7</td>
<td>57.37146</td>
<td>.1436187</td>
<td>57.18819</td>
<td>57.60937</td>
</tr>
</tbody>
</table>

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

Christopher F Baum  (BC / DIW)
Seemingly unrelated regression estimators

Ex ante forecasting

Predicted consumption share, real GDP per capita

Ex ante predictions

Christopher F Baum (BC / DIW)
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.
Cluster-robust standard errors

are 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.
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 \( k_c \) from the Penn World Tables data set, in which regressors are again specified as \( \text{openc} \) and \( \text{cgnp} \), 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

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IID</td>
<td>clCty</td>
<td>clCtyYr</td>
</tr>
<tr>
<td>openc</td>
<td>-0.036***</td>
<td>-0.036*</td>
<td>-0.036*</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.018)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>cgnp</td>
<td>0.800***</td>
<td>0.800***</td>
<td>0.800***</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.146)</td>
<td>(0.146)</td>
</tr>
<tr>
<td>N</td>
<td>4508</td>
<td>4508</td>
<td>4508</td>
</tr>
</tbody>
</table>

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.
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 $\epsilon_{it}$—contains contemporaneous values of $\epsilon$ for $t = 1 \ldots T$. The resulting correlation creates a bias in the estimate of the coefficient of the lagged dependent variable which is not mitigated by increasing $N$, the number of individual units.
The demeaning operation creates a regressor which cannot be distributed independently of the error term. Nickell demonstrates that the inconsistency of $\hat{\rho}$ as $N \to \infty$ is of order $1/T$, which may be quite sizable in a “small $T$” context. If $\rho > 0$, the bias is invariably negative, so that the persistence of $y$ will be underestimated.

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

$$y_{it} = \beta_1 + \rho y_{i,t-1} + X_{it}\beta_2 + u_i + \epsilon_{it} \quad (8)$$

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

$$\Delta y_{it} = \rho \Delta y_{i,t-1} + \Delta X_{it}\beta_2 + \Delta \epsilon_{it} \quad (9)$$

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 $\epsilon$ 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 $\epsilon$ 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).
The \textit{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 \texttt{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

\[
\begin{align*}
y_{it} & = X_{it} \beta_1 + W_{it} \beta_2 + v_{it} \\
v_{it} & = u_i + \epsilon_{it}
\end{align*}
\]

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 AB approach, and its extension to the ‘System GMM’ context, is an estimator designed for situations with:

- ‘small $T$, large $N$’ panels: few time periods and many individual units
- a linear functional relationship
- one left-hand variable that is dynamic, depending on its own past realisations
- right-hand variables that are not strictly exogenous: correlated with past and possibly current realisations of the error
- fixed individual effects, implying unobserved heterogeneity
- heteroskedasticity and autocorrelation within individual units’ errors, but not across them
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).

This estimator is available in Stata as \texttt{xtabond}. A more general version, allowing for autocorrelated errors, is available as \texttt{xtdpd}. An excellent alternative to Stata’s built-in commands is David Roodman’s \texttt{xtabond2}, available from SSC (\texttt{findit xtabond2}). It is very well documented in his paper “How to do \texttt{xtabond2}.” The \texttt{xtabond2} routine provides several additional features—such as the orthogonal deviations transformation discussed below—not available in official Stata’s commands.
In standard 2SLS, including the Anderson–Hsiao approach, the twice-lagged level appears in the instrument matrix as

$$Z_i = \begin{pmatrix} \cdot \\ y_{i,1} \\ \vdots \\ y_{i,T-2} \end{pmatrix}$$

where the first row corresponds to $t = 2$, given that the first observation is lost in applying the FD transformation. The missing value in the instrument for $t = 2$ causes that observation for each panel unit to be removed from the estimation.
If we also included the thrice-lagged level $y_{t-3}$ as a second instrument in the Anderson–Hsiao approach, we would lose another observation per panel:

$$Z_i = \begin{pmatrix} \cdots & \cdots \\ y_{i,1} & \cdots \\ y_{i,2} & y_{i,1} \\ \vdots & \vdots \\ y_{i,T-2} & y_{i,T-3} \end{pmatrix}$$

so that the first observation available for the regression is that dated $t = 4$. 
To avoid this loss of degrees of freedom, Holtz-Eakin et al. construct a set of instruments from the second lag of $y$, one instrument pertaining to each time period:

$$Z_i = \begin{pmatrix}
0 & 0 & \ldots & 0 \\
y_{i,1} & 0 & \ldots & 0 \\
0 & y_{i,2} & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & y_{i,T-2}
\end{pmatrix}$$

