Linear Regression Models for Panel Data Using SAS, Stata, LIMDEP, and SPSS*

Hun Myoung Park, Ph.D.

Last modified on September 2009

University Information Technology Services
Center for Statistical and Mathematical Computing
Indiana University
410 North Park Avenue Bloomington, IN 47408
(812) 855-4724 (317) 278-4740
http://www.indiana.edu/~statmath

* The citation of this document should read: "Park, Hun Myoung. 2009. *Linear Regression Models for Panel Data Using SAS, Stata, LIMDEP, and SPSS*. Working Paper. The University Information Technology Services (UITS) Center for Statistical and Mathematical Computing, Indiana University."
http://www.indiana.edu/~statmath/stat/all/panel

This document summarizes linear regression models for panel data and illustrates how to estimate each model using SAS 9.2, Stata 11, LIMDEP 9, and SPSS 17. This document does not address nonlinear models (i.e., logit and probit models) and dynamic models, but focuses on basic linear regression models.

1. Introduction

Panel (or longitudinal) data are cross-sectional and time-series. There are multiple entities, each of which has repeated measurements at different time periods. U.S. Census Bureau's Census 2000 data at the state or county level are cross-sectional but not time-series, while annual sales figures of Apple Computer Inc. for the past 20 years are time series but not cross-sectional. If annual sales data of IBM, LG, Siemens, Microsoft, and AT&T during the same periods are also available, they are panel data. The cumulative General Social Survey (GSS), American National Election Studies (ANES), and Current Population Survey (CPS) data are not panel data in the sense that individual respondents vary across survey years. Panel data may have group effects, time effects, or the both, which are analyzed by fixed effect and random effect models.

1.1 Data Arrangement

A panel data set contains n entities or subjects (e.g., firms and states), each of which includes T observations measured at 1 through t time period. Thus, the total number of observations is nT. Ideally, panel data are measured at regular time intervals (e.g., year, quarter, and month). Otherwise, panel data should be analyzed with caution. A *short panel data* set has many entities but few time periods (small T), while a *long panel* has many time periods (large T) but few entities (Cameron and Trivedi 2009: 230).

Panel data have a cross-section (entity or subject) variable and a time-series variable. In Stata, this arrangement is called the long form (as opposed to the wide form). While the long form has both group (individual level) and time variables, the wide form includes either group or time variable. Look at the following data set to see how panel data are arranged. There are 6 groups

(airlines) and 15 time periods (years). The `.use` command below loads a Stata data set through TCP/IP and `in 1/20` of the `.list` command displays the first 20 observations.

```
. use http://www.indiana.edu/~statmath/stat/all/panel/airline.dta, clear
(Cost of U.S. Airlines (Greene 2003))

. list airline year load cost output fuel in 1/20, sep(20)

     +-------------------------------------------------------------+
     | airline   year      load       cost      output       fuel |
     |-------------------------------------------------------------|
  1. |       1      1   .534487    13.9471   -.0483954   11.57731 |
  2. |       1      2   .532328   14.01082   -.0133315   11.61102 |
  3. |       1      3   .547736   14.08521    .0879925   11.61344 |
  4. |       1      4   .540846   14.22863    .1619318   11.71156 |
  5. |       1      5   .591167   14.33236    .1485665   12.18896 |
  6. |       1      6   .575417    14.4164    .1602123   12.48978 |
  7. |       1      7   .594495   14.52004    .2550375   12.48162 |
  8. |       1      8   .597409   14.65482    .3297856    12.6648 |
  9. |       1      9   .638522   14.78597    .4779284   12.85868 |
 10. |       1     10   .676287   14.99343    .6018211   13.25208 |
 11. |       1     11   .605735   15.14728    .4356969   13.67813 |
 12. |       1     12    .61436   15.16818    .4238942   13.81275 |
 13. |       1     13   .633366   15.20081    .5069381   13.75151 |
 14. |       1     14   .650117   15.27014    .6001049   13.66419 |
 15. |       1     15   .625603    15.3733    .6608616   13.62121 |
 16. |       2      1   .490851   13.25215    -.652706   11.55017 |
 17. |       2      2   .473449   13.37018    -.626186   11.62157 |
 18. |       2      3   .503013   13.56404   -.4228269   11.68405 |
 19. |       2      4   .512501    13.8148   -.2337306   11.65092 |
 20. |       2      5   .566782   14.00113   -.1708536   12.27989 |
     +-------------------------------------------------------------+
```

If data are structured in the wide form, you need to rearrange data first. Stata has the `.reshape` command to rearrange a data set back and forth between the long and wide form. The following command changes from the long form to wide one so that the wide form has only six observations that have a group variable and as many variables as the time period (4*15 year).

```
. keep airline year load cost output fuel

. reshape wide cost output fuel load, i(airline) j(year)
(note: j = 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15)

Data                               long   ->   wide
-----------------------------------------------------------------------------
Number of obs.                       90   ->      6
Number of variables                   6   ->     61
j variable (15 values)             year   ->   (dropped)
xij variables:
                                   cost   ->   cost1 cost2 ... cost15
                                 output   ->   output1 output2 ... output15
                                   fuel   ->   fuel1 fuel2 ... fuel15
                                   load   ->   load1 load2 ... load15
-----------------------------------------------------------------------------
```

If you wish to rearrange the data set back to the long form, run the following command.

```
. reshape long cost output fuel load, i(airline) j(year)
```

In balanced panel data, all entities have measurements in all time periods. In a contingency table of cross-sectional and time-series variables, each cell should have only one frequency. When each entity in a data set has different numbers of observations due to missing values, the panel data are not balanced. Some cells in the contingency table have zero frequency. In

unbalanced panel data, the total number of observations is not *nT*. Unbalanced panel data entail some computational and estimation issues although most software packages are able to handle both balanced and unbalanced data.

1.2 Fixed Effect versus Random Effect Models

Panel data models examine fixed and/or random effects of entity (individual or subject) or time. The core difference between fixed and random effect models lies in the role of dummy variables (Table 1.1). If dummies are considered as a part of the intercept, this is a fixed effect model. In a random effect model, the dummies act as an error term.

A fixed group effect model examines group differences in intercepts, assuming the same slopes and constant variance across entities or subjects. Since a group (individual specific) effect is time invariant and considered a part of the intercept, u_i is allowed to be correlated to other regressors. Fixed effect models use least squares dummy variable (LSDV) and within effect estimation methods. Ordinary least squares (OLS) regressions with dummies, in fact, are fixed effect models.

Table 1.1 Fixed Effect and Random Effect Models

	Fixed Effect Model	Random Effect Model
Functional form*	$y_{it} = (\alpha + u_i) + X_{it}^{'}\beta + v_{it}$	$y_{it} = \alpha + X_{it}^{'}\beta + (u_i + v_{it})$
Intercepts	Varying across groups and/or times	Constant
Error variances	Constant	Varying across groups and/or times
Slopes	Constant	Constant
Estimation	LSDV, within effect method	GLS, FGLS
Hypothesis test	Incremental F test	Breusch-Pagan LM test

* $v_{it} \sim IID(0, \sigma_v^2)$

A random effect model, by contrast, estimates variance components for groups (or times) and error, assuming the same intercept and slopes. u_i is a part of the errors and thus should not be correlated to any regressor; otherwise, a core OLS assumption is violated. The difference among groups (or time periods) lies in their variance of the error term, not in their intercepts. A random effect model is estimated by generalized least squares (GLS) when the Ω matrix, a variance structure among groups, is known. The feasible generalized least squares (FGLS) method is used to estimate the variance structure when Ω is not known. A typical example is the groupwise heteroscedastic regression model (Greene 2003). There are various estimation methods for FGLS including the maximum likelihood method and simulation (Baltagi and Cheng 1994).

Fixed effects are tested by the (incremental) F test, while random effects are examined by the Lagrange Multiplier (LM) test (Breusch and Pagan 1980). If the null hypothesis is not rejected, the pooled OLS regression is favored. The Hausman specification test (Hausman 1978) compares fixed effect and random effect models. If the null hypothesis that the individual effects are uncorrelated with the other regressors in the model is not rejected, a random effect model is better than its fixed counterpart.

If one cross-sectional or time-series variable is considered (e.g., country, firm, and race), this is called a one-way fixed or random effect model. Two-way effect models have two sets of dummy variables for group and/or time variables (e.g., state and year).

1.3 Estimation and Software Issues

The LSDV regression, within effect model, between effect model (group or time mean model), GLS, and FGLS are fundamentally based on OLS in terms of estimation. Thus, any procedure and command for OLS is good for linear panel data models (Table 1.2).

The REG procedure of SAS/STAT, Stata `.regress` (`.cnsreg`), LIMDEP `regress$`, and SPSS `regression` commands all fit LSDV1 by dropping one dummy and have options to suppress the intercept (LSDV2). SAS, Stata, and LIMDEP can estimate OLS with restrictions (LSDV3), but SPSS cannot. In Stata, `.cnsreg` command requires restrictions defined in the `.constraint` command.

Table 1.2 Procedures and Commands in SAS, Stata, LIMDEP, and SPSS

	SAS 9.2	Stata 11	LIMDEP 9	SPSS 17
Regression (OLS)	`PROC REG`	`.regress`	`Regress$`	`Regression`
LSDV1	w/o a dummy	w/o a dummy	w/o a dummy	w/o a dummy
LSDV2	`/NOINT`	`,noconstant`	w/o `One` in `Rhs`	`/Origin`
LSDV3	`RESTRICT`	`.cnsreg`	`Cls:`	N/A
One-way fixed effect (within)	`TSCSREG /FIXONE` `PANEL /FIXONE`	`.xtreg, fe` `.areg, abs`	`Regress;Panel;Str=;Fixed$`	N/A
Two-way fixed (within effect)	`TSCSREG /FIXTWO` `PANEL /FIXTWO`	N/A	`Regress;Panel;Str=;Period=;Fixed$`	N/A
Between effect	`PANEL /BTWNG` `PANEL /BTWNT`	`.xtreg, be`	`Regress;Panel;Str=;Means$`	N/A

One-way random effect	`TSCSREG /RANONE`
	`PANEL /RANONE`
	`MIXED /RANDOM`
		`.xtreg, re`
		`.xtgls`
		`.xtmixed`
			`Regress;Panel;Str=;Random$`	N/A

Two-way random	`TSCSREG /RANTWO` `PANEL /RANTWO`	`.xtmixed`	`Regress;Panel;Str=;Period=;Random$`	N/A
Random coefficient model	`MIXED /RANDOM`	`.xtmixed` `.xtrc`	`Regress;RPM=;Str=$`	N/A

SAS, Stata, and LIMDEP also provide the procedures and commands that estimate panel data models in a convenient way (Table 1.2). SAS/ETS has the TSCSREG and PANEL procedures to estimate one-way and two-way fixed/random effect models.[1] These procedures estimate the within effect model for a fixed effect model and by default employ the Fuller-Battese method (1974) to estimate variance components for group, time, and error for a random effect model. PROC TSCSREG and PROC PANEL also support other estimation methods such as Parks (1967) autoregressive model and Da Silva moving average method.

PROC TSCSREG can handle balanced data only, whereas PROC PANEL is able to deal with balanced and unbalanced data. PROC PANEL requires each entity (subject) has more than one observation. PROC TSCSREG provides one-way and two-way fixed and random effect models,

[1] PROC PANEL was an experimental procedure in 9.13 but becomes a regular procedure in 9.2. SAS 9.13 users need to download and install PROC PANEL from http://www.sas.com/apps/demosdownloads/setupintro.jsp.

while PROC PANEL supports the between effect model (/BTWNT and /BTWNG) and pooled OLS regression (/POOLED) as well. PROC PANEL has BP and BP2 options to conduct the Breusch-Pagen LM test for random effects, while PROC TSCSREG does not. [2] Despite advanced features of PROC PANEL, the output of the two procedures is similar. PROC MIXED is also able to fit random effect and random coefficient (parameter) models and supports maximum likelihood estimation that is not available in PROC PANEL and TSCSREG.

The Stata `.xtreg` command estimates a within effect (fixed effect) model with the `fe` option, a between effect model with `be`, and a random effect model with `re`. This command, however, does not directly fit two-way fixed and random effect models.[3] The `.areg` command with the `absorb` option, equivalent to the `.xtreg` with the `fe` option, fits the one-way within effect model that has a large dummy variable set. A random effect model can be also estimated using the `.xtmixed` command. Stata has `.xtgls` that fits panel data models with heteroscedasticity across groups and/or autocorrelation within groups.

The LIMDEP `Regress$` command with the `Panel` subcommand estimates panel data models. The `Fixed effect` subcommand fits a fixed effect model, `Random effect` estimates a random effect model, and `Means` is for a between effect model. SPSS has limited ability to analyze panel data.

1.4 Data Sets

This document uses two data sets. A cross-sectional data set contains research and development (R&D) expenditure data of the top 50 information technology firms presented in *OECD Information Technology Outlook 2004*. A panel data set has cost data for U.S. airlines (1970-1984), which are used in *Econometric Analysis* (Greene 2003). See the Appendix for the details.

[2] However, BP and BP2 produce invalid Breusch-Pagan statistics in cases of unbalanced data. http://support.sas.com/documentation/cdl/en/etsug/60372/HTML/default/etsug_panel_sect041.htm.

[3] You may fit the two-way fixed effect model by including a set of dummies and using the `fe` option. For the two-way random effect model, you need to use the `.xtmixed` command instead of `.xtreg`.

2. Least Squares Dummy Variable Regression

A dummy variable is a binary variable that is coded to either 1 or zero. It is commonly used to examine group and time effects in regression analysis. Consider a simple model of regressing R&D expenditure in 2002 on 2000 net income and firm type. The dummy variable `d1` is set to 1 for equipment and software firms and zero for telecommunication and electronics. The variable `d2` is coded in the opposite way. Take a look at the data structure (Figure 2.1).

Figure 2.1 Dummy Variable Coding for Firm Types

firm	rnd	income	type	d1	d2
LG Electronics	551	356	Electronics	0	1
AT&T	254	4,669	Telecom	0	1
IBM	4,750	8,093	IT Equipment	1	0
Ericsson	4,424	2,300	Comm. Equipment	1	0
Siemens	5,490	6,528	Electronics	0	1
Verizon	.	11,797	Telecom	0	1
Microsoft	3,772	9,421	Service & S/W	1	0
…	…	…	…	…	… …

2.1 Model 1 without a Dummy Variable: Pooled OLS

The ordinary least squares (OLS) regression without dummy variables, a pooled regression model, assumes a constant intercept and slope regardless of firm types. In the following regression equation, β_0 is the intercept; β_1 is the slope of net income in 2000; and ε_i is the error term.

Model 1: $R\&D_i = \beta_0 + \beta_1 income_i + \varepsilon_i$

The pooled model fits the data well at the .05 significance level (F=7.07, p<.0115). R^2 of .1604 says that this model accounts for 16 percent of the total variance. The model has the intercept of 1,482.697 and slope of .2231. For a $ one million increase in net income, a firm is likely to increase R&D expenditure by $ .2231 million (p<.012).

```
. use http://www.indiana.edu/~statmath/stat/all/panel/rnd2002.dta, clear
( R&D expenditure of IT firm (OECD 2002))

. regress rnd income

      Source |       SS       df       MS              Number of obs =      39
-------------+------------------------------           F(  1,    37) =    7.07
       Model |  15902406.5     1  15902406.5           Prob > F      =  0.0115
    Residual |  83261299.1    37  2250305.38           R-squared     =  0.1604
-------------+------------------------------           Adj R-squared =  0.1377
       Total |  99163705.6    38   2609571.2           Root MSE      =  1500.1

------------------------------------------------------------------------------
         rnd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   .2230523   .0839066     2.66   0.012     .0530414    .3930632
       _cons |   1482.697   314.7957     4.71   0.000     844.8599    2120.533
------------------------------------------------------------------------------
```

Pooled model: R&D = 1,482.697 + .2231*income

Despite moderate goodness of fit statistics such as F and t, this is a naïve model. R&D investment tends to vary across industries.

2.2 Model 2 with a Dummy Variable

You may assume that equipment and software firms have more R&D expenditure than other types of companies. Let us take this group difference into account.[4] We have to drop one of the two dummy variables in order to avoid perfect multicollinearity. That is, OLS does not work with both dummies in a model. The δ_1 in model 2 is the coefficient of equipment, service, and software companies.

Model 2: $R\ \&\ D_i = \beta_0 + \beta_1 income_i + \delta_1 d_{1i} + \varepsilon_i$

Model 2 fits the date better than Model 1 The p-value of the F test is .0054 (significant at the .01 level); R^2 is .2520, about .1 larger than that of Model 1; SSE (sum of squares due to error or residual) decreases from 83,261,299 to 74,175,757 and SEE (square root of MSE) also declines accordingly (1,500→1,435). The coefficient of `d1` is statistically discernable from zero at the .05 level (t=2.10, p<.043). Unlike Model 1, this model results in two different regression equations for two groups. The difference lies in the intercepts, but the slope remains unchanged.

```
. regress rnd income d1

      Source |       SS       df       MS              Number of obs =      39
-------------+------------------------------           F(  2,    36) =    6.06
       Model |  24987948.9     2  12493974.4           Prob > F      =  0.0054
    Residual |  74175756.7    36  2060437.69           R-squared     =  0.2520
-------------+------------------------------           Adj R-squared =  0.2104
       Total |  99163705.6    38   2609571.2           Root MSE      =  1435.4

------------------------------------------------------------------------------
         rnd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   .2180066   .0803248     2.71   0.010     .0551004    .3809128
          d1 |   1006.626   479.3717     2.10   0.043     34.41498    1978.837
       _cons |   1133.579   344.0583     3.29   0.002     435.7962    1831.361
------------------------------------------------------------------------------
```

d1=1: R&D = 2,140.2050 + .2180*income = 1,113.579 +1,006.6260*1 + .2180*income
d1=0: R&D = 1,133.5790 + .2180*income = 1,113.579 +1,006.6260*0 + .2180*income

The slope .2180 indicates a positive impact of two-year-lagged net income on a firm's R&D expenditure. Equipment and software firms on average spend $1,007 million (=2,140-1,134) more for R&D than telecommunication and electronics companies.

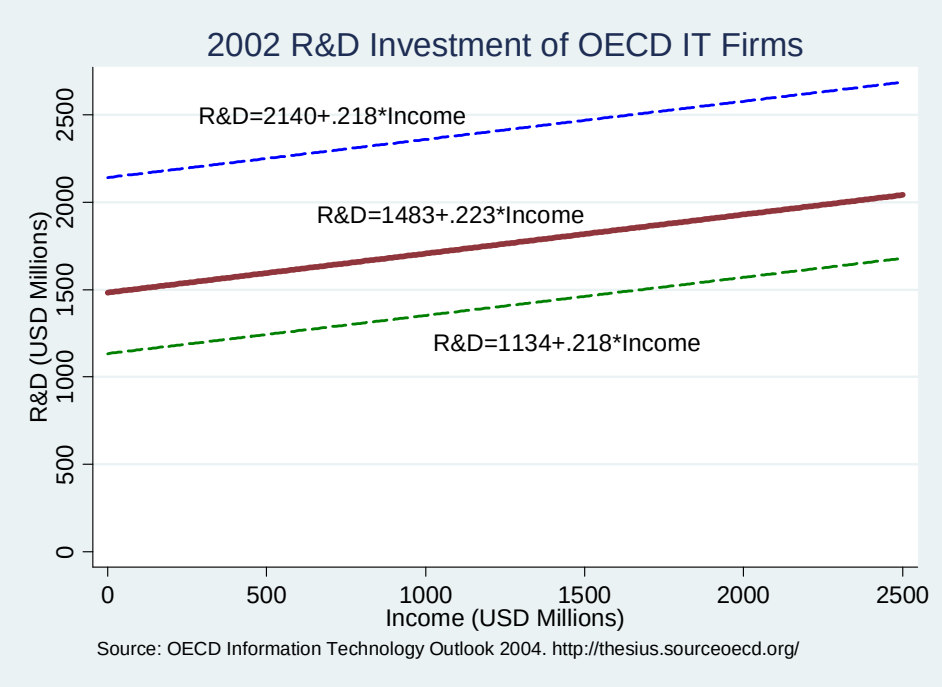
2.3 Visualization of Model 1 and 2

[4] The dummy variable (firm types) and regressors (net income) may or may not be correlated.

There is only a tiny difference in the slope (.2231 versus .2180) between Model 1 and Model 2. The intercept 1,483 of Model 1, however, is quite different from 1,134 for equipment and software companies and 2,140 for telecommunications and electronics in Model 2. This result appears to be supportive of Model 2.

Figure 2.2 highlights differences between Model 1 and 2 more clearly. The red line (pooled) in the middle is the regression line of Model 1; the dotted blue line at the top is one for equipment and software companies (`d1`=1) in Model 2; finally the dotted green line at the bottom is for telecommunication and electronics firms (`d2`=1 or `d1`=0).

Figure 2.2. Regression Lines of Model 1 and Model 2



This plot shows that Model 1 ignores the group difference, and thus reports the misleading intercept. The difference in the intercept between two groups of firms looks substantial. However, the two models have the similar slopes. Consequently, Model 2 considering a fixed group effect (i.e., firm type) seems better than the simple Model 1. Compare goodness of fit statistics (e.g., F, R^2, and SSE) of the two models. See Section 3.2.2 and 4.7 for formal hypothesis test.

2.4 Least Squares Dummy Variable Regression: LSDV1, LSDV2, and LSDV3

The least squares dummy variable (LSDV) regression is ordinary least squares (OLS) with dummy variables. Above Model 2 is a typical example of LSDV. The key issue in LSDV is how to avoid the perfect multicollinearity or so called "dummy variable trap." LSDV has three approaches to avoid getting caught in the trap. These approaches are different from each other with respect to model estimation and interpretation of dummy variable parameters (Suits 1984: 177). They produce different dummy parameter estimates, but their results are equivalent.

The first approach, LSDV1, drops a dummy variable as shown in Model 2 above. That is, the parameter of the eliminated dummy variable is set to zero and is used as a baseline (Table 3). A variable to be dropped, $d_{dropped}^{LSDV1}$ (*d2* in Model 2), needs to be carefully (as opposed to arbitrarily) selected so that it can play a role of the reference group effectively. LSDV2 includes all dummies and, in turn, suppresses the intercept (i.e., set the intercept to zero). Finally, LSDV3 includes the intercept and all dummies, and then impose a restriction that the sum of parameters of all dummies is zero. Each approach has a constraint (restriction) that reduces the number of parameters to be estimated by one and thus makes the model identified. The following functional forms compare these three LSDVs.

LSDV1: $R\&D_i = \beta_0 + \beta_1 income_i + \delta_1 d_{1i} + \varepsilon_i$ or $R\&D_i = \beta_0 + \beta_1 income_i + \delta_2 d_{2i} + \varepsilon_i$

LSDV2: $R\&D_i = \beta_1 income_i + \delta_1 d_{1i} + \delta_2 d_{2i} + \varepsilon_i$

LSDV3: $R\&D_i = \beta_0 + \beta_1 income_i + \delta_1 d_{1i} + \delta_2 d_{2i} + \varepsilon_i$, subject to $\delta_1 + \delta_2 = 0$

Table 2.1. Three Approaches of the Least Squares Dummy Variable Regression Model

	LSDV1	LSDV2	LSDV3
Dummies included	$d_1^{LSDV1} - d_d^{LSDV1}$ except for $d_{dropped}^{LSDV1}$	$d_1^* - d_d^*$	$d_1^{LSDV3} - d_d^{LSDV3}$
Intercept?	α^{LSDV1}	No	α^{LSDV3}
All dummies?	No (*d-1*)	Yes (*d*)	Yes (*d*)
Constraint (restriction)?	$\delta_{dropped}^{LSDV1} = 0$ (Drop one dummy)	$\alpha^{LSDV2} = 0$ (Suppress the intercept)	$\sum \delta_i^{LSDV3} = 0$ (Impose a restriction)
Actual dummy parameters	$\delta_i^* = \alpha^{LSDV1} + \delta_i^{LSDV1}$, $\delta_{dropped}^* = \alpha^{LSDV1}$	$\delta_1^*, \delta_2^*, \ldots \delta_d^*$	$\delta_i^* = \alpha^{LSDV3} + \delta_i^{LSDV3}$, $\alpha^{LSDV3} = \frac{1}{d}\sum \delta_i^*$
Meaning of a dummy coefficient	How far away from the reference group (dropped)?	Actual intercept	How far away from the average group effect?
H_0 of the t-test	$\delta_i^* - \delta_{dropped}^* = 0$	$\delta_i^* = 0$	$\delta_i^* - \frac{1}{d}\sum \delta_i^* = 0$

Source: Constructed from Suits (1984) and David Good's lecture (2004)

Three approaches end up fitting the same model but the coefficients of dummy variables in each approach have different meanings and thus are numerically different (Table 2.1). A parameter estimate in LSDV2, δ_d^*, is the actual intercept (Y-intercept) of group *d*. It is easy to interpret substantively. The t-test examines if δ_d^* is zero. In LSDV1, a dummy coefficient shows the extent to which the actual intercept of group *d* deviates from the reference point (the parameter of the dropped dummy variable), which is the intercept of LSDV1, $\delta_{dropped}^* = \alpha^{LSDV1}$.[5]

[5] In Model 2, $\hat{\delta}_1$ of 1,007 is the estimated (relative) distance between two types of firm (equipment and software versus telecommunications and electronics). In Figure 2.2, the Y-intercept of equipment and software (absolute distance from the origin) is 2,140 = 1,134+1,006. The Y-intercept of telecommunications and electronics is 1,134.

The null hypothesis holds that the deviation from the reference group is zero. In LSDV3, a dummy coefficient means how far its actual parameter is away from the average group effect (Suits 1984: 178). The average effect is the intercept of LSDV3: $\alpha^{LSDV3} = \frac{1}{d}\sum \delta_i^*$. Therefore, the null hypothesis is the deviation from the average is zero. In short, each approach has a different baseline and thus tests a different hypothesis but produces exactly the same parameter estimates of regressors. They all fit the same model; given one LSDV fitted, in other words, we can replicate the other two LSDVs. Table 2.1 summarizes differences in estimation and interpretation of the three LSDVs.

Which approach is better than the others? You need to consider both estimation and interpretation issues carefully. In general, LSDV1 is often preferred because of easy estimation in statistical software packages. Oftentimes researchers want to see how far dummy parameters deviate from the reference group rather than what are the actual intercept of each group. LSDV2 and LSDV3 involve some estimation problems; for example, LSDV2 reports a incorrect R^2.

2.5 Estimating Three LSDVs

The SAS REG procedure, Stata `.regress` command, LIMDEP `Regress$` command, and SPSS `Regression` command all fit OLS and LSDVs. Let us estimate three LSDVs using SAS, Stata, and LIMDEP.

2.5.1 LSDV 1 without a Dummy

LSDV 1 drops a dummy variable. The intercept is the actual parameter estimate (absolute distance from the origin) of the dropped dummy variable. The coefficient of a dummy included means how far its parameter estimate is away from the reference point or baseline (i.e., the intercept).

Here we include `d2` instead of `d1` to see how a different reference point changes the result. Check the sign of the dummy coefficient and the intercept.

```
PROC REG DATA=masil.rnd2002;
   MODEL rnd = income d2;
RUN;
```

```
                                The REG Procedure
                                  Model: MODEL1
                            Dependent Variable: rnd

                 Number of Observations Read                          50
                 Number of Observations Used                          39
                 Number of Observations with Missing Values           11


                                Analysis of Variance

                                      Sum of           Mean
```

Source	DF	Squares	Square	F Value	Pr > F
Model	2	24987949	12493974	6.06	0.0054
Error	36	74175757	2060438		
Corrected Total	38	99163706			

Root MSE	1435.42248	R-Square	0.2520
Dependent Mean	2023.56410	Adj R-Sq	0.2104
Coeff Var	70.93536		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2140.20468	434.48460	4.93	<.0001
income	1	0.21801	0.08032	2.71	0.0101
d2	1	-1006.62593	479.37174	-2.10	0.0428

d2=0: R&D = 2,140.2047 + .2180*income = 2,140.2047 - 1,006.6259*0 + .2180*income
d2=1: R&D = 1,133.5788 + .2180*income = 2,140.2047 - 1,006.6259*1 + .2180*income

The intercept 2,140 is the Y-intercept of equipment and software firms, whose dummy is dropped in the model (`d1`=1, `d2`=0). The coefficient -1,007 of telecommunications and electronics means that its Y-intercept is -1,007 smaller than 1,134 of equipment and software. That is, 1,134 = 2,140 (baseline) – 1,007. Therefore, this model is identical to Model 2 in Section 2.2. In short, dropping another dummy does not change the model although producing different dummy coefficients.

