The Stata Journal publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go “beyond the Stata manual” in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those teaching, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

For more information on the Stata Journal, including information for authors, see the web page

http://www.stata-journal.com

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

http://www.stata.com/bookstore/sj.html

Subscription rates:

Subscriptions mailed to US and Canadian addresses:

- 3-year subscription (includes printed and electronic copy) $153
- 2-year subscription (includes printed and electronic copy) $110
- 1-year subscription (includes printed and electronic copy) $ 59
- 1-year student subscription (includes printed and electronic copy) $ 35

Subscriptions mailed to other countries:

- 3-year subscription (includes printed and electronic copy) $225
- 2-year subscription (includes printed and electronic copy) $158
- 1-year subscription (includes printed and electronic copy) $ 83
- 1-year student subscription (includes printed and electronic copy) $ 59
- 3-year subscription (electronic only) $153

Back issues of the Stata Journal may be ordered online at

http://www.stata.com/bookstore/sj.html

The Stata Journal is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the Stata Journal, StataCorp, 4905 Lakeway Drive, College Station TX 77845, USA, or email sj@stata.com.
Estimation and testing of fixed-effect panel-data systems

J. Lloyd Blackwell, III
Department of Economics
University of North Dakota

Abstract. This paper describes how to specify, estimate, and test multiple-equation, fixed-effect, panel-data equations in Stata. By specifying the system of equations as seemingly unrelated regressions, Stata panel-data procedures worked seamlessly for estimation and testing of individual variable coefficients, but additional routines using test were needed for testing of individual equations and differences between equations.

Keywords: st0084, panel data, fixed effect, multiple equations, seemingly unrelated regressions, heteroskedasticity, autocorrelation, contemporaneous correlation, tests of linear combinations

1 Introduction

The motivation to develop a multiple-equation panel-data procedure came from a need to determine whether certain laboratory experiments with multiple panels of laboratory animals produced statistically different results. This suggested that each experiment could be treated as a separate equation in a system of seemingly unrelated regressions (SUR), although it was necessary to apply special restrictions prior to estimation. The SUR model described below is a slightly simplified version of one that was developed to estimate and test pairs of laboratory experiments.

2 Framework

The framework for this paper derives from seemingly unrelated regressions with error components (Baltagi 2001, 105–106). It is then assumed that coefficients of constant terms and quantitative independent variables require restriction across the panels in their equations, while fixed-effect dummies vary by panel (Judge et al. 1988, 456–459). Error structures of both equations below were assumed to be characterized by panel heteroskedasticity, panel autocorrelation, and contemporaneous correlation (HPAC). For simplicity, fixed effects were estimated with the dummy variable regression technique (Wooldridge 2002, 272–274). A simple model for exploration of this situation is

\[ y_{1it} = b_{10}x_{1it} + b_{11} + v_{1i} + \epsilon_{1it}; i = 1, 2, \ldots, N_1; t = 1, 2, \ldots, T \]  

\[ y_{2it} = b_{20}x_{2it} + b_{21} + v_{2i} + \epsilon_{2it}; i = N_1 + 1, \ldots, N_2; t = 1, 2, \ldots, T \]
Equations here are indicated by first subscripts, while other subscripts indicate panels and time periods, as shown. Thus (1) has $N_1$ panels, $T$ time periods, and $N_1$ fixed-effect coefficients, $v_{1,1}, v_{1,2}, \ldots, v_{1,N_1}$. Likewise, (2) has $N_2 - N_1$ panels, $T$ time periods, and the fixed-effect coefficients $v_{2,1}, v_{2,2}, \ldots, v_{2,N_2}$. Thus when (1) and (2) are pooled, we have a total of $N_2$ panels, with no repetition of time codes within any given panel. This is important because estimation fails when there is multiplicity of time codes within any panel.

The SUR matrix specification for (1) and (2) is

$$y = \begin{bmatrix} \mathbf{X}_1 & 0 \\ 0 & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \varepsilon$$

where

- $y = [y_{1,1}, 1, \ldots, y_{1,N_1}, T, y_{2,N_1+1}, 1, \ldots, y_{2,N_2}, T]'$
- $\mathbf{X}_1 = [c_1, x_1, d_{1,2}, \ldots, d_{1,N_1}]$
- $\mathbf{X}_2 = [c_2, x_2, d_{2,N_1+2}, \ldots, d_{2,N_2}]$
- $\beta_1 = [b_{1,0}^*, b_{1,1}, v_{1,2}^*, \ldots, v_{1,N_1}^*]'$
- $\beta_2 = [b_{2,0}^*, b_{2,1}, v_{2,2,N_1+2}^*, \ldots, v_{2,N_2}^*]'$
- $d_{1,2}, \ldots, d_{1,N_1}$ and $d_{2,N_1+2}, \ldots, d_{2,N_2}$ are fixed-effect dummies
- $\varepsilon = [\varepsilon_{1,1}, 1, \ldots, \varepsilon_{1,N_1}, T, \varepsilon_{2,N_1+1}, 1, \ldots, \varepsilon_{2,N_2}, T]'$
- Asterisks indicate composite variables

