Panel Data Models
Exercises

Fixed and Random Effects

Define the two basic approaches to modeling unobserved, time invariant effects in panel data. What are the different assumptions that are made in the two settings? What is the benefit of the fixed effects assumption? What is the cost? Same for the random effects specification. Now, consider the possibility that the unobserved effects are not time invariant.? How does your answer change?

Solution

The two approaches are fixed effects and random effects. In the "effects model,"

$$y_{it} = \mathbf{x}_{it}'\beta + \alpha_i + \varepsilon_{it}$$

where x_{it} is exogenous with respect to ε_{it}.

FE: α_i may be correlated with x_{it}.
Benefits: General approach,
Robust – estimator of β is consistent even if RE is the right model.
Cost: Many parameters, inefficient if RE is correct.
Precludes time invariant variables.

RE: α_i is uncorrelated with x_{it}
Benefits: Tight parameterization – only one new parameter
Efficient estimation – use GLS
Allows for time invariant parameters
Cost: Unreasonable orthogonality assumption
Inconsistent if FE is the right model.

Random parameters case. Replace the model statement with

$$y_{it} = \mathbf{x}_{it}'\beta_i + \alpha_i + \varepsilon_{it}$$

with $\beta_i = \beta + w_i$.

- **Case 1**: w_i may be correlated with x_{it}. This is the counterpart to FE. In this case, it is necessary to fit the equations one at a time. Requires that there be enough observations to do so, so $T > K$. The efficient estimator is equation by equation OLS. Same benefits (robustness) and costs (inefficiency) as FE.
- **Case 2**: w_i is uncorrelated with x_{it}. This RP model can be fit. An efficient estimator will be the matrix weighted FGLS estimator. (Swamy et al.) This would be a two step estimator, just like FGLS for the RE model. This model can also be fit by simulation.

If the unobserved heterogeneity is time varying, then taking deviations from means will not remove it from the model. Returning to the model specification, we now have

$$y_{it} = \beta'\mathbf{x}_{it} + \alpha_{it} + \varepsilon_{it}$$

If α_{it} is uncorrelated with x_{it} then it can be simply added to the disturbance in the model, and the model becomes a simple linear regression that can be fit by OLS. This is the RE case. In the FE case in which α_{it} is correlated with x_{it} we have a classic left out variable problem, and there is no way to proceed.

Dynamic Model

Consider the dynamic, linear, cross country, random effects regression model

$$y_{it} = \alpha + \mathbf{x}_{it}'\beta + \delta_i z_{it} + \gamma y_{i,t-1} + u_i + \varepsilon_{it},\ t = 1, 2, 3, 4$$

and $y_{i,0}$ is observed data, i is a country and t is a year; y_{it} is national income per capita, z_{it} is domestic investment and x_{it} is a measure of national labor input. You have 30 countries and 4 years of data. Note that the coefficient on z_{it} is allowed to differ across countries.

1. Assuming for the moment that δ_i is constant across countries, show that the pooled ordinary least squares estimator is inconsistent.

2. Continuing to assume that δ_i is the same for all countries, show two approaches, (i) Anderson and Hsiao and (i) Hausman and Taylor, could be used to obtain consistent estimators of β, δ and γ.

3. Let $w_{it} = (y_{it} - \alpha - \beta x_{it} - \delta z_{it} - \gamma y_{i,t-1})$. Consider the set of instruments $f_{it} = (1, x_{it}, z_{it}, x_{i,t-1}, z_{i,t-1})$.

   (a) Does the simple strategy of pooling the panel and simply using two stage least squares with $\mathbf{F}$ as the set of instruments produce a consistent estimator of the parameters? Explain.

   (b) I propose to use a GMM estimator based on the moment conditions corresponding to $E[f_{it}w_{it}] = 0,\ t = 2, 3, 4$. Describe in detail how the GMM estimator will proceed. How will this differ from the estimator in part (3a)?

   (c) Suppose I extend the strategy in (b) by further assuming "strict exogeneity," that is, $E[f_{it}w_{is}] = 0, t = 2, 3, 4$ and $s = 2, 3, 4$. How does this change the computations in (b)? (Note and hint: the constant term in f_{it} creates some redundant moment conditions. E.g., $(1/n)\sum_i f_{i4}w_{i4} = 0$ and $(1/n)\sum_i f_{i3}w_{i4} = 0$, both include a term that is $(1/30)\sum_i w_{i4} = 0$. For purposes of your answer to this question, ignore this fact – in practice, it would be necessary to reduce the set of moment conditions appropriately.)