Alternatively, you may use the GLM and MIXED procedures to get the same result.

```
PROC GLM DATA=masil.rnd2002;
   MODEL rnd = income d2 /SOLUTION;
RUN;

PROC MIXED DATA=masil.rnd2002;
   MODEL rnd = income d2 /SOLUTION;
RUN;
```

2.5.2 LSDV 2 without the Intercept

LSDV 2 includes all dummy variables and suppresses the intercept. The Stata `.regress` command has the `noconstant` option to fit LSDV2. The coefficients of dummies are actual parameter estimates; thus, you do not need to compute Y-intercepts of groups. This LSDV, however, reports incorrect (inflated) R^2 (.7135 > .2520) and F (29.88 > 6.06). This is because the X matrix does not have a column vector of 1 and produces incorrect sums of squares of model and total (Uyar and Erdem (1990: 298). However, the sum of squares of errors is correct in any LSDV.

```
. regress rnd income d1 d2, noconstant
```

```
      Source |       SS       df       MS              Number of obs =      39
-------------+------------------------------           F(  3,    36) =   29.88
       Model |   184685604     3  61561868.1           Prob > F      =  0.0000
    Residual |  74175756.7    36  2060437.69           R-squared     =  0.7135
-------------+------------------------------           Adj R-squared =  0.6896
       Total |   258861361    39  6637470.79           Root MSE      =  1435.4

------------------------------------------------------------------------------
         rnd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   .2180066   .0803248     2.71   0.010     .0551004    .3809128
          d1 |   2140.205   434.4846     4.93   0.000     1259.029     3021.38
          d2 |   1133.579   344.0583     3.29   0.002     435.7962    1831.361
------------------------------------------------------------------------------
```

d1=1: R&D = 2,140.205 + .2180*income
d2=1: R&D = 1,133.579 + .2180*income

2.5.3 LSDV 3 with a Restriction

LSDV 3 includes the intercept and all dummies and then imposes a restriction on the model. The restriction is that the sum of all dummy parameters is zero. The Stata `.constraint` command defines a constraint, while the `.cnsreg` command fits a constrained OLS using the `constraint()`option. The number in the parenthesis indicates the constraint number defined in the `.constraint` command.

```
. constraint 1 d1 + d2 = 0
. cnsreg rnd income d1 d2, constraint(1)

Constrained linear regression                   Number of obs   =         39
                                                F(   2,     36) =       6.06
                                                Prob > F        =     0.0054
                                                Root MSE        =  1435.4225

 ( 1)  d1 + d2 = 0
------------------------------------------------------------------------------
         rnd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   .2180066   .0803248     2.71   0.010     .0551004    .3809128
          d1 |    503.313   239.6859     2.10   0.043     17.20749    989.4184
          d2 |   -503.313   239.6859    -2.10   0.043    -989.4184   -17.20749
       _cons |   1636.892   310.0438     5.28   0.000     1008.094     2265.69
------------------------------------------------------------------------------
```

d1=1: R&D = 2,140.205 + .2180*income = 1,637 + 503*1 + (-503)*0 + .2180*income
d2=1: R&D = 1,133.579 + .2180*income = 1,637 + 503*0 + (-503)*1 + .2180*income

The intercept is the average of actual parameter estimates: 1,637 = (2,140+1,133)/2. Since there are two groups here, the coefficients of two dummies by definition share the same magnitude ($503) but have opposite directions. Equipment and software firms invest $2,140 millions for R&D expenditure, $503 millions MORE than the average expenditure of overall IT firms (=$2,140-$1,637), while telecommunications and electronics spend $503 millions LESS than the average (=$1,134-$1,637). In the SAS output below, the coefficient of RESTRICT is virtually zero and, in theory, should be zero.

```
PROC REG DATA=masil.rnd2002;
   MODEL rnd = income d1 d2;
   RESTRICT d1 + d2 = 0;
RUN;
```

```
                                The REG Procedure
                                  Model: MODEL1
                              Dependent Variable: rnd

NOTE: Restrictions have been applied to parameter estimates.

                  Number of Observations Read                          50
                  Number of Observations Used                          39
                  Number of Observations with Missing Values           11


                                Analysis of Variance

                                          Sum of           Mean
      Source                    DF        Squares         Square    F Value    Pr > F

      Model                      2       24987949       12493974       6.06    0.0054
      Error                     36       74175757        2060438
      Corrected Total           38       99163706


                Root MSE           1435.42248    R-Square     0.2520
                Dependent Mean     2023.56410    Adj R-Sq     0.2104
                Coeff Var            70.93536


                                Parameter Estimates

                                 Parameter       Standard
      Variable     DF             Estimate          Error    t Value    Pr > |t|

      Intercept     1          1636.89172      310.04381       5.28      <.0001
      income        1            0.21801        0.08032       2.71      0.0101
      d1            1           503.31297      239.68587       2.10      0.0428
      d2            1          -503.31297      239.68587      -2.10      0.0428
      RESTRICT     -1        1.81899E-12              0          .          .

                    * Probability computed using beta distribution.
```

Table 2.2 Estimating Three LSDVs Using SAS, Stata, LIMDEP, and SPSS

	LSDV 1	LSDV 2	LSDV 3
SAS	PROC REG; MODEL rnd = income d2; RUN;	PROC REG; MODEL rnd = income d1 d2 /NOINT; RUN;	PROC REG; MODEL rnd = income d1 d2; RESTRICT d1 + d2 = 0; RUN;
Stata	. regress ind income d2	. regress rnd income d1 d2, noconstant	. constraint 1 d1+ d2 = 0 . cnsreg rnd income d1 d2 const(1)
LIMDEP	REGRESS; Lhs=rnd; Rhs=ONE,income, d2$	REGRESS; Lhs=rnd; Rhs=income, d1, d2$	REGRESS; Lhs=rnd; Rhs=ONE,income, d1, d2; Cls: b(2)+b(3)=0$
SPSS	REGRESSION /MISSING LISTWISE /STATISTICS COEFF R ANOVA /CRITERIA=PIN(.05) POUT(.10) /NOORIGIN /DEPENDENT rnd /METHOD=ENTER income d2.	REGRESSION /MISSING LISTWISE /STATISTICS COEFF R ANOVA /CRITERIA=PIN(.05) POUT(.10) /ORIGIN /DEPENDENT rnd /METHOD=ENTER income d1 d2.	N/A

Table 2.2 compares how SAS, Stata, LIMDEP, and SPSS estimate LSDVs. SPSS is not able to fit the LSDV3. In LIMDEP, `ONE` indicates the intercept to be included. `Cls: b(2)+b(3)=0` fits the model under the condition that the sum of parameter estimates of `d1` (second parameter) and `d2` (third parameter) is zero. In SPSS, pay attention to the /ORIGIN option for LSDV2.

3. Panel Data Models

Panel data models examine group (individual-specific) effects, time effects, or both. These effects are either fixed effect or random effect. A *fixed effect model* examines if intercepts vary across groups or time periods, whereas a *random effect model* explores differences in error variances. A *one-way model* includes only one set of dummy variables (e.g., firm), while a *two-way model* considers two sets of dummy variables (e.g., firm and year). Model 2 in Chapter 2, in fact, is a one-way fixed group effect panel data model.

3.1 Functional Forms and Notation

The parameter estimate of a dummy variable is a part of the intercept in a fixed effect model and a component of error in the random effect model. Slopes remain the same across groups or time periods. The functional forms of one-way panel data models are as follows.

Fixed group effect model: $y_{it} = (\alpha + u_i) + X_{it}'\beta + v_{it}$, where $v_{it} \sim IID(0, \sigma_v^2)$
Random group effect model: $y_{it} = \alpha + X_{it}'\beta + (u_i + v_{it})$, where $v_{it} \sim IID(0, \sigma_v^2)$

Note that u_i is a fixed or random effect and errors are *independent identically distributed*, $v_{it} \sim IID(0, \sigma_v^2)$.

Notations used in this document include,

- $\bar{y}_{i\bullet}$: dependent variable (DV) mean of group *i*.
- $\bar{y}_{\bullet t}$: dependent variable (DV) mean at time *t*.
- $\bar{x}_{i\bullet}$: means of independent variables (IVs) of group *i*.
- $\bar{x}_{\bullet t}$: means of independent variables (IVs) at time *t*.
- $\bar{y}_{\bullet\bullet}$: overall means of the DV.
- $\bar{x}_{\bullet\bullet}$: overall means of the IVs.
- *n*: the number of groups or firms
- *T*: the number of time periods
- *N=nT*: total number of observations
- *k*: the number of regressors excluding dummy variables
- *K=k+1* (including the intercept)

3.2 Fixed Effect Models

There are several strategies for estimating fixed effect models. The *least squares dummy variable model (LSDV)* uses dummy variables, whereas the *within effect model* does not. These strategies, of course, produce the identical parameter estimates of non-dummy independent variables. The *between effect model* fits the model using group and/or time means of dependent and independent variables without dummies. Table 3.1 summarizes pros and cons of these models.

3.2.1 Estimations: LSDV, Within Effect, and Between Effect Models

As discussed in Chapter 2, LSDV is widely used because it is relatively easy to estimate and interpret substantively. This LSDV, however, becomes problematic when there are many groups or subjects in panel data. If T is fixed and $nT \to \infty$, only coefficients of regressors are consistent. The coefficients of dummy variables, $\alpha + u_i$, are not consistent since the number of these parameters increases as *nT* increases (Baltagi 2001). This is the so called *incidental parameter problem*. Under this circumstance, LSDV is useless and thus calls for another strategy, the within effect model.

A within group effect model does not need dummy variables, but it uses deviations from group means. Thus, this model is the OLS of $(y_{it} - \bar{y}_{i\bullet}) = (x_{it} - \bar{x}_{i\bullet})'\beta + (\varepsilon_{it} - \bar{\varepsilon}_{i\bullet})$ without an intercept.[6] The incidental parameter problem is no longer an issue. The parameter estimates of regressors in the within effect model are identical to those of LSDV. The within effect model in turn has several disadvantages.

Since this model does not report dummy coefficients, you need to compute them using the formula $d_i^* = \bar{y}_{i\bullet} - \bar{x}_{i\bullet}'\beta$ Since no dummy is used, the within effect model has larger degrees of freedom for error, resulting in small MSE (mean square error) and incorrect (smaller) standard errors of parameter estimates. Thus, you have to adjust the standard error using the formula $se_k^* = se_k \sqrt{\frac{df_{error}^{Within}}{df_{error}^{LSDV}}} = se_k \sqrt{\frac{nT-k}{nT-n-k}}$. Finally, R^2 of the within effect model is not correct because the intercept is suppressed.

Table 3.1 Comparison of Fixed Effect Models

	LSDV1	Within Effect	Between Effect
Functional form	$y_i = i\alpha_i + X_i\beta + \varepsilon_i$	$y_{it} - \bar{y}_{i\bullet} = x_{it} - \bar{x}_{i\bullet} + \varepsilon_{it} - \bar{\varepsilon}_{i\bullet}$	$\bar{y}_{i\bullet} = \alpha + \bar{x}_{i\bullet} + \varepsilon_i$
Dummy	Yes	No	No
Dummy coefficient	Presented	Need to be computed	N/A
Transformation	No	Deviation from the group means	Group means
Intercept (estimation)	Yes	No	Yes
R^2	Correct	Incorrect	
SSE	Correct	Correct	
MSE	Correct	Smaller	
Standard error of β	Correct	Incorrect (smaller)	
DF_{error}	*nT-n-k*	*nT-k* (n larger)	*n-K*
Observations	*nT*	*nT*	*n*

The between group effect model, so called the group mean regression, uses group means of the dependent and independent variables. This data aggregation reduces the number of

[6] You need to follow three steps: 1) compute group means of the dependent and independent variables; 2) transform variables to get deviations from the group means; 3) run OLS with the transformed variables without the intercept.

observations down to *n*. Then, run OLS of $\bar{y}_{i\bullet} = \alpha + \bar{x}_{i\bullet} + \varepsilon_i$. Table 3.1 contrasts LSDV, the within effect model, and the between group models.

3.2.2 Testing Group Effects

In a regression of $y_{it} = \alpha + \mu_i + X_{it}'\beta + \varepsilon_{it}$, the null hypothesis is that all dummy parameters except for one for the dropped are zero: $H_0 : \mu_1 = ... = \mu_{n-1} = 0$. This hypothesis is tested by the F test, which is based on loss of goodness-of-fit. The robust model in the following formula is LSDV (or within effect model) and the efficient model is the pooled regression.[7]

$$\frac{(e'e_{Efficient} - e'e_{Robust})/(n-1)}{(e'e_{Robust})/(nT-n-k)} = \frac{(R^2_{Robust} - R^2_{Efficient})/(n-1)}{(1-R^2_{Robust})/(nT-n-k)} \sim F(n-1, nT-n-k)$$

If the null hypothesis is rejected, you may conclude that the fixed group effect model is better than the pooled OLS model.

3.2.3 Fixed Time Effect and Two-way Fixed Effect Models

For the fixed time effects model, you need to switch *n* and *T*, and *i* and *t* in the formulas.

- Model: $y_{it} = \alpha + \tau_t + X_{it}'\beta + \varepsilon_{it}$
- Within effect model: $(y_{it} - \bar{y}_{\bullet t}) = (x_{it} - \bar{x}_{\bullet t})'\beta + (\varepsilon_{it} - \bar{\varepsilon}_{\bullet t})$
- Dummy coefficients: $d_t^* = \bar{y}_{\bullet t} - \bar{x}_{\bullet t}'\beta$
- Correct standard errors: $se_k^* = se_k \sqrt{\frac{df_{error}^{Within}}{df_{error}^{LSDV}}} = se_k \sqrt{\frac{Tn-k}{Tn-T-k}}$
- Between effect model: $\bar{y}_{\bullet t} = \alpha + \bar{x}_{\bullet t} + \varepsilon_t$
- $H_0 : \tau_1 = ... = \tau_{T-1} = 0$.
- F-test: $\frac{(e'e_{Pooled} - e'e_{Within})/(T-1)}{(e'e_{Within})/(Tn-T-k)} \sim F(T-1, Tn-T-k)$.

The fixed group and time effect model uses slightly different formulas. The within effect model of this two-way fixed model is estimated by five strategies (see Section 6.1).

- Model: $y_{it} = \alpha + \mu_i + \tau_t + X_{it}\beta + \varepsilon_{it}$.
- Within effect Model: $y_{it}^* = y_{it} - \bar{y}_{i\bullet} - \bar{y}_{\bullet t} + \bar{y}_{\bullet\bullet}$ and $x_{it}^* = x_{it} - \bar{x}_{i\bullet} - \bar{x}_{\bullet t} + \bar{x}_{\bullet\bullet}$.
- Dummy coefficients: $d_i^* = (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet}) - (\bar{x}_{i\bullet} - \bar{x}_{\bullet\bullet})'\beta$ and $d_t^* = (\bar{y}_{\bullet t} - \bar{y}_{\bullet\bullet}) - (\bar{x}_{\bullet t} - \bar{x}_{\bullet\bullet})'\beta$

[7] When comparing fixed effect and random effect models, the fixed effect estimates are considered as the robust estimates and random effect estimates as the efficient estimates.

- Correct standard errors: $se_k^* = se_k \sqrt{\frac{df_{error}^{Within}}{df_{error}^{LSDV}}} = se_k \sqrt{\frac{nT-k}{nT-n-T-k+1}}$
- $H_0 : \mu_1 = ... = \mu_{n-1} = 0$ and $\tau_1 = ... = \tau_{T-1} = 0$.
- F-test: $\frac{(e'e_{Efficient} - e'e_{Robust})/(n+T-2)}{(e'e_{Robust})/(nT-n-T-k+1)} \sim F[(n+T-2),(nT-n-T-k+1)]$

3.3 Random Effect Models

The one-way random group effect model is formulated as $y_{it} = \alpha + X_{it}'\beta + u_i + v_{it}$, $w_{it} = u_i + v_{it}$ where $u_i \sim IID(0,\sigma_u^2)$ and $v_{it} \sim IID(0,\sigma_v^2)$. The u_i are assumed independent of v_{it} and X_{it}, which are also independent of each other for all *i* and *t*. This assumption is not necessary in the fixed effect model. The components of $Cov(w_{it}, w_{js}) = E(w_{it}w_{js})$ are $\sigma_u^2 + \sigma_v^2$ if *i*=*j* and *t*=*s* and σ_u^2 if *i*=*j* and $t \neq s$.[8] Thus, the Ω matrix or the variance structure of errors looks like,

$$\underset{T\times T}{\Omega} = \begin{bmatrix} \sigma_u^2+\sigma_v^2 & \sigma_u^2 & ... & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2+\sigma_v^2 & ... & \sigma_u^2 \\ ... & ... & ... & ... \\ \sigma_u^2 & \sigma_u^2 & ... & \sigma_u^2+\sigma_v^2 \end{bmatrix}$$

A random effect model is estimated by generalized least squares (GLS) when the variance structure is known, and by feasible generalized least squares (FGLS) when the variance is unknown. Compared to fixed effect models, random effect models are relatively difficult to estimate. This document assumes panel data are balanced.

3.3.1 Generalized Least Squares (GLS)

When Ω is known (given), GLS based on the true variance components is BLUE and all the feasible GLS estimators considered are asymptotically efficient as either *n* or *T* approaches infinity (Baltagi 2001).

In GLS, you just need to compute θ using the Ω matrix: $\theta = 1 - \sqrt{\frac{\sigma_v^2}{T\sigma_u^2 + \sigma_v^2}}$.[9] Then transform variables as follows.

- $y_{it}^* = y_{it} - \theta\, \bar{y}_{i\bullet}$
- $x_{it}^* = x_{it} - \theta\, \bar{x}_{i\bullet}$ for all X_k
- $\alpha^* = 1 - \theta$

[8] This implies that $Corr(w_{it}, w_{js})$ is 1 if *i*=*j* and *t*=*s*, and $\sigma_u^2/(\sigma_u^2 + \sigma_v^2)$ if *i*=*j* and $t \neq s$.

[9] If $\theta = 0$, run pooled OLS. If $\theta = 1$ and $\sigma_v^2 = 0$, then run the within effect model.

Finally, run OLS on the transformed variables: $y_{it}^* = \alpha^* + x_{it}^{*\prime}\beta^* + \varepsilon_{it}^*$. Since Ω is often unknown, FGLS is more frequently used than GLS.

3.3.2 Feasible Generalized Least Squares (FGLS)

If Ω is unknown, first you have to estimate θ using $\hat{\sigma}_u^2$ and $\hat{\sigma}_v^2$:

$$\hat{\theta} = 1 - \sqrt{\frac{\hat{\sigma}_v^2}{T\hat{\sigma}_u^2 + \hat{\sigma}_v^2}} = 1 - \sqrt{\frac{\hat{\sigma}_v^2}{T\hat{\sigma}_{between}^2}}.$$

The $\hat{\sigma}_v^2$ is derived from the SSE (sum of squares due to error) of the within effect model or from the deviations of residuals from group means of residuals:

$$\hat{\sigma}_v^2 = \frac{SSE_{within}}{nT-n-k} = \frac{e'e_{within}}{nT-n-k} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(v_{it} - \bar{v}_{i\bullet})^2}{nT-n-k},$$ where v_{it} are the residuals of the LSDV1.

The $\hat{\sigma}_u^2$ comes from the between effect model (group mean regression):

$$\hat{\sigma}_u^2 = \hat{\sigma}_{between}^2 - \frac{\hat{\sigma}_v^2}{T}$$, where $\hat{\sigma}_{between}^2 = \frac{SSE_{between}}{n-K}$.

Next, transform variables using $\hat{\theta}$ and then run OLS: $y_{it}^* = \alpha^* + x_{it}^{*\prime}\beta^* + \varepsilon_{it}^*$.

- $y_{it}^* = y_{it} - \hat{\theta}\,\bar{y}_{i\bullet}$
- $x_{it}^* = x_{it} - \hat{\theta}\,\bar{x}_{i\bullet}$ for all X_k
- $\alpha^* = 1 - \hat{\theta}$

The estimation of the two-way random effect model is skipped here.

3.3.3 Testing Random Effects (LM test)

The null hypothesis is that cross-sectional variance components are zero, $H_0 : \sigma_u^2 = 0$. Breusch and Pagan (1980) developed the Lagrange multiplier (LM) test (Greene 2003). In the following formula, $\bar{e}$ is the *n X 1* vector of the group specific means of pooled regression residuals and $e'e$ is the SSE of the pooled OLS regression. The LM follows chi-squared distribution with one degree of freedom.

$$LM_u = \frac{nT}{2(T-1)}\left[\frac{e'DDe}{e'e} - 1\right]^2 = \frac{nT}{2(T-1)}\left[\frac{T^2\bar{e}'\bar{e}}{e'e} - 1\right]^2 \sim \chi^2(1).$$

Baltagi (2001) presents the same LM test in a different way.

$$LM_u = \frac{nT}{2(T-1)}\left[\frac{\sum\left(\sum e_{it}\right)^2}{\sum\sum e_{it}^2} - 1\right]^2 = \frac{nT}{2(T-1)}\left[\frac{\sum\left(T\bar{e}_{i\bullet}\right)^2}{\sum\sum e_{it}^2} - 1\right]^2 \sim \chi^2(1).$$

The two way random effect model has the null hypothesis of $H_0 : \sigma_{u1}^2 = 0$ and $\sigma_{u2}^2 = 0$. The LM test combines two one-way random effect models for group and time, $LM_{u12} = LM_{u1} + LM_{u2} \sim \chi^2(2)$.

3.4 Hausman Test: Fixed Effects versus Random Effects

The Hausman specification test compares the fixed versus random effects under the null hypothesis that the individual effects are uncorrelated with the other regressors in the model (Hausman 1978). If correlated (H_0 is rejected), a random effect model produces biased estimators, violating one of the Gauss-Markov assumptions; so a fixed effect model is preferred. Hausman's essential result is that the covariance of an efficient estimator with its difference from an inefficient estimator is zero (Greene 2003).

$$m = \left(b_{Robust} - b_{Efficient}\right)' \hat{\Sigma}^{-1} \left(b_{Robust} - b_{Efficient}\right) \sim \chi^2(k),$$

where, $\hat{\Sigma} = Var[b_{Robust} - b_{Efficient}] = Var(b_{Robust}) - Var(b_{Efficient})$ is the difference in the estimated covariance matrix of the parameter estimates between the LSDV model (robust) and the random effects model (efficient). It is notable that an intercept and dummy variables SHOULD be excluded in computation.

3.5 Poolability Test

What is poolability? Poolability tests whether or not slopes are the same across groups or over time. Thus, the null hypothesis of the poolability test is $H_0 : \beta_{ik} = \beta_k$. Remember that slopes remain constant in fixed and random effect models; only intercepts and error variances matter.

The poolability test is undertaken under the assumption of $\mu \sim N(0, s^2 I_{NT})$. This test uses the F statistic,

$$F_{obs} = \frac{(e'e - \sum e_i' e_i)/(n-1)K}{\sum e_i' e_i / n(T-K)} \sim F\left[(n-1)K, n(T-K)\right],$$

where $e'e$ is the SSE of the pooled OLS and $e_i' e_i$ is the SSE of the OLS regression for group *i*. If the null hypothesis is rejected, the panel data are not poolable. Under this circumstance, you may go to the random coefficient model or hierarchical regression model.

Similarly, the null hypothesis of the poolability test over time is $H_0 : \beta_{tk} = \beta_k$. The F-test is

$$F_{obs} = \frac{(e'e - \sum e_t' e_t)/(T-1)K}{\sum e_t' e_t / T(n-K)} = F\left[(T-1)K, T(n-K)\right],$$

where $e_t' e_t$ is SSE of the OLS regression at time *t*.

4. One-way Fixed Effect Models: Group Effects

A one-way fixed group model examines group differences in intercepts. The LSDV for this fixed model needs to create as many dummy variables as the number of entities or subjects. When many dummies are needed, the within effect model is useful since it transforms variables using group means to avoid dummies. The between effect model uses group means of variables.

The sample panel data set includes cost and its related data of six U.S. airlines measured at 15 different time points. The following `.use` command reads a data set `airline.dta` and `.describe` displays basic information of key variables.

```
. use http://www.indiana.edu/~statmath/stat/all/panel/airline.dta, clear

. describe airline year cost output fuel load
```

```
              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------------------------
airline         int    %8.0g                  Airline name
year            int    %8.0g                  Year
cost            float  %9.0g                  Total cost in $1,000
output          float  %9.0g                  Output in revenue passenger miles, index number
fuel            float  %9.0g                  Fuel price
load            float  %9.0g                  Load factor
```

You need to declare a cross-sectional (`airline`) and a time-series (`year`) variables using the `.tsset` command.

```
. tsset airline year
       panel variable:  airline (strongly balanced)
        time variable:  year, 1 to 15
                delta:  1 unit
```

Let us take a look at descriptive statistics of key variables using `.xtsum`.

```
. xtsum cost output fuel load
```

```
Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
cost     overall |  13.36561   1.131971   11.14154    15.3733 |     N =      90
         between |             .9978636   12.27441   14.67563 |     n =       6
         within  |             .6650252   12.11545   14.91617 |     T =      15
                 |                                            |
output   overall | -1.174309   1.150606  -3.278573   .6608616 |     N =      90
         between |             1.166556   -2.49898   .3192696 |     n =       6
         within  |             .4208405  -1.987984   .1339861 |     T =      15
                 |                                            |
fuel     overall |  12.77036   .8123749   11.55017     13.831 |     N =      90
         between |             .0237151    12.7318    12.7921 |     n =       6
         within  |             .8120832   11.56883    13.8513 |     T =      15
                 |                                            |
load     overall |  .5604602   .0527934    .432066    .676287 |     N =      90
         between |             .0281511   .5197756   .5971917 |     n =       6
         within  |             .0460361   .4368492   .6581019 |     T =      15
```

4.1 The Pooled OLS Regression Model

First, fit the pooled regression model without any dummy variable.

```
. regress cost output fuel load
```

```
      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  3,    86) = 2419.34
       Model |  112.705452     3  37.5684839           Prob > F      =  0.0000
    Residual |  1.33544153    86   .01552839           R-squared     =  0.9883
-------------+------------------------------           Adj R-squared =  0.9879
       Total |  114.040893    89  1.28135835           Root MSE      =  .12461

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |   .8827385   .0132545    66.60   0.000     .8563895    .9090876
        fuel |    .453977   .0203042    22.36   0.000     .4136136    .4943404
        load |   -1.62751    .345302    -4.71   0.000    -2.313948   -.9410727
       _cons |   9.516923   .2292445    41.51   0.000       9.0612    9.972645
------------------------------------------------------------------------------
```

The regression equation is cost = 9.5169 + .8827*output +.4540*fuel -1.6275*load. This model fits the data well (F=2419.34, p<.0000 and R^2=.9883). We may, however, suspect if there is a fixed group effect producing different intercepts across groups. Each airline may have a significantly different level of cost, its Y-intercept, when all regressors are set to zero. This difference is modeled as a fixed group effect.

As discussed in Chapter 2, there are three equivalent approaches of LSDV. They report the identical parameter estimates of regresors except for dummy coefficients. Let us begin with LSDV1.

4.2 LSDV1 without a Dummy

LSDV1 drops a dummy variable to get the model identified. LSDV1 produces correct ANOVA information, goodness of fit, parameter estimates, and standard errors. As a consequence, this approach is commonly used in practice. LSDV produces six regression equations for six airlines. How can we draw these equations using LSDV1?

Airline 1: cost = 9.7059 + .9193*output +.4175*fuel -1.0704*load
Airline 2: cost = 9.6647 + .9193*output +.4175*fuel -1.0704*load
Airline 3: cost = 9.4970 + .9193*output +.4175*fuel -1.0704*load
Airline 4: cost = 9.8905 + .9193*output +.4175*fuel -1.0704*load
Airline 5: cost = 9.7300 + .9193*output +.4175*fuel -1.0704*load
Airline 6: cost = 9.7930 + .9193*output +.4175*fuel -1.0704*load

In SAS, PROC REG fits the OLS regression model. Let us drop the last dummy `g6` and use it as the reference group. Of course, you may drop another dummy variable to get the equivalent result. LSDV1 fits the data better than does the pooled OLS. SSE decreases from 1.3354 to .2926, but R^2 increases from .9883 to .9974. Due to the dummies included, this model loses five degrees of freedom (from 86 to 81).

```
PROC REG DATA=masil.airline;
   MODEL cost = g1-g5 output fuel load;
RUN;
```

```
                                The REG Procedure
                                  Model: MODEL1
                            Dependent Variable: cost
```

```
                    Number of Observations Read          90
                    Number of Observations Used          90


                              Analysis of Variance

                                      Sum of           Mean
Source                    DF         Squares         Square    F Value    Pr > F

Model                      8       113.74827       14.21853    3935.79    <.0001
Error                     81         0.29262        0.00361
Corrected Total           89       114.04089


          Root MSE              0.06011    R-Square     0.9974
          Dependent Mean       13.36561    Adj R-Sq     0.9972
          Coeff Var             0.44970


                           Parameter Estimates

                        Parameter       Standard
Variable     DF          Estimate          Error    t Value    Pr > |t|

Intercept     1           9.79300        0.26366      37.14      <.0001
g1            1          -0.08706        0.08420      -1.03      0.3042
g2            1          -0.12830        0.07573      -1.69      0.0941
g3            1          -0.29598        0.05002      -5.92      <.0001
g4            1           0.09749        0.03301       2.95      0.0041
g5            1          -0.06301        0.02389      -2.64      0.0100
output        1           0.91928        0.02989      30.76      <.0001
fuel          1           0.41749        0.01520      27.47      <.0001
load          1          -1.07040        0.20169      -5.31      <.0001
```

The parameter estimate of `g6` is presented in the intercept (9.7930). Other dummy parameter estimates are computed using the reference point. The actual intercept of airline 1, for example, is computed as 9.7059 = 9.7930 + (-.0871)*1 + (-.1283)*0 + (-.2960)*0 + (.0975)*0 + (-.0630)*0 or simply 9.7930 + (-.0871), where 9.7930 is the reference point, the intercept of this model. The coefficient -.0871 says that the Y-intercept of airline 1 (9.7059) is .0871 smaller than that of airline 6 (reference point).