### 3 Data setup for estimation

In the experimental data described in the introduction, (1) and (2) each had seven panels. Thus the panels of the system were consecutively numbered 1–14, with panels 1–7 assigned to (1) and panels 8–14 assigned to (2). Deletion of one dummy from each equation then left

$$\mathbf{X}_1 = [c_1, x_1, d_2, \ldots, d_7] \quad \mathbf{X}_2 = [c_2, x_2, d_9, \ldots, d_{14}]$$

Extracting $x_1$ from $\mathbf{X}_1$ and $x_2$ from $\mathbf{X}_2$, the vector $\mathbf{x}$ was formed by stacking $x_1$ and $x_2$ to correspond with the elements of $y$. For purposes of data entry, however, it seemed advisable to unbold $y$ and $\mathbf{x}$, since Stata normally uses unbolded lowercase letters in variable names. Thus, the variables $y$ and $\mathbf{x}$, along with identifiers for panels, time codes, and equations were input into a Stata dataset. Remaining variables were then created and arranged by running the do-file:

- `tab eq, gen(eq)`
- `gen x1 = c1*x`  
- `gen x2 = c2*x`  
- `tab pnl, gen(pn)`
- `drop d1 d8`
Estimation and testing of fixed-effect panel-data systems

Table 1 shows a subset of the full dataset described above, created by selecting the first three panels and time periods from each equation, thus placing panels 1–3 in (1) and panels 8–10 in (2). Table 1 merits close scrutiny because the panel identifiers, time codes, equation codes, and contents of $X_1$ and $X_2$ accurately reflect the organization and features of the full dataset. For example, examination of $X_1$ and $X_2$ reveals that a fixed-effect dummy must be deleted from each to avert block-specific dummy variable traps. Thus estimation requires the nointercept option, as in the classic SUR specification (Judge et al. 1988, 470).

Table 1: sample dataset

<table>
<thead>
<tr>
<th></th>
<th>y</th>
<th>x</th>
<th>pnl</th>
<th>week</th>
<th>eq</th>
<th>c1</th>
<th>x1</th>
<th>d2</th>
<th>d3</th>
<th>c2</th>
<th>x2</th>
<th>d9</th>
<th>d10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>116</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>166</td>
<td>0.693</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.693</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>201</td>
<td>1.099</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1.099</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>149</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>197</td>
<td>0.693</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0.693</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>230</td>
<td>1.099</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1.099</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>126</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>151</td>
<td>0.693</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0.693</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>209</td>
<td>1.099</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1.099</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>114</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>164</td>
<td>0.693</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.693</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>225</td>
<td>1.099</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1.099</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>142</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>206</td>
<td>0.693</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.693</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>262</td>
<td>1.099</td>
<td>9</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1.099</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>120</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>160</td>
<td>0.693</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.693</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>218</td>
<td>1.099</td>
<td>10</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1.099</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

4 Estimation and testing of fixed-effect panel-data systems

Given that (1) and (2) include fixed effects, the user must choose among FGLS (xtgls), OLS with panel-corrected standard errors (PCSE) (xtpcse), or fixed-effects regression (areg or xtreg, fe). This topic is explored at length by Beck and Katz (1995), and
the key issues are neatly summarized by Wiggins (2001). However, only xtgls and xtpcse have all the options for the HPAC structure, so we will focus on those. See [XT] xtgls and [XT] xtpcse for details.