4. Now allowing δ_i to differ across countries, comment on the consistency of the estimator you used in part (3a). Is it consistent? Can you propose a consistent estimator of this model when δ_i varies across countries?

Solutions

1. Assuming δ_i is constant across countries, the regression is a linear model in which one of the independent variables, $y_{i,t-1}$ is correlated with the disturbance, $w_{it} = (u_i + \varepsilon_{it})$. u_i is part of the disturbance in the equation for $y_{i,t-1}$ as well. So, this is a familiar case of an endogenous variable – OLS is inconsistent.

2. In the Anderson and Hsiao approach, we can use an instrumental variable estimator, as usual. There are many available instruments using lagged values of x_{it} and z_{it}, say $(x_{i,t-1}, z_{i,t-1})$, or additional lags. A&H suggested taking first differences. $\Delta y_{it} = \beta(\Delta x_{it}) + \delta(\Delta z_{it}) + \gamma(\Delta y_{i,t-1}) + \Delta\varepsilon_{it}$. This eliminates u_i from the equation,

so in addition to the lags of x_{it} or lags of Δx_{it} we can use sufficiently lagged values of y_{it} or $\Delta y_{i,t-1}$. For example, if we go back to $y_{i,t-2}$, (or $\Delta y_{i,t-2}$) that is far enough that the instrument is not correlated with anything in the differenced equation. The model as stated is also a candidate for the Hausman and Taylor approach. The variable that is correlated with the effect is $y_{i,t-1}$. The rest of the model fits precisely into the HT framework.

3. (a) It does provide a consistent set of estimators. This is what was suggested at the begining of part (2) above.

3. (b) The estimator in (3)(a) is equivalent to using GMM while assuming homoscedasticity of the disturbances. The empirical moment condition is

$E[f_{it}w_{it}] = 0$ – note this is 5 equations in 4 unknowns
$E[1(y_{it} - \alpha - \beta x_{it} - \delta_i z_{it} - \gamma y_{i,t-1})] = 0,$
$E[z_{it}(y_{it} - \alpha - \beta x_{it} - \delta_i z_{it} - \gamma y_{i,t-1})] = 0,$
$E[x_{it}(y_{it} - \alpha - \beta x_{it} - \delta_i z_{it} - \gamma y_{i,t-1})] = 0$
$E[z_{i,t-1}(y_{it} - \alpha + \beta x_{it} - \delta_i z_{it} - \gamma y_{i,t-1})] = 0,$
$E[x_{i,t-1}(y_{it} - \alpha + \beta x_{it} - \delta_i z_{it} - \gamma y_{i,t-1})] = 0$

and the empirical moment proposed is simply $m(\beta) = \sum_i \sum_t f_{it}w_{it} = 0$. When we pool the data in this fashion and minimize $m(\beta)'m(\beta)$, the resulting estimator is simply 2SLS. The proposed estimator suggests that we use the moment conditions separately for three periods. You can think of this as if we were using periods 2, 3 and 4 separately to estimate the parameters, which we could do using 2SLS in each, then averaging the estimators. The suggestion is that we use the moments for the three periods separately. This would imply 15 moment equations,

$E[1(y_{i2} - \alpha - \beta x_{i2} - \delta_i z_{i2} - \gamma y_{i,1})] = 0,$
$E[x_{i2}(y_{i2} - \alpha - \beta x_{i2} - \delta_i z_{i2} - \gamma y_{i,1})] = 0,$
$E[z_{i2}(y_{i2} - \alpha - \beta x_{i2} - \delta_i z_{i2} - \gamma y_{i,1})] = 0,$
$E[x_{i1}(y_{i2} - \alpha - \beta x_{i2} - \delta_i z_{i2} - \gamma y_{i,1})] = 0,$
$E[z_{i1}(y_{i2} - \alpha - \beta x_{i2} - \delta_i z_{i2} - \gamma y_{i,1})] = 0,$
$E[1(y_{i3} - \alpha - \beta x_{i3} - \delta_i z_{i3} - \gamma y_{i,2})] = 0,$
$E[x_{i3}(y_{i3} - \alpha - \beta x_{i3} - \delta_i z_{i3} - \gamma y_{i,2})] = 0,$
$E[z_{i3}(y_{i3} - \alpha - \beta x_{i3} - \delta_i z_{i3} - \gamma y_{i,2})] = 0,$
$E[x_{i2}(y_{i3} - \alpha - \beta x_{i3} - \delta_i z_{i3} - \gamma y_{i,2})] = 0,$
$E[z_{i2}(y_{i3} - \alpha - \beta x_{i3} - \delta_i z_{i3} - \gamma y_{i,2})] = 0,$
$E[1(y_{i4} - \alpha - \beta x_{i4} - \delta_i z_{i4} - \gamma y_{i,3})] = 0,$
$E[x_{i4}(y_{i4} - \alpha - \beta x_{i4} - \delta_i z_{i4} - \gamma y_{i,3})] = 0,$
$E[z_{i4}(y_{i4} - \alpha - \beta x_{i4} - \delta_i z_{i4} - \gamma y_{i,3})] = 0,$
$E[x_{i3}(y_{i4} - \alpha - \beta x_{i4} - \delta_i z_{i4} - \gamma y_{i,3})] = 0.$
$E[z_{i3}(y_{i4} - \alpha - \beta x_{i4} - \delta_i z_{i4} - \gamma y_{i,3})] = 0.$

