A Fixed Effects Panel Data Model: Mathematics Achievement in the U.S.

Todd Sherron **Jeff M. Allen**
University of North Texas

Statistical models that combine cross section and time series data offer analysis and interpretation advantages over separate cross section or time series data analyses (Mátyás & Severstre, 1996). Time series and cross section designs have not been commonplace in the research community until the last 25 years (Tieslau, 1999). In this study, a fixed effects panel data model is applied to the National Education Longitudinal Study of 1988 (NELS:88) data to determine if educational process variables, teacher emphasis, student self-concept, and socio-economic status can account for variance in student mathematical achievement. A model that includes seven independent variables accounted for 25% of the variance in student mathematical achievement test score. The study provides educational researchers with an applied model for panel data analysis.

Time series and cross section designs have not been commonplace in the research community until the last 25 years (Tieslau, 1999). In fact, the U.S. Department of Education's National Center for Education Statistics (NCES) was not mandated to "collect and disseminate statistics and other data related to education in the United States" until the Education Amendments of 1974 (Public Law 93-380, Title V, Section 501, amending Part A of the General Education Provisions Act). Researchers commonly have termed data that contains time series and cross section units to be panel or longitudinal data. In this study, these terms are used interchangeable. Essentially, panel data is a set of individuals who are repeatedly sampled at different intervals in time, across a multitude of cross sectional variables. The term "individual" might be used loosely to imply a person, a household, a school, school districts, firms, or a geographical region. Figure 1 provides a typical Panel data structure. Schools have been used to represent the different "individuals". (Note: the individual unit could just as well have been different schools within a particular district, school districts within a state, or an aggregate representation by state).

Researchers who are interested in understanding, explaining, or predicting variation within longitudinal data are faced with complex stochastic specifications. The problem that occurs when measures exhibit two-dimensional variation—variation across time and cross section, in model specification. In other words, researchers need to specify a model that can capture individual differences in behavior across individuals and/or through time for estimation and inference purposes (Greene, 1997). In general, longitudinal (panel) data sets contain a large number of cross-section units and a relatively small number of time-series units.

The U.S. Department of Education began collecting data in 1988 about critical transitions experienced by students as they leave elementary schools and progress through high school and into postsecondary institutions or the work force. The National Education Longitudinal Study of 1988 (NELS:88) contains data about educational processes and outcomes pertaining to student learning, predictors of dropping out, and school effects on students' access to programs and equal opportunities to learn. The first follow-up was conducted with the same students, their teachers, and principals in 1990. The second follow up survey was conducted in 1992, and the third in 1994. Data from NELS:88 will be used in this study to determine if student perception of educational process variables can account for the variance in mathematical achievement.

Model Specification

When should a fixed effects or random effects model be utilized? The answer to this question is often debated. Some believe that it is dependent upon the underlying cause in the model. For example, if the individual effects are the result of a large number of non-observable stochastic variables, then the random effect interpretation is demanded. Others think the decision rests on the nature of the sample – that is when the sample is comprehensive or exhaustive, then fixed effects models are the natural choice to enhance the generalizability. On the contrary, if the sample does not contain a large percent of the population then the random effects model would be the model of choice. According to Hasiao (1985), it is ultimately, " up to the investigator to decide whether he wants to make an inference with respect to population characteristics or only with respect to effects that are in the sample" (*p*. 131). It is unlikely that this debate will ever be resolved per se, however, if the choice between the two underlying methods is clear, then the estimation method should be chosen accordingly. However, if the choice is not clear, then the decision should be based on the nature of the sample and statistical evidence. For example, if the individual effects are significant then this is a sign that a significant component of the model is accounted for