The inclusion of zeros in place of missing values prevents the loss of additional degrees of freedom, in that all observations dated $t = 2$ and later can now be included in the regression. Although the inclusion of zeros might seem arbitrary, the columns of the resulting instrument matrix will be orthogonal to the transformed errors. The resulting moment conditions correspond to an expectation we believe should hold: $E(y_{i,t-2}\epsilon^*_{it}) = 0$, where $\epsilon^*$ refers to the FD-transformed errors.
It would also be valid to ‘collapse’ the columns of this $\mathbf{Z}$ matrix into a single column, which embodies the same expectation, but conveys less information as it will only produce a single moment condition. In this context, the collapsed instrument set will be the same implied by standard IV, with a zero replacing the missing value in the first usable observation:

$$
\mathbf{Z}_i = \begin{pmatrix}
0 \\
y_{i,1} \\
\vdots \\
y_{i,T-2}
\end{pmatrix}
$$

This is specified in Roodman’s xtabond2 software by giving the collapse option.
Given this solution to the tradeoff between lag length and sample length, we can now adopt Holtz-Eakin et al.’s suggestion and include all available lags of the untransformed variables as instruments. For endogenous variables, lags 2 and higher are available. For predetermined variables that are not strictly exogenous, lag 1 is also valid, as its value is only correlated with errors dated $t - 2$ or earlier.

Using all available instruments gives rise to an instrument matrix such as

$$Z_i = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
y_{i,1} & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
0 & y_{i,2} & y_{i,1} & 0 & 0 & 0 & \cdots \\
0 & 0 & 0 & y_{i,3} & y_{i,2} & y_{i,1} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}$$
In this setup, we have different numbers of instruments available for each time period: one for $t = 2$, two for $t = 3$, and so on. As we move to the later time periods in each panel’s timeseries, additional orthogonality conditions become available, and taking these additional conditions into account improves the efficiency of the AB estimator.

One disadvantage of this strategy should be apparent. The number of instruments produced will be quadratic in $T$, the length of the timeseries available. If $T < 10$, that may be a manageable number, but for a longer timeseries, it may be necessary to restrict the number of past lags used.
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.
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 $xtdpdsys$. 
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 `reivreg2`) 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.
To illustrate the performance of the several estimators, we make use of the original AB dataset, available within Stata with `webuse abdata`. This is an unbalanced panel of annual data from 140 UK firms for 1976–1984. In their original paper, they modeled firms’ employment $n$ using a partial adjustment model to reflect the costs of hiring and firing, with two lags of employment.

Other variables included were the current and lagged wage level $w$, the current, once- and twice-lagged capital stock ($k$) and the current, once- and twice-lagged output in the firm’s sector ($y_s$). All variables are expressed as logarithms. A set of time dummies is also included to capture business cycle effects.
If we were to estimate this model ignoring its dynamic panel nature, we could merely apply `regress` with panel-clustered standard errors: **Try it out:**

```
regress n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*, cluster(id)
```

One obvious difficulty with this approach is the likely importance of firm-level unobserved heterogeneity. We have accounted for potential correlation between firms’ errors over time with the cluster-robust VCE, but this does not address the potential impact of unobserved heterogeneity on the conditional mean.

We can apply the within transformation to take account of this aspect of the data: **Try it out:**

```
xtreg n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*, fe cluster(id)
```
The fixed effects estimates will suffer from Nickell bias, which may be severe given the short timeseries available.

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>FE</th>
</tr>
</thead>
<tbody>
<tr>
<td>nL1</td>
<td>1.045***</td>
<td>0.733***</td>
</tr>
<tr>
<td></td>
<td>(20.17)</td>
<td>(12.28)</td>
</tr>
<tr>
<td>nL2</td>
<td>-0.0765</td>
<td>-0.139</td>
</tr>
<tr>
<td></td>
<td>(-1.57)</td>
<td>(-1.78)</td>
</tr>
<tr>
<td>w</td>
<td>-0.524**</td>
<td>-0.560***</td>
</tr>
<tr>
<td></td>
<td>(-3.01)</td>
<td>(-3.51)</td>
</tr>
<tr>
<td>k</td>
<td>0.343***</td>
<td>0.388***</td>
</tr>
<tr>
<td></td>
<td>(7.06)</td>
<td>(6.82)</td>
</tr>
<tr>
<td>ys</td>
<td>0.433*</td>
<td>0.469**</td>
</tr>
<tr>
<td></td>
<td>(2.42)</td>
<td>(2.74)</td>
</tr>
<tr>
<td>N</td>
<td>751</td>
<td>751</td>
</tr>
</tbody>
</table>

* t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
In the original OLS regression, the lagged dependent variable was positively correlated with the error, biasing its coefficient upward. In the fixed effects regression, its coefficient is biased downward due to the negative sign on $\nu_{t-1}$ in the transformed error. The OLS estimate of the first lag of $n$ is 1.045; the fixed effects estimate is 0.733.