The proposed GMM estimator would proceed as follows: We need a preliminary estimator of the parameters, which we computed before using 2SLS. We now need to compute the weighting matrix. We can simply compute $W = (1/30)\sum_i m_i m_i'$ where m_i is the 15×1 vector shown explicitly above. Then, the two step GMM estimator is

$$\widehat{\theta} = [X'ZW^{-1}Z'X]^{-1}[X'ZW^{-1}Z'y]$$

3. (c) The extended approach would add many additional moment equations. In addition to the preceding, consider just the equations added by $E[f_{i3}w_{i2}] = 0$. These would be

$$\begin{aligned}
&E[1(y_{i2} - \alpha - \beta x_{i2} - \delta_i z_{i2} - \gamma y_{i,1})] = 0,\\
&E[x_{i3}(y_{i2} - \alpha - \beta x_{i2} - \delta_i z_{i2} - \gamma y_{i,1})] = 0,\\
&E[z_{i3}(y_{i2} - \alpha - \beta x_{i2} - \delta_i z_{i2} - \gamma y_{i,1})] = 0,\\
&E[x_{i2}(y_{i2} - \alpha - \beta x_{i2} - \delta_i z_{i2} - \gamma y_{i,1})] = 0,\\
&E[z_{i2}(y_{i2} - \alpha - \beta x_{i2} - \delta_i z_{i2} - \gamma y_{i,1})] = 0.
\end{aligned}$$

Notice that the first of these is already in the set of 15 – it is the first one. But, this adds 4 new moment equations. If we do this for each pair (t, s), we have 4 new moment equations for each of $(t, s) = (2, 3), (2, 4), (3, 2), (3, 4), (4, 2), (4, 3)$, or 6 new sets of 4 moments, for a total of 39. In principle, this would now proceed exactly as we did before, using a 39×39 weighting matrix. There is a problem, however. We have only 30 observations. There are not enough observations to proceed in this fashion.

4. If δ differs across countries, then none of the GMM estimators suggested will be consistent, since they estimate only a single δ. The only hope is to estimate an equation for each country,

$$y_{it} = \alpha + \beta x_{it} + \delta_i z_{it} + \gamma y_{i,t-1} + u_i + \varepsilon_{it},\ t = 1, 2, 3, 4$$

and $y_{i,0}$ is observed data.

With only 4 observations, this does not look promising. Suppose you could assume that $\delta_i = \delta + w_i$ where w_i is orthogonal to the other variables in the model. Then,

$$y_{it} = \alpha + \beta x_{it} + \delta_i z_{it} + \gamma y_{i,t-1} + u_i + w_i z_i + \varepsilon_{it},\ t = 1, 2, 3, 4$$

and $y_{i,0}$ is observed data.

This is the same model as above, except there is now heteroscedasticity in the random effect. All the same problems as before exist, but the GMM estimators suggested do work – they may be inefficient – in the presence of heteroscedasticity. If, however, it cannot be assumed that w_i is uncorrelated with everything else, then the cause is lost. There is no consistent estimator.

Empirical Analysis of Panel Data

The following analysis is based on a panel of data on the Swiss railroad network. The data are a panel of observations on 50 railway companies, with numbers of observations per company ranging from 1 to 13.

(Frequencies are: 37:13 obs; 8:12 obs; 1:10 obs; 2:7 obs; 1: 3 obs; 1:1 obs.)