$$Y_{it} = \begin{bmatrix} Y_{\text{school\#1, year\#1}} \\ Y_{\text{school\#1, year\#2}} \\ \vdots \\ Y_{\text{school\#1, year\#}T} \\ \cdots \\ Y_{\text{school\#2, year\#1}} \\ Y_{\text{school\#2, year\#2}} \\ \vdots \\ Y_{\text{school\#2, year\#}T} \\ \cdots \\ Y_{\text{school\#}N\text{, year\#1}} \\ Y_{\text{school\#}N\text{, year\#2}} \\ \vdots \\ Y_{\text{school\#}N\text{, year\#}T} \end{bmatrix} \quad \boldsymbol{X}_{it} = \begin{bmatrix} X1_{\text{school\#1, year\#1}} & X2_{\text{school\#1, year\#1}} & X_K{}_{\text{school\#1, year\#1}} \\ X1_{\text{school\#1, year\#2}} & X2_{\text{school\#1, year\#2}} & X_K{}_{\text{school\#1, year\#2}} \\ \vdots & \vdots & \vdots \\ X1_{\text{school\#1, year\#}T} & X2_{\text{school\#1, year\#}T} & X_K{}_{\text{school\#1, year\#}T} \\ \cdots & \cdots & \cdots \\ X1_{\text{school\#2, year\#1}} & X2_{\text{school\#2, year\#1}} & X_K{}_{\text{school\#2, year\#1}} \\ X1_{\text{school\#2, year\#2}} & X2_{\text{school\#2, year\#2}} & X_K{}_{\text{school\#1, year\#2}} \\ \vdots & \vdots & \vdots \\ X1_{\text{school\#2, year\#}T} & X2_{\text{school\#2, year\#}T} & X_K{}_{\text{school\#1, year\#}T} \\ \cdots & \cdots & \cdots \\ X1_{\text{school\#}N\text{, year\#1}} & X2_{\text{school\#}N\text{, year\#1}} & X_K{}_{\text{school\#}N\text{, year\#1}} \\ X1_{\text{school\#}N\text{, year\#2}} & X2_{\text{school\#}N\text{, year\#2}} & X_K{}_{\text{school\#}N\text{ year\#2}} \\ \vdots & \vdots & \vdots \\ X1_{\text{school\#}N\text{, year\#}T} & X2_{\text{school\#}N\text{, year\#}T} & X_K{}_{\text{school\#}N\text{, year\#}T} \end{bmatrix}$$

Figure 1. Panel Data Structure

by the individual effects parameter and so fixed effects might be preferred over random effects. However, the Hausman test statistic, a statistic designed to test model fit, can be used to determine when a random effects model is preferred, i.e. a large Hausman test statistic indicates a random effects components (Greene, 1997).

The Fixed-Effects Model

The fixed effects (FE) model takes α_i to be a group specific constant term in the regression equation

$$Y_{it} = \alpha_i + \beta_1 X_{1it} + \beta_2 X_{2it} \ldots + \beta_K X_{Kit} + \varepsilon_{it} \quad (1)$$

or in matrix notation

$$Y_{it} = \alpha_i + \boldsymbol{X}_{it}'\beta + \varepsilon_{it} \quad (2)$$

where $\boldsymbol{X}_{it}' = [X_{1it}, X_{2it}, \ldots, X_{Kit}]$
and $\beta' = [\beta_1, \beta_2, \ldots, \beta_K]$.

The "*i*" indexes cross-section realizations so that $i = 1,2,3, \ldots, N$ and "*t*" indexes time-series realizations so that $t = 1, 2,3, \ldots, T$. The *individual effect* α_i, is regarding to be constant over time (*t)* and specific to the individual cross-sectional unit (*i)*. The term α_i is presumed to capture the unobservable, and non-measurable characteristics that differentiate individual units. Basically, this implies that all behavioral differences between individuals (e.g., schools in Figure 1) are fixed over time and are represented as parametric shifts of the regression function. Mátayás and Sevestre (1996, *p*. 34) state, "the intercept is allowed to vary from individual to individual while the slope parameters are assumed to be constants in both the individual and time dimensions".

The fundamental assumption of the fixed effects model are:

$$E[\varepsilon_{it}] = 0,$$

$$\text{Cov}(\varepsilon_{it}, \varepsilon_{jt}) = 0,$$

$$\text{Var}(\varepsilon_{it}) = E[\varepsilon^2_{it}] = \sigma^2_e,$$

$E[\varepsilon_{it}, X_{1it}] = E[\varepsilon_{it}, X_{2it}] = \ldots = E[\varepsilon_{it}, X_{Kit}] = 0,$
and X_{kit} is not invariant.

Under these assumptions, the ordinary least squared estimator (OLS) can be use to obtain unbiased, consistent, and efficient (BLUE) parameter estimates.

The Random Effects Model

The random effects (RE) model—also know as the error component model, includes a non-measurable stochastic variable, which differentiates individuals. It is written as:

$$Y_{it} = \alpha_i + \beta_1 X_{1it} + \beta_2 X_{2it} \ldots + \beta_K X_{Kit} + u_i + \varepsilon_{it} \quad (3)$$

or in matrix notation

$$Y_{it} = \alpha_i + \boldsymbol{X}_{it}'\beta + u_i + \varepsilon_{it} \quad (4)$$

where $\boldsymbol{X}_{it}' = [X_{1it}, X_{2it}, \ldots, X_{Kit}]$

and $\beta' = [\beta_1, \beta_2, \ldots, \beta_K]$.