Given the opposite directions of bias present in these estimates, consistent estimates should lie between these values, which may be a useful check. As the coefficient on the second lag of $n$ cannot be distinguished from zero, the first lag coefficient should be below unity for dynamic stability.
To deal with these two aspects of the estimation problem, we might use the Anderson–Hsiao estimator to the first-differenced equation, instrumenting the lagged dependent variable with the twice-lagged level: **Try it out:**

```
ivregress 2sls D.n (D.nL1 = nL2) D.(nL2 w wL1 k kL1 kL2 ///
```
Although these results should be consistent, they are quite disappointing. The coefficient on lagged $n$ is outside the bounds of its OLS and FE counterparts, and much larger than unity, a value consistent with dynamic stability. It is also very imprecisely estimated.
The difference GMM approach deals with this inherent endogeneity by transforming the data to remove the fixed effects. The standard approach applies the first difference (FD) transformation, which as discussed earlier removes the fixed effect at the cost of introducing a correlation between $\Delta y_{i,t-1}$ and $\Delta \nu_{it}$, both of which have a term dated $(t - 1)$. This is preferable to the application of the within transformation, as that transformation makes every observation in the transformed data endogenous to every other for a given individual.

The one disadvantage of the first difference transformation is that it magnifies gaps in unbalanced panels. If some value of $y_{it}$ is missing, then both $\Delta y_{it}$ and $\Delta y_{i,t-1}$ will be missing in the transformed data. This motivates an alternative transformation: the forward orthogonal deviations (FOD) transformation, proposed by Arellano and Bover (J. Econometrics, 1995).
In contrast to the within transformation, which subtracts the average of all observations’ values from the current value, and the FD transformation, that subtracts the previous value from the current value, the FOD transformation subtracts the average of all available future observations from the current value. While the FD transformation drops the first observation on each individual in the panel, the FOD transformation drops the last observation for each individual. It is computable for all periods except the last period, even in the presence of gaps in the panel.

The FOD transformation is not available in any of official Stata’s DPD commands, but it is available in David Roodman’s xtabond2 implementation of the DPD estimator, available from SSC.
To illustrate the use of the AB estimator, we may reestimate the model with `xtabond2`, assuming that the only endogeneity present is that involving the lagged dependent variable. **Try it out:**

```bash
xtabond2 n L(1/2).n L(0/1).w L(0/2).(k ys) yr*, gmm(L.n) ///
iv(L(0/1).w L(0/2).(k ys) yr*) nolevel robust small
```

Note that in `xtabond2` syntax, every right-hand variable generally appears twice in the command, as instruments must be explicitly specified when they are instrumenting themselves. In this example, all explanatory variables except the lagged dependent variable are taken as “IV-style” instruments, entering the $Z$ matrix as a single column. The lagged dependent variable is specified as a “GMM-style” instrument, where all available lags will be used as separate instruments. The `nolevel` option is needed to specify the AB estimator.
### A-B

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>L.n</td>
<td>0.686***</td>
<td>(4.67)</td>
</tr>
<tr>
<td>L2.n</td>
<td>-0.0854</td>
<td>(-1.50)</td>
</tr>
<tr>
<td>w</td>
<td>-0.608**</td>
<td>(-3.36)</td>
</tr>
<tr>
<td>k</td>
<td>0.357***</td>
<td>(5.95)</td>
</tr>
<tr>
<td>ys</td>
<td>0.609***</td>
<td>(3.47)</td>
</tr>
<tr>
<td>N</td>
<td>611</td>
<td></td>
</tr>
</tbody>
</table>

$t$ statistics in parentheses

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

In these results, 41 instruments have been created, with 17 corresponding to the “IV-style” regressors and the rest computed from lagged values of $n$. Note that the coefficient on the lagged dependent variable now lies within the range for dynamic stability. In contrast to that produced by the Anderson–Hsiao estimator, the coefficient is quite precisely estimated.
There are 25 overidentifying restrictions in this instance, as shown in the first column below. The *hansen_df* represents the degrees of freedom for the Hansen $J$ test of overidentifying restrictions. The $p$-value of that test is shown as *hansenp*.

