1 Introduction

Panel data, cross-sectional timeseries or longitudinal data are observations on a panel of i units or cases over t time periods. Most panel data commands start with xt For an overview of panel data type help xt. A typical panel data might record data on the income and expenditure of a group of individuals repeated over a number of years.

These notes present the annotated log of a STATA session demonstrating the use of many of these commands. The data sets used are those used in the STATA cross-sectional time series reference manual. This note should be regarded as an introduction to that manual and to the STATA on-line help files which give comprehensive descriptions of the facilities in STATA for cross-sectional time series analysis.

To obtain the optimum benefit from these notes I would recommend that one should work through the STATA session with a copy of Wooldridge (2002) (or Cameron and Trivedi
The references at the end of this note are to books on panel data analysis or on the use of Stata in econometrics. These panel data books are not always easy going and are suitable for those undertaking quantitative research using panel data. If you have problems you might consult one of the more general texts recommended for your course as an introduction to the topic.
xtpoisson Fixed-effects, random-effects, & population-averaged Poisson models
xtnbreg Fixed-effects, random-effects, & population-averaged negative binomial models
xtnbreg Fixed-effects, random-effects, & population-averaged negative binomial models
xtmelogit Multilevel mixed-effects logistic regression
xtmepoisson Multilevel mixed-effects Poisson regression
xtgee Population-averaged panel-data models using GEE

Panel datasets have the form $x_{it}$, where $x_{it}$ is a vector of observations for unit $i$ and time $t$. The particular commands (such as xtdescribe, xtsum, and xtreg) are documented in their own help file entries. This entry deals with concepts common across commands.

The xtset command sets the panel variable and the time variable. Most xt commands require that the panel variable be specified, and some require that the time variable also be specified. Once you xtset your data, you need not do it again. The xtset information is stored with your data.

If you have previously tsset your data by using both a panel and a time variable, these settings will be recognized by xtset, and you need not xtset your data.

Example

An xt dataset:

<table>
<thead>
<tr>
<th>pid</th>
<th>yr_visit</th>
<th>fev</th>
<th>age</th>
<th>sex</th>
<th>height</th>
<th>smokes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1071</td>
<td>1991</td>
<td>1.21</td>
<td>25</td>
<td>1</td>
<td>69</td>
<td>0</td>
</tr>
<tr>
<td>1071</td>
<td>1992</td>
<td>1.52</td>
<td>26</td>
<td>1</td>
<td>69</td>
<td>0</td>
</tr>
<tr>
<td>1071</td>
<td>1993</td>
<td>1.32</td>
<td>28</td>
<td>1</td>
<td>68</td>
<td>0</td>
</tr>
<tr>
<td>1072</td>
<td>1991</td>
<td>1.33</td>
<td>18</td>
<td>1</td>
<td>71</td>
<td>1</td>
</tr>
<tr>
<td>1072</td>
<td>1992</td>
<td>1.18</td>
<td>20</td>
<td>1</td>
<td>71</td>
<td>1</td>
</tr>
<tr>
<td>1072</td>
<td>1993</td>
<td>1.19</td>
<td>21</td>
<td>1</td>
<td>71</td>
<td>0</td>
</tr>
</tbody>
</table>

The other xt commands need to know the identities of the variables identifying patient and time. You could type

```
. xtset pid yr_visit
```

Also see

Manual: [XT] xt

Online: [XT] xtset

Load the data set nlswork.dta.

```
. use nlswork, clear
. describe
```

Contains data

National Longitudinal Survey, Young Women 14-26 years of age in 1968
To start one must set the indices i (units) and t (time). As already described this can be done in Stata 10 using the `xtset` command.\(^2\) Examples of the commands follow.

```
. tsset idcode year
    panel variable:  idcode, 1 to 5159
    time variable:  year, 68 to 88, but with gaps

. tsset
    panel variable:  idcode, 1 to 5159
    time variable:  year, 68 to 88, but with gaps
```

### 2 Examining Panel Data

The instructions in this section may be used to extract various panel properties of a panel data set. The temptation may be to skip this material and move to the estimation instructions in the next section. You should, at least, have a look at it as these instructions may be very useful in examining the kind of panel data found in real examples.