Stata has the `.regress` command for OLS regression (LSDV). The output is identical to that of PROC REG.

```
. regress cost g1-g5 output fuel load

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  8,    81) = 3935.79
       Model |   113.74827     8  14.2185338           Prob > F      =  0.0000
    Residual |  .292622872    81  .003612628           R-squared     =  0.9974
-------------+------------------------------           Adj R-squared =  0.9972
       Total |  114.040893    89  1.28135835           Root MSE      =  .06011

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          g1 |  -.0870617   .0841995    -1.03   0.304    -.2545924     .080469
          g2 |  -.1282976   .0757281    -1.69   0.094    -.2789728    .0223776
```

```
          g3 |  -.2959828   .0500231    -5.92   0.000     -.395513   -.1964526
          g4 |    .097494   .0330093     2.95   0.004     .0318159    .1631721
          g5 |   -.063007   .0238919    -2.64   0.010    -.1105443   -.0154697
      output |   .9192846   .0298901    30.76   0.000     .8598126    .9787565
        fuel |   .4174918   .0151991    27.47   0.000     .3872503    .4477333
        load |  -1.070396     .20169    -5.31   0.000    -1.471696   -.6690963
       _cons |   9.793004   .2636622    37.14   0.000     9.268399    10.31761
------------------------------------------------------------------------------
```

In LIMDEP, run the `Regress$` command to fit the LSDV1. Do not forget to include ONE for the intercept in the `Rhs` subcommand.

```
--> REGRESS;Lhs=COST;Rhs=ONE,G1,G2,G3,G4,G5,OUTPUT,FUEL,LOAD$
```

```
+----------------------------------------------------+
| Ordinary    least squares regression               |
| Model was estimated Aug 27, 2009 at 03:51:23PM     |
| LHS=COST     Mean                 =   13.36561     |
|              Standard deviation   =   1.131971     |
| WTS=none     Number of observs.   =         90     |
| Model size   Parameters           =          9     |
|              Degrees of freedom   =         81     |
| Residuals    Sum of squares       =   .2926208     |
|              Standard error of e  =   .6010493E-01 |
| Fit          R-squared            =   .9974341     |
|              Adjusted R-squared   =   .9971806     |
| Model test   F[  8,    81] (prob) =3935.82 (.0000) |
| Diagnostic   Log likelihood       =   130.0865     |
|              Restricted(b=0)      =  -138.3581     |
|              Chi-sq [  8]  (prob) = 536.89 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =  -5.528017     |
|              Akaike Info. Criter. =  -5.528687     |
| Autocorrel   Durbin-Watson Stat.  =  1.0264504     |
|              Rho = cor[e,e(-1)]   =   .4867748     |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    9.79302127       .26366104    37.142   .0000
 G1      |    -.08707202       .08419916    -1.034   .3042    .16666667
 G2      |    -.12830600       .07572778    -1.694   .0940    .16666667
 G3      |    -.29598860       .05002285    -5.917   .0000    .16666667
 G4      |     .09749253       .03300915     2.954   .0041    .16666667
 G5      |    -.06300770       .02389180    -2.637   .0100    .16666667
 OUTPUT  |     .91928814       .02988997    30.756   .0000  -1.17430918
 FUEL    |     .41749105       .01519907    27.468   .0000   12.7703592
 LOAD    |   -1.07039502       .20168924    -5.307   .0000    .56046016
```

What if we drop a different dummy variable, say `g1`, instead of `g6`? Since the different reference point is applied, you will get different dummy coefficients. As shown in the above, the intercept 9.7059 in this model is the actual parameter estimate (Y-intercept) of `g1`, which was excluded from the model. The Y-intercept of airline 2 is computed to get 9.6647=9.7059-.0412. The Y-intercept of airline 2 (9.6647) is .0412 smaller than the reference point of 9.7059. Actual Y-intercepts of other dummies are computed in this manner. The other statistics such as parameter estimates of regressors and goodness-of-fit measures remain unchanged. That is, choice of a dummy variable to be dropped does not change a model.

```
. regress cost g2-g6 output fuel load
```

```
      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  8,    81) = 3935.79
       Model |   113.74827     8  14.2185338           Prob > F      =  0.0000
    Residual |  .292622872    81  .003612628           R-squared     =  0.9974
-------------+------------------------------           Adj R-squared =  0.9972
       Total |  114.040893    89  1.28135835           Root MSE      =  .06011
```

```
------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          g2 |  -.0412359   .0251839    -1.64   0.105    -.0913441    .0088722
          g3 |  -.2089211   .0427986    -4.88   0.000    -.2940769   -.1237652
          g4 |   .1845557   .0607527     3.04   0.003     .0636769    .3054345
          g5 |   .0240547   .0799041     0.30   0.764    -.1349293    .1830387
          g6 |   .0870617   .0841995     1.03   0.304     -.080469    .2545924
      output |   .9192846   .0298901    30.76   0.000     .8598126    .9787565
        fuel |   .4174918   .0151991    27.47   0.000     .3872503    .4477333
        load |  -1.070396     .20169    -5.31   0.000    -1.471696   -.6690963
       _cons |   9.705942    .193124    50.26   0.000     9.321686     10.0902
------------------------------------------------------------------------------
```

When you have not created dummy variables, take advantage of the `.xi` prefix command (interaction expansion) to obtain the identical result. The Stata `.xi`, like `.bysort`, is used either as an ordinary command or a prefix command. `.xi` creates dummies from a categorical variable specified in the term `i.` and then run the command following the colon. Stata by default drops the first dummy variable, while PROC TSCSREG and PROC PANEL in Section 4.5.2 drop the last dummy.

```
. xi: regress cost i.airline output fuel load

i.airline         _Iairline_1-6       (naturally coded; _Iairline_1 omitted)

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  8,    81) = 3935.79
       Model |   113.74827     8  14.2185338           Prob > F      =  0.0000
    Residual |  .292622872    81  .003612628           R-squared     =  0.9974
-------------+------------------------------           Adj R-squared =  0.9972
       Total |  114.040893    89  1.28135835           Root MSE      =  .06011

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 _Iairline_2 |  -.0412359   .0251839    -1.64   0.105    -.0913441    .0088722
 _Iairline_3 |  -.2089211   .0427986    -4.88   0.000    -.2940769   -.1237652
 _Iairline_4 |   .1845557   .0607527     3.04   0.003     .0636769    .3054345
 _Iairline_5 |   .0240547   .0799041     0.30   0.764    -.1349293    .1830387
 _Iairline_6 |   .0870617   .0841995     1.03   0.304     -.080469    .2545924
      output |   .9192846   .0298901    30.76   0.000     .8598126    .9787565
        fuel |   .4174918   .0151991    27.47   0.000     .3872503    .4477333
        load |  -1.070396     .20169    -5.31   0.000    -1.471696   -.6690963
       _cons |   9.705942    .193124    50.26   0.000     9.321686     10.0902
------------------------------------------------------------------------------
```

4.3 LSDV2 without the Intercept

LSDV2 reports actual parameter estimates of the dummies. You do not need to compute actual Y-intercept any more. Because LSDV2 suppresses the intercept, you will get incorrect F and R^2 statistics. However, the SSE of LSDV2 is correct.

In PROC REG, you need to use the /NOINT option to suppress the intercept. Obviously, the F value of 497,985 and R^2 of 1 are not likely. However, SSE, parameter estimates of regressors, and their standard errors are correct. Make sure that the intercepts presented in the beginning of Section 4.2 are what we got here using LSDV2.

```
PROC REG DATA=masil.airline;
   MODEL cost = g1-g6 output fuel load /NOINT;
RUN;
```

```
                              The REG Procedure
                                Model: MODEL1
                            Dependent Variable: cost

                        Number of Observations Read          90
                        Number of Observations Used          90


                 NOTE: No intercept in model. R-Square is redefined.

                                Analysis of Variance

                                          Sum of           Mean
Source                     DF            Squares         Square    F Value    Pr > F

Model                       9              16191     1799.03381     497985    <.0001
Error                      81            0.29262        0.00361
Uncorrected Total          90              16192


                  Root MSE              0.06011    R-Square     1.0000
                  Dependent Mean       13.36561    Adj R-Sq     1.0000
                  Coeff Var             0.44970


                                Parameter Estimates

                              Parameter       Standard
Variable       DF              Estimate          Error    t Value    Pr > |t|

g1              1               9.70594        0.19312      50.26      <.0001
g2              1               9.66471        0.19898      48.57      <.0001
g3              1               9.49702        0.22496      42.22      <.0001
g4              1               9.89050        0.24176      40.91      <.0001
g5              1               9.73000        0.26094      37.29      <.0001
g6              1               9.79300        0.26366      37.14      <.0001
output          1               0.91928        0.02989      30.76      <.0001
fuel            1               0.41749        0.01520      27.47      <.0001
load            1              -1.07040        0.20169      -5.31      <.0001
```

Stata uses the noconstant option to suppress the intercept. Notice that noc is its abbreviation.

```
. regress cost g1-g6 output fuel load, noc

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  9,    81) =       .
       Model |  16191.3043     9  1799.03381           Prob > F      =  0.0000
    Residual |  .292622872    81  .003612628           R-squared     =  1.0000
-------------+------------------------------           Adj R-squared =  1.0000
       Total |  16191.5969    90  179.906633           Root MSE      =  .06011

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          g1 |   9.705942    .193124    50.26   0.000     9.321686     10.0902
          g2 |   9.664706    .198982    48.57   0.000     9.268794    10.06062
          g3 |   9.497021   .2249584    42.22   0.000     9.049424    9.944618
          g4 |   9.890498   .2417635    40.91   0.000     9.409464    10.37153
          g5 |   9.729997   .2609421    37.29   0.000     9.210804    10.24919
```

```
          g6 |   9.793004   .2636622    37.14   0.000     9.268399    10.31761
      output |   .9192846   .0298901    30.76   0.000     .8598126    .9787565
        fuel |   .4174918   .0151991    27.47   0.000     .3872503    .4477333
        load |  -1.070396     .20169    -5.31   0.000    -1.471696   -.6690963
------------------------------------------------------------------------------
```

In LIMDEP, you need to drop `ONE` out of the `Rhs` subcommand to suppress the intercept. Unlike SAS and Stata, LIMDEP reports correct R^2 (.9974) and F (3,936) even in LSDV2.

```
REGRESS;Lhs=COST;Rhs=G1,G2,G3,G4,G5,G6,OUTPUT,FUEL,LOAD$
```

```
+----------------------------------------------------+
| Ordinary    least squares regression               |
| Model was estimated Aug 27, 2009 at 03:53:24PM     |
| LHS=COST     Mean                 =   13.36561     |
|              Standard deviation   =   1.131971     |
| WTS=none     Number of observs.   =         90     |
| Model size   Parameters           =          9     |
|              Degrees of freedom   =         81     |
| Residuals    Sum of squares       =   .2926208     |
|              Standard error of e  =   .6010493E-01 |
| Fit          R-squared            =   .9974341     |
|              Adjusted R-squared   =   .9971806     |
| Model test   F[  8,    81] (prob) =3935.82 (.0000) |
| Diagnostic   Log likelihood       =   130.0865     |
|              Restricted(b=0)      =  -138.3581     |
|              Chi-sq [  8]  (prob) = 536.89 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =  -5.528017     |
|              Akaike Info. Criter. =  -5.528687     |
| Autocorrel   Durbin-Watson Stat.  =  1.0264504     |
|              Rho = cor[e,e(-1)]   =   .4867748     |
| Not using OLS or no constant. Rsqd & F may be < 0. |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 G1      |    9.70594925       .19312325    50.258   .0000    .16666667
 G2      |    9.66471527       .19898117    48.571   .0000    .16666667
 G3      |    9.49703267       .22495746    42.217   .0000    .16666667
 G4      |    9.89051381       .24176245    40.910   .0000    .16666667
 G5      |    9.73001357       .26094094    37.288   .0000    .16666667
 G6      |    9.79302127       .26366104    37.142   .0000    .16666667
 OUTPUT  |     .91928814       .02988997    30.756   .0000  -1.17430918
 FUEL    |     .41749105       .01519907    27.468   .0000   12.7703592
 LOAD    |   -1.07039502       .20168924    -5.307   .0000    .56046016
```

4.4 LSDV3 with Restrictions

LSDV3 imposes a restriction that the sum of the dummy parameters is zero. PROC REG has the RESTRICT statement to impose restrictions. LSDV3 reports the correct ANOVA table and parameter estimates of regressors but produces different, compared to those of LSDV1 and LSDV2, dummy coefficients due to the different baseline (group average) used.

```
PROC REG DATA=masil.airline;
   MODEL cost = g1-g6 output fuel load;
   RESTRICT g1 + g2 + g3 + g4 + g5 + g6 = 0;
RUN;
```

```
                                The REG Procedure
                                  Model: MODEL1
                            Dependent Variable: cost
```

```
NOTE: Restrictions have been applied to parameter estimates.

                      Number of Observations Read          90
                      Number of Observations Used          90


                              Analysis of Variance

                                        Sum of           Mean
Source                    DF          Squares         Square    F Value    Pr > F

Model                      8        113.74827       14.21853    3935.79    <.0001
Error                     81          0.29262        0.00361
Corrected Total           89        114.04089


            Root MSE              0.06011    R-Square     0.9974
            Dependent Mean       13.36561    Adj R-Sq     0.9972
            Coeff Var             0.44970


                              Parameter Estimates

                         Parameter       Standard
Variable      DF          Estimate          Error    t Value    Pr > |t|

Intercept      1           9.71353        0.22964      42.30      <.0001
g1             1          -0.00759        0.04562      -0.17      0.8683
g2             1          -0.04882        0.03798      -1.29      0.2023
g3             1          -0.21651        0.01606     -13.48      <.0001
g4             1           0.17697        0.01942       9.11      <.0001
g5             1           0.01647        0.03669       0.45      0.6547
g6             1           0.07948        0.04050       1.96      0.0532
output         1           0.91928        0.02989      30.76      <.0001
fuel           1           0.41749        0.01520      27.47      <.0001
load           1          -1.07040        0.20169      -5.31      <.0001
RESTRICT      -1     3.01674E-15    7.82306E-11       0.00      1.0000*

                * Probability computed using beta distribution.
```

A dummy coefficient means the deviation from the averaged group effect (9.714). The actual intercept of airline 2, for example, is 9.6647 =9.7135+ (-.0488). Notice that the 3.01674E-15 of RESTRICT is virtually zero.

In Stata, you have to use the `.cnsreg` command in stead of `.regress`. The command, however, does not provide an ANOVA table and goodness-of-fit statistics other than F and SEE (standard error of residual--error term, square root of MSE).

```
. constraint define 1 g1 + g2 + g3 + g4 + g5 + g6 = 0
. cnsreg cost g1-g6 output fuel load, constraint(1)

Constrained linear regression                   Number of obs   =         90
                                                F(   8,     81) =    3935.79
                                                Prob > F        =     0.0000
                                                Root MSE        =     0.0601
```

```
 ( 1)  g1 + g2 + g3 + g4 + g5 + g6 = 0
------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          g1 |  -.0075859   .0456178    -0.17   0.868    -.0983509    .0831792
          g2 |  -.0488218   .0379787    -1.29   0.202    -.1243875    .0267439
          g3 |  -.2165069   .0160624   -13.48   0.000    -.2484661   -.1845478
          g4 |   .1769698   .0194247     9.11   0.000     .1383208    .2156189
          g5 |   .0164689   .0366904     0.45   0.655    -.0565335    .0894712
          g6 |   .0794759   .0405008     1.96   0.053     -.001108    .1600597
      output |   .9192846   .0298901    30.76   0.000     .8598126    .9787565
        fuel |   .4174918   .0151991    27.47   0.000     .3872503    .4477333
        load |  -1.070396     .20169    -5.31   0.000    -1.471696   -.6690963
       _cons |   9.713528    .229641    42.30   0.000     9.256614    10.17044
------------------------------------------------------------------------------
```

LIMDEP has the `Cls` subcommand to impose restrictions. Again, do not forget to include ONE in `Rhs`. `b(2)` in `Cls:` indicates the parameter of the second variable, `g1`, listed in `Rhs`.

```
REGRESS;Lhs=COST;Rhs=ONE,G1,G2,G3,G4,G5,G6,OUTPUT,FUEL,LOAD;
        Cls:b(2)+b(3)+b(4)+b(5)+b(6)+b(7)=0$
```

```
+-----------------------------------------------------+
| Linearly restricted regression                      |
| Ordinary    least squares regression                |
| Model was estimated Aug 31, 2009 at 06:39:21PM      |
| LHS=COST     Mean                 =   13.36561      |
|              Standard deviation   =   1.131971      |
| WTS=none     Number of observs.   =         90      |
| Model size   Parameters           =          9      |
|              Degrees of freedom   =         81      |
| Residuals    Sum of squares       =   .2926208      |
|              Standard error of e  =   .6010493E-01 |
| Fit          R-squared            =   .9974341      |
|              Adjusted R-squared   =   .9971806      |
| Model test   F[  8,    81] (prob) =3935.82 (.0000)  |
| Diagnostic   Log likelihood       =   130.0865      |
|              Restricted(b=0)      =  -138.3581      |
|              Chi-sq [  8]  (prob) = 536.89 (.0000)  |
| Info criter. LogAmemiya Prd. Crt. =  -5.528017      |
|              Akaike Info. Criter. =  -5.528687      |
| Autocorrel   Durbin-Watson Stat.  =  1.0264504      |
|              Rho = cor[e,e(-1)]   =   .4867748      |
| Restrictns.  F[  1,    80] (prob) =    .00 (*****)  |
| Not using OLS or no constant. Rsqd & F may be < 0.  |
| Note, with restrictions imposed,  Rsqd may be < 0.  |
+-----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    9.71354097       .22964002    42.299   .0000
 G1      |    -.00759172       .04561756     -.166   .8682    .16666667
 G2      |    -.04882570       .03797853    -1.286   .2023    .16666667
 G3      |    -.21650830       .01606233   -13.479   .0000    .16666667
 G4      |     .17697283       .01942459     9.111   .0000    .16666667
 G5      |     .01647259       .03669023      .449   .6547    .16666667
 G6      |     .07948030       .04050059     1.962   .0532    .16666667
 OUTPUT  |     .91928814       .02988997    30.756   .0000  -1.17430918
 FUEL    |     .41749105       .01519907    27.468   .0000   12.7703592
 LOAD    |   -1.07039502       .20168924    -5.307   .0000    .56046016
```

LSDV3 in LIMDEP reports different dummy coefficients. But you may compute actual intercepts of groups in a manner similar to what you would do in SAS and Stata. The actual intercept of airline 5, for example, is 9.7300 = 12.1221 + (-2.3920).

4.5 Within Group Effect Model

The within effect model does not use dummy variables and thus has larger degrees of freedom, smaller MSE, and smaller standard errors of parameters than those of LSDV. As a consequence,

you need to adjust standard errors. This model does not report individual dummy coefficients either; you need to compute them if really needed. The SAS TSCSREG and PANEL procedures and LIMDEP `Regress$` command report the adjusted (correct) MSE, SEE (square root of MSE), R^2, and standard errors.

4.5.1 Estimating the Within Effect Model

First, let us manually estimate the within group effect model with Stata. You need to compute group means.

```
. quietly egen gm_cost=mean(cost), by(airline)
. quietly egen gm_output=mean(output), by(airline)
. quietly egen gm_fuel=mean(fuel), by(airline)
. quietly egen gm_load=mean(load), by(airline)
```

You will get the following group means of variables.

```
  +-----------------------------------------------------+
  | airline    gm_cost   gm_output    gm_fuel    gm_load |
  |-----------------------------------------------------|
  |       1   14.67563    .3192696    12.7318   .5971917 |
  |       2   14.37247    -.033027   12.75171   .5470946 |
  |       3   13.37231   -.9122626   12.78972   .5845358 |
  |       4    13.1358   -1.635174   12.77803   .5476773 |
  |       5   12.36304   -2.285681    12.7921   .5664859 |
  |       6   12.27441    -2.49898    12.7788   .5197756 |
  +-----------------------------------------------------+
```

Then transform dependent and independent variables to compute deviations from group means.

```
. quietly gen gw_cost = cost - gm_cost
. quietly gen gw_output = output - gm_output
. quietly gen gw_fuel = fuel - gm_fuel
. quietly gen gw_load = load - gm_load
```

Now, we are ready to run the within effect model. Keep in mind that you have to suppress the intercept. The within effect model reports correct SSE and parameter estimates of regressors but incorrect R^2 and standard errors of parameter estimates. Notice that the degrees of freedom increase from 81 (LSDV) to 87 since six dummy variables are not used.

```
. regress gw_cost gw_output gw_fuel gw_load, noc

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  3,    87) = 3871.82
       Model |  39.0683861     3  13.0227954           Prob > F      =  0.0000
    Residual |  .292622861    87  .003363481           R-squared     =  0.9926
-------------+------------------------------           Adj R-squared =  0.9923
       Total |   39.361009    90  .437344544           Root MSE      =    .058

------------------------------------------------------------------------------
     gw_cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   gw_output |   .9192846    .028841    31.87   0.000       .86196    .9766092
     gw_fuel |   .4174918   .0146657    28.47   0.000     .3883422    .4466414
     gw_load |  -1.070396   .1946109    -5.50   0.000    -1.457206   -.6835858
------------------------------------------------------------------------------
```

You may compute group intercepts using $d_i^* = \bar{y}_{i\bullet} - \beta' \bar{x}_{i\bullet}$. For example, the intercept of airline 5 is computed as 9.730 = 12.3630 – {.9193*(-2.2857) + .4175*12.7921 + (-1.0704)*.5665}. In order to get the correct standard errors, you need to adjust them using the ratio of degrees of

freedom of the within effect model and LSDV. For example, the standard error of the logged output is computed as .0299=.0288*sqrt(87/81).

4.5.2 Using SAS: PROC TSCSREG and PROC PANEL

PROC TSCSREG and PROC PANEL of SAS/ETS allows users to fit the within effect model conveniently. They, in fact, report LSDV1, but you do not need to create dummy variables and compute deviations from group means.

```
PROC SORT DATA=masil.airline;
   BY airline year;
```

A data set needs to be sorted in advance by the variables, which will appear in the ID statement of PROC TSCSREG and PROC PANEL. These time-series and cross-sectional variables may be numeric or string in SAS. /FIXONE of the MODEL statement fits a one-way fixed effect model.

```
PROC TSCSREG DATA=masil.airline;
   ID airline year;
   MODEL cost = output fuel load /FIXONE;
RUN;
```

```
                                The TSCSREG Procedure
                               Fixed One Way Estimates

Dependent Variable: cost

                                  Model Description

                          Estimation Method            FixOne
                          Number of Cross Sections          6
                          Time Series Length               15


                                   Fit Statistics

                    SSE             0.2926    DFE                 81
                    MSE             0.0036    Root MSE        0.0601
                    R-Square        0.9974


                          F Test for No Fixed Effects

                         Num DF      Den DF    F Value    Pr > F

                              5          81      57.73    <.0001


                                  Parameter Estimates

                                    Standard
    Variable        DF    Estimate      Error    t Value    Pr > |t|    Label

    CS1              1    -0.08706     0.0842      -1.03      0.3042    Cross Sectional
```

```
                                                                  Effect    1
    CS2               1     -0.1283      0.0757      -1.69      0.0941    Cross Sectional
                                                                  Effect    2
    CS3               1    -0.29598      0.0500      -5.92      <.0001    Cross Sectional
                                                                  Effect    3
    CS4               1    0.097494      0.0330       2.95      0.0041    Cross Sectional
                                                                  Effect    4
    CS5               1    -0.06301      0.0239      -2.64      0.0100    Cross Sectional
                                                                  Effect    5
    Intercept         1    9.793004      0.2637      37.14      <.0001    Intercept
    output            1    0.919285      0.0299      30.76      <.0001
    fuel              1    0.417492      0.0152      27.47      <.0001
    load              1     -1.0704      0.2017      -5.31      <.0001
```

The following PANEL procedure returns the same output.

```
PROC PANEL DATA=masil.airline;
   ID airline year;
   MODEL cost = output fuel load /FIXONE;
RUN;
```

Both PROC TSCSREG and PROC PANEL report correct (adjusted) MSE, SEE, R^2, and standard errors, and conduct the F test for fixed group effect as well. They have strong advantages over other software packages in this respect.

4.5.3 Using Stata

The Stata `.xtreg` command fits the within group effect model without creating dummy variables. `.xtreg` should follow the `.tsset` command that specifies cross-sectional and time-series variables. Both variables should be numeric in Stata; string variables are not allowed in `.tsset`.

```
. quietly tsset airline year
```

The `fe` option of `.xtreg` indicates the within effect model and `i(airline)` specifies `airline` as the independent unit. Once `.tsset` is executed, `i(airline)` is redundant. This command report incorrect F 3,604 and R^2 of .9926.

```
. xtreg cost output fuel load, fe i(airline)

Fixed-effects (within) regression               Number of obs      =        90
Group variable: airline                         Number of groups   =         6

R-sq:  within  = 0.9926                         Obs per group: min =        15
       between = 0.9856                                        avg =      15.0
       overall = 0.9873                                        max =        15

                                                F(3,81)            =   3604.80
corr(u_i, Xb)  = -0.3475                        Prob > F           =    0.0000

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |   .9192846   .0298901    30.76   0.000     .8598126    .9787565
        fuel |   .4174918   .0151991    27.47   0.000     .3872503    .4477333
        load |  -1.070396     .20169    -5.31   0.000    -1.471696   -.6690963
       _cons |   9.713528    .229641    42.30   0.000     9.256614    10.17044
```

```
-------------+----------------------------------------------------------------
     sigma_u |   .1320775
     sigma_e |  .06010514
         rho |  .82843653   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(5, 81) =    57.73              Prob > F = 0.0000
```

Like PROC PANEL, `.xtreg` reports correct standard errors and the F test for a fixed group effect. But this command does not provide an analysis of variance (ANOVA) table. R^2 and F statistic are not correct. The last line of the output tests the null hypothesis that five dummy parameters in LSDV1 are zero (e.g., $\mu_1=0$, $\mu_2=0$, $\mu_3=0$, $\mu_4=0$, and $\mu_5=0$). Notice that the intercept of 9.7135 is that of LSDV3.