<table>
<thead>
<tr>
<th></th>
<th>All lags</th>
<th>lags 2-5</th>
<th>lags 2-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>L.n</td>
<td>0.686***</td>
<td>0.835*</td>
<td>1.107***</td>
</tr>
<tr>
<td></td>
<td>(4.67)</td>
<td>(2.59)</td>
<td>(3.94)</td>
</tr>
<tr>
<td>L2.n</td>
<td>-0.0854</td>
<td>0.262</td>
<td>0.231</td>
</tr>
<tr>
<td></td>
<td>(-1.50)</td>
<td>(1.56)</td>
<td>(1.32)</td>
</tr>
<tr>
<td>w</td>
<td>-0.608**</td>
<td>-0.671**</td>
<td>-0.709**</td>
</tr>
<tr>
<td></td>
<td>(-3.36)</td>
<td>(-3.18)</td>
<td>(-3.26)</td>
</tr>
<tr>
<td>k</td>
<td>0.357***</td>
<td>0.325***</td>
<td>0.309***</td>
</tr>
<tr>
<td></td>
<td>(5.95)</td>
<td>(4.95)</td>
<td>(4.55)</td>
</tr>
<tr>
<td>ys</td>
<td>0.609***</td>
<td>0.640**</td>
<td>0.698***</td>
</tr>
<tr>
<td></td>
<td>(3.47)</td>
<td>(3.07)</td>
<td>(3.45)</td>
</tr>
<tr>
<td>hansen_df</td>
<td>25</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>hansenp</td>
<td>0.177</td>
<td>0.676</td>
<td>0.714</td>
</tr>
</tbody>
</table>

$t$ statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
In this table, we can examine the sensitivity of the results to the choice of “GMM-style” lag specification. In the first column, all available lags of the level of $n$ are used. In the second column, the \texttt{lag(2 5)} option is used to restrict the maximum lag to 5 periods, while in the third column, the maximum lag is set to 4 periods. Fewer instruments are used in those instances, as shown by the smaller values of \texttt{sar_df}.

The $p$-value of Hansen’s $J$ is also considerably larger for the restricted-lag cases. On the other hand, the estimate of the lagged dependent variable’s coefficient appears to be quite sensitive to the choice of lag length.
We illustrate estimating this equation with both the FD transformation and the forward orthogonal deviations (FOD) transformation:

<table>
<thead>
<tr>
<th></th>
<th>First diff</th>
<th>FOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>L.n</td>
<td>0.686***</td>
<td>0.737***</td>
</tr>
<tr>
<td></td>
<td>(4.67)</td>
<td>(5.14)</td>
</tr>
<tr>
<td>L2.n</td>
<td>-0.0854</td>
<td>-0.0960</td>
</tr>
<tr>
<td></td>
<td>(-1.50)</td>
<td>(-1.38)</td>
</tr>
<tr>
<td>w</td>
<td>-0.608**</td>
<td>-0.563***</td>
</tr>
<tr>
<td></td>
<td>(-3.36)</td>
<td>(-3.47)</td>
</tr>
<tr>
<td>k</td>
<td>0.357***</td>
<td>0.384***</td>
</tr>
<tr>
<td></td>
<td>(5.95)</td>
<td>(6.85)</td>
</tr>
<tr>
<td>ys</td>
<td>0.609***</td>
<td>0.469**</td>
</tr>
<tr>
<td></td>
<td>(3.47)</td>
<td>(2.72)</td>
</tr>
<tr>
<td>hansen_df</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>hansenp</td>
<td>0.177</td>
<td>0.170</td>
</tr>
</tbody>
</table>

_t statistics in parentheses_

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

The results appear reasonably robust to the choice of transformation, with slightly more precise estimates for most coefficients when the FOD transformation is employed.
We might reasonably consider, as did Blundell and Bond (J. Econometrics, 1998), that wages and the capital stock should not be taken as strictly exogenous in this context, as we have in the above models.