When xtgls is indicated, the command for HPAC SUR estimation is

```
.xtgls y c1 x1 d2-d7 c2 x2 d9-d14, i(pnl) t(week) p(c) c(psar1) nocon
```

When xtpcse is indicated, the commands for HPAC estimation are

```
.tsset pnl week
.xtpcse y c1 x1 d2-d7 c2 x2 d9-d14, c(psar1) nocon
```

Since our dataset has 14 panels and 9 time periods, FGLS is of course ruled out (Beck and Katz 1995, 637), so the output below (slightly edited) was produced with xtpcse:

```
 Prais-Winsten regression, correlated panels corrected standard errors (PCSEs)

Group variable: pnl  Number of obs = 126
Time variable: week  Number of groups = 14
Panels: correlated (balanced)  Obs per group: min = 9
Autocorrelation: panel-specific AR(1)  avg = 9
                   max = 9
Estimated covariances = 105  R-squared = X
Estimated autocorrelations = 14  Wald chi2(8) = X
Estimated coefficients = 16  Prob > chi2 = X

                      Panel-corrected       Std. Err.     z    P>|z|    [95% Conf. Interval]
                   Coef.  Std. Err.       z    P>|z|    [95% Conf. Interval]

 c1    111.5311    6.208398   17.96    0.000     99.36284   123.6993
 x1   88.67886    3.943791   22.49    0.000     80.94918   96.40855
 d2   27.81103    1.400501   19.86    0.000     25.0661   30.55596
 d3   4.477261    3.969675    1.13    0.259    -3.303159   12.25768
 d4  15.006611    4.364777    3.45    0.001     6.511306   3.620923
 d5  -39.5333    3.116503  -12.69    0.000    -45.64154  -33.42507
 d6  -23.79004    2.0549  -11.58    0.000    -27.81757  -19.76251
 d7   9.415222    5.07866    1.85    0.064    -5.387687   19.36921
 c2  111.6787    8.215256   13.59    0.000     95.57709   127.7803
 x2  115.4025    5.231997   22.06    0.000     105.148   125.657
 d9   89.29659    3.15385   28.66    0.000     83.1596   95.43358
 d10  -22.60235    5.72391   -3.95    0.000    -33.8210  -11.38369
 d11  -42.56708    5.499978  -7.74    0.000    -53.33684  -31.77322
 d12  -28.38979    5.906044  -4.81    0.000    -39.97442  -16.82315
 d13   29.44482    1.958789   15.03    0.000     25.60566   33.28397
 d14   78.11019    58.09246    1.34    0.179    -35.74894  191.9693

 rhos = .249791  .0870025  .1838984  .3643183  -.2943893  -.0062851
```

The R-squared and Wald statistics in the top section of the above output are obliterated for two reasons. First, they include the influences of the fixed-effect dummies, which serve only to control for the influences of unobserved variables ([R] areg, Stata 8, 87 or Stata 9, 95; Wooldridge 2003, 465–467). Second, the R-squared and Wald statistics in the xtpcse output are computed for the entire system of equations, which is meaning-
Estimation and testing of fixed-effect panel-data systems

less for either equation individually. Fortunately, individual equations can be properly evaluated with \texttt{test}, as discussed below.

The coefficients and panel-corrected standard errors in the main body of the \texttt{xtpcse} output are correctly reported and consistent, but inefficient. The reported $z$ statistics assume zero nulls and two-sided alternatives. See Baltagi (2001, 78) for a comparison of the properties of FGLS and OLS with PCSEs.

We can easily obtain Wald statistics for any linear combination of coefficients in the above output (Judge et al. 1988, 456–459). For any particular equation, we can use \texttt{test} to obtain the Wald statistic to determine the joint influence of all explanatory variables pertaining to that equation, but not including the influences of the fixed-effect dummies, as explained above. In (1), however, $x_1$ is the only explanatory variable for which we wish to know the effect on $y$, so the Wald statistic is simply the square of the $z$-score for $x_1$, and the Wald statistic and the $z$-score are equivalent. Likewise, the $z$-score for $x_2$ provides the test for (2). Again, both $x_1$ and $x_2$ are tested against zero nulls and two-sided alternatives.

The test for a structural difference requires a joint test of all corresponding coefficients from the two equations in the system, in this case including the constant terms, since these too can vary by equation. For our equations, we thus require two linear restrictions to be jointly tested against two-sided alternatives:

\begin{verbatim}
. test (c1-c2 = 0) (x1-x2 = 0)
   ( 1)  c1 - c2 = 0
   ( 2)  x1 - x2 = 0
   chi2(  2) = 137.10
   Prob > chi2 = 0.0000
\end{verbatim}

5 Acknowledgments

Sincere thanks to Brian Poi and Gustavo Sanchez for support in the development of the ideas presented here. Any remaining errors are, of course, mine alone.

6 References


**About the Author**

J. Lloyd Blackwell, III is Professor of Economics, Director of the Bureau of Business and Economic Research, and Director of the Master of Science in Applied Economics, University of North Dakota, Grand Forks, ND.