`xtdescribe` describes the participation pattern in panel data. We have 4711 women in the survey. The maximum number of years over which any women is observed is 15. The most common pattern is participation in only the first year (136 or 2.89% are observed in this pattern). The bottom line of the table give the totals for participation patterns not observed. The `pattern(#)` option allows one to increase the number of patterns shown.

\(^2\)In earlier versions of Stata one used the `iis tis` commands, the `i()` `t()` options or the `tsset` command. While these can still be used it is recommended that one use the newer version 10 methods.
The table below shows the distribution of $T_i$ with the specified frequency, percent, and cumulative percentages. The patterns are also listed for each frequency range.

<table>
<thead>
<tr>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>136</td>
<td>2.89</td>
<td>2.89</td>
<td>1....................</td>
</tr>
<tr>
<td>114</td>
<td>2.42</td>
<td>5.31</td>
<td>....................1</td>
</tr>
<tr>
<td>89</td>
<td>1.89</td>
<td>7.20</td>
<td>.................1.11</td>
</tr>
<tr>
<td>87</td>
<td>1.85</td>
<td>9.04</td>
<td>....................1</td>
</tr>
<tr>
<td>86</td>
<td>1.83</td>
<td>10.87</td>
<td>111111.1.11.11.11</td>
</tr>
<tr>
<td>61</td>
<td>1.29</td>
<td>12.16</td>
<td>....................11</td>
</tr>
<tr>
<td>56</td>
<td>1.19</td>
<td>13.35</td>
<td>11...................</td>
</tr>
<tr>
<td>54</td>
<td>1.15</td>
<td>14.50</td>
<td>....................1.11</td>
</tr>
<tr>
<td>3974</td>
<td>84.36</td>
<td>100.00</td>
<td>(other patterns)</td>
</tr>
<tr>
<td>4711</td>
<td>100.00</td>
<td></td>
<td>XXXXX.X.XX.XXX.XXX</td>
</tr>
</tbody>
</table>

- xtdes

\[
\text{idcode: } 1, 2, \ldots, 5159 \quad n = 4711
\]
\[
\text{year: } 68, 69, \ldots, 88 \quad T = 15
\]
\[
\Delta(\text{year}) = 1; (88-68)+1 = 21
\]
(idcode*year uniquely identifies each observation)

**Distribution of $T_i$:**

<table>
<thead>
<tr>
<th>min</th>
<th>5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>9</td>
<td>13</td>
<td>15</td>
</tr>
</tbody>
</table>

- xtdes, pattern(20)

\[
\text{idcode: } 1, 2, \ldots, 5159 \quad n = 4711
\]
\[
\text{year: } 68, 69, \ldots, 88 \quad T = 15
\]
\[
\Delta(\text{year}) = 1; (88-68)+1 = 21
\]
(idcode*year uniquely identifies each observation)

**Distribution of $T_i$:**

<table>
<thead>
<tr>
<th>min</th>
<th>5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>9</td>
<td>13</td>
<td>15</td>
</tr>
</tbody>
</table>

- xtdes, pattern(20)

\[
\text{idcode: } 1, 2, \ldots, 5159 \quad n = 4711
\]
\[
\text{year: } 68, 69, \ldots, 88 \quad T = 15
\]
\[
\Delta(\text{year}) = 1; (88-68)+1 = 21
\]
(idcode*year uniquely identifies each observation)

**Distribution of $T_i$:**

<table>
<thead>
<tr>
<th>min</th>
<th>5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>9</td>
<td>13</td>
<td>15</td>
</tr>
</tbody>
</table>

- xtdes, pattern(20)

\[
\text{idcode: } 1, 2, \ldots, 5159 \quad n = 4711
\]
\[
\text{year: } 68, 69, \ldots, 88 \quad T = 15
\]
\[
\Delta(\text{year}) = 1; (88-68)+1 = 21
\]
(idcode*year uniquely identifies each observation)