Reestimate the equation producing “GMM-style” instruments for all three variables, with both one-step and two-step VCE:

**Try it out:**

```bash
xtabond2 n L(1/2).n L(0/1).w L(0/2).(k ys) yr*, gmm(L.(n w k)) ///
iv(L(0/2).ys yr*) nolevel robust small
```
The results from both one-step and two-step estimation appear reasonable. Interestingly, only the coefficient on $ys$ appears to be more precisely estimated by the two-step VCE. With no restrictions on the instrument set, 74 overidentifying restrictions are defined, with 90 instruments in total.
To illustrate system GMM, we follow Blundell and Bond, who used the same `abdata` dataset on a somewhat simpler model, dropping the second lags and removing sectoral demand. We consider wages and capital as potentially endogenous, with GMM-style instruments.

Estimate the one-step BB model.

**Try it out:**

```plaintext
xtabond2 n L.n L(0/1).(w k) yr*, gmm(L.(n w k)) iv(yr*, equation(level)) ///
robust small
```

We indicate here with the `equation(level)` suboption that the year dummies are only to be considered instruments in the level equation. As the default for `xtabond2` is the BB estimator, we omit the `noleveleq` option that has called for the AB estimator in earlier examples.
We find that the $\alpha$ coefficient is much higher than in the AB estimates, although it may be distinguished from unity. 113 instruments are created, with 100 degrees of freedom in the test of overidentifying restrictions.
A second empirical exercise

We also illustrate DPD estimation using the Penn World Table cross-country panel. We specify a model for \( k_c \) (the consumption share of real GDP per capita) depending on its own lag, \( c_{gnp} \), and a set of time fixed effects, which we compute with the \( xi \) command, as \texttt{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 \texttt{testparm _I*} after estimation to evaluate the joint significance of time effects (listing of which has been suppressed).
Dynamic panel data estimators

A second empirical exercise

```
i.year
_Iyear_1991-2007 (naturally coded; _Iyear_1991 omitted)
.xtabond2 kc L.kc cgnp _I*, gmm(L.kc open 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
```

<table>
<thead>
<tr>
<th>Group variable: iso</th>
<th>Number of obs = 1485</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time variable: year</td>
<td>Number of groups = 99</td>
</tr>
<tr>
<td>Number of instruments = 283</td>
<td>Obs per group: min = 15</td>
</tr>
<tr>
<td>Wald chi2(17) = 94.96</td>
<td>avg = 15.00</td>
</tr>
<tr>
<td>Prob &gt; chi2 = 0.000</td>
<td>max = 15</td>
</tr>
</tbody>
</table>

| kc    | Coef.   | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|-------|---------|-----------|-------|-------|----------------------|
| L1.   | 0.6478636 | 0.1041122 | 6.22  | 0.000 | 0.4438075 to 0.8519197 |
| cgnp  | 0.233404  | 0.1080771 | 2.16  | 0.031 | 0.0215768 to 0.4452312 |
| ...   |         |           |       |       |                      |
Instruments for first differences equation

Standard

 _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 $\text{openc}$ as a GMM-style instrument. The autoregressive coefficient is 0.648, and the $\text{cgnp}$ 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 $\text{openc}$ as a IV-style instrument in this specification.
Dynamic panel data estimators
A second empirical exercise

.xtabond2 kc L.kc cgnp _I*, gmm(L.kc cgnp, lag(2 8)) iv(_I* L.openc) ///
twostep robust nodiffsaran

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

|         | Coef.    | Std. Err. | z   | P>|z|   | [95% Conf. Interval] |
|---------|----------|-----------|-----|-------|----------------------|
|         | Corrected|           |     |       |                      |
| kc      |          |           |     |       |                      |
| L1.     | 0.9452696| 0.0191167 | 49.45| 0.000 | 0.9078014             | 0.9827377   |
| cgnp    | 0.097109 | 0.0436338 | 2.23 | 0.026 | 0.0115882             | 0.1826297   |
| _cons   | -6.091674| 3.45096   | -1.77| 0.078 | -12.85543             | 0.672083    |
A second empirical exercise

Instruments for first differences equation

**Standard**

  _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_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:** $z = -3.29 \quad Pr > z = 0.001$

**Arellano-Bond test for AR(2) in first differences:** $z = 0.42 \quad Pr > z = 0.677$

**Sargan test of overid. restrictions:** $\chi^2(189) = 353.99 \quad Prob > \chi^2 = 0.000$

(Not robust, but not weakened by many instruments.)