The "*i*" indexes cross-section realizations so that $i=1,2,3,\ldots, N$ and "*t*" indexes time-series realizations so that $t = 1,2,3,\ldots, T$. The term "u_i" is a stochastic

variable that embodies the *unobservable* or *non-measurable* disturbances that accounts for individual differences. Essentially, the effect is thought to be a random individual effect rather that fixed parameter. For example, a researcher might try to discern whether there is difference in achievement between districts in the State of Texas. Instead of including every school district in the equation (as we would have in the fixed effects model using dummy variables) one can randomly sample school districts and assumes that the effect is random distributed across "individuals" but constant through time.

The fundamental assumptions of the random effects model are as such:

$$E[u_i, X_{1it}] = E[u_i, X_{2it}] = \ldots = E[u_i, X_{Kit}] = 0,$$

$$E[\varepsilon_{it}] = [u_i] = 0,$$

$$\text{Var}(u_i) = E[u^2_i] = \sigma^2_u, \text{ and}$$

$$\text{Cov}(u_i, \varepsilon_{it}) = E[u_i, \varepsilon_{it}] = \sigma_{\varepsilon,u},$$

Assuming normality $u_i \sim N(0, \sigma^2_u)$, $\varepsilon_{it} \sim N(0, \sigma^2_e)$, both "$u_i$" and "$\varepsilon_{it}$" are stochastic variables, but form one composite error term-called omega $(u_i+\varepsilon_{it}) \equiv \omega_{it}$,

$$\text{where} \quad \omega_{it} = \begin{bmatrix} u_1 + \varepsilon_{11} \\ u_1 + \varepsilon_{12} \\ \vdots \\ u_1 + \varepsilon_{1T} \\ \ldots \\ u_N + \varepsilon_{N1} \\ u_N + \varepsilon_{N2} \\ \vdots \\ u_N + \varepsilon_{NT} \end{bmatrix}$$

The error term now consist of two components: (1) the error disturbance ε_{it}, and (2) the individual specific disturbance u_i,. The RE model now takes the form of

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} \ldots + \beta_K X_{Kit} + \omega_{it} \quad (5)$$

or in matrix notation

$$Y_{it} = \alpha + \boldsymbol{X}_{it}'\beta + \omega_{it} \quad (6)$$

The error term in the model now exhibits the following characteristics:

$$\text{Var}(\omega_{it}) = \begin{bmatrix} \sigma^2_e & \sigma_{\varepsilon,u} \\ \sigma_{u,\varepsilon} & \sigma^2_u \end{bmatrix}$$

The OLS estimator can not be applied to equation 6 because the error term not longer possess ideal properties (constant variance and zero covariance) thus the estimate would be inefficient and, hence generalized least squares (GLS) is appropriate. However, the nature of data in behavioral sciences does not permit the variance components σ^2_u and σ^2_e to be known, therefore, alternative estimation methods must be utilized. One common estimation method that can deal with the unknown variance components feasible generalized least squares (FGLS). FGLS takes an estimate of the variance components and then estimates the equation.

The individual effect in the random effect model may, too, be tested with the following hypotheses:

$$H_0: u_i = 0, \text{ or equivalently, } \sigma^2_u = 0$$
$$H_a: \sigma^2_u \neq 0.$$

After correcting the error term (ω_{it}) the *t* and *F*-test are reliable, thus inference can be regarded as valid. Based on statistical evidence, a FE model will be used in this paper.

National Education Longitudinal Data Set:88

The NELS:88 database is divided into two sections: (1) N2P, and (2) N4P. In this study, data were extracted from N2P. A representative sample of students (N=16,749) enrolled in tenth grade in the spring of 1990, who completed a questionnaire in both the first follow-up and second follow-up, were identified and used in the analysis. The LIMDEP program (Greene, 1992) and output are in the appendix.

Seven independent variables are included in the specified model. They are listed as entered into the model: (1) Review Work (F2S19BA), (2) ListenLecture (F2S19BB), (3) CopyNotes (F2S19BC), (4) Calculators (F2S19BF), (5) Think Problem (F2S20D), (6) SES (F2SES1),and (7) Self concept (F2CNCPT). The first four variables, (ReviewWork, ListenLecture, Copynotes, Calculators), are frequency measures of student educational processes and are scaled as followed: (1) Never/Rarely, (2) 1-2 Times/Month, (3) 1-2 Times/Week, (4) Almost each day, (5) Every Day. For example, the variable ReviewWork is a measure of how frequently students review their work for the previous day. The variable ListenLecture is a measure of how frequently students listen to the teacher's lecture. The variable CopyNotes is a measure of how frequently students take notes. The variable Calculator is a measure of how frequently students use calculators. The variable ThinkProblem measures student perception of teachers emphases on mathematical objectives and is scaled, (0) none, (1) minor emphasis, (2) moderate emphasis, (3) major emphasis. SES is a continuous variable