**Distribution of $T_i$:**

<table>
<thead>
<tr>
<th>min</th>
<th>5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>9</td>
<td>13</td>
<td>15</td>
</tr>
</tbody>
</table>
xtsum generalizes summarize by reporting means and standard for panel data. It differs from summarize in that it decomposes the standard deviation into between and within components. The between figures refers to the standard deviation, minimum and maximum of the averages for each individual ($\bar{x}_i$) (4710 individuals). The within figure calculates the statistics for the deviations of each individual from his own average ($x_{ij} - \bar{x}_i + \bar{x}$) 28467 observations. Thus is for each person you calculate the average hours worked and then calculate the standard deviation of these averages you will get 7.846585. If, for each time period and each person you calculate his deviation from his own average and then calculate the standard deviation of all these 28467 adjusted observations yo get 7.520712. (The addition of an overall average to the individual deviations does not change the variance but does impinge on the explanation of the maximum and minimum. If you find the meaning of between and within a bit difficult leave it and return to it later)

```
. summ hours
      Variable | Obs  Mean  Std. Dev.  Min  Max
-----------------+------------------------------------------
hours | 28467  36.55956  9.869623   1  168
```

```
. xtab
      Variable | Mean  Std. Dev.  Min  Max | Observations
-----------------+--------------------------------------------+----------------
hours overall | 36.55956  9.869623   1   168 | N = 28467
between | 7.846585   1   83.5 | n = 4710
within | 7.520712   2.154726  130.0596 | T-bar = 6.04395
```

tabulate by performing one-way tabulations and giving details of between and within frequencies. 3113 in the between category is the number of women who showed a 0 in some year in the sample while 3643 is the number who showed a 1. As there were only 4711 in the sample we can see that many changed status during the sample period. The 55.06% is the fraction of time, on average, a women recorded 0 given that she recorded 0 in some period. Similarly 71.90 is the fraction of time, on average, that a woman recorded 1 given that she recorded 1 in some period. For comparison purposes xtab is repeated for race, a characteristic which does not change over the period.

```
. tab msp

  1 if | married, spouse
```
xttab msp

<table>
<thead>
<tr>
<th>msp</th>
<th>Overall</th>
<th>Between</th>
<th>Within</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Freq.</td>
<td>Percent</td>
<td>Freq.</td>
</tr>
<tr>
<td>0</td>
<td>11,324</td>
<td>39.71</td>
<td>3113</td>
</tr>
<tr>
<td>1</td>
<td>17,194</td>
<td>60.29</td>
<td>3643</td>
</tr>
<tr>
<td>Total</td>
<td>28,518</td>
<td>100.00</td>
<td>6756</td>
</tr>
</tbody>
</table>

(n = 4711)

**xttrans** looks at transitions from one state to another over time. Thus 80.49% of the women who were 0 in one year were also zero in the next recorded year while 19.51% changed to 1. As we have seen above this data set is not balanced (i.e. There are not observations for all persons for all years. Thus some of the transitions may be over a period of years. To solve this problem we need to fill in the missing observations with NAs. This is accomplished by the **fillin** command. Rerunning the **xttrans** command gives appropriate estimates of the transition probabilities.

* xttrans msp

<table>
<thead>
<tr>
<th>married, 1 if married, spouse present</th>
<th>0</th>
<th>1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>80.49</td>
<td>19.51</td>
<td>100.00</td>
</tr>
<tr>
<td>1</td>
<td>7.96</td>
<td>92.04</td>
<td>100.00</td>
</tr>
<tr>
<td>Total</td>
<td>37.11</td>
<td>62.89</td>
<td>100.00</td>
</tr>
</tbody>
</table>

* xttrans msp, freq /* Does not normalize for missing time periods */

<table>
<thead>
<tr>
<th>married, 1 if married, spouse present</th>
<th>0</th>
<th>1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7,697</td>
<td>1,866</td>
<td>9,563</td>
</tr>
<tr>
<td>1</td>
<td>80.49</td>
<td>19.51</td>
<td>100.00</td>
</tr>
<tr>
<td>1</td>
<td>1,133</td>
<td>13,100</td>
<td>14,233</td>
</tr>
<tr>
<td>1</td>
<td>7.96</td>
<td>92.04</td>
<td>100.00</td>
</tr>
<tr>
<td>Total</td>
<td>8,830</td>
<td>14,966</td>
<td>23,796</td>
</tr>
<tr>
<td>37.11</td>
<td>62.89</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>