Alternatively, you may use `.areg` to get the same result except for R^2, which is correct. The intercept 9.7135 is the average of six airlines, the intercept of LSDV3.

```
. areg cost output fuel load, absorb(airline)

Linear regression, absorbing indicators                Number of obs =      90
                                                       F(  3,    81) = 3604.80
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.9974
                                                       Adj R-squared =  0.9972
                                                       Root MSE      =  .06011

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |   .9192846   .0298901    30.76   0.000     .8598126    .9787565
        fuel |   .4174918   .0151991    27.47   0.000     .3872503    .4477333
        load |  -1.070396     .20169    -5.31   0.000    -1.471696   -.6690963
       _cons |   9.713528    .229641    42.30   0.000     9.256614    10.17044
-------------+----------------------------------------------------------------
     airline |          F(5, 81) =     57.732   0.000           (6 categories)
```

4.5.4 Using LIMDEP

In LIMDEP, the `Panel` and `Fixed` subcommands in the `Regress$` command fit a fixed effect panel data model. The `Str` subcommand specifies a stratification variable.

```
REGRESS;Lhs=COST;Rhs=ONE,OUTPUT,FUEL,LOAD;Panel;Str=AIRLINE;Fixed$
```

```
+---------------------------------------------------+
| OLS Without Group Dummy Variables                 |
| Ordinary    least squares regression              |
| Model was estimated Aug 27, 2009 at 03:56:52PM    |
| LHS=COST     Mean                 =   13.36561    |
|              Standard deviation   =   1.131971    |
| WTS=none     Number of observs.   =         90    |
| Model size   Parameters           =          4    |
|              Degrees of freedom   =         86    |
| Residuals    Sum of squares       =   1.335450    |
|              Standard error of e  =   .1246133   |
| Fit          R-squared            =   .9882897    |
|              Adjusted R-squared   =   .9878812    |
| Model test   F[  3,    86] (prob) =2419.33 (.0000)|
| Diagnostic   Log likelihood       =   61.76991    |
|              Restricted(b=0)      =  -138.3581    |
|              Chi-sq [  3]  (prob) = 400.26 (.0000)|
| Info criter. LogAmemiya Prd. Crt. =  -4.121594    |
|              Akaike Info. Criter. =  -4.121653    |
+---------------------------------------------------+
```

```
+----------------------------------------------------+
| Panel Data Analysis of COST       [ONE way]        |
|          Unconditional ANOVA (No regressors)       |
| Source      Variation   Deg. Free.     Mean Square |
| Between       74.6799           5.     14.9360     |
| Residual      39.3611          84.     .468584     |
| Total         114.041          89.     1.28136     |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 OUTPUT  |     .88273863       .01325455    66.599   .0000  -1.17430918
 FUEL    |     .45397771       .02030424    22.359   .0000   12.7703592
 LOAD    |   -1.62750780       .34530293    -4.713   .0000    .56046016
 Constant|    9.51691223       .22924522    41.514   .0000

+----------------------------------------------------+
| Least Squares with Group Dummy Variables           |
| Ordinary    least squares regression               |
| Model was estimated Aug 27, 2009 at 03:56:52PM     |
| LHS=COST     Mean                 =   13.36561     |
|              Standard deviation   =   1.131971     |
| WTS=none     Number of observs.   =         90     |
| Model size   Parameters           =          9     |
|              Degrees of freedom   =         81     |
| Residuals    Sum of squares       =   .2926208     |
|              Standard error of e  =   .6010493E-01 |
| Fit          R-squared            =   .9974341     |
|              Adjusted R-squared   =   .9971806     |
| Model test   F[  8,    81] (prob) =3935.82 (.0000) |
| Diagnostic   Log likelihood       =   130.0865     |
|              Restricted(b=0)      =  -138.3581     |
|              Chi-sq [  8]  (prob) = 536.89 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =  -5.528017     |
|              Akaike Info. Criter. =  -5.528687     |
| Estd. Autocorrelation of e(i,t)     .573531        |
+----------------------------------------------------+

+----------------------------------------------------+
| Panel:Groups   Empty       0,   Valid data       6 |
|                Smallest   15,   Largest         15 |
|                Average group size            15.00 |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 OUTPUT  |     .91928814       .02988997    30.756   .0000  -1.17430918
 FUEL    |     .41749105       .01519907    27.468   .0000   12.7703592
 LOAD    |   -1.07039502       .20168924    -5.307   .0000    .56046016

+---------------------------------------------------------------------+
|             Test Statistics for the Classical Model                 |
+---------------------------------------------------------------------+
|       Model            Log-Likelihood  Sum of Squares    R-squared |
|(1)  Constant term only     -138.35814  .1140409821D+03    .0000000 |
|(2)  Group effects only      -90.48804  .3936109461D+02    .6548513 |
|(3)  X - variables only       61.76991  .1335449522D+01    .9882897 |
|(4)  X and group effects     130.08647  .2926207777D+00    .9974341 |
+---------------------------------------------------------------------+
|                        Hypothesis Tests                             |
|         Likelihood Ratio Test           F Tests                     |
|         Chi-squared   d.f.  Prob.       F    num. denom.   P value |
|(2) vs (1)    95.740      5  .00000   31.875     5      84   .00000 |
|(3) vs (1)   400.256      3  .00000 2419.329     3      86   .00000 |
|(4) vs (1)   536.889      8  .00000 3935.818     8      81   .00000 |
|(4) vs (2)   441.149      3  .00000 3604.832     3      81   .00000 |
|(4) vs (3)   136.633      5  .00000   57.733     5      81   .00000 |
+---------------------------------------------------------------------+
```

LIMDEP reports both the pooled OLS regression under the label `OLS Without Group Dummy Variables` and the within effect model under `Least Squares with Group Dummy Variables`. Like the SAS TSCSREG procedure, LIMDEP provides correct MSE, SEE, R^2, and standard errors of the fixed effect model. LIMDEP also conducts the F test for checking a fixed group effect (see the last line of the LIMDEP output above to get 57.733).

4.6 Between Group Effect Model: Group Mean Regression

A between effect model uses aggregate information, group means of variables. In other words, the unit of analysis is not an individual observation, but entity or subject. The number of observations jumps down to *n* from *nT*. This group mean regression produces different goodness-of-fit measures and parameter estimates compared to those of LSDV and the within effect model.

Let us compute group means and run OLS with them. The `.collapse` command computes aggregate information and stores into a new data set. This model fits data relatively well but its t-tests report insignificant parameters. Note that /// links two command lines.

```
. collapse (mean) gm_cost=cost (mean) gm_output=output (mean) gm_fuel=fuel (mean) ///
  gm_load=load, by(airline)

. regress gm_cost gm_output gm_fuel gm_load

      Source |       SS       df       MS              Number of obs =       6
-------------+------------------------------           F(  3,     2) =  104.12
       Model |  4.94698124     3  1.64899375           Prob > F      =  0.0095
    Residual |  .031675926     2  .015837963           R-squared     =  0.9936
-------------+------------------------------           Adj R-squared =  0.9841
       Total |  4.97865717     5  .995731433           Root MSE      =  .12585

------------------------------------------------------------------------------
     gm_cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   gm_output |   .7824568   .1087646     7.19   0.019     .3144803    1.250433
     gm_fuel |  -5.523904   4.478718    -1.23   0.343    -24.79427    13.74647
     gm_load |  -1.751072   2.743167    -0.64   0.589    -13.55397    10.05182
       _cons |    85.8081   56.48199     1.52   0.268    -157.2143    328.8305
------------------------------------------------------------------------------
```

The SAS PANEL procedure has the /BTWNG and /BTWNT option to estimate the between effect model, but PROC TSCSREG does not. /BTWNG and /BTWNT fit the between group and time effect models, respectively.

```
PROC PANEL DATA=masil.airline;
   ID airline year;
   MODEL cost = output fuel load /BTWNG;
RUN;
                                   The PANEL Procedure
                                Between Groups Estimates

Dependent Variable: cost

                                   Model Description

                             Estimation Method           BtwGrps
                             Number of Cross Sections          6
```

```
                          Time Series Length                15


                                 Fit Statistics

                    SSE             0.0317    DFE                    2
                    MSE             0.0158    Root MSE          0.1258
                    R-Square        0.9936


                               Parameter Estimates

                                      Standard
    Variable        DF    Estimate       Error    t Value    Pr > |t|    Label

    Intercept        1    85.80901     56.4830       1.52      0.2681    Intercept
    output           1    0.782455      0.1088       7.19      0.0188
    fuel             1    -5.52398      4.4788      -1.23      0.3427
    load             1    -1.75102      2.7432      -0.64      0.5886
```

The Stata `.xtreg` command has the be option to fit the between effect model but does not report the ANOVA table.

```
. xtreg cost output fuel load, be i(airline)

Between regression (regression on group means)  Number of obs      =        90
Group variable: airline                         Number of groups   =         6

R-sq:  within  = 0.8808                         Obs per group: min =        15
       between = 0.9936                                        avg =      15.0
       overall = 0.1371                                        max =        15

                                                F(3,2)             =    104.12
sd(u_i + avg(e_i.))=  .1258491                  Prob > F           =    0.0095

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |   .7824552   .1087663     7.19   0.019     .3144715    1.250439
        fuel |  -5.523978   4.478802    -1.23   0.343    -24.79471    13.74675
        load |  -1.751016    2.74319    -0.64   0.589    -13.55401    10.05198
       _cons |   85.80901   56.48302     1.52   0.268    -157.2178    328.8358
------------------------------------------------------------------------------
```

LIMDEP has the `Means` subcommand to fit the between effect model.

```
REGRESS;Lhs=COST;Rhs=ONE,OUTPUT,FUEL,LOAD;Panel;Str=AIRLINE;Means$
```

```
+----------------------------------------------------+
| Group Means Regression                             |
| Ordinary    least squares regression               |
| Model was estimated Aug 27, 2009 at 04:04:12PM     |
| LHS=YBAR(i.) Mean                 =   13.36561     |
|              Standard deviation   =   .9978636     |
| WTS=NTi/Nobs Number of observs.   =          6     |
| Model size   Parameters           =          4     |
|              Degrees of freedom   =          2     |
| Residuals    Sum of squares       =   .3167277E-01 |
|              Standard error of e  =   .1258427     |
| Fit          R-squared            =   .9936383     |
|              Adjusted R-squared   =   .9840957     |
| Model test   F[  3,     2] (prob) = 104.13 (.0095) |
| Diagnostic   Log likelihood       =   7.218541     |
|              Restricted(b=0)      =  -7.953835     |
|              Chi-sq [  3]  (prob) =  30.34 (.0000) |
```

```
| Info criter. LogAmemiya Prd. Crt. =  -3.634619     |
|              Akaike Info. Criter. =  -3.910724     |
+----------------------------------------------------+
+--------+-------------+----------------+--------+--------+----------+
|Variable| Coefficient | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+----------------+--------+--------+----------+
 OUTPUT  |     .78244727       .10876126     7.194   .0000  .230256D-11
 FUEL    |   -5.52443747      4.47865187    -1.234   .2174    .18642891
 LOAD    |   -1.75094765      2.74304702     -.638   .5233    .32541105
 Constant|    85.8148317      56.4811479     1.519   .1287
```

SAS, Stata, and LIMDEP all report the same result: SSE .0317, SEE .1258, F 104.12 (p<.0095), and R^2 .9936.

4.7 Testing Fixed Group Effects (F-test)

How do we know whether there is a significant fixed group effect? The null hypothesis is that all dummy parameters except for one are zero: $H_0 : \mu_1 = ... = \mu_{n-1} = 0$.

In order to conduct a F-test, let us obtain the SSE (e'e) of 1.3354 from the pooled OLS regression and .2926 from the LSDVs (LSDV1 through LSDV3) or the within effect model. Alternatively, you may draw R^2 of .9974 from LSDV1 or LSDV3 and .9883 from the pooled OLS. Do not, however, use LSDV2 and the within effect model for R^2.

The F statistic is computed as $\frac{(1.3354-.2926)/(6-1)}{(.2926)/(90-6-3)} = \frac{(.9974-.9883)/(6-1)}{(1-.9974)/(90-6-3)} \sim 57.7319[5,81]$.

The large F statistic rejects the null hypothesis in favor of the fixed group effect model (p<.0000). There is a fixed group effect in these panel data.

The SAS TSCSREG and PANEL procedures, Stata `.xtreg` command, and LIMDEP `Regress$` command by default conduct the F test. Alternatively, you may conduct the same test in LSDV1. In SAS, add the TEST statement in PROC REG and then run the procedure again (ANOVA table and parameter estimates are skipped).

```
PROC REG DATA=masil.airline;
   MODEL cost = g1-g5 output fuel load;
   TEST g1 = g2 = g3 = g4 = g5 = 0;
RUN;
```

```
                          The REG Procedure
                            Model: MODEL1

                Test 1 Results for Dependent Variable cost

                                          Mean
Source                DF                Square    F Value    Pr > F

Numerator              5               0.20856      57.73    <.0001
Denominator           81               0.00361
```

In Stata, run the `.test` command, a follow-up command for the Wald test, right after estimating the model.

```
. quietly regress cost g1-g5 output fuel load
. test g1 g2 g3 g4 g5

 ( 1)  g1 = 0
 ( 2)  g2 = 0
 ( 3)  g3 = 0
 ( 4)  g4 = 0
 ( 5)  g5 = 0

       F(  5,    81) =   57.73
            Prob > F =    0.0000
```

4.8 Summary

Table 4.1 summarizes the estimation of a fixed effect model in SAS, Stata, and LIMDEP. The SAS PANEL procedure is generally preferred to Stata and LIMDEP counterparts since it produces correct statistics and conducts various hypothesis tests conveniently.

Table 4.1 Comparison of the Fixed Effect Model in SAS, Stata, LIMDEP*

	SAS 9	Stata 11	LIMDEP 9
OLS estimation	`PROC REG;`	`.regress, .cnsreg`	`Regress$`
LSDV1	Correct	Correct	Correct (slightly different F)
LSDV2	Incorrect F, (adjusted) R^2	Incorrect F, (adjusted) R^2	Correct (slightly different F)
LSDV3	Correct	`.cnsreg` No ANOVA table and R^2	Correct (slightly different F) Different dummy coefficients
Panel Estimation	`PROC TSCSREG;` `PROC PANEL;`	`.xtreg, .areg`	`Regress; Panel$`
Estimation type	LSDV1	Within effect	Within effect
SSE (e'e)	Correct	No	Correct
MSE or SEE	Correct (adjusted)	No	Correct (adjusted) SEE
Model test (F)	No	Incorrect	Slightly different F
(adjusted) R^2	Correct	Incorrect (correct in `.areg`)	Correct
Intercept	Correct	LSDV3 intercept	No
Coefficients	Correct	Correct	Correct
Standard errors	Correct (adjusted)	Correct (adjusted)	Correct (adjusted)
Effect test (F)	Yes	Yes	Yes
Between effect	/BTWNG, /BTWNT	`,be`	`Means;`

* "Yes/No" means whether the software reports the statistics. "Correct/incorrect" indicates whether the statistics are different from those of the least squares dummy variable (LSDV) 1 without a dummy variable.

5. One-way Fixed Effect Models: Time Effects

A fixed time effect model investigates how time affects the intercept using time dummy variables. The logic and method are the same as those of the fixed group effect model.

5.1 Least Squares Dummy Variable Models

The least squares dummy variable (LSDV) model produces the following fifteen regression equations

Time 01: cost = 20.4959 + .8677*output - .4845*fuel -1.9544*load
Time 02: cost = 20.5782 + .8677*output - .4845*fuel -1.9544*load
Time 03: cost = 20.6559 + .8677*output - .4845*fuel -1.9544*load
Time 04: cost = 20.7409 + .8677*output - .4845*fuel -1.9544*load
Time 05: cost = 21.2000 + .8677*output - .4845*fuel -1.9544*load
Time 06: cost = 21.4118 + .8677*output - .4845*fuel -1.9544*load
Time 07: cost = 21.5035 + .8677*output - .4845*fuel -1.9544*load
Time 08: cost = 21.6542 + .8677*output - .4845*fuel -1.9544*load
Time 09: cost = 21.8397 + .8677*output - .4845*fuel -1.9544*load
Time 10: cost = 22.1140 + .8677*output - .4845*fuel -1.9544*load
Time 11: cost = 22.4655 + .8677*output - .4845*fuel -1.9544*load
Time 12: cost = 22.6515 + .8677*output - .4845*fuel -1.9544*load
Time 13: cost = 22.6167 + .8677*output - .4845*fuel -1.9544*load
Time 14: cost = 22.5524 + .8677*output - .4845*fuel -1.9544*load
Time 15: cost = 22.5369 + .8677*output - .4845*fuel -1.9544*load

5.1.1 LSDV1 without a Dummy

In SAS REG procedure, include time dummy variables instead of group dummies. You need to exclude one of time dummies, say `t15` here, in LSDV1.

```
PROC REG DATA=masil.airline;
   MODEL cost = t1-t14 output fuel load;
RUN;
```

```
                                The REG Procedure
                                  Model: MODEL1
                              Dependent Variable: cost

                        Number of Observations Read          90
                        Number of Observations Used          90


                                Analysis of Variance

                                           Sum of           Mean
        Source                     DF      Squares         Square    F Value    Pr > F

        Model                      17    112.95270        6.64428     439.62    <.0001
        Error                      72      1.08819        0.01511
```

```
Corrected Total          89      114.04089


                Root MSE             0.12294    R-Square     0.9905
                Dependent Mean      13.36561    Adj R-Sq     0.9882
                Coeff Var            0.91981


                              Parameter Estimates

                           Parameter       Standard
      Variable     DF       Estimate          Error    t Value    Pr > |t|

      Intercept     1       22.53677        4.94053       4.56      <.0001
      t1            1       -2.04096        0.73469      -2.78      0.0070
      t2            1       -1.95873        0.72275      -2.71      0.0084
      t3            1       -1.88103        0.72036      -2.61      0.0110
      t4            1       -1.79601        0.69882      -2.57      0.0122
      t5            1       -1.33693        0.50604      -2.64      0.0101
      t6            1       -1.12514        0.40862      -2.75      0.0075
      t7            1       -1.03341        0.37642      -2.75      0.0076
      t8            1       -0.88274        0.32601      -2.71      0.0085
      t9            1       -0.70719        0.29470      -2.40      0.0190
      t10           1       -0.42296        0.16679      -2.54      0.0134
      t11           1       -0.07144        0.07176      -1.00      0.3228
      t12           1        0.11457        0.09841       1.16      0.2482
      t13           1        0.07979        0.08442       0.95      0.3477
      t14           1        0.01546        0.07264       0.21      0.8320
      output        1        0.86773        0.01541      56.32      <.0001
      fuel          1       -0.48448        0.36411      -1.33      0.1875
      load          1       -1.95440        0.44238      -4.42      <.0001
```

In Stata and LIMDEP, execute following commands to fit the same LSDV1 (output is skipped).

```
. regress cost t1-t14 output fuel load
```

```
REGRESS;Lhs=COST;Rhs=ONE,T1,T2,T3,T4,T5,T6,T7,T8,T9,T10,T11,T12,T13,T14,OUTPUT,FUEL,LOAD$
```

5.1.2 LSDV2 without the Intercept

In LIMDEP, take `ONE` out to fit LSDV2 by suppressing the intercept. Unlike SAS and Stata, LIMDEP reports correct, although slightly different, F and R^2 statistics.

```
REGRESS;Lhs=COST;Rhs=T1,T2,T3,T4,T5,T6,T7,T8,T9,T10,T11,T12,T13,T14,T15,OUTPUT,FUEL,LOAD$
```

```
+-----------------------------------------------------+
| Ordinary    least squares regression                |
| Model was estimated Aug 27, 2009 at 04:15:08PM      |
| LHS=COST     Mean                 =   13.36561      |
|              Standard deviation   =   1.131971      |
| WTS=none     Number of observs.   =         90      |
| Model size   Parameters           =         18      |
|              Degrees of freedom   =         72      |
| Residuals    Sum of squares       =   1.088193      |
|              Standard error of e  =   .1229382      |
| Fit          R-squared            =   .9904579      |
|              Adjusted R-squared   =   .9882049      |
| Model test   F[ 17,    72] (prob) = 439.62 (.0000)  |
| Diagnostic   Log likelihood       =   70.98362      |
|              Restricted(b=0)      =  -138.3581      |
```

```
|              Chi-sq [ 17]  (prob) = 418.68 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =  -4.009826     |
|              Akaike Info. Criter. =  -4.015291     |
| Autocorrel   Durbin-Watson Stat.  =   .2363289     |
|              Rho = cor[e,e(-1)]   =   .8818355     |
| Not using OLS or no constant. Rsqd & F may be < 0. |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 T1      |    20.4959389      4.20954636     4.869   .0000    .06666667
 T2      |    20.5781713      4.22154389     4.875   .0000    .06666667
 T3      |    20.6558664      4.22419549     4.890   .0000    .06666667
 T4      |    20.7408923      4.24576770     4.885   .0000    .06666667
 T5      |    21.1999763      4.44035103     4.774   .0000    .06666667
 T6      |    21.4117634      4.53864000     4.718   .0000    .06666667
 T7      |    21.5034994      4.57141663     4.704   .0000    .06666667
 T8      |    21.6541766      4.62290530     4.684   .0000    .06666667
 T9      |    21.8297215      4.65692608     4.688   .0000    .06666667
 T10     |    22.1139553      4.79266903     4.614   .0000    .06666667
 T11     |    22.4654855      4.94992975     4.539   .0000    .06666667
 T12     |    22.6514956      5.00861379     4.523   .0000    .06666667
 T13     |    22.6167135      4.98616006     4.536   .0000    .06666667
 T14     |    22.5523879      4.95596262     4.551   .0000    .06666667
 T15     |    22.5369251      4.94055238     4.562   .0000    .06666667
 OUTPUT  |     .86772681       .01540818    56.316   .0000  -1.17430918
 FUEL    |    -.48449467       .36410984    -1.331   .1875   12.7703592
 LOAD    |   -1.95441438       .44237791    -4.418   .0000    .56046016
```

In SAS and Stata, use /NOINT and `noconstant`, respectively, to suppress the intercept and estimate the same LSDV2 (output is skipped).

```
PROC REG DATA=masil.airline;
   MODEL cost = t1-t15 output fuel load /NOINT;
RUN;
```

```
. regress cost t1-t15 output fuel load, noc
```

5.1.3 LSDV3 with a Restriction

In PROC REG, you need to impose a restriction using the RESTRICT statement.

```
PROC REG DATA=masil.airline;
   MODEL cost = t1-t15 output fuel load;
   RESTRICT t1 + t2 + t3 + t4 + t5 + t6 + t7 + t8 + t9 + t10 + t11 + t12 + t13 + t14 + t15 = 0;
RUN;
```

```
                                The REG Procedure
                                  Model: MODEL1
                            Dependent Variable: cost

NOTE: Restrictions have been applied to parameter estimates.


                          Number of Observations Read          90
                          Number of Observations Used          90


                                Analysis of Variance

                                           Sum of           Mean
        Source                     DF      Squares         Square    F Value    Pr > F
```

```
Model                    17      112.95270        6.64428     439.62    <.0001
Error                    72        1.08819        0.01511
Corrected Total          89      114.04089


              Root MSE              0.12294    R-Square     0.9905
              Dependent Mean       13.36561    Adj R-Sq     0.9882
              Coeff Var             0.91981


                              Parameter Estimates

                           Parameter       Standard
     Variable     DF       Estimate          Error    t Value    Pr > |t|

     Intercept     1       21.66698        4.62405       4.69      <.0001
     t1            1       -1.17118        0.41783      -2.80      0.0065
     t2            1       -1.08894        0.40586      -2.68      0.0090
     t3            1       -1.01125        0.40323      -2.51      0.0144
     t4            1       -0.92622        0.38177      -2.43      0.0178
     t5            1       -0.46715        0.19076      -2.45      0.0168
     t6            1       -0.25536        0.09856      -2.59      0.0116
     t7            1       -0.16363        0.07190      -2.28      0.0258
     t8            1       -0.01296        0.04862      -0.27      0.7907
     t9            1        0.16259        0.06271       2.59      0.0115
     t10           1        0.44682        0.17599       2.54      0.0133
     t11           1        0.79834        0.32940       2.42      0.0179
     t12           1        0.98435        0.38756       2.54      0.0132
     t13           1        0.94957        0.36537       2.60      0.0113
     t14           1        0.88524        0.33549       2.64      0.0102
     t15           1        0.86978        0.32029       2.72      0.0083
     output        1        0.86773        0.01541      56.32      <.0001
     fuel          1       -0.48448        0.36411      -1.33      0.1875
     load          1       -1.95440        0.44238      -4.42      <.0001
     RESTRICT     -1    -3.9462E-15              .          .          .

                  * Probability computed using beta distribution.
```

In Stata, define the restriction with the `.constraint` command and specify the restriction using the `constraint()` option of the `.cnsreg` command.

```
. constraint define 3 t1+t2+t3+t4+t5+t6+t7+t8+t9+t10+t11+t12+t13+t14+t15=0
. cnsreg cost t1-t15 output fuel load, constraint(3)

Constrained linear regression                   Number of obs   =         90
                                                F(  17,     72) =     439.62
                                                Prob > F        =     0.0000
                                                Root MSE        =     0.1229

 ( 1)  t1 + t2 + t3 + t4 + t5 + t6 + t7 + t8 + t9 + t10 + t11 + t12 + t13 + t14 + t15 = 0
------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          t1 |  -1.171179   .4178338    -2.80   0.007    -2.004115   -.3382422
          t2 |  -1.088945   .4058579    -2.68   0.009    -1.898008   -.2798816
          t3 |  -1.011252   .4032308    -2.51   0.014    -1.815078   -.2074266
          t4 |  -.9262249   .3817675    -2.43   0.018    -1.687265   -.1651852
          t5 |  -.4671515   .1907596    -2.45   0.017    -.8474239   -.0868791
          t6 |  -.2553627   .0985615    -2.59   0.012    -.4518415   -.0588839
          t7 |  -.1636326   .0718969    -2.28   0.026    -.3069564   -.0203088
```

```
          t8 |  -.0129552   .0486249    -0.27   0.791    -.1098872    .0839768
          t9 |   .1625876   .0627099     2.59   0.012     .0375776    .2875976
         t10 |   .4468191    .175994     2.54   0.013     .0959814    .7976568
         t11 |   .7983439   .3294027     2.42   0.018     .1416916    1.454996
         t12 |   .9843536   .3875583     2.54   0.013     .2117702    1.756937
         t13 |   .9495716   .3653675     2.60   0.011     .2212248    1.677918
         t14 |   .8852448   .3354912     2.64   0.010     .2164554    1.554034
         t15 |   .8697821   .3202933     2.72   0.008     .2312891    1.508275
      output |   .8677268   .0154082    56.32   0.000     .8370111    .8984424
        fuel |  -.4844835   .3641085    -1.33   0.188    -1.210321    .2413535
        load |  -1.954404   .4423777    -4.42   0.000    -2.836268    -1.07254
       _cons |   21.66698   4.624053     4.69   0.000      12.4491    30.88486
------------------------------------------------------------------------------
```

In LIMDEP, run the following command to fit the same LSDV3.

```
REGRESS;Lhs=COST;Rhs=ONE,T1,T2,T3,T4,T5,T6,T7,T8,T9,T10,T11,T12,T13,T14,T15,OUTPUT,FUEL,LOAD;
       Cls:b(1)+b(2)+b(3)+b(4)+b(5)+b(6)+b(7)+b(8)+b(9)+b(10)+b(11)+b(12)+b(13)+b(14)+b(15)=0$
```

```
+----------------------------------------------------+
| Linearly restricted regression                     |
| Ordinary    least squares regression               |
| Model was estimated Aug 27, 2009 at 04:16:47PM     |
| LHS=COST     Mean                 =   13.36561     |
|              Standard deviation   =   1.131971     |
| WTS=none     Number of observs.   =         90     |
| Model size   Parameters           =         18     |
|              Degrees of freedom   =         72     |
| Residuals    Sum of squares       =   1.088193     |
|              Standard error of e  =   .1229382     |
| Fit          R-squared            =   .9904579     |
|              Adjusted R-squared   =   .9882049     |
| Model test   F[ 17,    72] (prob) = 439.62 (.0000) |
| Diagnostic   Log likelihood       =   70.98362     |
|              Restricted(b=0)      =  -138.3581     |
|              Chi-sq [ 17]  (prob) = 418.68 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =  -4.009826     |
|              Akaike Info. Criter. =  -4.015291     |
| Autocorrel   Durbin-Watson Stat.  =   .2363289     |
|              Rho = cor[e,e(-1)]   =   .8818355     |
| Restrictns.  F[  1,    71] (prob) =    .00 (*****) |
| Not using OLS or no constant. Rsqd & F may be < 0. |
| Note, with restrictions imposed,  Rsqd may be < 0. |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 T1      |   -1.17119233       .41783540    -2.803   .0065    .06666667
 T2      |   -1.08895999       .40585988    -2.683   .0091    .06666667
 T3      |   -1.01126486       .40323211    -2.508   .0144    .06666667
 T4      |    -.92623900       .38176914    -2.426   .0178    .06666667
 T5      |    -.46715493       .19075952    -2.449   .0168    .06666667
 T6      |    -.25536788       .09856234    -2.591   .0116    .06666667
 T7      |    -.16363186       .07189683    -2.276   .0259    .06666667
 T8      |    -.01295461       .04862498     -.266   .7907    .06666667
 T9      |     .16259020       .06271009     2.593   .0116    .06666667
 T10     |     .44682406       .17599505     2.539   .0133    .06666667
 T11     |     .79835421       .32940389     2.424   .0179    .06666667
 T12     |     .98436437       .38755999     2.540   .0133    .06666667
 T13     |     .94958221       .36536879     2.599   .0114    .06666667
 T14     |     .88525662       .33549236     2.639   .0102    .06666667
 T15     |     .86979380       .32029396     2.716   .0083    .06666667
 OUTPUT  |     .86772681       .01540818    56.316   .0000  -1.17430918
 FUEL    |    -.48449467       .36410984    -1.331   .1876   12.7703592
 LOAD    |   -1.95441438       .44237791    -4.418   .0000    .56046016
 Constant|    21.6671313      4.62407240     4.686   .0000
```

5.2 Within Time Effect Model

The within effect model for a fixed time effect needs to compute deviations from time means. Keep in mind that the intercept should be suppressed.