Table 1. Estimated Fixed Effects

Group	Coefficient	Standard Error	*t*-ratio
North East	45.84312*	0.42968	106.69
North Central	45.48305*	0.44005	103.36
South	43.06207*	0.42588	101.11
West	44.10279*	0.44472	99.17

Note. *Statistically significant at the $p < .01$.

indicating socioeconomic status. This measure is based on Duncan's (1961) socioeconomic index for all occupations. It was derived from the parent questionnaire data, the student questionnaire data, or the first follow-up or second follow-up New Student Supplement data. This variable has been standardized to have a mean of 0 and standard deviation of 1.

Selfconcept is a composite measure of all of the self-concept items (question 66) in the student questionnaire. Essentially, this variable measures students' self concept on a four point scale with: (1) strongly agree ,(2) agree ,(3) disagree,(4) strongly disagree. It should be noted that this variable was reverse scaled before a composite score was created and was standardized to have mean of zero and a standard deviation of 1. MathScore is the dependent variable and it was derived by Item response theory (IRT) to have a mean of 50 and standard deviation of 10.

Empirical Results

The FE model below was specified and estimated.

$$(Mathscore)_{it} = \beta_0 + \beta_1(ReviewWork)_{it} + \beta_2(ListenLecture)_{it} + \beta_3(CopyNotes)_{it} + \beta_4(Calculators)_{it} + \beta_5(ThinkProblem)_{it} + \beta_6(SES)_{it} + \beta_7(Selfconcept)_{it} + \varepsilon_{ty}$$

Table 1 provides evidence that the FE model is indeed the correct choice over the RE model; all *t*-values are significant. Region is the cross section unit (i = 1, 2, 3, 4) indicating which of the four US Census regions (1) Northeast, (2) Midwest,(3) South, or (4) West.

Table 2 provides the descriptive statistics, measure of central tendency, measure of dispersion, minimum and maximum, and number of cases. Table 3 provides the correlation coefficients for all of the variables used in the analysis.

Six out of the seven independent variables were statistically significant at the $p < 0.0001$ alpha level accounting for 25% of the variance in the dependent variable (mathematics achievement score). See Table 4 for parameter estimates.

The variable ReviewWork is statistically significant ($t = 3.27$, $p < .001$). As ReviewWork increases by one unit, MathScore increases by 0.234 points. In other words, as students increase the frequency in which they review their work, holding all else constant, their math score increases by 0.234 points.

Table 2. Descriptive Statistics

Variable	Mean	SD	Min	Max
MathScore (N = 12,992)	51.81	9.93	29.50	71.49
ReviewWork (N = 13,577	3.87	1.21	1.00	5.00
ListenLecture (N = 13,565)	4.24	1.03	1.00	5.00
CopyNotes (N = 13,565)	4.01	1.27	1.00	5.00
Calculators (N = 13,560)	3.69	1.38	1.00	5.00
ThinkProblem (N = 13,568)	2.23	0.84	0	3.00
SES (N = 16,563)	0.04	0.81	-3.24	2.75
Selfconcept (N = 15,123)	0.01	0.70	-3.69	1.24
Region (N = 16,426)	2.56	1.01	1.00	4.00

The variable ListenLecture is statistically significant ($t = 3.924$, $p < .001$). As ListenLecture increases by one unit, MathScore increases by 0.370 points. Or put differently, the more attentive the student is to the teacher's lecture, their math score increases by 0.370 points. The variable CopyNotes is not statistically significant ($t = -1.488$, $p < .1367$).

The variable Calculators is statistically significant ($t = 11.10$, $p < .001$). As Calculators increases by one unit, MathScore increases by 0.667 points. Essentially, this estimate is showing that students math score will increase with the use of a calculator.

The variable ThinkProblem is statistically significantly ($t = 14.304$, $p < .001$). Recall, this variable measures student perception of teachers emphasis on mathematical objectives and is scaled, (0) none, (1) minor emphasis, (2) moderate emphasis, (3) major emphasis. As ThinkProblm increases by 1 unit, MathScore increases by 1.387 point. Abstracted differently, the more teachers' emphasize "thinking about what a problem means and ways it might be solved", holding all else constant, students math score increases by 1.387 points.

The variable SES is statistically significant ($t = 47.333$, $p < 0.001$). As SES increases by 1 unit, holding all else constant, MathScore increases by 4.937 points. Recall SES is a continuous variable indicating member's socioeconomic status. This measure is based on Duncan's 1961 socioeconomic Index for all Occupations.