* * Rectangularize the data
  fillin idcode year
At this stage we might mention the \texttt{reshape} command. To illustrate this we load a small artificial data set which is listed in the computer output below. This is in what is known as long format. In long format each record has two identifiers. Here the identifiers are the \texttt{idcode} and the year. There are two variables, wage and tax, and 12 observations on each variable. In wide format as shown in the second list command there are three records indexed by \texttt{idcode}. There are four wage (wage2001, wage2002, wage2003 and wage2004) and four tax variables. This allows one to summarise data for each year if required. On some occasions one may find panel data on an excel spreadsheet in this format and may need to transform it to long format for analysis. Transfer from long to wide format is accomplished by \texttt{reshape wide} and the reverse transfer by \texttt{reshape long} in Stata 10 details of the \texttt{reshape} command are given in the Data manual.

\texttt{xtline} draws line plots for panel data. detail of wide and long

```
//
insheet using long.csv, comma clear
xtset idcode year
describe
summarize
list
// Switch to wide format
reshape wide wage tax , i(idcode) j(year)
describe
summarize
list
// Switch back to narrow format
reshape long wage tax , i(idcode) j(year)
describe
summarize
list
//
```

3 Estimation using \texttt{xtreg}

3.1 Introduction

The basic linear unobserved effects panel data model is

\[ y_{it} = X_{it} \beta + c_i + u_{it} \] (1)
(For a full explanation of the symbols see Wooldridge page 251, etc.). In equation (1) \( c_i \) is the unit specific residual and differs between units but not across time within units. Averaging equation 1 over time we get

\[
\bar{y}_i = \bar{X}_i \beta + c_i + \bar{u}_i
\]  

(2)

Subtracting equation (2) from equation (1) gives equation (3) which does not include the unit specific effect.

\[
(y_{it} - \bar{y}_i) = (X_{it} - \bar{X}_i) \beta + (u_{it} - \bar{u}_i)
\]  

(3)

These three equations form the basis for the various ways of estimating \( \beta \).

- \( \text{xtreg ...,fe} \) gives the fixed effects or within estimator of \( \beta \) and is derived from equation (3). It is equivalent to performing \( OLS \) on equation (3).

- \( \text{xtreg ...,be} \) gives the between effects and corresponds to \( OLS \) estimation of equation (2).

- \( \text{xtreg ...,re} \) gives the random effects estimator and is a weighted average of the within and between effects estimator. The random effects estimator is equivalent to estimating

\[
(y_{it} - \theta \bar{y}_i) = (X_{it} - \theta \bar{X}_i) \beta + (1 - \theta)c_i + (u_{it} - \theta \bar{u}_i)
\]  

(4)

where \( \theta \) is a function of \( \sigma^2_c \) and \( \sigma^2_u \).

- \( \text{xtreg ...,mle} \) produces maximum likelihood estimates of the random effects estimator.

For additional options available with the \( \text{xtreg} \) command see the on-line help files or the \text{STATA} manuals.

\[
.y \ /
. // Using xtreg
. //
. use nlswork, clear
(National Longitudinal Survey. Young Women 14-26 years of age in 1968)

. generate age2 = age^2
(24 missing values generated)

. generate ttl_exp2 = ttl_exp^2

. generate tenure2 = tenure^2
(433 missing values generated)

. generate byte black = race==2

3.2 Pooled of Stacked OLS

For comparison purposes, we first estimate the model using OLS. Some might refer to this model as a stacked or pooled OLS estimate.

\[
.y \ /
. // OLS
. //
. regress ln_wage grade age* ttl_exp* tenure* black not_smsa south

9
3.3 Fixed Effects Estimator

Next we have a fixed effects estimator first assuming heteroscedasticity and then with a robust estimator of the variance covariance matrix. Note at the bottom of the the first table the F-test that all the fixed effects are zero. This test is not given in the table containing the results of the robust estimation procedure as the equivalent robust statistic is difficult to calculate. Finally there is a table summarising the results of the OLS and fixed effects estimator. There is also a Stata command `areg` which may be used for OLS estimation when one has a large number of dummy variables. `areg` is designed for datasets with many groups, but not a number of groups that increases with the sample size. `xtreg, fe` can handle the case in which the number of groups increases with the sample size. With the addition of the `vce()` options to the `xtreg, fe` command access to the `areg` command is not now required for that purpose.
### 3.4 Between Effects Estimator