**Hansen test of overid. restrictions:** $\chi^2(189) = 88.59 \quad Prob > \chi^2 = 1.000$

(Robust, but can be weakened by many instruments.)
Note that the autoregressive coefficient is much larger: 0.945 in this context. The \( c_{gnp} \) 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

<table>
<thead>
<tr>
<th>Group variable: iso</th>
<th>Number of obs = 1584</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time variable: year</td>
<td>Number of groups = 99</td>
</tr>
<tr>
<td>Number of instruments = 207</td>
<td>Obs per group: min = 16</td>
</tr>
<tr>
<td>Wald chi2(17) = 8904.24</td>
<td>avg = 16.00</td>
</tr>
<tr>
<td>Prob &gt; chi2 = 0.000</td>
<td>max = 16</td>
</tr>
</tbody>
</table>

| kc       | Coef. | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|----------|-------|-----------|------|-----|---------------------|
| kc       | 0.9550247 | 0.0142928 | 66.82 | 0.000 | 0.9270114, 0.983038 |
| L1.      | 0.0723786 | 0.0339312 | 2.13 | 0.033 | 0.0058746, 0.1388825|
| cgnp     |        |           |      |      |                     |
| ...      |        |           |      |      |                     |
| _cons    | -4.329945 | 2.947738  | -1.47 | 0.142 | -10.10741, 1.447515 |
Instruments for orthogonal deviations equation

Standard


GMM-type (missing=0, separate instruments for each period unless collapsed)
L(2/8). (L.kc cgnp)

Instruments for levels equation

Standard

_cons 
 _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 \( c_{gnp} \) 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)
Dynamic panel data estimators

Ex ante forecasting

Graphs by ISO country code

Consumption / Real GDP per capita

Fitted Values

Christopher F Baum (BC / DIW)
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. A very useful reference for DPD modeling is David Roodman’s paper “How to do $\text{xtabond2}$” paper, available from http://ideas.repec.org.
A very significant addition to Stata 13 is the forecast suite of commands that support the definition of a model, containing a number of estimated equations and identities, and the ability to produce static and dynamic forecasts from the possibly nonlinear structure. For instance, a model might predict the percentage growth rate of GDP, but contain the national income identity in the level of GDP.

Simulation methods may be used to obtain prediction intervals, and scenario analysis may easily be performed, allowing the comparison of a baseline forecast and a ‘what-if’ scenario involving alternate exogenous factors. For many Stata users, this suite of commands will greatly enhance Stata’s usefulness in their work with time series.
Let’s look at a very simple example of the capabilities of forecast. Stata can handle a much larger model, which may involve either time series or panel data.

We construct our simple macro model from the usmacrol dataset. It contains three estimated, or stochastic equations: for the percentage change in real consumption, the log of real gross investment, and the banking sector’s prime rate.
In order to compute out-of-sample (ex ante) forecasts, we end the estimation sample at 2006q4. By default, forecasts for 2007q1 and following quarters will be dynamic forecasts, using the prior values of the endogenous variables computed from the forecast procedure.

. use usmacro1
. g termspread = tr10yr - tr3yr
. g dlrconsump = D.lrconsump
   (1 missing value generated)
. loc endest 2006q4
. loc begfc 2007q1
We could estimate the consumption equation with $D. \text{lrcConsump}$ as the dependent variable, but for simplicity we create the variable $dlr\text{consump}$ as the percentage change of real consumption. It is modeled as depending on the contemporaneous percentage changes in real GDP ($D. \text{lrgdp}$) and the prime rate.

As GDP growth is directly influenced by consumption growth, we estimate the equation with IV-GMM with a HAC VCE, using the Federal funds rate ($f\text{frate}$) and the percentage change in government consumption ($D. \text{lgovt}$) as instruments.
Model specification, solution and dynamic forecasting

```
.system

. ivreg2 dlrconsump (D.lrgdp = ffrate D.lrgovt) D.primerate ///
> if tin(,'`endest'), gmm2s robust bw(5)

2-Step GMM estimation

Estimates efficient for arbitrary heteroskedasticity and autocorrelation
Statistics robust to heteroskedasticity and autocorrelation
  kernel=Bartlett; bandwidth=5
  time variable (t): yq

Number of obs = 191
F( 2, 188) = 11.34
Prob > F = 0.0000

Total (centered) SS = .0087107485 Centered R2 = 0.3930
Total (uncentered) SS = .0234863835 Uncentered R2 = 0.7749
Residual SS = .0052874007 Root MSE = .005261

| dlrconsump | Robust Coef. | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|------------|--------------|-----------|------|-----|-------------------|
| lrgdp      |              |           |      |     |                   |
| D1.        | .667709      | .1394246  | 4.79 | 0.000| .3944418          | .9409763 |
| primerate  |              |           |      |     |                   |
| D1.        | -.0014486    | .0004658  | -3.11| 0.002| -.0023616         | -.0005355 |
| _cons      | .0033298     | .0012235  | 2.72 | 0.006| .0009317          | .0057279 |

Underidentification test (Kleibergen-Paap rk LM statistic): 6.292
Chi-sq(2) P-val = 0.0430

Instrumented: D.lrgdp
Included instruments: D.primerate
Excluded instruments: ffrate D.lrgovt
```