The variable SelfConcept is statistically significant ($t = 10.071$, $p < 0.001$). As SelfConcept

Table 3. Correlation Matrix

	Math Score	Review Work	Listen Lecture	CopyNotes	Calculators	Think Problem	SES	Self Concept
Review Work	.099	1.000						
ListenLecture	.119	.369	1.000					
CopyNotes	.103	.278	.535*	1.000				
Calculators	.148	.135	.140	.095	1.000			
Think Problem	.181	.224	.256	.237	.099	1.000		
SES	.470	.042	.044	.084	.103	.050	1.000	
Self Concept	.144	.085	.085	.076	.053	.134	.081	1.000
Region	-.103	.009	.022	-.031	.074	-.009	-.082	.002

Note. The moderate correlation between the two variables, CopyNotes and ListenLecture $r = .535$ is indicative of multicolinearity. This correlation gives reason to question the inference drawn from the *t*-ratio values on these two variables, however, the parameter estimates for CopyNotes and ListenLecture are still the best least square estimates.

increases by 1 unit, MathScore increases by 1.137 points. Essentially, students who have a more positive self perception, are scoring higher on the standardized math test.

Conclusions

In this study, a fixed effects panel data model were applied to the National Education Longitudinal Study of 1988 (NELS:88). The empirical evidence presented here suggests that student mathematics test score is influenced by educational process variables, teacher emphasis, student self-concept, and socio-economic status. Specifically, a model that included seven independent variables accounted for 25% of the variance in student mathematical achievement test score.

Caveat

The NELS:88 data set does not have a means of extracting the time component in the data. Although, models for analyzing time effects were not discussed in the study, it is an important aspect of panel data that should be coded when the data file is constructed. In addition, the time series unit should be measured in smaller periods of time. Residual analysis should be performed on the error term. That is, the error term should be analyzed for heteroscedasticity and autocorrelation.

Correspondence should be directed to:
Todd Sherron
Educational Research Lab
University of North Texas
Denton, Texas 76203-1337
E-mail: sherron@coefs.coe.unt.edu

Table 4. Fixed Effects Estimates

Variable	Coefficient	SE	*t*	*p*-value
ReviewWork	0.234	0.072	3.27	0.0011
ListenLecture	0.370	0.094	3.92	0.0001
CopyNotes	-0.190	0.073	-1.49	0.1367
Calculators	0.667	0.601	11.10	0.0000
ThinkProblem	1.388	0.097	14.30	0.0000
SES	4.938	0.103	47.73	0.0000
Selfconcept	1.137	0.113	10.07	0.0000
$R^2 = 0.246$			$R^2adj = 0.245$	

References

Green, W. (Ed.). (1997). Econometric analysis (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Greene, W. (1992). *LIMDEP User's Manual and Reference Guide*. New York: Econometric Software.

Griffiths, W., Hill, R., & Judge G., (1993). *Learning and Practicing Econometrics*. New York: Wiley.

Hasiao, C. (1985). Benefits and limitations of panel data, *Econometric Reviews, 4*(1), 121-174.

Judge, G., Griffiths, W., Hill, R., Lütkepohl, H., & Lee, T. (1985). *The theory and practice of econometrics* (2nd ed.). New York: Wiley.

Mátayás, L., & Sevestre, P. (Eds.). (1996). *The econometrics of panel data: A handbook of the theory with applications* (2nd ed.). Dordrecht, Netherlands: Kluwer Academic Publishers.

Tieslau, M. (1999) *Empirical Econometrics 5670*. Unpublished manuscript, University of North Texas.

Appendix

LimDep Code

```
READ;File=C:\WINDOWS\Program Files\ES Limdep\PROGRAM\nels6.lpj;
   Nobs=16749;
   Nvar=23;
   Names=x1,x2,x3,x4,x5,x6,x7,x8,x9,x10,x11,x12,x13,x14,x15,x16,x17,x18,
      x19,x20,x21,x22,x23$

SKIP$

DSTATS; RHS = X2,X3,X4,X7,X13,X19,X20,X23; OUTPUT = 2 $

REGRESS;Lhs=X22
    ;Rhs=X2,X3,X4,X7,X13,X19,X20
       ;Str=X23
       ;Wts=X16
       ;Panel $

REGRESS;Lhs=X22
    ;Rhs=X2,X3,X4,X7,X13,X19,X20
       ;Str=X23
       ;Panel
    ;Output=2
       ;Wts=X16
       ;Fixed $
```