5.2.1 Estimating the Fixed Time Effect Model

Let us manually estimate the fixed time effect model first.

```
. quietly egen tm_cost = mean(cost), by(year)
. quietly egen tm_output = mean(output), by(year)
. quietly egen tm_fuel = mean(fuel), by(year)
. quietly egen tm_load = mean(load), by(year)
```

year	tm_cost	tm_output	tm_fuel	tm_load
1	12.36897	-1.790283	11.63606	.4788587
2	12.45963	-1.744389	11.66868	.4868322
3	12.60706	-1.577767	11.67494	.52358
4	12.77912	-1.443695	11.73193	.5244486
5	12.94143	-1.398122	12.26843	.5635266
6	13.0452	-1.393002	12.53826	.5541809
7	13.15965	-1.302416	12.62714	.5607425
8	13.29884	-1.222963	12.76768	.5670587
9	13.4651	-1.067003	12.86104	.6179098
10	13.70187	-.9023156	13.23183	.6233943
11	13.91324	-.9205539	13.66246	.5802577
12	14.05984	-.8641667	13.82315	.5856243
13	14.12841	-.7923916	13.75979	.5803183
14	14.23517	-.6428015	13.67403	.5804528
15	14.32062	-.5527684	13.62997	.5797168

Once time means are ready, transform the dependent and independent variables and then run OLS with the intercept suppressed.

```
. quietly gen tw_cost = cost - tm_cost
. quietly gen tw_output = output - tm_output
. quietly gen tw_fuel = fuel - tm_fuel
. quietly gen tw_load = load - tm_load

. regress tw_cost tw_output tw_fuel tw_load, noc

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  3,    87) = 2015.95
       Model |  75.6459391     3   25.215313           Prob > F      =  0.0000
    Residual |  1.08819023    87  .012507934           R-squared     =  0.9858
-------------+------------------------------           Adj R-squared =  0.9853
       Total |  76.7341294    90  .852601437           Root MSE      =  .11184

------------------------------------------------------------------------------
     tw_cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   tw_output |   .8677268   .0140171    61.90   0.000     .8398663    .8955873
     tw_fuel |  -.4844836   .3312359    -1.46   0.147    -1.142851    .1738836
     tw_load |  -1.954404   .4024388    -4.86   0.000    -2.754295   -1.154514
------------------------------------------------------------------------------
```

If you want to get intercepts of years, use $d_t^* = \bar{y}_{\bullet t} - \beta' \bar{x}_{\bullet t}$. For example, the intercept of year 7 is 21.5035=13.1597-{.8677*(-1.3024) + (-.4845)*12.6271 + (-1.9544)*.5607}. As discussed previously, standard errors of a within effect model need to be adjusted. For instance, the correct standard error of fuel price is computed as .3641= .3312*sqrt(87/72).

```
. sum cost output fuel load if year==7

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        cost |         6    13.15965    1.071738   11.88492   14.52004
      output |         6   -1.302416    1.272691  -2.865108   .2550375
        fuel |         6    12.62714    .0747646   12.48162   12.68725
        load |         6    .5607425     .029541    .510342    .594495
```

5.2.2 Using SAS: PROC TSCSREG and PROC PANEL

You need to sort the data set by variables (i.e., `year` and `airline`), which will appear in the ID statement of PROC TSCSREG and PROC PANEL. The output is very similar to that of LSDV1 in Section 5.1.1.

```
PROC SORT DATA=masil.airline;
   BY year airline;
RUN;

PROC TSCSREG DATA=masil.airline;
   ID year airline;
   MODEL cost = output fuel load /FIXONE;
RUN;

(output is skipped)
```

The F test does not reject the null hypothesis of no fixed time effect (F=1.17, p<.3178); that is, there is no fixed time effect in these panel data.

```
PROC PANEL DATA=masil.airline;
   ID year airline;
   MODEL cost = output fuel load /FIXONE;
RUN;

                                The PANEL Procedure
                              Fixed One Way Estimates

Dependent Variable: cost

                                Model Description

                        Estimation Method            FixOne
                        Number of Cross Sections         15
                        Time Series Length                6


                                 Fit Statistics

                  SSE             1.0882    DFE                72
                  MSE             0.0151    Root MSE       0.1229
                  R-Square        0.9905


                          F Test for No Fixed Effects

                        Num DF      Den DF    F Value    Pr > F
```

```
                                14          72       1.17    0.3178


                                  Parameter Estimates

                                             Standard
    Variable        DF    Estimate           Error    t Value    Pr > |t|    Label

    CS1              1    -2.04096          0.7347      -2.78      0.0070    Cross Sectional
                                                                            Effect    1
    CS2              1    -1.95873          0.7228      -2.71      0.0084    Cross Sectional
                                                                            Effect    2
    CS3              1    -1.88103          0.7204      -2.61      0.0110    Cross Sectional
                                                                            Effect    3
    CS4              1    -1.79601          0.6988      -2.57      0.0122    Cross Sectional
                                                                            Effect    4
    CS5              1    -1.33693          0.5060      -2.64      0.0101    Cross Sectional
                                                                            Effect    5
    CS6              1    -1.12514          0.4086      -2.75      0.0075    Cross Sectional
                                                                            Effect    6
    CS7              1    -1.03341          0.3764      -2.75      0.0076    Cross Sectional
                                                                            Effect    7
    CS8              1    -0.88274          0.3260      -2.71      0.0085    Cross Sectional
                                                                            Effect    8
    CS9              1    -0.70719          0.2947      -2.40      0.0190    Cross Sectional
                                                                            Effect    9
    CS10             1    -0.42296          0.1668      -2.54      0.0134    Cross Sectional
                                                                            Effect   10
    CS11             1    -0.07144          0.0718      -1.00      0.3228    Cross Sectional
                                                                            Effect   11
    CS12             1    0.114571          0.0984       1.16      0.2482    Cross Sectional
                                                                            Effect   12
    CS13             1    0.079789          0.0844       0.95      0.3477    Cross Sectional
                                                                            Effect   13
    CS14             1    0.015463          0.0726       0.21      0.8320    Cross Sectional
                                                                            Effect   14
    Intercept        1    22.53677          4.9405       4.56      <.0001    Intercept
    output           1    0.867727          0.0154      56.32      <.0001
    fuel             1    -0.48448          0.3641      -1.33      0.1875
    load             1     -1.9544          0.4424      -4.42      <.0001
```

5.2.3 Using Stata

In Stata `.xtreg` command, the `fe` option fits the fixed effect model. The following `.iis` command specifies `year` as a panel identification variable. In this case, `i(year)` is redundant.

```
. iis year

. xtreg cost output fuel load, fe i(year)

Fixed-effects (within) regression               Number of obs      =        90
Group variable: year                            Number of groups   =        15

R-sq:  within  = 0.9858                         Obs per group: min =         6
       between = 0.4812                                        avg =       6.0
       overall = 0.5265                                        max =         6

                                                F(3,72)            =   1668.37
corr(u_i, Xb)  = -0.1503                        Prob > F           =    0.0000
```

```
------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |   .8677268   .0154082    56.32   0.000     .8370111    .8984424
        fuel |  -.4844835   .3641085    -1.33   0.188    -1.210321    .2413535
        load |  -1.954404   .4423777    -4.42   0.000    -2.836268    -1.07254
       _cons |   21.66698   4.624053     4.69   0.000      12.4491    30.88486
-------------+----------------------------------------------------------------
     sigma_u |   .8027907
     sigma_e |  .12293801
         rho |  .97708602   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(14, 72) =     1.17              Prob > F = 0.3178
```

Again, the intercept 21.6670 is the intercept of LSDV3 (see 5.1.3).

5.2.4 Using LIMDEP

In LIMDEP, specify a time-series variable for stratification in the `Str=` subcommand. The pooled OLS part of the output is skipped. Do not forget to include `ONE` for the intercept.

```
REGRESS;Lhs=COST;Rhs=ONE,OUTPUT,FUEL,LOAD;Panel;Str=YEAR;Fixed$
```

```
+-----------------------------------------------------+
| Least Squares with Group Dummy Variables            |
| Ordinary    least squares regression                |
| Model was estimated Aug 27, 2009 at 04:19:57PM      |
| LHS=COST     Mean                 =   13.36561      |
|              Standard deviation   =   1.131971      |
| WTS=none     Number of observs.   =         90      |
| Model size   Parameters           =         18      |
|              Degrees of freedom   =         72      |
| Residuals    Sum of squares       =   1.088193      |
|              Standard error of e  =   .1229382      |
| Fit          R-squared            =   .9904579      |
|              Adjusted R-squared   =   .9882049      |
| Model test   F[ 17,    72] (prob) = 439.62 (.0000)  |
| Diagnostic   Log likelihood       =   70.98362      |
|              Restricted(b=0)      =  -138.3581      |
|              Chi-sq [ 17]  (prob) = 418.68 (.0000)  |
| Info criter. LogAmemiya Prd. Crt. =  -4.009826      |
|              Akaike Info. Criter. =  -4.015291      |
| Estd. Autocorrelation of e(i,t)     .881836         |
+-----------------------------------------------------+

+-----------------------------------------------------+
| Panel:Groups   Empty       0,   Valid data      15 |
|                Smallest    6,   Largest          6 |
|                Average group size             6.00 |
+-----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 OUTPUT  |     .86772681       .01540818    56.316   .0000  -1.17430918
 FUEL    |    -.48449467       .36410984    -1.331   .1868   12.7703592
 LOAD    |   -1.95441438       .44237791    -4.418   .0000    .56046016

+---------------------------------------------------------------------+
|             Test Statistics for the Classical Model                 |
+---------------------------------------------------------------------+
|       Model            Log-Likelihood  Sum of Squares    R-squared |
|(1)  Constant term only     -138.35814  .1140409821D+03    .0000000 |
|(2)  Group effects only     -120.52864  .7673414157D+02    .3271354 |
|(3)  X - variables only       61.76991  .1335449522D+01    .9882897 |
|(4)  X and group effects      70.98362  .1088193393D+01    .9904579 |
+---------------------------------------------------------------------+
|                       Hypothesis Tests                              |
```

```
|         Likelihood Ratio Test           F Tests                     |
|         Chi-squared   d.f.  Prob.       F    num. denom.   P value |
|(2) vs (1)    35.659     14  .00117    2.605    14      75   .00404 |
|(3) vs (1)   400.256      3  .00000 2419.329     3      86   .00000 |
|(4) vs (1)   418.684     17  .00000  439.617    17      72   .00000 |
|(4) vs (2)   383.025      3  .00000 1668.364     3      72   .00000 |
|(4) vs (3)    18.427     14  .18800    1.169    14      72   .31776 |
+-------------------------------------------------------------------+
```

You may find F statistic 1.169 at the last line of the output and do not reject the null hypothesis of no fixed time effect.

5.3 Between Time Effect Model

The between effect model regresses time means of dependent variables on those of independent variables. See Sections 3.2 and 4.6.

```
. collapse (mean) tm_cost=cost (mean) tm_output=output (mean) tm_fuel=fuel ///
  (mean) tm_load=load, by(year)

. regress tm_cost tm_output tm_fuel tm_load

      Source |       SS       df       MS              Number of obs =      15
-------------+------------------------------           F(  3,    11) = 4074.33
       Model |  6.21220479     3  2.07073493           Prob > F      =  0.0000
    Residual |  .005590631    11  .000508239           R-squared     =  0.9991
-------------+------------------------------           Adj R-squared =  0.9989
       Total |  6.21779542    14  .444128244           Root MSE      =  .02254

------------------------------------------------------------------------------
     tm_cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   tm_output |   1.133337   .0512898    22.10   0.000     1.020449    1.246225
     tm_fuel |   .3342486   .0228284    14.64   0.000     .2840035    .3844937
     tm_load |  -1.350727   .2478264    -5.45   0.000    -1.896189   -.8052644
       _cons |   11.18505   .3660016    30.56   0.000     10.37949    11.99062
------------------------------------------------------------------------------
```

PROC PANEL has the /BTWNT option to estimate the between effect model.

```
PROC PANEL DATA=masil.airline;
   ID airline year;
   MODEL cost = output fuel load /BTWNT;
RUN;
```

```
                                The PANEL Procedure
                            Between Time Periods Estimates

Dependent Variable: cost

                                  Model Description

                          Estimation Method           BtwTime
                          Number of Cross Sections          6
                          Time Series Length               15


                                  Fit Statistics

                     SSE             0.0056    DFE                 11
```

```
                    MSE             0.0005    Root MSE         0.0225
                    R-Square        0.9991


                                 Parameter Estimates

                                 Standard
    Variable        DF    Estimate       Error    t Value    Pr > |t|    Label

    Intercept        1    11.18504      0.3660      30.56      <.0001    Intercept
    output           1    1.133335      0.0513      22.10      <.0001
    fuel             1    0.334249      0.0228      14.64      <.0001
    load             1    -1.35073      0.2478      -5.45      0.0002
```

Alternatively, use the be option in the Stata `.xtreg` command and the `Means` subcommand in LIMDEP `Regress$` command to get the same result.

```
. xtreg cost output fuel load, be i(year)

Between regression (regression on group means)  Number of obs      =        90
Group variable: year                            Number of groups   =        15

R-sq:  within  = 0.9840                         Obs per group: min =         6
       between = 0.9991                                        avg =       6.0
       overall = 0.9749                                        max =         6

                                                F(3,11)           =   4074.35
sd(u_i + avg(e_i.))=  .0225441                  Prob > F          =    0.0000

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |   1.133335   .0512897    22.10   0.000     1.020447    1.246223
        fuel |   .3342494   .0228284    14.64   0.000     .2840044    .3844943
        load |   -1.35073   .2478257    -5.45   0.000    -1.896191   -.8052695
       _cons |   11.18504   .3660008    30.56   0.000     10.37948     11.9906
------------------------------------------------------------------------------
```

```
REGRESS;Lhs=COST;Rhs=ONE,OUTPUT,FUEL,LOAD;Panel;Str=YEAR;Means$

+----------------------------------------------------+
| Group Means Regression                             |
| Ordinary    least squares regression               |
| Model was estimated Aug 27, 2009 at 04:23:24PM     |
| LHS=YBAR(i.) Mean                =   13.36561      |
|              Standard deviation   =   .6664301      |
| WTS=NTi/Nobs Number of observs.   =         15      |
| Model size   Parameters           =          4      |
|              Degrees of freedom   =         11      |
| Residuals    Sum of squares       =   .5590461E-02  |
|              Standard error of e  =   .2254382E-01  |
| Fit          R-squared            =   .9991009      |
|              Adjusted R-squared   =   .9988557      |
| Model test   F[  3,    11] (prob) =4074.46 (.0000)  |
| Diagnostic   Log likelihood       =   37.92650      |
|              Restricted(b=0)      =  -14.67933      |
|              Chi-sq [  3]  (prob) = 105.21 (.0000)  |
| Info criter. LogAmemiya Prd. Crt. =  -7.348200      |
|              Akaike Info. Criter. =  -7.361410      |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 OUTPUT  |    1.13334032       .05128905    22.097   .0000  .111879D-13
 FUEL    |     .33424795       .02282811    14.642   .0000  .111879D-13
```

```
 LOAD     |   -1.35072980       .24782272    -5.450   .0000  .141312D-06
 Constant|    11.1850651       .36599619    30.561   .0000
```

5.4 Testing Fixed Time Effects.

The null hypothesis of the fixed time effect model is that all time dummy parameters except one are zero: $H_0 : \tau_1 = ... = \tau_{t-1} = 0$. The F statistic is $\frac{(1.3354 - 1.0882)/(15-1)}{(1.0882)/(6*15-15-3)} \sim 1.1683[14,72]$. The small F statistic does not reject the null hypothesis of no fixed time effect (p<.3180).

SAS PROC PANEL, LIMDEP, and Stata `.xtreg` by default conduct the F test. You may conduct the same test using the TEST statement in LSDV1 and the Stata `.test` command.

```
PROC REG DATA=masil.airline;
   MODEL cost = t1-t14 output fuel load;
   TEST t1=t2=t3=t4=t5=t6=t7=t8=t9=t10=t11=t12=t13=t14=0;
RUN;

(output is skipped)

. quietly regress cost t1-t14 output fuel load
. test t1 t2 t3 t4 t5 t6 t7 t8 t9 t10 t11 t12 t13 t14

 ( 1)  t1 = 0
 ( 2)  t2 = 0
 ( 3)  t3 = 0
 ( 4)  t4 = 0
 ( 5)  t5 = 0
 ( 6)  t6 = 0
 ( 7)  t7 = 0
 ( 8)  t8 = 0
 ( 9)  t9 = 0
 (10)  t10 = 0
 (11)  t11 = 0
 (12)  t12 = 0
 (13)  t13 = 0
 (14)  t14 = 0

       F( 14,    72) =    1.17
            Prob > F =    0.3178
```

6. Two-way Fixed Effect Models

A two-way fixed model explores fixed effects of two group variables, two time variables, or one group or one time variables. This chapter investigates fixed group and time effects. This model thus needs two sets of group and time dummy variables (i.e., `airline` and `year`).

6.1 Strategies of the Least Squares Dummy Variable Models

You may combine LSDV1, LSDV2, and LSDV3 to avoid perfect multicollinearity or the dummy variable trap in a two-way fixed effect model. There are five strategies when combining three LSDVs. Since `.cnsreg` does not allow suppressing the intercept, strategy 4 does not work in Stata. The first strategy of dropping two dummies is generally recommended because of its convenience of model estimation and interpretation.

1. Drop one cross-section and one time-series dummy variables.
2. Drop one cross-section dummy and suppress the intercept. Alternatively, drip one time dummy and suppress the intercept
3. Drop one cross-section dummy and impose a restriction on the time-series dummy parameters: $\sum \tau_t = 0$. Alternatively, drop one time-series dummy and impose a restriction on the cross-section dummy parameters: $\sum \mu_i = 0$
4. Suppress the intercept and impose a restriction on the cross-section dummy parameters: $\sum \mu_i = 0$. Alternatively, suppress the intercept and impose a restriction on the time-series dummy parameters: $\sum \tau_t = 0$.
5. Include all dummy variables and impose two restrictions on the cross-section and time-series dummy parameters: $\sum \mu_i = 0$ and $\sum \tau_t = 0$

Each strategy produces different dummy coefficients but returns exactly same parameter estimates of regressors. In general, dummy coefficients are not of primary interest in panel data models.

6.2 LSDV1 without Two Dummies

The first strategy excludes two dummy variables, one dummy from each set of dummy variables. Let us exclude `g6` for the sixth airline and `t15` for the last time period.

```
. regress cost g1-g5 t1-t14 output fuel load

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F( 22,    67) = 1960.82
       Model |  113.864044    22  5.17563838           Prob > F      =  0.0000
    Residual |  .176848775    67  .002639534           R-squared     =  0.9984
-------------+------------------------------           Adj R-squared =  0.9979
       Total |  114.040893    89  1.28135835           Root MSE      =  .05138

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          g1 |   .1742825   .0861201     2.02   0.047     .0023861     .346179
          g2 |   .1114508   .0779551     1.43   0.157    -.0441482    .2670499
```

```
          g3 |   -.143511   .0518934    -2.77   0.007    -.2470907   -.0399313
          g4 |   .1802087   .0321443     5.61   0.000     .1160484    .2443691
          g5 |  -.0466942   .0224688    -2.08   0.042    -.0915422   -.0018463
          t1 |  -.6931382   .3378385    -2.05   0.044    -1.367467   -.0188098
          t2 |  -.6384366   .3320802    -1.92   0.059    -1.301271    .0243983
          t3 |  -.5958031   .3294473    -1.81   0.075    -1.253383    .0617764
          t4 |  -.5421537   .3189139    -1.70   0.094    -1.178708    .0944011
          t5 |  -.4730429   .2319459    -2.04   0.045    -.9360088   -.0100769
          t6 |  -.4272042     .18844    -2.27   0.027    -.8033319   -.0510764
          t7 |  -.3959783   .1732969    -2.28   0.025    -.7418804   -.0500762
          t8 |  -.3398463   .1501062    -2.26   0.027    -.6394596    -.040233
          t9 |  -.2718933   .1348175    -2.02   0.048    -.5409901   -.0027964
         t10 |  -.2273857   .0763495    -2.98   0.004      -.37978   -.0749914
         t11 |  -.1118032   .0319005    -3.50   0.001     -.175477   -.0481295
         t12 |   -.033641   .0429008    -0.78   0.436    -.1192713    .0519893
         t13 |  -.0177346   .0362554    -0.49   0.626    -.0901007    .0546315
         t14 |  -.0186451    .030508    -0.61   0.543    -.0795393     .042249
      output |   .8172487    .031851    25.66   0.000     .7536739    .8808235
        fuel |      .16861    .163478     1.03   0.306    -.1576935    .4949135
        load |  -.8828142   .2617373    -3.37   0.001    -1.405244   -.3603843
       _cons |   12.94004   2.218231     5.83   0.000     8.512434    17.36765
------------------------------------------------------------------------------
```

In SAS, run the following script to get the same result.

```
PROC REG DATA=masil.airline;
   MODEL cost = g1-g5 t1-t14 output fuel load;
RUN;
```

The REG Procedure
Model: MODEL1
Dependent Variable: cost

Number of Observations Read 90
Number of Observations Used 90

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	22	113.86404	5.17564	1960.82	<.0001
Error	67	0.17685	0.00264		
Corrected Total	89	114.04089			

Root MSE	0.05138	R-Square	0.9984
Dependent Mean	13.36561	Adj R-Sq	0.9979
Coeff Var	0.38439		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	12.94004	2.21823	5.83	<.0001
g1	1	0.17428	0.08612	2.02	0.0470
g2	1	0.11145	0.07796	1.43	0.1575
g3	1	-0.14351	0.05189	-2.77	0.0073

```
g4            1        0.18021        0.03214       5.61      <.0001
g5            1       -0.04669        0.02247      -2.08      0.0415
t1            1       -0.69314        0.33784      -2.05      0.0441
t2            1       -0.63844        0.33208      -1.92      0.0588
t3            1       -0.59580        0.32945      -1.81      0.0750
t4            1       -0.54215        0.31891      -1.70      0.0938
t5            1       -0.47304        0.23195      -2.04      0.0454
t6            1       -0.42720        0.18844      -2.27      0.0266
t7            1       -0.39598        0.17330      -2.28      0.0255
t8            1       -0.33985        0.15011      -2.26      0.0268
t9            1       -0.27189        0.13482      -2.02      0.0477
t10           1       -0.22739        0.07635      -2.98      0.0040
t11           1       -0.11180        0.03190      -3.50      0.0008
t12           1       -0.03364        0.04290      -0.78      0.4357
t13           1       -0.01773        0.03626      -0.49      0.6263
t14           1       -0.01865        0.03051      -0.61      0.5432
output        1        0.81725        0.03185      25.66      <.0001
fuel          1        0.16861        0.16348       1.03      0.3061
load          1       -0.88281        0.26174      -3.37      0.0012
```

In LIMDEP, the following command fits the same model (output is skipped).

```
REGRESS;Lhs=COST;
       Rhs=ONE,G1,G2,G3,G4,G5,T1,T2,T3,T4,T5,T6,T7,T8,T9,T10,T11,T12,T13,T14,OUTPUT,FUEL,LOAD$
```

6.3 LSDV1 + LSDV2: Drop a Dummy and Suppress the Intercept

The second strategy combines LSDV1 and LSDV2 to drop a dummy and suppress the intercept. Let us drop a dummy `g6` and suppress the intercept. Keep in mind that SSE is still correct but F and R^2 are not.

```
. regress cost g1-g5 t1-t15 output fuel load, noc

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F( 23,    67) =       .
       Model |  16191.4201    23  703.974786           Prob > F      =  0.0000
    Residual |  .176848775    67  .002639534           R-squared     =  1.0000
-------------+------------------------------           Adj R-squared =  1.0000
       Total |  16191.5969    90  179.906633           Root MSE      =  .05138

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          g1 |   .1742825   .0861201     2.02   0.047     .0023861     .346179
          g2 |   .1114508   .0779551     1.43   0.157    -.0441482    .2670499
          g3 |   -.143511   .0518934    -2.77   0.007    -.2470907   -.0399313
          g4 |   .1802087   .0321443     5.61   0.000     .1160484    .2443691
          g5 |  -.0466942   .0224688    -2.08   0.042    -.0915422   -.0018463
          t1 |    12.2469   1.885399     6.50   0.000      8.48363    16.01018
          t2 |    12.3016   1.891045     6.51   0.000     8.527062    16.07615
          t3 |   12.34424    1.89341     6.52   0.000     8.564976     16.1235
          t4 |   12.39789   1.903395     6.51   0.000     8.598694    16.19708
          t5 |     12.467   1.991503     6.26   0.000     8.491942    16.44206
          t6 |   12.51284   2.035334     6.15   0.000     8.450294    16.57538
          t7 |   12.54406    2.05038     6.12   0.000     8.451487    16.63664
          t8 |   12.60019   2.073782     6.08   0.000     8.460909    16.73948
          t9 |   12.66815   2.090527     6.06   0.000     8.495438    16.84086
         t10 |   12.71266   2.151893     5.91   0.000     8.417458    17.00785
         t11 |   12.82824   2.221401     5.77   0.000     8.394303    17.26217
         t12 |    12.9064   2.247972     5.74   0.000      8.41943    17.39337
         t13 |   12.92231   2.237999     5.77   0.000     8.455241    17.38937
         t14 |    12.9214   2.224893     5.81   0.000     8.480492     17.3623
```

```
         t15 |   12.94004   2.218231     5.83   0.000     8.512434    17.36765
      output |   .8172487    .031851    25.66   0.000     .7536739    .8808235
        fuel |     .16861    .163478     1.03   0.306    -.1576935    .4949135
        load |  -.8828142   .2617373    -3.37   0.001    -1.405244   -.3603843
------------------------------------------------------------------------------
```

Alternatively, you may drop one of time dummies and suppress the intercept. The dummy coefficients are different from those above but parameter estimates of regressors remained unchanged.

```
. regress cost g1-g6 t1-t14 output fuel load, noc

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F( 23,    67) =       .
       Model |  16191.4201    23  703.974786           Prob > F      =  0.0000
    Residual |  .176848775    67  .002639534           R-squared     =  1.0000
-------------+------------------------------           Adj R-squared =  1.0000
       Total |  16191.5969    90  179.906633           Root MSE      =  .05138

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          g1 |   13.11432   2.229552     5.88   0.000      8.66412    17.56453
          g2 |   13.05149   2.229864     5.85   0.000     8.600665    17.50232
          g3 |   12.79653   2.230546     5.74   0.000     8.344341    17.24872
          g4 |   13.12025   2.223638     5.90   0.000      8.68185    17.55865
          g5 |   12.89335   2.222204     5.80   0.000      8.45781    17.32888
          g6 |   12.94004   2.218231     5.83   0.000     8.512434    17.36765
          t1 |  -.6931382   .3378385    -2.05   0.044    -1.367467   -.0188098
          t2 |  -.6384366   .3320802    -1.92   0.059    -1.301271    .0243983
          t3 |  -.5958031   .3294473    -1.81   0.075    -1.253383    .0617764
          t4 |  -.5421537   .3189139    -1.70   0.094    -1.178708    .0944011
          t5 |  -.4730429   .2319459    -2.04   0.045    -.9360088   -.0100769
          t6 |  -.4272042     .18844    -2.27   0.027    -.8033319   -.0510764
          t7 |  -.3959783   .1732969    -2.28   0.025    -.7418804   -.0500762
          t8 |  -.3398463   .1501062    -2.26   0.027    -.6394596    -.040233
          t9 |  -.2718933   .1348175    -2.02   0.048    -.5409901   -.0027964
         t10 |  -.2273857   .0763495    -2.98   0.004      -.37978   -.0749914
         t11 |  -.1118032   .0319005    -3.50   0.001     -.175477   -.0481295
         t12 |   -.033641   .0429008    -0.78   0.436    -.1192713    .0519893
         t13 |  -.0177346   .0362554    -0.49   0.626    -.0901007    .0546315
         t14 |  -.0186451    .030508    -0.61   0.543    -.0795393     .042249
      output |   .8172487    .031851    25.66   0.000     .7536739    .8808235
        fuel |     .16861    .163478     1.03   0.306    -.1576935    .4949135
        load |  -.8828142   .2617373    -3.37   0.001    -1.405244   -.3603843
------------------------------------------------------------------------------
```

In SAS, execute the following script that has /NOINT to suppress the intercept.

```
PROC REG DATA=masil.airline;
   MODEL cost = g1-g5 t1-t15 output fuel load /NOINT;
   MODEL cost = g1-g6 t1-t14 output fuel load /NOINT;
RUN;

(output is skippted)
```

In LIMDEP, `ONE` should be taken out to suppress the intercept.

```
REGRESS;Lhs=COST;
        Rhs=G1,G2,G3,G4,G5,T1,T2,T3,T4,T5,T6,T7,T8,T9,T10,T11,T12,T13,T14,T15, OUTPUT,FUEL,LOAD$

(output is skippted)

REGRESS;Lhs=COST;
```

```
Rhs=G1,G2,G3,G4,G5,G6,T1,T2,T3,T4,T5,T6,T7,T8,T9,T10,T11,T12,T13,T14,OUTPUT,FUEL,LOAD$
```

```
+----------------------------------------------------+
| Ordinary    least squares regression               |
| Model was estimated Aug 30, 2009 at 03:58:13PM     |
| LHS=COST     Mean                 =   13.36561     |
|              Standard deviation   =   1.131971     |
| WTS=none     Number of observs.   =         90     |
| Model size   Parameters           =         23     |
|              Degrees of freedom   =         67     |
| Residuals    Sum of squares       =   .1768479     |
|              Standard error of e  =   .5137627E-01 |
| Fit          R-squared            =   .9984493     |
|              Adjusted R-squared   =   .9979401     |
| Model test   F[ 22,    67] (prob) =1960.83 (.0000) |
| Diagnostic   Log likelihood       =   152.7479     |
|              Restricted(b=0)      =  -138.3581     |
|              Chi-sq [ 22]  (prob) = 582.21 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =  -5.709580     |
|              Akaike Info. Criter. =  -5.721164     |
| Autocorrel   Durbin-Watson Stat.  =   .6035047     |
|              Rho = cor[e,e(-1)]   =   .6982476     |
| Not using OLS or no constant. Rsqd & F may be < 0. |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 G1      |    13.1139819      2.22955625     5.882   .0000    .16666667
 G2      |    13.0511515      2.22986828     5.853   .0000    .16666667
 G3      |    12.7961914      2.23055043     5.737   .0000    .16666667
 G4      |    13.1199153      2.22364115     5.900   .0000    .16666667
 G5      |    12.8930131      2.22220692     5.802   .0000    .16666667
 G6      |    12.9397087      2.21823375     5.833   .0000    .16666667
 T1      |    -.69308729       .33783938    -2.052   .0441    .06666667
 T2      |    -.63838795       .33208126    -1.922   .0588    .06666667
 T3      |    -.59575348       .32944797    -1.808   .0750    .06666667
 T4      |    -.54210773       .31891465    -1.700   .0938    .06666667
 T5      |    -.47300784       .23194606    -2.039   .0454    .06666667
 T6      |    -.42717813       .18844068    -2.267   .0266    .06666667
 T7      |    -.39595152       .17329717    -2.285   .0255    .06666667
 T8      |    -.33982426       .15010661    -2.264   .0268    .06666667
 T9      |    -.27187359       .13481769    -2.017   .0477    .06666667
 T10     |    -.22737840       .07634935    -2.978   .0040    .06666667
 T11     |    -.11180525       .03190046    -3.505   .0008    .06666667
 T12     |    -.03364915       .04290088     -.784   .4356    .06666667
 T13     |    -.01774030       .03625541     -.489   .6262    .06666667
 T14     |    -.01864714       .03050793     -.611   .5431    .06666667
 OUTPUT  |     .81725242       .03185102    25.659   .0000   -1.17430918
 FUEL    |     .16863516       .16347826     1.032   .3060    12.7703592
 LOAD    |    -.88281516       .26173663    -3.373   .0012    .56046016
```

Notice that LIMDEP reports correct F (1960.83), and R^2 (.9984).