We now calculate a between effects estimator.
Between Effects Estimator

.xtreg ln_wage grade age* ttl_exp* tenure* black not_smsa south, be

Between regression (regression on group means) Number of obs = 28091
Group variable: idcode Number of groups = 4697
R-sq: within = 0.1591 Obs per group: min = 1
between = 0.4900 avg = 6.0
overall = 0.3695 max = 15
F(10,4686) = 450.23 Prob > F = 0.0000

sd(u_i + avg(e_i.))=.3036114

<table>
<thead>
<tr>
<th>ln_wage</th>
<th>Coef. Std. Err. t P&gt;</th>
<th>t</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>grade</td>
<td>.0607602 .0020006 30.37 0.000 .0568382 .0646822</td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>.0323158 .0087251 3.70 0.000 .0152105 .0494211</td>
<td></td>
<td></td>
</tr>
<tr>
<td>age2</td>
<td>-.0005997 .0001429 -4.20 0.000 -.0027598 .00250108</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ttl_exp</td>
<td>.0138853 .0056749 2.45 0.014 .0027598 .0250108</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ttl_exp2</td>
<td>.0007342 .0003267 2.25 0.025 .0000936 .0013747</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tenure</td>
<td>.0698419 .0060729 11.50 0.000 .0579361 .0817476</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tenure2</td>
<td>-.0028756 .0004098 -7.02 0.000 -.0036789 -.0020722</td>
<td></td>
<td></td>
</tr>
<tr>
<td>black</td>
<td>-.0564167 .0105131 -5.37 0.000 -.1770272 -.0358061</td>
<td></td>
<td></td>
</tr>
<tr>
<td>not_smsa</td>
<td>-.1860406 .0112495 -16.54 0.000 -.2080949 -.1639862</td>
<td></td>
<td></td>
</tr>
<tr>
<td>south</td>
<td>-.0993378 .010136 -9.80 0.000 -.1192091 -.0794665</td>
<td></td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>.3339113 .1210434 2.76 0.006 .0966093 .5712133</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.5 Random Effects Estimator

Four examples of the random effects estimator follow — random effects, random effects
with robust errors, Maximum likelihood. Note that the maximum likelihood estimator give
a test for zero variance of the random effects.

Random Effects Estimators

.xtreg ln_wage grade age* ttl_exp* tenure* black not_smsa south, re

Random-effects GLS regression Number of obs = 28091
Group variable: idcode Number of groups = 4697
R-sq: within = 0.1715 Obs per group: min = 1
between = 0.4784 avg = 6.0
overall = 0.3708 max = 15
Random effects u_i ~ Gaussian Wald chi2(10) = 9244.87
corr(u_i, X) = 0 (assumed) Prob > chi2 = 0.0000

<table>
<thead>
<tr>
<th>ln_wage</th>
<th>Coef. Std. Err. z P&gt;</th>
<th>z</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>grade</td>
<td>.0646499 .0020006 30.37 0.000 .0568382 .0646822</td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>.0323158 .0087251 3.70 0.000 .0152105 .0494211</td>
<td></td>
<td></td>
</tr>
<tr>
<td>age2</td>
<td>-.0005997 .0001429 -4.20 0.000 -.0027598 .00250108</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ttl_exp</td>
<td>.0138853 .0056749 2.45 0.014 .0027598 .0250108</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ttl_exp2</td>
<td>.0007342 .0003267 2.25 0.025 .0000936 .0013747</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

12
tenure | .039252   .001755      22.36   0.000   .0358114   .0426927
tenure2 | -.0020035     .0001193  -16.80  0.000  -.0022373  -.0017697
black   | -.0530532    .0099924   -5.31   0.000  -.0728397  -.0332667
not_smsa| -.1308263     .0071751  -18.23  0.000  -.1448891  -.1167634
south   | -.0868927    .0073031   -11.90  0.000  -.1012066  -.0725788
_cons   | .2387209     .0494688    4.83   0.000  .1417639  .335678

 sigma_u  | .25790313
 sigma_e  | .29069544
rho      | .44043812 (fraction of variance due to u_i)