Data Output

```
--> SKIP$
--> DSTATS; RHS = X2,X3,X4,X7,X13,X19,X20,X23; OUTPUT = 2 $

                          Descriptive Statistics
             All results based on nonmissing observations.
Variable        Mean         Std.Dev.        Minimum         Maximum      Cases
------------------------------------------------------------------------------
X2        3.87029535      1.21401018      1.00000000      5.00000000      13577
X3        4.24061924      1.02966250      1.00000000      5.00000000      13565
X4        4.00906612      1.27206705      1.00000000      5.00000000      13567
X7        3.69041298      1.37809060      1.00000000      5.00000000      13560
X13       2.23349057      .839799327      .000000000      3.00000000      13568
X19       .485648735E-01  .811172698     -3.24000000      2.75000000      16563
X20       .111955300E-01  .701701693     -3.69000000      1.24000000      15123
X23       2.55777426      1.01482683      1.00000000      4.00000000      16426


Correlation Matrix for Listed Variables

              X2       X3       X4       X7      X13      X19      X20      X23
     X2  1.00000   .38402   .28129   .13514   .22252   .04671   .08881   .00758
     X3   .38402  1.00000   .54066   .13636   .26014   .04523   .08992   .01662
     X4   .28129   .54066  1.00000   .09404   .23307   .07737   .07659  -.03424
     X7   .13514   .13636   .09404  1.00000   .10102   .10769   .05500   .07762
    X13   .22252   .26014   .23307   .10102  1.00000   .05100   .14062  -.01871
    X19   .04671   .04523   .07737   .10769   .05100  1.00000   .08907  -.08566
    X20   .08881   .08992   .07659   .05500   .14062   .08907  1.00000   .01745
    X23   .00758   .01662  -.03424   .07762  -.01871  -.08566   .01745  1.00000
```

```
--> REGRESS;Lhs=X22
       ;Rhs=X2,X3,X4,X7,X13,X19,X20
    ;Str=X23
    ;Wts=X16
    ;Panel $
```

```
+----------------------------------------------------------------------------+
| OLS Without Group Dummy Variables                                          |
| Ordinary    least squares regression    Weighting variable = X16           |
| Dep. var. = X22      Mean=   52.55493863    , S.D.=   9.343603189          |
| Model size: Observations =   10895, Parameters =   8, Deg.Fr.=  10887       |
| Residuals:  Sum of squares= 731652.8662    , Std.Dev.=        8.19782       |
| Fit:        R-squared=  .230712, Adjusted R-squared =          .23022       |
| Model test: F[  7,  10887] =  466.44,    Prob value =          .00000       |
| Diagnostic: Log-L = -38376.9800, Restricted(b=0) Log-L =  -39805.8044       |
|             LogAmemiyaPrCrt.=    4.208, Akaike Info. Crt.=      7.046       |
| Panel Data Analysis of X22        [ONE way]                                |
|           Unconditional ANOVA (No regressors)                              |
| Source      Variation        Deg. Free.     Mean Square                    |
| Between       20751.0               3.         6917.00                    |
| Residual      930327.           10891.         85.4216                    |
| Total         951078.           10894.         87.3029                    |
+----------------------------------------------------------------------------+
+---------+-------------+----------------+---------+----------+------------+
|Variable | Coefficient | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X |
+---------+-------------+----------------+---------+----------+------------+
 X2        .2150986215      .72295926E-01    2.975   .0029  3.8895953
 X3        .3746630995      .95235646E-01    3.934   .0001  4.2207458
 X4       -.1516385665      .73563678E-01   -2.061   .0393  3.9716586
 X7        .7025006784      .59425942E-01   11.821   .0000  3.7239903
 X13       1.398697309      .97973819E-01   14.276   .0000  2.2173183
 X19       5.040076855      .10413200       48.401   .0000  .86701895E-01
 X20       1.035652953      .11377009        9.103   .0000  .42022957E-01
 Constant  44.54122268      .41506161      107.312   .0000
```

```
+----------------------------------------------------------------------------+
| Least Squares with Group Dummy Variables                                   |
| Ordinary    least squares regression    Weighting variable = X16           |
| Dep. var. = X22      Mean=   52.55493863    , S.D.=   9.343603189          |
| Model size: Observations =   10895, Parameters =  11, Deg.Fr.=  10884       |
| Residuals:  Sum of squares= 716924.2581    , Std.Dev.=        8.11601       |
| Fit:        R-squared=  .246198, Adjusted R-squared =          .24551       |
| Model test: F[ 10,  10884] =  355.48,    Prob value =          .00000       |
| Diagnostic: Log-L = -38266.1998, Restricted(b=0) Log-L =  -39805.8044       |
|             LogAmemiyaPrCrt.=    4.189, Akaike Info. Crt.=      7.027       |