6.4 LSDV1 + LSDV3: Drop a Dummy and Impose a Restriction

The third strategy excludes one dummy from a set of dummy variables and imposes a restriction on another set of dummy parameters. Let us drop a time dummy here and then impose a restriction on group dummy parameters.

```
PROC REG DATA=masil.airline;
   MODEL cost = g1-g6 t1-t14 output fuel load;
   RESTRICT g1 + g2 + g3 + g4 + g5 + g6 = 0;
RUN;
```

```
                                The REG Procedure
```

Model: MODEL1
Dependent Variable: cost

NOTE: Restrictions have been applied to parameter estimates.

Number of Observations Read	90
Number of Observations Used	90

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	22	113.86404	5.17564	1960.82	<.0001
Error	67	0.17685	0.00264		
Corrected Total	89	114.04089			

Root MSE	0.05138	R-Square	0.9984
Dependent Mean	13.36561	Adj R-Sq	0.9979
Coeff Var	0.38439		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	12.98600	2.22540	5.84	<.0001
g1	1	0.12833	0.04601	2.79	0.0069
g2	1	0.06549	0.03897	1.68	0.0975
g3	1	-0.18947	0.01561	-12.14	<.0001
g4	1	0.13425	0.01832	7.33	<.0001
g5	1	-0.09265	0.03731	-2.48	0.0155
g6	1	-0.04596	0.04161	-1.10	0.2733
t1	1	-0.69314	0.33784	-2.05	0.0441
t2	1	-0.63844	0.33208	-1.92	0.0588
t3	1	-0.59580	0.32945	-1.81	0.0750
t4	1	-0.54215	0.31891	-1.70	0.0938
t5	1	-0.47304	0.23195	-2.04	0.0454
t6	1	-0.42720	0.18844	-2.27	0.0266
t7	1	-0.39598	0.17330	-2.28	0.0255
t8	1	-0.33985	0.15011	-2.26	0.0268
t9	1	-0.27189	0.13482	-2.02	0.0477
t10	1	-0.22739	0.07635	-2.98	0.0040
t11	1	-0.11180	0.03190	-3.50	0.0008
t12	1	-0.03364	0.04290	-0.78	0.4357
t13	1	-0.01773	0.03626	-0.49	0.6263
t14	1	-0.01865	0.03051	-0.61	0.5432
output	1	0.81725	0.03185	25.66	<.0001
fuel	1	0.16861	0.16348	1.03	0.3061
load	1	-0.88281	0.26174	-3.37	0.0012
RESTRICT	-1	-1.9387E-16	.	.	.

* Probability computed using beta distribution.

In Stata, you need to run the `.cnsreg` command with a constraint on the group dummy parameters. `.cnsreg` with the `.constraint(1)` option fits OLS under constraint 1 defined in `.constraint`.

```
. constraint define 1 g1 + g2 + g3 + g4 + g5 + g6 = 0
. cnsreg cost g1-g6 t1-t14 output fuel load, constraint(1)

Constrained linear regression                   Number of obs   =         90
                                                F(  22,     67) =    1960.82
                                                Prob > F        =     0.0000
                                                Root MSE        =     0.0514

 ( 1)  g1 + g2 + g3 + g4 + g5 + g6 = 0
------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          g1 |   .1283264   .0460126     2.79   0.007     .0364849    .2201679
          g2 |   .0654947   .0389685     1.68   0.097    -.0122867    .1432761
          g3 |  -.1894671   .0156096   -12.14   0.000     -.220624   -.1583102
          g4 |   .1342526   .0183163     7.33   0.000      .097693    .1708121
          g5 |  -.0926504   .0373085    -2.48   0.016    -.1671184   -.0181824
          g6 |  -.0459561   .0416069    -1.10   0.273    -.1290038    .0370916
          t1 |  -.6931382   .3378385    -2.05   0.044    -1.367467   -.0188098
          t2 |  -.6384366   .3320802    -1.92   0.059    -1.301271    .0243983
          t3 |  -.5958031   .3294473    -1.81   0.075    -1.253383    .0617764
          t4 |  -.5421537   .3189139    -1.70   0.094    -1.178708    .0944011
          t5 |  -.4730429   .2319459    -2.04   0.045    -.9360088   -.0100769
          t6 |  -.4272042     .18844    -2.27   0.027    -.8033319   -.0510764
          t7 |  -.3959783   .1732969    -2.28   0.025    -.7418804   -.0500762
          t8 |  -.3398463   .1501062    -2.26   0.027    -.6394596    -.040233
          t9 |  -.2718933   .1348175    -2.02   0.048    -.5409901   -.0027964
         t10 |  -.2273857   .0763495    -2.98   0.004      -.37978   -.0749914
         t11 |  -.1118032   .0319005    -3.50   0.001     -.175477   -.0481295
         t12 |   -.033641   .0429008    -0.78   0.436    -.1192713    .0519893
         t13 |  -.0177346   .0362554    -0.49   0.626    -.0901007    .0546315
         t14 |  -.0186451    .030508    -0.61   0.543    -.0795393     .042249
      output |   .8172487    .031851    25.66   0.000     .7536739    .8808235
        fuel |      .16861    .163478     1.03   0.306    -.1576935    .4949135
        load |  -.8828142   .2617373    -3.37   0.001    -1.405244   -.3603843
       _cons |     12.986   2.225402     5.84   0.000     8.544076    17.42792
------------------------------------------------------------------------------
```

In LIMDEP, run a `Regress$` command with the `Cls:` subcommand. `b(2)` in the subcommand indicates the second parameter estimate listed in the `Rhs=` subcommand. Therefore, LIMDEP fits the LSDV1 under the constraint that the sum of all group dummy parameters, b(2) for `g1` through b(7) for `g6`, is zero.

```
REGRESS;Lhs=COST;
        Rhs=ONE,G1,G2,G3,G4,G5,G6,T1,T2,T3,T4,T5,T6,T7,T8,T9,T10,T11,T12,T13,T14,OUTPUT,FUEL,LOAD;
        Cls:b(2)+b(3)+b(4)+b(5)+b(6)+b(7)=0$

+-----------------------------------------------------+
| Linearly restricted regression                      |
| Ordinary    least squares regression                |
| Model was estimated Aug 30, 2009 at 04:24:35PM      |
| LHS=COST     Mean                 =   13.36561      |
|              Standard deviation   =   1.131971      |
| WTS=none     Number of observs.   =         90      |
| Model size   Parameters           =         23      |
|              Degrees of freedom   =         67      |
| Residuals    Sum of squares       =   .1768479      |
|              Standard error of e  =   .5137627E-01 |
| Fit          R-squared            =   .9984493      |
|              Adjusted R-squared   =   .9979401      |
| Model test   F[ 22,    67] (prob) =1960.83 (.0000) |
```

```
| Diagnostic   Log likelihood       =   152.7479     |
|              Restricted(b=0)      =  -138.3581     |
|              Chi-sq [ 22]  (prob) = 582.21 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =  -5.709580     |
|              Akaike Info. Criter. =  -5.721164     |
| Autocorrel   Durbin-Watson Stat.  =   .6035047     |
|              Rho = cor[e,e(-1)]   =   .6982476     |
| Restrictns.  F[  1,    66] (prob) =    .00 (*****) |
| Not using OLS or no constant. Rsqd & F may be < 0. |
| Note, with restrictions imposed,  Rsqd may be < 0. |
+----------------------------------------------------+
+--------+-------------+---------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+-------------+---------------+--------+--------+----------+
 Constant|    12.9856603      2.22540616     5.835   .0000
 G1      |     .12832155       .04601257     2.789   .0069    .16666667
 G2      |     .06549116       .03896849     1.681   .0976    .16666667
 G3      |    -.18946893       .01560965   -12.138   .0000    .16666667
 G4      |     .13425504       .01831636     7.330   .0000    .16666667
 G5      |    -.09264719       .03730846    -2.483   .0156    .16666667
 G6      |    -.04595164       .04160692    -1.104   .2734    .16666667
 T1      |    -.69308729       .33783938    -2.052   .0442    .06666667
 T2      |    -.63838795       .33208126    -1.922   .0589    .06666667
 T3      |    -.59575348       .32944797    -1.808   .0751    .06666667
 T4      |    -.54210773       .31891465    -1.700   .0939    .06666667
 T5      |    -.47300784       .23194606    -2.039   .0454    .06666667
 T6      |    -.42717813       .18844068    -2.267   .0267    .06666667
 T7      |    -.39595152       .17329717    -2.285   .0255    .06666667
 T8      |    -.33982426       .15010661    -2.264   .0269    .06666667
 T9      |    -.27187359       .13481769    -2.017   .0478    .06666667
 T10     |    -.22737840       .07634935    -2.978   .0041    .06666667
 T11     |    -.11180525       .03190046    -3.505   .0008    .06666667
 T12     |    -.03364915       .04290088     -.784   .4356    .06666667
 T13     |    -.01774030       .03625541     -.489   .6262    .06666667
 T14     |    -.01864714       .03050793     -.611   .5432    .06666667
 OUTPUT  |     .81725242       .03185102    25.659   .0000  -1.17430918
 FUEL    |     .16863516       .16347826     1.032   .3061   12.7703592
 LOAD    |    -.88281516       .26173663    -3.373   .0012    .56046016
```

Alternatively, you may drop one group dummy and imposes a restriction on time dummy variables. In LIMDEP, `b(7)` indicates the seventh parameter estimate for `t1`. The output is skipped.

```
PROC REG DATA=masil.airline;
   MODEL cost = g1-g5 t1-t15 output fuel load;
   RESTRICT t1+t2+t3+t4+t5+t6+t7+t8+t9+t10+t11+t12+t13+t14+t15=0;
RUN;
```

```
. constraint define 3 t1+t2+t3+t4+t5+t6+t7+t8+t9+t10+t11+t12+t13+t14+t15=0
. cnsreg cost g1-g5 t1-t15 output fuel load, constraint(3)
```

```
REGRESS;Lhs=COST;
        Rhs=ONE,G1,G2,G3,G4,G5,T1,T2,T3,T4,T5,T6,T7,T8,T9,T10,T11,T12,T13,T14,T15,OUTPUT,FUEL,LOAD;
        Cls:b(7)+b(8)+b(9)+b(10)+b(11)+b(12)+b(13)+b(14)+b(15)+b(16)+b(17)+b(18)+b(19)+b(20)+b(21)=0$
```

6.5 LSDV2 + LSDV3: Suppress the Intercept and Impose a Restriction

The strategy of LSDV2 + LSDV3 includes all two sets of dummy variables and instead suppresses the intercept and imposes a restriction. Stata does not support this approach. The following procedure has a constraint on the group variable. Since the intercept is suppressed, F (703.9748) and R^2 are incorrect.

```
PROC REG DATA=masil.airline;
   MODEL cost = g1-g6 t1-t15 output fuel load /NOINT;
```

```
   RESTRICT g1 + g2 + g3 + g4 + g5 + g6 = 0;
RUN;
```

The REG Procedure
Model: MODEL1
Dependent Variable: cost

NOTE: Restrictions have been applied to parameter estimates.

Number of Observations Read	90
Number of Observations Used	90

NOTE: No intercept in model. R-Square is redefined.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	23	16191	703.97479	266704	<.0001
Error	67	0.17685	0.00264		
Uncorrected Total	90	16192			

Root MSE	0.05138	R-Square	1.0000
Dependent Mean	13.36561	Adj R-Sq	1.0000
Coeff Var	0.38439		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
g1	1	0.12833	0.04601	2.79	0.0069
g2	1	0.06549	0.03897	1.68	0.0975
g3	1	-0.18947	0.01561	-12.14	<.0001
g4	1	0.13425	0.01832	7.33	<.0001
g5	1	-0.09265	0.03731	-2.48	0.0155
g6	1	-0.04596	0.04161	-1.10	0.2733
t1	1	12.29286	1.89169	6.50	<.0001
t2	1	12.34756	1.89736	6.51	<.0001
t3	1	12.39019	1.89982	6.52	<.0001
t4	1	12.44384	1.90989	6.52	<.0001
t5	1	12.51295	1.99808	6.26	<.0001
t6	1	12.55879	2.04195	6.15	<.0001
t7	1	12.59002	2.05706	6.12	<.0001
t8	1	12.64615	2.08052	6.08	<.0001
t9	1	12.71410	2.09734	6.06	<.0001
t10	1	12.75861	2.15883	5.91	<.0001
t11	1	12.87419	2.22838	5.78	<.0001
t12	1	12.95236	2.25499	5.74	<.0001
t13	1	12.96826	2.24505	5.78	<.0001
t14	1	12.96735	2.23202	5.81	<.0001
t15	1	12.98600	2.22540	5.84	<.0001

```
output        1        0.81725        0.03185      25.66     <.0001
fuel          1        0.16861        0.16348       1.03     0.3061
load          1       -0.88281        0.26174      -3.37     0.0012
RESTRICT     -1    5.89339E-14    1.250165E-9       0.00     1.0000*

                * Probability computed using beta distribution.
```

You may impose an alternative restriction on the time variable to obtain the equivalent result despite different dummy coefficients. The output is skipped.

```
PROC REG DATA=masil.airline;
   MODEL cost = g1-g6 t1-t15 output fuel load /NOINT;
   RESTRICT t1 + t2 + t3 + t4 + t5 + t6 + t7 + t8 + t9 + t10 + t11 + t12 + t13 + t14 + t15 = 0;
RUN;
```

In LIMDEP, following commands are supposed to work, but they return different parameter estimates and goodness-of-fit measures probably due to its estimation method.

```
REGRESS;Lhs=COST;
        Rhs=G1,G2,G3,G4,G5,G6,T1,T2,T3,T4,T5,T6,T7,T8,T9,T10,T11,T12,T13,T14,T15,OUTPUT,FUEL,LOAD;
        Cls:b(1)+b(2)+b(3)+b(4)+b(5)+b(6)=0$

(output is skipped)

REGRESS;Lhs=COST;
        Rhs=G1,G2,G3,G4,G5,G6,T1,T2,T3,T4,T5,T6,T7,T8,T9,T10,T11,T12,T13,T14,T15,OUTPUT,FUEL,LOAD;
        Cls:b(7)+b(8)+b(9)+b(10)+b(11)+b(12)+b(13)+b(14)+b(15)+b(16)+b(17)+b(18)+b(19)+b(20)+b(21)=0$
```

```
+----------------------------------------------------+
| Linearly restricted regression                     |
| Ordinary    least squares regression               |
| Model was estimated Aug 30, 2009 at 04:47:10PM     |
| LHS=COST     Mean                 =   13.36561     |
|              Standard deviation   =   1.131971     |
| WTS=none     Number of observs.   =         90     |
| Model size   Parameters           =         23     |
|              Degrees of freedom   =         67     |
| Residuals    Sum of squares       =   .1790783     |
|              Standard error of e  =   .5169924E-01 |
| Fit          R-squared            =   .9984297     |
|              Adjusted R-squared   =   .9979141     |
| Model test   F[ 22,    67] (prob) =1936.37 (.0000) |
| Diagnostic   Log likelihood       =   152.1839     |
|              Restricted(b=0)      =  -138.3581     |
|              Chi-sq [ 22]  (prob) = 581.08 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =  -5.697046     |
|              Akaike Info. Criter. =  -5.708630     |
| Autocorrel   Durbin-Watson Stat.  =   .6164424     |
|              Rho = cor[e,e(-1)]   =   .6917788     |
| Restrictns.  F[  1,    66] (prob) =    .68 (.4113) |
| Not using OLS or no constant. Rsqd & F may be < 0. |
| Note, with restrictions imposed,  Rsqd may be < 0. |
+----------------------------------------------------+
+--------+-------------+---------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+-------------+---------------+--------+--------+----------+
 G1      |    13.0058594    .....(Fixed Parameter).......
 G2      |    12.9453125      216842.319      .000  1.0000    .16666667
 G3      |    12.6894531      216842.319      .000  1.0000    .16666667
 G4      |    13.0117188      216842.319      .000  1.0000    .16666667
 G5      |    12.7812500    .....(Fixed Parameter).......
 G6      |    12.8261719    .....(Fixed Parameter).......
 T1      |    -.39453125      306661.348      .000  1.0000    .06666667
 T2      |    -.33203125      433684.637      .000  1.0000    .06666667
 T3      |    -.29101563      216842.319      .000  1.0000    .06666667
```

```
T4       |    -.24414063      306661.348      .000  1.0000    .06666667
T5       |    -.16406250    ......(Fixed Parameter).......
T6       |    -.10742188    ......(Fixed Parameter).......
T7       |    -.07421875    ......(Fixed Parameter).......
T8       |    -.02148438    ......(Fixed Parameter).......
T9       |     .05859375      216842.319      .000  1.0000    .06666667
T10      |     .10351563      216842.319      .000  1.0000    .06666667
T11      |     .22070313      216842.319      .000  1.0000    .06666667
T12      |     .30468750      216842.319      .000  1.0000    .06666667
T13      |     .31250000      216842.319      .000  1.0000    .06666667
T14      |     .31835938      216842.319      .000  1.0000    .06666667
T15      |     .33203125    ......(Fixed Parameter).......
OUTPUT   |     .81399272       .03205125    25.397   .0000  -1.17430918
FUEL     |     .15204518       .16450594      .924   .3587   12.7703592
LOAD     |    -.88619366       .26338199    -3.365   .0013    .56046016
```

6.6 LSDV3 with Two Restrictions

The last strategy includes all group and time dummies and then imposes two restrictions on group and time dummy parameters. Pay attention to the two RESTRICT statements in the following PROC REG.

```
PROC REG DATA=masil.airline;
   MODEL cost = g1-g6 t1-t15 output fuel load;
   RESTRICT g1 + g2 + g3 + g4 + g5 + g6 = 0;
   RESTRICT t1 + t2 + t3 + t4 + t5 + t6 + t7 + t8 + t9 + t10 + t11 + t12 + t13 + t14 + t15 = 0;
RUN;
```

```
                                The REG Procedure
                                  Model: MODEL1
                              Dependent Variable: cost

NOTE: Restrictions have been applied to parameter estimates.


                        Number of Observations Read          90
                        Number of Observations Used          90


                                 Analysis of Variance

                                            Sum of           Mean
        Source                   DF        Squares         Square    F Value    Pr > F

        Model                    22      113.86404        5.17564    1960.82    <.0001
        Error                    67        0.17685        0.00264
        Corrected Total          89      114.04089


                  Root MSE              0.05138    R-Square     0.9984
                  Dependent Mean       13.36561    Adj R-Sq     0.9979
                  Coeff Var             0.38439


                                 Parameter Estimates

                              Parameter       Standard
            Variable     DF    Estimate          Error    t Value    Pr > |t|

            Intercept     1    12.66688        2.08107       6.09     <.0001
            g1            1     0.12833        0.04601       2.79     0.0069
```

```
g2            1        0.06549        0.03897       1.68     0.0975
g3            1       -0.18947        0.01561     -12.14     <.0001
g4            1        0.13425        0.01832       7.33     <.0001
g5            1       -0.09265        0.03731      -2.48     0.0155
g6            1       -0.04596        0.04161      -1.10     0.2733
t1            1       -0.37402        0.19187      -1.95     0.0554
t2            1       -0.31932        0.18609      -1.72     0.0908
t3            1       -0.27669        0.18335      -1.51     0.1360
t4            1       -0.22304        0.17297      -1.29     0.2017
t5            1       -0.15393        0.08644      -1.78     0.0795
t6            1       -0.10809        0.04486      -2.41     0.0187
t7            1       -0.07686        0.03193      -2.41     0.0188
t8            1       -0.02073        0.02045      -1.01     0.3143
t9            1        0.04722        0.02908       1.62     0.1091
t10           1        0.09173        0.08115       1.13     0.2624
t11           1        0.20731        0.14914       1.39     0.1691
t12           1        0.28547        0.17564       1.63     0.1088
t13           1        0.30138        0.16603       1.82     0.0740
t14           1        0.30047        0.15362       1.96     0.0546
t15           1        0.31911        0.14749       2.16     0.0341
output        1        0.81725        0.03185      25.66     <.0001
fuel          1        0.16861        0.16348       1.03     0.3061
load          1       -0.88281        0.26174      -3.37     0.0012
RESTRICT     -1    -2.5962E-16    4.04547E-11      -0.00     1.0000*
RESTRICT     -1    -2.3598E-16              .          .          .

                  * Probability computed using beta distribution.
```

In Stata, execute the following command to get the same result. Notice that constraints 1 and 3 were defined above.

```
. cnsreg cost g1-g6 t1-t15 output fuel load, constraint(1 3)

Constrained linear regression                   Number of obs   =         90
                                                F(  22,     67) =    1960.82
                                                Prob > F        =     0.0000
                                                Root MSE        =     0.0514

 ( 1)  g1 + g2 + g3 + g4 + g5 + g6 = 0
 ( 2)  t1 + t2 + t3 + t4 + t5 + t6 + t7 + t8 + t9 + t10 + t11 + t12 + t13 + t14 + t15 = 0
------------------------------------------------------------------------------
       cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------+-----------------------------------------------------------------
         g1 |   .1283264   .0460126     2.79   0.007     .0364849    .2201679
         g2 |   .0654947   .0389685     1.68   0.097    -.0122867    .1432761
         g3 |  -.1894671   .0156096   -12.14   0.000     -.220624   -.1583102
         g4 |   .1342526   .0183163     7.33   0.000      .097693    .1708121
         g5 |  -.0926504   .0373085    -2.48   0.016    -.1671184   -.0181824
         g6 |  -.0459561   .0416069    -1.10   0.273    -.1290038    .0370916
         t1 |  -.3740245    .191872    -1.95   0.055    -.7570026    .0089536
         t2 |  -.3193228   .1860877    -1.72   0.091    -.6907554    .0521097
         t3 |  -.2766893   .1833501    -1.51   0.136    -.6426576    .0892789
         t4 |  -.2230399   .1729671    -1.29   0.202    -.5682837    .1222038
         t5 |  -.1539291   .0864404    -1.78   0.079    -.3264649    .0186066
         t6 |  -.1080904   .0448591    -2.41   0.019    -.1976296   -.0185513
         t7 |  -.0768646   .0319336    -2.41   0.019    -.1406043   -.0131248
         t8 |  -.0207326   .0204506    -1.01   0.314     -.061552    .0200869
         t9 |   .0472205   .0290822     1.62   0.109    -.0108278    .1052688
        t10 |   .0917281   .0811525     1.13   0.262    -.0702531    .2537092
        t11 |   .2073105   .1491443     1.39   0.169    -.0903829    .5050039
        t12 |   .2854727   .1756365     1.63   0.109    -.0650993    .6360447
        t13 |   .3013791   .1660294     1.82   0.074     -.030017    .6327752
        t14 |   .3004686   .1536212     1.96   0.055    -.0061606    .6070978
```

```
         t15 |   .3191137   .1474883     2.16   0.034     .0247259    .6135015
      output |   .8172487    .031851    25.66   0.000     .7536739    .8808235
        fuel |     .16861    .163478     1.03   0.306    -.1576935    .4949135
        load |  -.8828142   .2617373    -3.37   0.001    -1.405244   -.3603843
       _cons |   12.66688   2.081068     6.09   0.000     8.513054    16.82071
------------------------------------------------------------------------------
```

In LIMDEP, the following command returns the same result (output is skipped). Notice that two restrictions in `Cls:` are separated by a comma.

```
REGRESS;Lhs=COST;
        Rhs=One,G1,G2,G3,G4,G5,G6,T1,T2,T3,T4,T5,T6,T7,T8,T9,T10,T11,T12,T13,T14,T15,OUTPUT,FUEL,LOAD;
        Cls:b(2)+b(3)+b(4)+b(5)+b(6)+b(7)=0,
            b(8)+b(9)+b(10)+b(11)+b(12)+b(13)+b(14)+b(15)+b(16)+b(17)+b(18)+b(19)+b(20)+b(21)+b(22)=0$
```

6.7 Two-way Within Effect Model

The two-way fixed effect model requires a transformation of dependent and independent variables using group means. $y_{it}^* = y_{it} - \bar{y}_{i\bullet} - \bar{y}_{\bullet t} + \bar{y}_{\bullet\bullet}$ and $x_{it}^* = x_{it} - \bar{x}_{i\bullet} - \bar{x}_{\bullet t} + \bar{x}_{\bullet\bullet}$.

```
. gen w_cost = cost - gm_cost - tm_cost + m_cost
. gen w_output = output - gm_output - tm_output + m_output
. gen w_fuel = fuel - gm_fuel - tm_fuel + m_fuel
. gen w_load = load - gm_load - tm_load + m_load
```

Once data are transformed, run the OLS with the transformed variables. Do not forget to suppress the intercept.

```
. regress w_cost w_output w_fuel w_load, noc

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  3,    87) =  307.86
       Model |  1.87739643     3  .625798811           Prob > F      =  0.0000
    Residual |  .176848774    87  .002032745           R-squared     =  0.9139
-------------+------------------------------           Adj R-squared =  0.9109
       Total |  2.05424521    90  .022824947           Root MSE      =  .04509

------------------------------------------------------------------------------
      w_cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    w_output |   .8172487   .0279512    29.24   0.000     .7616927    .8728048
      w_fuel |     .16861   .1434621     1.18   0.243    -.1165364    .4537565
      w_load |  -.8828142   .2296907    -3.84   0.000    -1.339349    -.426279
------------------------------------------------------------------------------
```

Remember that F, R^2, standard errors, and DF_{error} are not correct. Standard errors need to be adjusted; for instance, the standard error of the load factor is .2617=.2297*sqrt(87/67).