. estimates store m_re
. qui xtreg ln_wage grade age* ttl_exp* tenure* black not_smsa south, re vce(ro bust)
. estimates store m_re_ro
. xtreg ln_wage grade age* ttl_exp* tenure* black not_smsa south, mle

Fitting constant-only model:
Iteration 0:  log likelihood = -13690.161
Iteration 1:  log likelihood = -12819.317
Iteration 2:  log likelihood = -12662.039
Iteration 3:  log likelihood = -12649.744
Iteration 4:  log likelihood = -12649.614
Iteration 5:  log likelihood = -12649.614

Fitting full model:
Iteration 0:  log likelihood = -8922.145
Iteration 1:  log likelihood = -8853.6409
Iteration 2:  log likelihood = -8853.4255
Iteration 3:  log likelihood = -8853.4254

Random-effects ML regression
Number of obs  = 28091
Group variable: idcode
Number of groups = 4697
Random effects u_i ~ Gaussian
Obs per group: min  =  1
               avg  =  6.0
               max  =  15
LR chi2(10)    =  7592.38
Prob > chi2     =  0.0000

------------------------------------------------------------------------------
|     ln_wage  | Coef.      Std. Err.  z    P>|z|    [95% Conf. Interval]
-------------+----------------------------------------------
  grade       |  .0646093   .0017372  37.19  0.000   .0612044  .0680142
  age         |  .0368531   .0031226  11.80  0.000   .030733   .0429732
  age2        | -.0007132   .000501   -14.24  0.000  -.0010133  -.0004132
  ttl_exp     |  .0288196   .0024143  11.94  0.000   .0240877   .0335515
  ttl_exp2    |  .000309    .0001163   2.66  0.010   .0000811   .0005369
  tenure      |  .0394371   .0017604  22.40  0.000   .0359868   .0428875
  tenure2     | -.0020052   .0001195  -16.77  0.000  -.0022395  -.0017709
  black       | -.0533394   .0097338  -5.48  0.000   -.0724172  -.0342615
  not_smsa    | -.1308263   .0071751  -18.23  0.000  -.1448891  -.1167634
  south       | -.0868927   .0073031  -11.90  0.000  -.1012066  -.0725788
  _cons       |  .2387209   .0494688   4.83  0.000   .1417639   .335678
-------------+----------------------------------------------
 /sigma_u     |  .2485556   .0035017  .2417863  .2555144
 /sigma_e     |  .2906954   .001352   .289208   .2945076
------------------------------------------------------------------------------
<table>
<thead>
<tr>
<th>Variable</th>
<th>m_re</th>
<th>m_re_ro</th>
<th>m_mle</th>
</tr>
</thead>
<tbody>
<tr>
<td>grade</td>
<td>.0646</td>
<td>.0646</td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>.0368</td>
<td>.0368</td>
<td></td>
</tr>
<tr>
<td>age2</td>
<td>-.00071</td>
<td>-.00071</td>
<td></td>
</tr>
<tr>
<td>ttl_exp</td>
<td>.00176</td>
<td>.00185</td>
<td></td>
</tr>
<tr>
<td>ttl_exp2</td>
<td>.00012</td>
<td>.00013</td>
<td></td>
</tr>
<tr>
<td>tenure</td>
<td>.0012</td>
<td>.0012</td>
<td></td>
</tr>
<tr>
<td>tenure2</td>
<td>.000099</td>
<td>.000983</td>
<td></td>
</tr>
<tr>
<td>black</td>
<td>-.0531</td>
<td>-.0531</td>
<td></td>
</tr>
<tr>
<td>not_smsa</td>
<td>-.131</td>
<td>-.131</td>
<td></td>
</tr>
<tr>
<td>south</td>
<td>-.0869</td>
<td>-.0869</td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>.239</td>
<td>.239</td>
<td></td>
</tr>
<tr>
<td>ln_wage</td>
<td>grade</td>
<td>.0646</td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>.0369</td>
<td>.0369</td>
<td></td>
</tr>
<tr>
<td>age2</td>
<td>-.00071</td>
<td>-.00071</td>
<td></td>
</tr>
<tr>
<td>ttl_exp</td>
<td>.00176</td>
<td>.00176</td>
<td></td>
</tr>
<tr>
<td>ttl_exp2</td>
<td>.00012</td>
<td>.00012</td>
<td></td>
</tr>
<tr>
<td>tenure</td>
<td>.000099</td>
<td>.000099</td>
<td></td>
</tr>
<tr>
<td>tenure2</td>
<td>-.0531</td>
<td>-.0531</td>
<td></td>
</tr>
<tr>
<td>black</td>
<td>-.0531</td>
<td>-.0531</td>
<td></td>
</tr>
<tr>
<td>not_smsa</td>
<td>-.131</td>
<td>-.131</td>
<td></td>
</tr>
<tr>
<td>south</td>
<td>-.0869</td>
<td>-.0869</td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>.239</td>
<td>.239</td>
<td></td>
</tr>
</tbody>
</table>