| Estd. Autocorrelation of e(i,t)    -.000540                                |
+----------------------------------------------------------------------------+
+---------+-------------+----------------+---------+----------+------------+
|Variable | Coefficient | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X |
+---------+-------------+----------------+---------+----------+------------+
 X2        .2342486655      .71596894E-01    3.272   .0011  3.8895953
 X3        .3704126069      .94402933E-01    3.924   .0001  4.2207458
 X4       -.1089542352      .73215840E-01   -1.488   .1367  3.9716586
 X7        .6668605394      .60076332E-01   11.100   .0000  3.7239903
 X13       1.387686017      .97011428E-01   14.304   .0000  2.2173183
 X19       4.937094032      .10343209       47.733   .0000  .86701895E-01
 X20       1.136968803      .11289363       10.071   .0000  .42022957E-01
```

```
+---------------------------------------------------------------------------+
|                Test Statistics for the Classical Model                    |
|                                                                           |
|        Model            Log-Likelihood    Sum of Squares    R-squared     |
| (1)  Constant term only   -39805.80420   .9510780165D+06     .0000000     |
| (2)  Group effects only   -39685.63264   .9303270109D+06     .0218184     |
| (3)  X - variables only   -38376.97986   .7316528662D+06     .2307120     |
| (4)  X and group effects  -38266.19962   .7169242581D+06     .2461983     |
|                                                                           |
|                              Hypothesis Tests                             |
|              Likelihood Ratio Test                F Tests                 |
|          Chi-squared   d.f.  Prob.         F    num. denom. Prob value   |
| (2) vs (1)   240.343      3     .00000    80.975    3 10891     .00000     |
| (3) vs (1)  2857.649      7     .00000   466.435    7 10887     .00000     |
| (4) vs (1)  3079.209     10     .00000   355.481   10 10884     .00000     |
| (4) vs (2)  2838.866      7     .00000   462.825    7 10884     .00000     |
| (4) vs (3)   221.560      3     .00000    74.534    3 10884     .00000     |
+---------------------------------------------------------------------------+
REGR;PANEL. Could not invert VC matrix for Hausman test.


            +-------------------------------------------------------+
            | Random Effects Model: v(i,t) = e(i,t) + u(i)          |
            | Estimates:  Var[e]              =   .658696D+02       |
            |             Var[u]              =   .489455D+01       |
            |             Corr[v(i,t),v(i,s)] =   .069167           |
            | Lagrange Multiplier Test vs. Model (3) = 6971.10      |
            | ( 1 df, prob value =  .000000)                        |
            | (High values of LM favor FEM/REM over CR model.)      |
            | Fixed vs. Random Effects (Hausman)     =     .23      |
            | ( 7 df, prob value = 1.000000)                        |
            | (High (low) values of H favor FEM (REM).)             |
            | Reestimated using GLS coefficients:                   |
            | Estimates:  Var[e]              =   .663837D+02       |
            |             Var[u]              =   .499445D+01       |
            |             Sum of Squares          .729365D+06       |
            |             R-squared               .233117D+00       |
            +-------------------------------------------------------+
+---------+-------------+----------------+---------+----------+------------+
|Variable | Coefficient | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X  |
+---------+-------------+----------------+---------+----------+------------+
 X2        .2341713422      .71596797E-01    3.271   .0011    3.8895953
 X3        .3703833036      .94402278E-01    3.923   .0001    4.2207458
 X4       -.1090676626      .73213889E-01   -1.490   .1363    3.9716586
 X7        .6669200976      .60069733E-01   11.102   .0000    3.7239903
 X13       1.387746607      .97011341E-01   14.305   .0000    2.2173183
 X19       4.937585087      .10343032       47.738   .0000    .86701895E-01
 X20       1.136524709      .11289247       10.067   .0000    .42022957E-01
 Constant  44.62217113      1.1804686       37.800   .0000

--> REGRESS;Lhs=X22
    ;Rhs=X2,X3,X4,X7,X13,X19,X20
    ;Str=X23
    ;Panel
       ;Output=2
    ;Wts=X16
    ;Fixed $
```

```
+----------------------------------------------------------------------------+
| OLS Without Group Dummy Variables                                          |
| Ordinary    least squares regression    Weighting variable = X16           |
| Dep. var. = X22      Mean=   52.55493863    , S.D.=   9.343603189          |