The dummy variable coefficients are computed as $d_i^* = (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet}) - (\bar{x}_{i\bullet} - \bar{x}_{\bullet\bullet})'\beta$ and $d_t^* = (\bar{y}_{\bullet t} - \bar{y}_{\bullet\bullet}) - (\bar{x}_{\bullet t} - \bar{x}_{\bullet\bullet})'\beta$. We need to compute overall means and group specific, say airline 3, means.

```
. sum cost output fuel load

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        cost |        90    13.36561    1.131971   11.14154    15.3733
      output |        90   -1.174309    1.150606  -3.278573   .6608616
        fuel |        90    12.77036    .8123749   11.55017     13.831
        load |        90    .5604602    .0527934    .432066    .676287
```

```
. sum cost output fuel load if airline==3

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        cost |        15    13.37231    .5220657   12.56479   13.99694
      output |        15   -.9122625    .2435335  -1.337794  -.6169364
        fuel |        15    12.78972    .8177211    11.6851     13.831
        load |        15    .5845359    .0324437    .524334    .654256
```

The actual (absolute) intercept of airline 3 is -.1895 =(13.3723-13.3656)-(-.9123-(-1.1743))*(.8172) -(12.7897-12.7704)*(.1686)- (.5845-.5605)*(-.8828). The actual intercept of time period 9 is .0472=(13.4651-13.3656)-(-1.0670-(-1.1743))*(.8172) -(12.8610-12.7704)*(.1686)- (.6179-.5605)*(-.8828). See the SAS output in Section 6.6 to cross-check the computation.

```
. sum cost output fuel load if year==9

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        cost |         6     13.4651    1.042032   12.20495   14.78597
      output |         6   -1.067003    1.278931  -2.673258   .4779284
        fuel |         6    12.86104    .0212523   12.83356   12.89337
        load |         6    .6179098    .0376737    .546723    .654256
```

6.8 Using SAS: PROC TSCSREG and PROC PANEL

PROC TSCSREG and PROC PANEL have the /FIXTWO option to fit the two-way fixed effect model. The data set needs to be sorted by the group and time variables that will be declared in the ID statement in PROC PANEL.

```
PROC SORT DATA=masil.airline;
   BY airline year;

PROC PANEL DATA=masil.airline;
   ID airline year;
   MODEL cost = output fuel load /FIXTWO;
RUN;
```

```
                                The PANEL Procedure
                              Fixed Two Way Estimates

Dependent Variable: cost

                                Model Description

                        Estimation Method            FixTwo
                        Number of Cross Sections          6
                        Time Series Length               15


                                 Fit Statistics

                  SSE             0.1768    DFE                67
                  MSE             0.0026    Root MSE       0.0514
                  R-Square        0.9984
```

F Test for No Fixed Effects

Num DF	Den DF	F Value	Pr > F
19	67	23.10	<.0001

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Pr > |t|	Label
CS1	1	0.174283	0.0861	2.02	0.0470	Cross Sectional Effect 1
CS2	1	0.111451	0.0780	1.43	0.1575	Cross Sectional Effect 2
CS3	1	-0.14351	0.0519	-2.77	0.0073	Cross Sectional Effect 3
CS4	1	0.180209	0.0321	5.61	<.0001	Cross Sectional Effect 4
CS5	1	-0.04669	0.0225	-2.08	0.0415	Cross Sectional Effect 5
TS1	1	-0.69314	0.3378	-2.05	0.0441	Time Series Effect 1
TS2	1	-0.63844	0.3321	-1.92	0.0588	Time Series Effect 2
TS3	1	-0.5958	0.3294	-1.81	0.0750	Time Series Effect 3
TS4	1	-0.54215	0.3189	-1.70	0.0938	Time Series Effect 4
TS5	1	-0.47304	0.2319	-2.04	0.0454	Time Series Effect 5
TS6	1	-0.4272	0.1884	-2.27	0.0266	Time Series Effect 6
TS7	1	-0.39598	0.1733	-2.28	0.0255	Time Series Effect 7
TS8	1	-0.33985	0.1501	-2.26	0.0268	Time Series Effect 8
TS9	1	-0.27189	0.1348	-2.02	0.0477	Time Series Effect 9
TS10	1	-0.22739	0.0763	-2.98	0.0040	Time Series Effect 10
TS11	1	-0.1118	0.0319	-3.50	0.0008	Time Series Effect 11
TS12	1	-0.03364	0.0429	-0.78	0.4357	Time Series Effect 12
TS13	1	-0.01773	0.0363	-0.49	0.6263	Time Series Effect 13
TS14	1	-0.01865	0.0305	-0.61	0.5432	Time Series Effect 14
Intercept	1	12.94004	2.2182	5.83	<.0001	Intercept
output	1	0.817249	0.0319	25.66	<.0001	
fuel	1	0.16861	0.1635	1.03	0.3061	
load	1	-0.88281	0.2617	-3.37	0.0012	

6.9 Using Stata and LIMDEP

The Stata `.xtreg` command does not have an option for two-way fixed or two-way random effect models. However, this command is able to fit the two-way fixed effect model by including a set of dummies for a group (LSDV1) and using the `fe` option.

```
. xtreg cost t1-t14 output fuel load, fe i(airline)

Fixed-effects (within) regression               Number of obs      =        90
Group variable: airline                         Number of groups   =         6

R-sq:  within  = 0.9955                         Obs per group: min =        15
       between = 0.9859                                        avg =      15.0
       overall = 0.9885                                        max =        15

                                                F(17,67)           =    873.24
corr(u_i, Xb)  = 0.3361                         Prob > F           =    0.0000

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          t1 |  -.6931382   .3378385    -2.05   0.044    -1.367467   -.0188098
          t2 |  -.6384366   .3320802    -1.92   0.059    -1.301271    .0243983
          t3 |  -.5958031   .3294473    -1.81   0.075    -1.253383    .0617764
          t4 |  -.5421537   .3189139    -1.70   0.094    -1.178708    .0944011
          t5 |  -.4730429   .2319459    -2.04   0.045    -.9360088   -.0100769
          t6 |  -.4272042     .18844    -2.27   0.027    -.8033319   -.0510764
          t7 |  -.3959783   .1732969    -2.28   0.025    -.7418804   -.0500762
          t8 |  -.3398463   .1501062    -2.26   0.027    -.6394596    -.040233
          t9 |  -.2718933   .1348175    -2.02   0.048    -.5409901   -.0027964
         t10 |  -.2273857   .0763495    -2.98   0.004      -.37978   -.0749914
         t11 |  -.1118032   .0319005    -3.50   0.001     -.175477   -.0481295
         t12 |   -.033641   .0429008    -0.78   0.436    -.1192713    .0519893
         t13 |  -.0177346   .0362554    -0.49   0.626    -.0901007    .0546315
         t14 |  -.0186451    .030508    -0.61   0.543    -.0795393     .042249
      output |   .8172487    .031851    25.66   0.000     .7536739    .8808235
        fuel |      .16861    .163478     1.03   0.306    -.1576935    .4949135
        load |  -.8828142   .2617373    -3.37   0.001    -1.405244   -.3603843
       _cons |     12.986   2.225402     5.84   0.000     8.544076    17.42792
-------------+----------------------------------------------------------------
     sigma_u |   .1306712
     sigma_e |  .05137639
         rho |  .86611203   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(5, 67) =    69.05                 Prob > F = 0.0000
```

The F statistic of 69.05 tests only if parameters of `g1` through `g5` are all zero. You may double-check this test by running the following commands.

```
. quietly regress cost g1-g5 t1-t14 output fuel load
. test g1=g2=g3=g4=g5=0

 ( 1)  g1 - g2 = 0
 ( 2)  g1 - g3 = 0
 ( 3)  g1 - g4 = 0
 ( 4)  g1 - g5 = 0
 ( 5)  g1 = 0

       F(  5,    67) =   69.05
            Prob > F =    0.0000
```

The following LIMDEP command fits the two-way fixed model. This command has `Str` and `Period` to specify stratification and time variables. This command presents the pooled model and one-way group effect model as well, but reports the incorrect intercept in the two-way fixed model, 12.667 (2.081). The pooled OLS and fixed group effect parts of the entire output is skipped below since they are redundant.

```
REGRESS;Lhs=COST;Rhs=ONE,OUTPUT,FUEL,LOAD;Panel;Str=AIRLINE;Period=YEAR;Fixed$
```

```
+-----------------------------------------------------+
| Least Squares with Group and Period Effects         |
| Ordinary    least squares regression                |
| Model was estimated Aug 27, 2009 at 04:27:40PM      |
| LHS=COST     Mean                 =   13.36561      |
|              Standard deviation   =   1.131971      |
| WTS=none     Number of observs.   =         90      |
| Model size   Parameters           =         23      |
|              Degrees of freedom   =         67      |
| Residuals    Sum of squares       =   .1768479      |
|              Standard error of e  =   .5137627E-01 |
| Fit          R-squared            =   .9984493      |
|              Adjusted R-squared   =   .9979401      |
| Model test   F[ 22,    67] (prob) =1960.83 (.0000) |
| Diagnostic   Log likelihood       =   152.7479      |
|              Restricted(b=0)      =  -138.3581      |
|              Chi-sq [ 22]  (prob) = 582.21 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =  -5.709580      |
|              Akaike Info. Criter. =  -5.721164      |
| Estd. Autocorrelation of e(i,t)     .651825         |
+-----------------------------------------------------+

+-----------------------------------------------------+
| Panel:Groups   Empty       0,   Valid data       6 |
|                Smallest   15,   Largest         15 |
|                Average group size            15.00 |
| Panel: Prds:   Empty       0,   Valid data      15 |
|                Smallest    0,   Largest          6 |
|                Average group size             6.00 |
+-----------------------------------------------------+
+--------+-------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+-------------+----------------+--------+--------+----------+
 OUTPUT  |     .81725242       .03185102    25.659   .0000  -1.17430918
 FUEL    |     .16863516       .16347826     1.032   .3052   12.7703592
 LOAD    |    -.88281516       .26173663    -3.373   .0011    .56046016
 Constant|    12.6665675      2.08107166     6.087   .0000

+----------------------------------------------------------------------+
|             Test Statistics for the Classical Model                  |
+----------------------------------------------------------------------+
|       Model            Log-Likelihood  Sum of Squares    R-squared |
|(1)  Constant term only     -138.35814  .1140409821D+03    .0000000 |
|(2)  Group effects only      -90.48804  .3936109461D+02    .6548513 |
|(3)  X - variables only       61.76991  .1335449522D+01    .9882897 |
|(4)  X and group effects     130.08647  .2926207777D+00    .9974341 |
|(5)  X ind.&time effects     152.74790  .1768479062D+00    .9984493 |
+----------------------------------------------------------------------+
|                       Hypothesis Tests                               |
|         Likelihood Ratio Test           F Tests                      |
|         Chi-squared   d.f.  Prob.       F    num. denom.   P value |
|(2) vs (1)    95.740      5  .00000   31.875     5      84   .00000 |
|(3) vs (1)   400.256      3  .00000 2419.329     3      86   .00000 |
|(4) vs (1)   536.889      8  .00000 3935.818     8      81   .00000 |
|(4) vs (2)   441.149      3  .00000 3604.832     3      81   .00000 |
|(4) vs (3)   136.633      5  .00000   57.733     5      81   .00000 |
|(5) vs (4)    45.323     14  .00004    3.133    14      67   .00085 |
|(5) vs (3)   181.956     20  .00000   21.947    20      67   .00000 |
+----------------------------------------------------------------------+
```

6.10 Testing Two-way Fixed Effects

The null hypothesis is that parameters of group and time dummies are zero:
$H_0 : \mu_1 = ... = \mu_{n-1} = 0$ and $\tau_1 = ... = \tau_{T-1} = 0$. The F test compares the pooled regression and

two-way fixed group and time effect model. The F statistic of 23.1085 rejects the null hypothesis at the .01 significance level (p<.0000).

$$\frac{(1.3354-.1768)/(6+15-2)}{(.1768)/(6*15-6-15-3+1)} \sim 23.1085[19,67]$$

The SAS TSCSREG and PANEL procedures conduct this F-test for the group and time effects. You may also run the following SAS REG procedure and Stata `.regress` command to perform the same test. The Stata output is skipped.

```
PROC REG DATA=masil.airline;
   MODEL cost = g1-g5 t1-t14 output fuel load;
   TEST g1=g2=g3=g4=g5=t1=t2=t3=t4=t5=t6=t7=t8=t9=t10=t11=t12=t13=t14=0;
RUN;
```

Test 1 Results for Dependent Variable cost

Source	DF	Mean Square	F Value	Pr > F
Numerator	19	0.06098	23.10	<.0001
Denominator	67	0.00264		

```
. quietly regress cost g1-g5 t1-t14 output fuel load
. test g1 g2 g3 g4 g5 t1 t2 t3 t4 t5 t6 t7 t8 t9 t10 t11 t12 t13 t14
```

7. Random Effect Models

A random effect model examines how group and/or time affect error variances. This model is appropriate for *n* individuals who were drawn randomly from a large population. This chapter focuses on the feasible generalized least squares (FGLS) with variance component estimation methods.[10]

7.1 One-way Random Group Effect Model

When the omega matrix is not known, you have to estimate θ using the SSEs of the between group effect model (.0317) and the fixed group effect model (.2926).

The variance component of error $\hat{\sigma}_v^2$ is .00361263 = .292622872/(6*15-6-3)
The variance component of group $\hat{\sigma}_u^2$ is .01559712 =.031675926/(6-4) - .00361263/15

Thus, $\hat{\theta}$ is $.87668488 = 1 - \sqrt{\frac{\hat{\sigma}_v^2}{T\hat{\sigma}_u^2 + \hat{\sigma}_v^2}} = 1 - \sqrt{\frac{\hat{\sigma}_v^2}{T\hat{\sigma}_{between}^2}} = 1 - \sqrt{\frac{.00361263}{15 * .031675926/(6-4)}}$,

where $\hat{\sigma}_{between}^2 = \frac{SSE_{between}}{n-K} = \frac{.031675926}{6-4} = .01583796$.

Next, transform the dependent and independent variables including the intercept using $\hat{\theta}$.

```
. gen rg_cost = cost - .87668488*gm_cost
. gen rg_output = output - .87668488*gm_output
. gen rg_fuel = fuel - .87668488*gm_fuel
. gen rg_load = load - .87668488*gm_load
. gen rg_int = 1 - .87668488 // for the intercept
```

Finally, run the OLS with the transformed variables. Do not forget to suppress the intercept. This is the groupwise heteroscedastic regression model (Greene 2003).

```
. regress rg_cost rg_int rg_output rg_fuel rg_load, noc

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  4,    86) =19642.72
       Model |  284.670313     4  71.1675783           Prob > F      =  0.0000
    Residual |  .311586777    86  .003623102           R-squared     =  0.9989
-------------+------------------------------           Adj R-squared =  0.9989
       Total |    284.9819    90  3.16646556           Root MSE      =  .06019

------------------------------------------------------------------------------
     rg_cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      rg_int |   9.627911   .2101638    45.81   0.000     9.210119     10.0457
```

[10] Baltagi and Cheng (1994) introduce various ANOVA estimation methods, such as a modified Wallace and Hussain method, the Wansbeek and Kapteyn method, the Swamy and Arora method, and Henderson's method III. They also discuss maximum likelihood (ML) estimators, restricted ML estimators, minimum norm quadratic unbiased estimators (MINQUE), and minimum variance quadratic unbiased estimators (MIVQUE). Based on a Monte Carlo simulation, they argue that ANOVA estimators are Best Quadratic Unbiased estimators of the variance components for the balanced model, whereas ML, restricted ML, MINQUE, and MIVQUE are recommended for the unbalanced models.

```
   rg_output |   .9066808   .0256249    35.38   0.000     .8557401    .9576215
     rg_fuel |   .4227784   .0140248    30.15   0.000      .394898    .4506587
     rg_load |    -1.0645   .2000703    -5.32   0.000    -1.462226   -.6667731
------------------------------------------------------------------------------
```

7.2 Estimations in SAS, Stata, and LIMDEP

In SAS, the TSCSREG and PANEL procedures have the /RANONE option to fit the one-way random effect model. These procedures by default use the Fuller and Battese (1974) estimation method, which produces slightly different estimates from FGLS.

PROC PANEL has the /VCOMP=WK option for the Wansbeek and Kapteyn (1989) method, which is the groupwise heteroscedastic regression. The BP option of the MODEL statement, not available in PROC TSCSREG, conducts the Breusch-Pagen LM test for random effects. Unlike PROC PANEL, PROC TSCSREG does not have VCOMP= to specify the type of variance component estimation.

```
PROC PANEL DATA=masil.airline;
   ID airline year;
   MODEL cost = output fuel load /RANONE BP VCOMP=WK;
RUN;
```

```
                              The PANEL Procedure
                    Wansbeek and Kapteyn Variance Components (RanOne)

Dependent Variable: cost

                                  Model Description

                          Estimation Method            RanOne
                          Number of Cross Sections          6
                          Time Series Length               15


                                   Fit Statistics

                    SSE             0.3111    DFE                  86
                    MSE             0.0036    Root MSE         0.0601
                    R-Square        0.9923


                             Variance Component Estimates

                     Variance Component for Cross Sections    0.016015
                     Variance Component for Error             0.003613


                                  Hausman Test for
                                   Random Effects

                                DF    m Value    Pr > m

                                 2       1.63    0.4429
```

```
Breusch Pagan Test for Random
      Effects (One Way)

     DF     m Value     Pr > m

      1      334.85     <.0001


          Parameter Estimates

                                Standard
Variable        DF    Estimate     Error    t Value    Pr > |t|

Intercept        1    9.629513    0.2107      45.71      <.0001
output           1    0.906918    0.0257      35.30      <.0001
fuel             1    0.422676    0.0140      30.11      <.0001
load             1    -1.06452    0.2000      -5.32      <.0001
```

PROC PANEL and PROC TSCSREG estimate the same variance component for error (.0036) but a different variance component for groups (.0160 versus .4744). Notice that there are some differences in the output of PROC TSCSREG (variance component estimates and Hausman test) between SAS 9.2 and 9.13.

```
PROC TSCSREG DATA=masil.airline;
   ID airline year;
   MODEL cost = output fuel load /RANONE;
RUN;

(output is skipped)
```

Alternatively, you may use PROC MIXED to get the same results. The following script returns a set of random effect estimates. Unlike SAS 9.13, SAS 9.2 requires the CLASS statement to explicitly specify an effect variable, `airline` in this case.

```
PROC MIXED DATA=masil.airline;
       CLASS airline;
       MODEL cost = output fuel load /SOLUTION;
       RANDOM INTERCEPT / SUBJECT=airline TYPE=UN SOLUTION;
RUN;
```

```
The Mixed Procedure

 Covariance Parameter Estimates

Cov Parm     Subject    Estimate

UN(1,1)      airline     0.01674
Residual                0.003609


         Fit Statistics

-2 Res Log Likelihood          -210.4
AIC (smaller is better)        -206.4
AICC (smaller is better)       -206.3
```

```
                    BIC (smaller is better)        -206.8


                      Null Model Likelihood Ratio Test

                        DF    Chi-Square      Pr > ChiSq

                         1        107.49          <.0001


                         Solution for Fixed Effects

                                  Standard
      Effect       Estimate       Error      DF    t Value    Pr > |t|

      Intercept      9.6322      0.2116       5      45.53      <.0001
      output         0.9073     0.02581      81      35.16      <.0001
      fuel           0.4225     0.01406      81      30.05      <.0001
      load          -1.0646      0.1998      81      -5.33      <.0001


                        Solution for Random Effects

                                        Std Err
    Effect       airline    Estimate        Pred      DF    t Value    Pr > |t|

    Intercept    1           0.01012     0.06594      81       0.15      0.8784
    Intercept    2          -0.03450     0.06239      81      -0.55      0.5818
    Intercept    3           -0.2106     0.05507      81      -3.82      0.0003
    Intercept    4            0.1691     0.05581      81       3.03      0.0033
    Intercept    5          0.002981     0.06180      81       0.05      0.9616
    Intercept    6           0.06291     0.06349      81       0.99      0.3247


                      Type 3 Tests of Fixed Effects

                           Num     Den
                Effect      DF      DF    F Value    Pr > F

                output       1      81    1235.88    <.0001
                fuel         1      81     903.03    <.0001
                load         1      81      28.40    <.0001
```

In Stata, the `.xtreg` command has the `re` option to produce FGLS estimates. Let us specify `airline` as a panel identification variable using the `.iis` command. The `theta` option reports an estimated theta (.8767).

```
. iis airline

. xtreg cost output fuel load, re theta

Random-effects GLS regression                   Number of obs      =        90
Group variable: airline                         Number of groups   =         6

R-sq:  within  = 0.9925                         Obs per group: min =        15
       between = 0.9856                                        avg =      15.0
       overall = 0.9876                                        max =        15

Random effects u_i ~ Gaussian                   Wald chi2(3)       =  11091.33
```

```
corr(u_i, X)       = 0 (assumed)               Prob > chi2        =    0.0000
theta              = .87668503

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |   .9066805    .025625    35.38   0.000     .8564565    .9569045
        fuel |   .4227784   .0140248    30.15   0.000     .3952904    .4502665
        load |  -1.064499   .2000703    -5.32   0.000    -1.456629    -.672368
       _cons |   9.627909    .210164    45.81   0.000     9.215995    10.03982
-------------+----------------------------------------------------------------
     sigma_u |  .12488859
     sigma_e |  .06010514
         rho |  .81193816   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

The `sigma_u` and `sigma_e` are square roots of the variance components for groups and errors (.0156=.1249^2, .0036=.0601^2).

Alternatively, `.xtmixed` fits the same model, the random-intercept model. The `|| airline:,` option tells Stata to fit the model using the subject variable `airline`. Variance components for groups and errors are reported under the labels `sd(_cons)` and `sd(Residual)`.

```
. xtmixed cost output fuel load || airline:,

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log restricted-likelihood =  105.20458
Iteration 1:   log restricted-likelihood =  105.20458

Computing standard errors:

Mixed-effects REML regression                   Number of obs      =        90
Group variable: airline                         Number of groups   =         6

                                                Obs per group: min =        15
                                                               avg =      15.0
                                                               max =        15


                                                Wald chi2(3)       =  11114.85
Log restricted-likelihood =  105.20458          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |   .9073166    .025809    35.16   0.000      .856732    .9579013
        fuel |   .4225032   .0140598    30.05   0.000     .3949465      .45006
        load |  -1.064572   .1997763    -5.33   0.000    -1.456126   -.6730179
       _cons |   9.632212    .211559    45.53   0.000     9.217564    10.04686
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
airline: Identity            |
                   sd(_cons) |   .1293723   .0429029      .0675403    .2478107
-----------------------------+------------------------------------------------
                sd(Residual) |   .0600715   .0047138       .051508    .0700588
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =   107.49 Prob >= chibar2 = 0.0000
```

You may use the maximum likelihood estimation to fit random effect (or random intercept) model. In SAS, add METHOD=ML to PROC MIXED. PROC PANEL and TSCSREG do not have such option.

```
PROC MIXED DATA=masil.airline METHOD=ML;
       CLASS airline;
       MODEL cost = output fuel load /SOLUTION;
       RANDOM INTERCEPT / SUBJECT=airline TYPE=UN SOLUTION;
RUN;
```

The Mixed Procedure

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)	airline	0.01302
Residual		0.003494

Fit Statistics

-2 Log Likelihood	-229.5
AIC (smaller is better)	-217.5
AICC (smaller is better)	-216.4
BIC (smaller is better)	-218.7

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	105.92	<.0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	9.6186	0.2026	5	47.47	<.0001
output	0.9053	0.02466	81	36.72	<.0001
fuel	0.4234	0.01364	81	31.05	<.0001
load	-1.0645	0.1962	81	-5.42	<.0001

Solution for Random Effects

Effect	airline	Estimate	Std Err Pred	DF	t Value	Pr > |t|
Intercept	1	0.01306	0.05994	81	0.22	0.8281
Intercept	2	-0.03211	0.05640	81	-0.57	0.5707
Intercept	3	-0.2094	0.04900	81	-4.27	<.0001
Intercept	4	0.1676	0.04976	81	3.37	0.0012
Intercept	5	0.000761	0.05580	81	0.01	0.9892
Intercept	6	0.06008	0.05750	81	1.04	0.2992

```
                    Type 3 Tests of Fixed Effects

                          Num     Den
Effect                     DF      DF    F Value    Pr > F

output                      1      81    1348.19    <.0001
fuel                        1      81     963.88    <.0001
load                        1      81      29.43    <.0001
```

In Stata, the `mle` option is used in `.xtreg` and `.xtmixed` commands to produce the same result. You may also try `.xtgls` that fits panel data models with heteroscedasticity across and within groups. Notice that error variance components are computed as .0130=1141^2 and .0035 = .0591^2. Compare the output of PROC MIXED above and `.xtreg` below.

```
. xtreg cost output fuel load, re mle

Random-effects ML regression                    Number of obs      =        90
Group variable: airline                         Number of groups   =         6

Random effects u_i ~ Gaussian                   Obs per group: min =        15
                                                               avg =      15.0
                                                               max =        15

                                                LR chi2(3)         =    436.32
Log likelihood  =  114.72896                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |   .9053099   .0253759    35.68   0.000     .8555741    .9550458
        fuel |   .4233757    .013888    30.48   0.000     .3961557    .4505957
        load |  -1.064456    .196231    -5.42   0.000    -1.449062   -.6798506
       _cons |   9.618648    .206622    46.55   0.000     9.213677    10.02362
-------------+----------------------------------------------------------------
    /sigma_u |   .1140843   .0345293                      .0630373    .2064687
    /sigma_e |   .0591072   .0045701                      .0507956    .0687787
         rho |   .7883772   .1047419                      .5365302    .9344669
------------------------------------------------------------------------------
Likelihood-ratio test of sigma_u=0: chibar2(01)=  105.92 Prob>=chibar2 = 0.000

. xtmixed cost output fuel load || airline:, mle
(output is skipped)

. xtgls cost output fuel load, i(airline) panels(hetero) corr(independent)
(output is skipped)
```

In LIMDEP, you have to specify `Panel`, `Random Effect`, and `Het=` subcommands for the groupwise heteroscedastic model. LIMDEP estimates a slightly different variance component for groups (.0119), thus producing different parameter estimates.

```
REGRESS;Lhs=COST;Rhs=ONE,OUTPUT,FUEL,LOAD;Panel;Str=AIRLINE;Het=AIRLINE;Random Effect$

+-----------------------------------------------------+
| OLS Without Group Dummy Variables                   |
| Ordinary    least squares regression                |
| Model was estimated Aug 30, 2009 at 08:26:15PM      |
| LHS=COST     Mean                 =   13.36561      |
|              Standard deviation   =   1.131971      |
| WTS=none     Number of observs.   =         90      |
| Model size   Parameters           =          4      |
|              Degrees of freedom   =         86      |
| Residuals    Sum of squares       =   1.335450      |
|              Standard error of e  =   .1246133     |
```

```
| Fit          R-squared            =   .9882897     |
|              Adjusted R-squared   =   .9878812     |
| Model test   F[  3,    86] (prob) =2419.33 (.0000) |
| Diagnostic   Log likelihood       =   61.76991     |
|              Restricted(b=0)      =  -138.3581     |
|              Chi-sq [  3]  (prob) = 400.26 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =  -4.121594     |
|              Akaike Info. Criter. =  -4.121653     |
+----------------------------------------------------+

+----------------------------------------------------+
| Panel Data Analysis of COST       [ONE way]        |
|          Unconditional ANOVA (No regressors)       |
| Source      Variation   Deg. Free.     Mean Square |
| Between       74.6799           5.     14.9360     |
| Residual      39.3611          84.     .468584     |
| Total         114.041          89.     1.28136     |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 OUTPUT  |     .88273863       .01325455    66.599   .0000  -1.17430918
 FUEL    |     .45397771       .02030424    22.359   .0000   12.7703592
 LOAD    |   -1.62750780       .34530293    -4.713   .0000    .56046016
 Constant|    9.51691223       .22924522    41.514   .0000

+----------------------------------------------------+
| Panel:Groups   Empty       0,   Valid data       6 |
|                Smallest   15,   Largest         15 |
|                Average group size            15.00 |
+----------------------------------------------------+

+----------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)       |
| Estimates:  Var[e]              =   .361260D-02    |
|             Var[u]              =   .119159D-01    |
|             Corr[v(i,t),v(i,s)] =   .767356        |
| Lagrange Multiplier Test vs. Model (3) =  334.85   |
| ( 1 df, prob value =  .000000)                     |
| (High values of LM favor FEM/REM over CR model.)   |
| Baltagi-Li form of LM Statistic =         334.85   |
|             Sum of Squares          .147779D+01    |
|             R-squared               .987042D+00    |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 OUTPUT  |     .90412380       .02461548    36.730   .0000  -1.17430918
 FUEL    |     .42389869       .01374650    30.837   .0000   12.7703592
 LOAD    |   -1.06455866       .19933132    -5.341   .0000    .56046016
 Constant|    9.61063438       .20277404    47.396   .0000
```

7.3 One-way Random Time Effect Model

Let us compute $\hat{\theta}$ using the SSEs of the between time effect model (.0056) and the fixed time effect model (1.0882).

The variance component for error $\hat{\sigma}_v^2$ is .01511375 = 1.08819022/(15*6-15-3)

The variance component for time $\hat{\sigma}_u^2$ is -.00201072 =.005590631/(15-4)- .01511375/6

The $\hat{\theta}$ is -1.226263 $= 1 - \sqrt{\dfrac{\hat{\sigma}_v^2}{n\hat{\sigma}_{between}^2}} = 1 - \sqrt{\dfrac{.01511375}{6 * .005590631/(15-4)}}$

```
. gen rt_cost = cost - (-1.226263)*tm_cost
. gen rt_output = output - (-1.226263)*tm_output
. gen rt_fuel = fuel - (-1.226263)*tm_fuel
. gen rt_load = load - (-1.226263)*tm_load
. gen rt_int = 1 - (-1.226263) // for the intercept
```

```
. regress rt_cost rt_int rt_output rt_fuel rt_load, noc

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  4,    86) =       .
       Model |  79944.1804     4  19986.0451           Prob > F      =  0.0000
    Residual |  1.79271995    86  .020845581           R-squared     =  1.0000
-------------+------------------------------           Adj R-squared =  1.0000
       Total |  79945.9732    90  888.288591           Root MSE      =  .14438

------------------------------------------------------------------------------
     rt_cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      rt_int |   9.516098   .1489281    63.90   0.000     9.220038    9.812157
   rt_output |   .8883838   .0143338    61.98   0.000     .8598891    .9168785
     rt_fuel |   .4392731   .0129051    34.04   0.000     .4136186    .4649277
     rt_load |  -1.279176   .2482869    -5.15   0.000    -1.772754   -.7855982
------------------------------------------------------------------------------
```

However, the negative value of the variance component for time is not likely.

In SAS, use the TSCSREG or PANEL procedure with the /RANONE option. Notice that the data are sorted by `year` and `airline`. The /VCOMP=WH option in the MODEL statement employs Wallace and Hussian's method to estimating variance components and produces the same parameter estimates.