Christopher F Baum (BC / DIW) Panel data models NCER/QUT, 2014 114 / 126
Weak identification test (Cragg-Donald Wald F statistic): 9.587
    (Kleibergen-Paap rk Wald F statistic): 5.291
Stock-Yogo weak ID test critical values: 10% maximal IV size 19.93
    15% maximal IV size 11.59
    20% maximal IV size 8.75
    25% maximal IV size 7.25
NB: Critical values are for Cragg-Donald F statistic and i.i.d. errors.

Hansen J statistic (overidentification test of all instruments): 0.000
    Chi-sq(1) P-val = 0.9924

Instrumented:        D.lrgdp
Included instruments: D.primerate
Excluded instruments: ffrate D.lrgovt

. est store c

The estimated coefficients’ signs and significance are plausible.
Following estimation, we use estimates store setname to make
this equation available to the forecast command.
We specify the investment equation in terms of the log of real investment, \( \text{lrgrossinv} \). It is modeled as a function of its own first lag, the lagged change in the log of real GDP (\( \text{LD.lrgdp} \)) and the log of the real wage (\( \text{lrwage} \)). The GDP term is the investment accelerator, where investment spending responds to the (percentage) change in GDP rather than its (log) level. The real wage accounts for substitution between capital and labour: higher wages encourage firms to employ more physical capital.

We estimate this equation with OLS with a HAC VCE, and store its estimates as setname \( i \). Its coefficient estimates have plausible signs and significance.
Model specification, solution and dynamic forecasting

```
. ivreg2 lrgrossinv L.lrgrossinv LD.lrgdp lrwage ///
> if tin(``endest``), robust bw(5)
```

OLS estimation

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity and autocorrelation
  kernel=Bartlett; bandwidth=5
time variable (t):  yq

Number of obs = 190
F( 3, 186) = 37639.64
Prob > F = 0.0000

Total (centered) SS = 35.13194784
Centered R2 = 0.9983
Total (uncentered) SS = 9639.577385
Uncentered R2 = 1.0000
Residual SS = .0588143245
Root MSE = .01759

| Variable   | Coef.    | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|------------|----------|-----------|-------|------|----------------------|
| lrgrossinv |          |           |       |      |                      |
|            | L1.      | .9797325  | .0122183 | 80.19 | 0.000 | .955785   | 1.00368 |
| lrgdp      |          |           |       |      |                      |
|            | LD.      | .8336934  | .1678427 | 4.97  | 0.000 | .5047278  | 1.162659 |
| lrwage     |          |           |       |      |                      |
|            | _cons    | -.2666804 | .1427258 | -1.87 | 0.062 | -.5464177 | 0.013057 |

Included instruments: L.lrgrossinv LD.lrgdp lrwage
The last estimated equation, that for the prime rate, includes the Federal funds rate \((\text{ffrate})\) and the ‘term spread’ on Treasury securities \((\text{termspread})\) as explanatory factors. The term spread, or term premium, was defined as the difference between the 10-year and 3-year bond rates on US Treasuries. This factor, effectively the slope of the term structure, is normally positive, as investors demand a premium (in terms of a higher interest rate) to hold longer-term securities of the same credit quality.

We estimate the equation with OLS with a HAC VCE, and store its estimates as \(r\). Its coefficient estimates have plausible signs and significance.
. ivreg2 primerate ffrate termspread if tin(,`endest`), robust bw(5)

OLS estimation

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity and autocorrelation
  kernel=Bartlett; bandwidth=5
  time variable (t): yq

|                | Coef. | Std. Err. |    z  |   P>|z| |    [95% Conf. Interval] |
|----------------|-------|-----------|-------|-------|--------------------------|
|    primerate   |       |           |       |       |                          |
|    ffrate      | 1.001963 |  .0447049 | 22.41 | 0.000 | .9143433 – 1.089583     |
|    termspread  | .8547286 |  .2078859 |  4.11 | 0.000 | .4472797 – 1.262178     |
|     _cons      | 1.52236  |  .3787312 |  4.02 | 0.000 | .7800599 – 2.264659     |

Included instruments: ffrate termspread

. est store r
Now that we have estimated all of the model’s stochastic equations, we are ready to set up the model. The `forecast create modelname` command defines the model, and `forecast estimates setname` commands are used to add each set of stored estimates to the model.