| Model size: Observations =   10895, Parameters =   8, Deg.Fr.=  10887       |
| Residuals:  Sum of squares= 731652.8662    , Std.Dev.=        8.19782      |
| Fit:        R-squared=  .230712, Adjusted R-squared =          .23022      |
| Model test: F[  7,  10887] =  466.44,    Prob value =          .00000      |
| Diagnostic: Log-L = -38376.9800, Restricted(b=0) Log-L =  -39805.8044      |
|             LogAmemiyaPrCrt.=    4.208, Akaike Info. Crt.=      7.046      |
| Panel Data Analysis of X22        [ONE way]                                |
|           Unconditional ANOVA (No regressors)                              |
| Source      Variation        Deg. Free.     Mean Square                    |
| Between       20751.0               3.         6917.00                    |
| Residual      930327.           10891.         85.4216                    |
| Total         951078.           10894.         87.3029                    |
+----------------------------------------------------------------------------+
+---------+-------------+----------------+---------+----------+------------+
|Variable | Coefficient | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X |
+---------+-------------+----------------+---------+----------+------------+
 X2        .2150986215      .72295926E-01    2.975   .0029   3.8895953
 X3        .3746630995      .95235646E-01    3.934   .0001   4.2207458
 X4       -.1516385665      .73563678E-01   -2.061   .0393   3.9716586
 X7        .7025006784      .59425942E-01   11.821   .0000   3.7239903
 X13       1.398697309      .97973819E-01   14.276   .0000   2.2173183
 X19       5.040076855      .10413200       48.401   .0000   .86701895E-01
 X20       1.035652953      .11377009        9.103   .0000   .42022957E-01
 Constant  44.54122268      .41506161      107.312   .0000

+----------------------------------------------------------------------------+
| Least Squares with Group Dummy Variables                                   |
| Ordinary    least squares regression    Weighting variable = X16           |
| Dep. var. = X22      Mean=   52.55493863    , S.D.=   9.343603189          |
| Model size: Observations =   10895, Parameters =  11, Deg.Fr.=  10884       |
| Residuals:  Sum of squares= 716924.2581    , Std.Dev.=        8.11601      |
| Fit:        R-squared=  .246198, Adjusted R-squared =          .24551      |
| Model test: F[ 10,  10884] =  355.48,    Prob value =          .00000      |
| Diagnostic: Log-L = -38266.1998, Restricted(b=0) Log-L =  -39805.8044      |
|             LogAmemiyaPrCrt.=    4.189, Akaike Info. Crt.=      7.027      |
| Estd. Autocorrelation of e(i,t)    -.000540                                |
+----------------------------------------------------------------------------+
+---------+-------------+----------------+---------+----------+------------+
|Variable | Coefficient | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X |
+---------+-------------+----------------+---------+----------+------------+
 X2        .2342486655      .71596894E-01    3.272   .0011   3.8895953
 X3        .3704126069      .94402933E-01    3.924   .0001   4.2207458
 X4       -.1089542352      .73215840E-01   -1.488   .1367   3.9716586
 X7        .6668605394      .60076332E-01   11.100   .0000   3.7239903
 X13       1.387686017      .97011428E-01   14.304   .0000   2.2173183
 X19       4.937094032      .10343209       47.733   .0000   .86701895E-01
 X20       1.136968803      .11289363       10.071   .0000   .42022957E-01

        Estimated Fixed Effects
        Group       Coefficient       Standard Error       t-ratio
            1          45.84312               .42968     106.69106
            2          45.48305               .44005     103.35884
            3          43.06207               .42588     101.11422
            4          44.10279               .44472      99.16970
```

Test Statistics for the Classical Model			
Model	Log-Likelihood	Sum of Squares	R-squared
(1) Constant term only	-39805.80420	.9510780165D+06	.0000000
(2) Group effects only	-39685.63264	.9303270109D+06	.0218184
(3) X - variables only	-38376.97986	.7316528662D+06	.2307120
(4) X and group effects	-38266.19962	.7169242581D+06	.2461983

Hypothesis Tests						
	Likelihood Ratio Test			F Tests		
	Chi-squared	d.f.	Prob.	F	num. denom.	Prob value
(2) vs (1)	240.343	3	.00000	80.975	3 10891	.00000
(3) vs (1)	2857.649	7	.00000	466.435	7 10887	.00000
(4) vs (1)	3079.209	10	.00000	355.481	10 10884	.00000
(4) vs (2)	2838.866	7	.00000	462.825	7 10884	.00000
(4) vs (3)	221.560	3	.00000	74.534	3 10884	.00000

Multiple Linear Regression Viewpoints **needs your submissions. See the inside Back cover for submission details and for information on how to join the MLR: GLM SIG and get *MLRV*.**