```
PROC SORT DATA=masil.airline;
   BY year airline;

PROC TSCSREG DATA=masil.airline;
   ID year airline;
   MODEL cost = output fuel load /RANONE;
RUN;
(Output is skipped)

PROC PANEL DATA=masil.airline;
   ID year airline;
   MODEL cost = output fuel load /RANONE BP VCOMP=WH;
RUN;
```

```
                          The PANEL Procedure
                Wallace and Hussain Variance Components (RanOne)

Dependent Variable: cost

                          Model Description

                Estimation Method            RanOne
                Number of Cross Sections         15
                Time Series Length                6


                           Fit Statistics

           SSE             1.3354    DFE                86
           MSE             0.0155    Root MSE       0.1246
           R-Square        0.9883
```

Variance Component Estimates

Variance Component for Cross Sections	0
Variance Component for Error	0.016437

Hausman Test for Random Effects

DF	m Value	Pr > m
2	12.17	0.0023

Breusch Pagan Test for Random Effects (One Way)

DF	m Value	Pr > m
1	1.55	0.2135

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	9.516923	0.2292	41.51	<.0001
output	1	0.882739	0.0133	66.60	<.0001
fuel	1	0.453977	0.0203	22.36	<.0001
load	1	-1.62751	0.3453	-4.71	<.0001

PROC MIXED fits the same random time effect model although /SOLUTION in the RANDOM statement does not work to produce random effect parameter estimates in this case.

```
PROC MIXED DATA=masil.airline;
       CLASS airline;
       MODEL cost = output fuel load /SOLUTION;
       RANDOM INTERCEPT / SUBJECT=airline TYPE=UN;
RUN;
```

The Mixed Procedure

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)	year	0
Residual		0.01553

Fit Statistics

-2 Res Log Likelihood	-102.9
AIC (smaller is better)	-100.9
AICC (smaller is better)	-100.9
BIC (smaller is better)	-100.2

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
0	0.00	1.0000

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	9.5169	0.2292	14	41.51	<.0001
output	0.8827	0.01325	72	66.60	<.0001
fuel	0.4540	0.02030	72	22.36	<.0001
load	-1.6275	0.3453	72	-4.71	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
output	1	72	4435.44	<.0001
fuel	1	72	499.92	<.0001
load	1	72	22.22	<.0001

In Stata, you have to switch group and time variables using the `.tsset` command.

```
. tsset year airline
       panel variable:  year (strongly balanced)
        time variable:  airline, 1 to 6
                delta:  1 unit

. xtreg cost output fuel load, re i(year) theta

Random-effects GLS regression                   Number of obs      =        90
Group variable: year                            Number of groups   =        15

R-sq:  within  = 0.9843                         Obs per group: min =         6
       between = 0.9966                                        avg =       6.0
       overall = 0.9883                                        max =         6

Random effects u_i ~ Gaussian                   Wald chi2(3)       =   7258.03
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000
theta              = 0

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |   .8827385   .0132545    66.60   0.000     .8567602    .9087169
        fuel |    .453977   .0203042    22.36   0.000     .4141815    .4937724
        load |   -1.62751    .345302    -4.71   0.000     -2.30429   -.9507309
       _cons |   9.516923   .2292445    41.51   0.000     9.067612    9.966233
-------------+----------------------------------------------------------------
     sigma_u |          0
     sigma_e |  .12293801
         rho |          0   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

You may runt the following command to get the same result.

```
. xtmixed cost output fuel load || year:,
(output is skipped)
```

In LIMDEP, you need to use the `Str=` and `Random` subcommands. The output below includes only the random effect part. You may find that parameter estimates of SAS, Stata, and LIMDEP are slightly different each other.

```
REGRESS;Lhs=COST;Rhs=ONE,OUTPUT,FUEL,LOAD;Panel;Str=YEAR;Het=YEAR;Random$

+----------------------------------------------------+
| Panel:Groups   Empty       0,   Valid data      15 |
|                Smallest    6,   Largest          6 |
|                Average group size             6.00 |
+----------------------------------------------------+

+------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)   |
| Estimates:  Var[e]              =   .151138D-01  |
|             Var[u]              =   .414686D-03  |
|             Corr[v(i,t),v(i,s)] =   .026705      |
| Lagrange Multiplier Test vs. Model (3) =    1.55 |
| ( 1 df, prob value =  .213557)                   |
| (High values of LM favor FEM/REM over CR model.) |
| Baltagi-Li form of LM Statistic =           1.55 |
|             Sum of Squares          .133564D+01  |
|             R-squared               .988288D+00  |
+------------------------------------------------+
+--------+-------------+---------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+---------------+--------+--------+----------+
 OUTPUT  |     .88285277       .01314515    67.162   .0000  -1.17430918
 FUEL    |     .45500533       .02122856    21.434   .0000   12.7703592
 LOAD    |   -1.66267268       .35084190    -4.739   .0000    .56046016
 Constant|    9.52363173       .24108843    39.503   .0000
```

7.4 Two-way Random Effect Model in SAS

The random group and time effect model is formulated as $y_{it} = \alpha + \beta' X_{ti} + u_i + \gamma_t + \varepsilon_{it}$. Let us first estimate the two way FGLS using the SAS PANEL procedure with the /RANTWO option. The BP2 option conducts the Breusch-Pagan LM test for the two-way random effect model.

```
PROC TSCSREG DATA=masil.airline;
   ID airline year;
   MODEL cost = output fuel load /RANTWO;
RUN;
(Output is skipped)

PROC PANEL DATA=masil.airline;
   ID airline year;
   MODEL cost = output fuel load /RANTWO BP2;
RUN;

                                The PANEL Procedure
                    Fuller and Battese Variance Components (RanTwo)

Dependent Variable: cost

                                Model Description

                          Estimation Method            RanTwo
```

```
                Number of Cross Sections          6
                Time Series Length               15


                        Fit Statistics

SSE              0.2322    DFE                  86
MSE              0.0027    Root MSE         0.0520
R-Square         0.9829


                Variance Component Estimates

 Variance Component for Cross Sections    0.017439
 Variance Component for Time Series       0.001081
 Variance Component for Error              0.00264


                        Hausman Test for
                         Random Effects

                       DF    m Value    Pr > m

                        3       6.93    0.0741

                    Breusch Pagan Test for Random
                          Effects (Two Way)

                       DF    m Value    Pr > m

                        2     336.40    <.0001

                     Parameter Estimates

                                    Standard
Variable        DF    Estimate       Error    t Value    Pr > |t|

Intercept        1    9.362677      0.2440      38.38      <.0001
output           1    0.866448      0.0255      33.98      <.0001
fuel             1    0.436163      0.0172      25.41      <.0001
load             1    -0.98053      0.2235      -4.39      <.0001
```

The following `.xtmixed` command suffers from convergence problem in this case and LIMDEP command produces different results (output is skipped).

```
. xtmixed cost output fuel load || airline:  || year:, mle

REGRESS;Lhs=COST;Rhs=ONE,OUTPUT,FUEL,LOAD;Panel;Str=AIRLINE;Period=YEAR;Random Effect$
```

7.5 Testing Random Effect Models

The Breusch-Pagan Lagrange multiplier (LM) test is designed to test random effects. The null hypothesis of the one-way random group effect model is that individual-specific or time-series error variances are zero: $H_0 : \sigma_u^2 = 0$. If the null hypothesis is not rejected, the pooled

regression model is appropriate. The e'e of the pooled OLS is 1.33544153 and $\bar{e}'\bar{e}$ is .0665147.

LM is 334.8496= $\frac{6*15}{2(15-1)}\left[\frac{15^2*.0665}{1.3354}-1\right]^2 \sim \chi^2(1)$ with p <.0000.

With the large chi-squared of 334.8496, we reject the null hypothesis in favor of the random group effect model. The SAS PANEL procedure with the /BP option and the LIMDEP `Panel` and `Het` subcommands report the same LM statistic (see 7.2). In Stata, run the `.xttest0` command right after estimating the one-way random group effect model.

```
. quietly xtreg cost output fuel load, re i(airline)

. xttest0

Breusch and Pagan Lagrangian multiplier test for random effects

        cost[airline,t] = Xb + u[airline] + e[airline,t]

        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
                    cost |   1.281358       1.131971
                       e |   .0036126       .0601051
                       u |   .0155972       .1248886

        Test:   Var(u) = 0
                             chi2(1) =   334.85
                          Prob > chi2 =     0.0000
```

The null hypothesis of the one-way random time effect is that variance components for time are zero, $H_0 : \sigma_u^2 = 0$. The following LM test uses Baltagi's formula. The small chi-squared of 1.5472 does not reject the null hypothesis at the .01 level. SAS and LIMDEP return the same LM statistic (see 7.3).

LM is $1.5472 = \frac{Tn}{2(n-1)}\left[\frac{\sum(n\bar{e}_{\bullet t})^2}{\sum\sum e_{it}^2}-1\right]^2 = \frac{15*6}{2(6-1)}\left[\frac{.7817}{1.3354}-1\right]^2 \sim \chi^2(1)$ with p<.2135

```
. quietly xtreg cost output fuel load, re i(year)

. xttest0

Breusch and Pagan Lagrangian multiplier test for random effects

        cost[year,t] = Xb + u[year] + e[year,t]

        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
                    cost |   1.281358       1.131971
                       e |   .0151138        .122938
                       u |          0             0

        Test:   Var(u) = 0
                             chi2(1) =     1.55
                          Prob > chi2 =     0.2135
```

The two way random effects model has the null hypothesis that variance components for groups and time are all zero. The LM statistic with two degrees of freedom is 336.3968 = 334.8496 + 1.5472 (p<.0001).

7.6 Fixed Effects versus Random Effects

How do we compare a fixed effect model and its counterpart random effect model? The Hausman specification test examines if the individual effects are uncorrelated with the other regressors in the model. Since computation is complicated, let us conduct the test in Stata.

```
. tsset airline year
       panel variable:  airline (strongly balanced)
        time variable:  year, 1 to 15
                delta:  1 unit

. quietly xtreg cost output fuel load, fe

. estimates store fixed_group

. quietly xtreg cost output fuel load, re

. hausman fixed_group .

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |  fixed_group       .          Difference          S.E.
-------------+------------------------------------------------------------
      output |    .9192846     .9066805        .0126041        .0153877
        fuel |    .4174918     .4227784       -.0052867        .0058583
        load |   -1.070396    -1.064499       -.0058974        .0255088
------------------------------------------------------------------------------
                           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(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        2.12
                Prob>chi2 =      0.5469
                (V_b-V_B is not positive definite)
```

The Hausman statistic 2.12 is different from PROC PANEL's 1.63 and Greene (2003)'s 4.16. It is because SAS, Stata, and LIMDEP use different estimation methods to produce slightly different parameter estimates. These tests, however, do not reject the null hypothesis in favor of the random effect model.

7.7 Summary

Table 7.1 summarizes random effect estimations in SAS, Stata, and LIMDEP. PROC PANEL is highly recommended.

Table 7.1 Comparison of the Random Effect Model in SAS, Stata, LIMDEP*

	SAS 9.2		Stata 11	LIMDEP 9
Procedure/Command	PROC TSCSREG	PROC PANEL	.xtreg	Regress; Panel$
One-way	/RANONE	/RANONE WK	re	Str=;Random$
Two-way	/RANTWO	/RANTWO	No	Str=;Period;Random$
SSE (e'e)	Slightly different	Correct	No	Incorrect
MSE or SEE	Slightly different	Correct	No	No

Model test (F) (adjusted) R^2	No Slightly different	No Slightly different	Wald test Incorrect	No Incorrect
Intercept	Slightly different	Correct	Correct	Slightly different
Coefficients	Slightly different	Correct	Correct	Slightly different
Standard errors	Slightly different	Correct	Correct	Slightly different
Variance for group	Slightly different	Correct	Correct (sigma)	Slightly different
Variance for error	Correct	Correct	Correct (sigma)	Correct
Theta	No	No	`theta`	No
Breusch-Pagan (LM)	No	`BP, BP2`	`.xttest0`	Yes
Hausman Test (H)	Incorrect	Yes	`.hausman`	Yes (unstable)

* "Yes/No" means whether a software package reports the statistic. "Correct/incorrect" indicates whether the statistics are different from those of the groupwise heteroscedastic regression.

8. Poolability Test

Table 8.1 summarizes the results of pooled OLS, fixed effect, and random effect model. We may ask, "Which model is better than the others?" Do we have to consider individual-specific or time effect? Are these effects are fixed or random?

Table 8.1 Summary of Pooled, Fixed Effect, and Random Effect Models

Model	Output	Fuel	Load	SSE/SEE	DF	F	R^2 (Adj.)
Pooled	.8827**	.4540**	-1.6275**	1.3354	86	2419.34	.9883
	(.0133)	(.0203)	(.3453)	(.1246)		(p<.0000)	(.9879)
Between group	.7825*	-5.5239	-1.7511	.0317	2	104.12	.9936
	(.1088)	(4.4787)	(2.7432)	(.1259)		(p<.0095)	(.9841)
Between time	1.1333**	.3342**	-1.3507**	.0056	11	4074.33	.9991
	(.0513)	(.0228)	(.2478)	(.0225)		(p<.0000)	(.9989)
Fixed group	.9193**	.4175**	-1.0704**	.2926	81	3935.79	.9974
	(.0299)	(.0152)	(.2017)	(.0601)		(p<.0000)	(.9972)
Fixed time	.8677**	-.4845	-1.9544**	1.0882	72	439.62	.9905
	(.0154)	(.3641)	(.4424)	(.1229)		(p<.0001)	(.9882)
Two-way fixed	.8173**	.1686	-.8828**	.1769	67	1960.82	.9984
	(.0319)	(.1635)	(.2617)	(.0514)		(p<.0000)	(.9979)
Random group	.9069**	.4227**	-1.0645**	.3111	86		.9923
	(.0257)	(.0140)	(.2000)	(.0601)			
Random time	.8820**	.2749+	-2.0050**	1.1722	86		.9848
	(.0134)	(.0568)	(.4184)	(.1167)			
Two-way random	.8664**	.4362**	-.9805**	.2322	86		.9829
	(.0255)	(.0172)	(.2235)	(.0520)			

The poolability test examine if data are poolable so that individual entities or time periods have the same constant slopes of regressors. For poolability test, you need to run group by group OLS regressions and/or time by time OLS regressions. If the null hypothesis is rejected, the panel data are not poolable. In this case, you may consider the random coefficient model and hierarchical regression model.

8.1 Group by Group OLS Regression

In SAS, use the BY statement in PROC REG. Do not forget to sort the data set in advance.

```
PROC SORT DATA=masil.airline;
   BY airline;

PROC REG DATA=masil.airline;
   MODEL cost = output fuel load;
   BY airline;
RUN;
```

In Stata, the `if` qualifier makes it easy to run group by group regressions.

```
forvalues i= 1(1)6 { // run group by group regression
    display "OLS regression for group " `i'
    regress cost output fuel load if airline==`i'
}

OLS regression for group 1
```

```
      Source |       SS       df       MS              Number of obs =      15
-------------+------------------------------           F(  3,    11) = 1843.46
       Model |  3.41824348     3  1.13941449           Prob > F      =  0.0000
    Residual |  .006798918    11  .000618083           R-squared     =  0.9980
-------------+------------------------------           Adj R-squared =  0.9975
       Total |   3.4250424    14  .244645886           Root MSE      =  .02486

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |    1.18318   .0968946    12.21   0.000     .9699164    1.396444
        fuel |   .3865867   .0181946    21.25   0.000     .3465406    .4266329
        load |  -2.461629   .4013571    -6.13   0.000     -3.34501   -1.578248
       _cons |     10.846   .2972551    36.49   0.000     10.19174    11.50025
------------------------------------------------------------------------------
OLS regression for group 2

      Source |       SS       df       MS              Number of obs =      15
-------------+------------------------------           F(  3,    11) = 3129.50
       Model |  6.47622084     3  2.15874028           Prob > F      =  0.0000
    Residual |  .007587838    11  .000689803           R-squared     =  0.9988
-------------+------------------------------           Adj R-squared =  0.9985
       Total |  6.48380868    14  .463129191           Root MSE      =  .02626

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |   1.459104   .0792856    18.40   0.000     1.284597     1.63361
        fuel |   .3088958   .0272443    11.34   0.000     .2489315      .36886
        load |  -2.724785   .2376522   -11.47   0.000    -3.247854   -2.201716
       _cons |   11.97243   .4320951    27.71   0.000     11.02139    12.92346
------------------------------------------------------------------------------
OLS regression for group 3

      Source |       SS       df       MS              Number of obs =      15
-------------+------------------------------           F(  3,    11) =  608.10
       Model |  3.79286673     3  1.26428891           Prob > F      =  0.0000
    Residual |  .022869767    11   .00207907           R-squared     =  0.9940
-------------+------------------------------           Adj R-squared =  0.9924
       Total |   3.8157365    14  .272552607           Root MSE      =   .0456

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |   .7268305   .1554418     4.68   0.001     .3847054    1.068956
        fuel |   .4515127   .0381103    11.85   0.000     .3676324    .5353929
        load |  -.7513069   .6105989    -1.23   0.244    -2.095226    .5926122
       _cons |   8.699815   .8985786     9.68   0.000     6.722057    10.67757
------------------------------------------------------------------------------
OLS regression for group 4

      Source |       SS       df       MS              Number of obs =      15
-------------+------------------------------           F(  3,    11) =  777.86
       Model |  7.37252558     3  2.45750853           Prob > F      =  0.0000
    Residual |  .034752343    11  .003159304           R-squared     =  0.9953
-------------+------------------------------           Adj R-squared =  0.9940
       Total |  7.40727792    14   .52909128           Root MSE      =  .05621

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |   .9353749   .0759266    12.32   0.000     .7682616    1.102488
        fuel |   .4637263    .044347    10.46   0.000     .3661192    .5613333
        load |  -.7756708   .4707826    -1.65   0.128    -1.811856    .2605148
       _cons |   9.164608   .6023241    15.22   0.000     7.838902    10.49031
------------------------------------------------------------------------------
OLS regression for group 5

      Source |       SS       df       MS              Number of obs =      15
-------------+------------------------------           F(  3,    11) = 1999.89
       Model |  7.08313716     3  2.36104572           Prob > F      =  0.0000
```

```
    Residual |  .012986435    11  .001180585           R-squared     =  0.9982
-------------+------------------------------           Adj R-squared =  0.9977
       Total |  7.09612359    14  .506865971           Root MSE      =  .03436

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |   1.076299   .0771255    13.96   0.000     .9065471    1.246051
        fuel |   .2920542   .0434213     6.73   0.000     .1964845    .3876239
        load |  -1.206847   .3336308    -3.62   0.004    -1.941163   -.4725305
       _cons |   11.77079   .7430078    15.84   0.000     10.13544    13.40614
------------------------------------------------------------------------------
OLS regression for group 6

      Source |       SS       df       MS              Number of obs =      15
-------------+------------------------------           F(  3,    11) = 2602.49
       Model |  11.1173565     3  3.70578551           Prob > F      =  0.0000
    Residual |  .015663323    11  .001423938           R-squared     =  0.9986
-------------+------------------------------           Adj R-squared =  0.9982
       Total |  11.1330199    14  .795215705           Root MSE      =  .03774

------------------------------------------------------------------------------
        cost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |   .9673393   .0321728    30.07   0.000     .8965275    1.038151
        fuel |   .3023258   .0308235     9.81   0.000     .2344839    .3701678
        load |   .1050328   .4767508     0.22   0.830    -.9442886    1.154354
       _cons |   10.77381   .4095921    26.30   0.000     9.872309    11.67532
------------------------------------------------------------------------------
```

8.2 Poolability Test across Groups

The null hypothesis of the poolability test across groups is $H_0 : \beta_{ik} = \beta_k$. The $e'e$ is 1.3354, the SSE of the pooled OLS regression. The $e_i'e_i$ is .1007 = .0068 + .0076 + .0229 + .0348 + .0130 + .0157.

The F statistic is $\dfrac{(1.3354 - .1007/(6-1)4}{.1007/6(15-4)} \sim 40.4812[20,66]$

The large 40.4812 rejects the null hypothesis of poolability (p< .0000). We conclude that the panel data are not poolable with respect to airline.

8.3 Poolability Test over Time

The null hypothesis of the poolability test over time is $H_0 : \beta_{tk} = \beta_k$. The sum of $e_t'e_t$ is computed from the 15 time by time regression.

```
forvalues i= 1(1)15 { // run year by year regression
    display "OLS regression for year " `i'
    regress cost output fuel load if year==`i'
}

(output is skipped)

. di .044807673 + .023093978 + .016506613 + .012170358 + .014104542 + ///
     .000469826 + .063648817 + .085430285 + .049329439 + .077112957 + ///
     .029913538 + .087240016 + .143348297 + .066075346 + .037256216

.7505079
```

The F statistic is $.4175[84,30] = \dfrac{(1.3354 - .7505)/(15-1)4}{.7505/15(6-4)}$

The small F statistic does not reject the null hypothesis in favor of poolable panel data with respect to time ($p<.9991$).

9. Conclusion

Panel data are analyzed to investigate group and time effects using fixed effect and random effect models. The fixed effect model asks how group and/or time affect the intercept, while the random effect model analyzes error variance structures affected by group and/or time. Slopes are assumed unchanged in both fixed effect and random effect models.

A panel data set needs to be arranged in the long format as shown in Section 1.1. If the number of groups (subjects) or time periods is extremely large, panel data models may be less useful because the null hypothesis of F test is too strong. Then, you may consider categorizing subjects to reduce the number of groups. If data are severely unbalanced, read output with caution and consider dropping subjects with many missing data points. This document assumes that data are balanced without missing values.

Fixed effect models are estimated by the least squares dummy variable (LSDV) regression and within effect model. LSDV has three approaches to avoid perfect multicollinearity. LSDV1 drops a dummy, LSDV2 suppresses the intercept, and LSDV3 includes all dummies and imposes restrictions instead. LSDV1 is commonly used since it produces correct statistics. LSDV2 provides actual parameter estimates of groups (Y-intercepts), but reports incorrect R^2 and F statistic. Notice that the dummy parameters of three LSDV approaches have different meanings and thus conduct different t-tests.

The within effect model does not use dummy variables but deviations from group means. Thus, this model is useful when there are many groups and/or time periods in the panel data set since it is able to avoid the incidental parameter problem. The dummy parameter estimates need to be computed afterward. Because of its larger degrees of freedom, the within effect model produces incorrect MSE and standard errors of parameters. As a result, you need to adjust the standard errors to conduct correct t-tests.

Random effect models are estimated by the generalized least squares (GLS) and the feasible generalization least squares (FGLS). When the variance structure is known, GLS is used. If unknown, FGLS estimates theta. Parameter estimates vary depending on estimation methods.

Fixed effects are tested by the F-test and random effects by the Breusch-Pagan Lagrange multiplier test. The Hausman specification test compares a fixed effect model and a random effect model. If the null hypothesis of uncorrelation is rejected, the fixed effect model is preferred. Poolabiltiy is tested by running group by group or time by time regressions.

Among the four statistical packages addressed in this document, I would recommend SAS and Stata. In particular, PROC PANEL provides various ways of analyzing panel data and report correct (adjusted) statistics (see Table 4.1 and 7.1). Stata is very handy to manipulate panel data reports incorrect F-test and R^2. LIMDEP is able to estimate various panel data models but does not good at data management. SPSS is least recommended for panel data models.

Extensions to these basic linear panel data models include dynamic models with autocorrelation, random coefficient model, and hierarchical linear model, and logit/probit models.

Appendix: Data Sets

Data set 1: Data of the top 50 information technology firms presented in *OECD Information Technology Outlook 2004* (http://thesius.sourceoecd.org/).

URL: http://www.indiana.edu/~statmath/stat/all/panel/rnd2002.csv
http://www.indiana.edu/~statmath/stat/all/panel/rnd2002.dta

firm = IT company name
type = type of IT firm
rnd = 2002 R&D investment in current USD millions
income = 2000 net income in current USD millions
d1 = 1 for equipment and software firms and 0 for telecommunication and electronics

```
. tab type d1
```

```
                |          d1
   Type of Firm |         0          1 |     Total
----------------+----------------------+----------
        Telecom |        18          0 |        18
    Electronics |        17          0 |        17
   IT Equipment |         0          6 |         6
Comm. Equipment |         0          5 |         5
  Service & S/W |         0          4 |         4
----------------+----------------------+----------
          Total |        35         15 |        50
```

```
. sum rnd income
```

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         rnd |        39    2023.564    1615.417          0       5490
      income |        50     2509.78    3104.585       -732      11797
```

Data set 2: Cost data for U.S. airlines (1970-1984) presented in Greene (2003).

URL: http://pages.stern.nyu.edu/~wgreene/Text/tables/tablelist5.htm
http://www.indiana.edu/~statmath/stat/all/panel/airline.dta

airline = airline (six airlines)
year = year (fifteen years)
output0 = output in revenue passenger miles, index number
cost0 = total cost in $1,000
fuel0 = fuel price
load = load factor, the average capacity utilization of the fleet

```
. sum output0 cost0 fuel0 load
```

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     output0 |        90    .5449946    .5335865    .037682    1.93646
       cost0 |        90     1122524     1192075      68978    4748320
       fuel0 |        90      471683    329502.9     103795    1015610
        load |        90    .5604602    .0527934    .432066    .676287
```

References

Baltagi, Badi H. 2001. *Econometric Analysis of Panel Data*. Wiley, John & Sons.

Baltagi, Badi H., and Young-Jae Chang. 1994. "Incomplete Panels: A Comparative Study of Alternative Estimators for the Unbalanced One-way Error Component Regression Model." *Journal of Econometrics*, 62(2): 67-89.

Breusch, T. S., and A. R. Pagan. 1980. "The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics." *Review of Economic Studies*, 47(1):239-253.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Cameron, A. Colin, and Pravin K. Trivedi. 2009. *Microeconometrics Using Stata*. TX: Stata Press.

Freund, Rudolf J., and Ramon C. Littell. 2000. *SAS System for Regression*, 3rd ed. Cary, NC: SAS Institute.

Fuller, Wayne A. and George E. Battese. 1973. "Transformations for Estimation of Linear Models with Nested-Error Structure." *Journal of the American Statistical Association*, 68(343) (September): 626-632.

Fuller, Wayne A. and George E. Battese. 1974. "Estimation of Linear Models with Crossed-Error Structure." *Journal of Econometrics*, 2: 67-78.

Greene, William H. 2003. *Econometric Analysis*, 5th ed. Upper Saddle River, NJ: Prentice Hall.

Greene, William H. 2007. *LIMDEP Version 9.0 Econometric Modeling Guide 1*. Plainview, New York: Econometric Software.

Hausman, J. A. 1978. "Specification Tests in Econometrics." *Econometrica*, 46(6):1251-1271.

SAS Institute. 2004. *SAS/ETS 9.1 User's Guide*. Cary, NC: SAS Institute.

SAS Institute. 2004. *SAS/STAT 9.1 User's Guide*. Cary, NC: SAS Institute.

SPSS Inc. 2007. *SPSS 16.0 Command Syntax Reference*. Chicago, IL: SPSS Inc.

Stata Press. 2007. *Stata Base Reference Manual, Release 10*. College Station, TX: Stata Press.

Stata Press. 2007. *Stata Longitudinal/Panel Data Reference Manual, Release 10*. College Station, TX: Stata Press.

Stata Press. 2007. *Stata Time-Series Reference Manual, Release 10*. College Station, TX: Stata Press.

Suits, Daniel B. 1984. "Dummy Variables: Mechanics V. Interpretation." *Review of Economics & Statistics* 66 (1):177-180.

Uyar, Bulent, and Orhan Erdem. 1990. "Regression Procedures in SAS: Problems?" *American Statistician* 44(4): 296-301.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Acknowledgements

I have to thank Dr. Heejoon Kang of the Kelley School of Business and Dr. David H. Good of the School of Public and Environmental Affairs, Indiana University at Bloomington. I am also grateful to Jeremy Albright, Dani Marinova, and Kevin Wilhite at the UITS Center for Statistical and Mathematical Computing for comments and suggestions. A special thanks to many readers around the world who have eagerly provided constructive feedback and encouraged me to keep improving this document.

Revision History

- 2005.11 First draft
- 2008.04, 11 Corrected some errors and added Stata examples
- 2009.09 Second draft (updated LSDV section and analysis output)