| sigma_u | _cons | .249 |
Note that STATA has no direct command for two way fixed effects. If you wish to also introduce a second set of fixed effects for, say, time periods create a set of appropriate dummy variables for inclusion in your regressions and use a one way estimator.

4 Testing after xtreg

Note that the post estimation commands test, testnl, estimates, lincom, lrtest, mfx, nlncom, predict, predictnl and hausman are also available after xtreg. The command xttest0 is the Breusch and Pagan LM test for random effects. We can also do a Hausman test. When the common effects are orthogonal to the regressors both fixed and random effect estimators are consistent (but fixed effects are not efficient) under the alternative fixed effects are consistent whereas random effects are not. The Hausman test looks at the difference between the coefficients estimated using a random effects estimator and a fixed effects estimator. Roughly speaking in both estimates are similar we can use a random effects estimator. If they are different one may use the fixed effects estimator. The format of the hausman instruction is

```
hausman m_consistent m_efficient
```

where m_consistent and m_efficient are estimates of two models that have been estimated and saved (estimates store). m_consistent is the estimate that is consistent in both cases but not efficient and m_efficient is efficient where the first model is not efficient. In this case the estimates are significantly different and we reject the random effects estimator.

```
/* After xtreg, re */
// After the random effects estimate

. * Breusch & Pagan score test for random effects
. xttest0

Breusch and Pagan Lagrangian multiplier test for random effects:

\[ \ln_wage[idcode, t] = Xb + u[idcode] + e[idcode, t] \]

Estimated results:

<table>
<thead>
<tr>
<th>Var</th>
<th>sd = sqrt(Var)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\ln_wage)</td>
<td>.2283326</td>
</tr>
<tr>
<td>e</td>
<td>.0845038</td>
</tr>
<tr>
<td>u</td>
<td>.066514</td>
</tr>
</tbody>
</table>

Test: Var(u) = 0
\[ \chi^2(1) = 14779.98 \]
\[ \text{Prob} > \chi^2 = 0.0000 \]
. * Hausman specification test (compares fe and re)
qui xtreg ln_wage grade age age2 ttl_exp ttl_exp2 tenure tenure2 not_smsa south, fe
    F(4696, 23386) = 5.19         Prob > F = 0.0000
. estimates store fe
qui xtreg ln_wage grade age age2 ttl_exp ttl_exp2 tenure tenure2 not_smsa south, re
. estimates store re
. hausman fe re

<table>
<thead>
<tr>
<th></th>
<th>(b)</th>
<th>(B)</th>
<th>(b-B)</th>
<th>sqrt(diag(V_b-V_B))</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>.0359987</td>
<td>.0363062</td>
<td>-.0003075</td>
<td>.0013183</td>
<td></td>
</tr>
<tr>
<td>age2</td>
<td>-.000723</td>
<td>-.000705</td>
<td>-.000018</td>
<td>.0000184</td>
<td></td>
</tr>
<tr>
<td>ttl_exp</td>
<td>.0334668</td>
<td>.0292321</td>
<td>.0042347</td>
<td>.0017085</td>
<td></td>
</tr>
<tr>
<td>ttl_exp2</td>
<td>.0002163</td>
<td>.0002946</td>
<td>.0000783</td>
<td>.0000529</td>
<td></td>
</tr>
<tr>
<td>tenure</td>
<td>.0357539</td>
<td>.0390983</td>
<td>-.0033444</td>
<td>.0005789</td>
<td></td>
</tr>
<tr>
<td>tenure2</td>
<td>-.0019701</td>
<td>-.0020014</td>
<td>.0000313</td>
<td>.0000372</td>
<td></td>
</tr>
<tr>
<td>not_smsa</td>
<td>-.0890108</td>
<td>-.1268961</td>
<td>.0378853</td>
<td>.0063038</td>
<td></td>
</tr>
<tr>
<td>south</td>