We define identities that allow us to compute endogenous variables of interest. For instance, we want to forecast the levels of real consumption, investment and GDP. The first four `forecast identity` commands define those levels and express the national income identity for this (hypothetically closed) economy, \( Y = C + I + G \). We also define the `termspread` variable in terms of its components. With these five identities, Stata considers that we now have eight endogenous variables defined by the model.
Model specification, solution and dynamic forecasting

. forecast create modela
Forecast model modela started.

. forecast est c
Added estimation results from *ivreg2*.
Forecast model modela now contains 1 endogenous variable.

. forecast est i
Added estimation results from *ivreg2*.
Forecast model modela now contains 2 endogenous variables.

. forecast est r
Added estimation results from *ivreg2*.
Forecast model modela now contains 3 endogenous variables.

. forecast identity rconsump = exp(L.1rconsump + dlrconsump)
Forecast model modela now contains 4 endogenous variables.

. forecast identity rgrossinv = exp(lrgrossinv)
Forecast model modela now contains 5 endogenous variables.

. forecast identity rgovt = exp(lrgovt)
Forecast model modela now contains 6 endogenous variables.

. forecast identity rgdp = rconsump + rgrossinv + rgovt
Forecast model modela now contains 7 endogenous variables.

. forecast identity termspread = tr10yr - tr3yr
Forecast model modela now contains 8 endogenous variables.
Finally, we define five variables as being exogenous to the model. Those variables could be altered to produce scenario analysis: for instance, how the model would respond to a decrease in (log) government consumption, or an increase in the Federal funds rate caused by the monetary authorities.

- `forecast exog ffrate`
  Forecast model modela now contains 1 declared exogenous variable.
- `forecast exog lgovt`
  Forecast model modela now contains 2 declared exogenous variables.
- `forecast exog tr10yr`
  Forecast model modela now contains 3 declared exogenous variables.
- `forecast exog tr3yr`
  Forecast model modela now contains 4 declared exogenous variables.
- `forecast exog lrwage`
  Forecast model modela now contains 5 declared exogenous variables.
We’re ready to solve the model, providing a suffix for the forecast variables, and indicating that the dynamic forecasts should begin in 2007q1:

```
. forecast solve, suf(_modela) begin(tq(`begfc`))
```

Computing dynamic forecasts for model modela.

Starting period: 2007q1
Ending period: 2010q3
Forecast suffix: _modela

2007q1: ..............
2007q2: ..............
2007q3: ..............
2007q4: ..............
2008q1: ..............
2008q2: ..............
2008q3: ..............
2008q4: ..............
2009q1: ..............
2009q2: ..............
2009q3: ..............
2009q4: ..............
2010q1: ..............
2010q2: ..............
2010q3: ..............

Forecast 8 variables spanning 15 periods.
To illustrate the forecast results, we produce shorter labels for the variables of interest and use `tsline` to graph the actual values and forecast values, combining these into a single graph:

```
. lab var rconsump "C"
. lab var rconsump_modela "C_pred"
. lab var rgrossinv "I"
. lab var rgrossinv_modela "I_pred"
. lab var rgdp "GDP"
. lab var rgdp_modela "GDP_pred"
. lab var primerate "r"
. lab var primerate_modela "r_pred"
. loc tograph rconsump rgrossinv rgdp primerate
. foreach v of loc tograph {
  2. tsline `v´ `v´_modela if tin(`begfc´,), nodraw ///
  > ylab(,angle(0)) scheme(s2mono) name(`v´, replace)
  3. }
. graph combine `tograph´, ti("ModelA baseline forecasts")
```
Model specification, solution and dynamic forecasting

Model A baseline forecasts

- **C** (solid line)
- **C_pred** (dashed line)

- **I** (solid line)
- **I_pred** (dashed line)

- **GDP** (solid line)
- **GDP_pred** (dashed line)

- **r** (solid line)
- **r_pred** (dashed line)
The model’s dynamic forecasts fail to capture the significant decline in investment spending that occurred during the financial crisis, and as a result predict that GDP would be considerably higher than that experienced during the *ex ante* forecast period. Its forecasts of real consumption spending are considerably more accurate, as are forecasts of the prime rate after early 2008.