<td>-.0606309</td>
<td>-.094716</td>
<td>.0340851</td>
<td>.008259</td>
<td></td>
</tr>
</tbody>
</table>

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(8) = (b-B)'[(V_b-V_B)^(-1)](b-B)
    = 142.53
    Prob>chi2 = 0.0000

5 Prediction after xtreg

After xtreg the predict command has the following options

- **xb** $x_j b$, fitted values
- **stdp** standard error of the fitted values
- **ue** $\hat{c}_i + \hat{u}_{it}$, combined residual
- **xbu** $x_j b + \hat{c}_i$, prediction including effect
- **u** $c_i$, The fixed or random effect component
- **e** $u_{it}$, the error component

The last three of these are only available within sample
6 Other Stata Panel estimators

Here we mention some more detail regarding other Stata panel data commands. A complete list has already been given in Section (1). Details are given in the manuals ot the help files. Before using any of these commands one should be familiar with the theory involved.

6.1 Faster estimation of alternative models using \texttt{xtdata}

\texttt{xtdata varlist ...} produces a converted data set of the variables specified or, if varlist is not specified, all the variables in the data. Once converted, Stata’s ordinary regress command may be used to perform various panel data regressions more quickly than using \texttt{xtreg}. Before using \texttt{xdata} you must eliminate any variables that you do not intend to use and that have missing values. After converting the data, with \texttt{xdata} you may form linear transformations of the regressors. All nonlinear transformations of the data must be done before conversion. The gain in this instruction is where one needs to do a large specification search. The following stata commands illustrate this procedure. Output is suppressed.

\begin{verbatim}
. qui xtdata ln\_wage grade age* ttl\_exp* tenure* black not\_smsa south, fe clear
. qui regress ln\_wage grade age ttl\_exp tenure black not\_smsa south
. qui regress ln\_wage grade age* ttl\_exp* tenure* black not\_smsa south
. qui xtdata ln\_wage grade age* ttl\_exp* tenure* black not\_smsa south, re ratio(.95) clear
. qui regress ln\_wage constant grade age ttl\_exp tenure black not\_smsa south, noconstant
. qui xtdata ln\_wage grade age* ttl\_exp* tenure* black not\_smsa south, fe clear
. qui regress ln\_wage grade age* ttl\_exp* tenure* black not\_smsa south
\end{verbatim}

6.2 More general error structures

\texttt{xtregar} fits fixed affects and random effects models where the disturbance follows an AR(1) process i.e.

\[
\gamma_{it} = X_{it}\beta + c_{i} + u_{it}
\]

\[
u_{it} = \rho u_{i,t-1} + \eta_{it}
\]

and \(\eta_{it}\) are independent \(N(0,\sigma^2_\eta)\).

\texttt{xtpcse} and \texttt{xtgls} estimate panel models under various assumptions about heterogeneity of variances across panels and possible serial correlation.

6.3 Dynamic panel data

The commands of interest here are \texttt{xtabond}, \texttt{xtdpsys} and \texttt{xtdpd}. There are considerable changes here relative to earlier versions of Stata.

6.4 Limited Dependent Variables in Panel Data

There are a variety of LDV estimation commands corresponding to the standard methods. These include \texttt{xttobit}, \texttt{xtintreg}, \texttt{xtlogit}, \texttt{xtprobit}, \texttt{xtclogit}, \texttt{xtpoisson}, \texttt{xtnreg}, \texttt{xtmelogit} and \texttt{xtmepoisson}.
References