The variables in the data set that are used in the regressions below are as follows:

- ID: Company ID from 1 to 51 (50 companies, 605 obs)
- YEAR: Year (1985 to 1997)
- TOTCOST: Total cost (in 1000 CHF)

- NI: Number of years for each company
- CT: Total costs adjusted for inflation (1000 CHF)
- Q1: Total output in train-kilometers
- Q2: Total passenger-output in passenger-kilometers
- Q3: Total goods-output in ton-kilometers
- PL: Labor price adjusted for inflation (
- PK: Capital price using the total number of seats as a proxy for capital stock (CHF per seat)
- PE: Price of electricity (CHF per kWh)
- STOPS: Number of stations on the network
- NARROW_TRACK: Dummy for the networks with narrow track (1 m wide) The usual width is 1.435m.
- RACK: Dummy for the networks with RACK RAIL (cremaillere) in at least some part (used to maintain a
- slow movement of the train on high slopes)
- TUNNEL: Dummy for networks that have tunnels with an average length of more than 300 meters.
- VIRAGE: Dummy for the networks whose minimum radius of curvature is 100 meters or less.
- In the regressions below,
- lnCT = log(totcost/pE)
- lnpk = log(pK/pE)
- lnpl = log(pL/pE)
- lnq2 = log(q2)
- lnq3 = log(q3)
- t = time trend for year, coded Year – 1984 = 1,2,...

The essential model is

$$\begin{aligned}\log C_{it} &= \beta_1 \log Q_{2,it} + \beta_2 \log Q_{3,it} + \beta_3 \log PK_{it} + \beta_4 \log PL_{it} + \beta_5 \log PE_{it} + \beta_6 t \\ &\quad \gamma_1 Virage_i + \gamma_2 Tunnel_i + \gamma_3 \text{Narrow\_T}_i + \gamma_4 Rack_i + \alpha_i + \varepsilon_{it}\end{aligned}$$

The constraint that $\beta_3 + \beta_4 + \beta_5 = 1$ has been built into the estimated model by dividing C_{it}, PK_{it} and PL_{it} all by PE_{it} then using logs of the normalized variables in the regression. This is the constraint that imposes linear homogeneity in the input prices on the cost function.

Two sets of results are given below. The first set is based on a restricted model in which $\gamma_1, ..., \gamma_4$ all equal zero. That is, the time invariant variables are not included in the model. The second set of results includes the time invariant variables.

1. How would you test the restriction of linear homogeneity in the input prices in the context of the pooled linear regression model? Do the results given below provide the statistics you need to carry out the test? If yes, show how to do it. If not, explain why not – i.e., what you need that is not provided.

The results given do not provide any way to test this hypothesis. We would need either (1) The unrestricted regression, which would give us the sum of squares or R2 so we could carry out an F test (2) The covariance matrix for the unrestricted regression so we could carry out a Wald test or (3) enough results to carry out an LM test using the restricted regression, which we do not have either.

2. Using the pooled least squares results, test the hypothesis that $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0$. Can you carry out this test using the fixed effects results? Explain? How would you carry out this test using the random effects results?

With the restriction imposed, the R^2 is .9151612. In the unrestricted regression, R^2 =0.9546219. So, F = [(0.9546219 - 0.9151612)/4]/[(1 - 0.9546219)/(605 – 11)] = 129.135. The critical F statistic with (4,594) degrees of freedom is about 2.39, so the hypothesis is rejected.

3. Based on the results given, which model do you think the analyst should report as their best estimates, the pooled least squares results, the fixed effects results or the random effects results? Justify your answer with the statistical evidence.

In the model without the time invariant regressors, the LM statistic reported is 2941.2 = chi squared with 1 degree of freedom. This is very large and would rule out OLS – the classical model. Then, the Hausman statistic is 63.52 = chi squared with 6 degrees of freedom. This is large. The critical value is about 12.59, so this supports the fixed effects approach.

4. Notice that in the first set of results, the sum of squared residuals for the fixed effects estimator is 3.097795. In the second set of results, where the time invariant variables are added to the regression, the sum of squared residuals given for the fixed effects regression is 3.097795 again!!. Shouldn't the sum of squared residuals decline when variables are added to the regression? Can you explain this strange outcome?

The time invariant variables cannot contribute to the fit of a fixed effects model, since they are all linear combinations of variables that are already in the model – the fixed effects. So, their coefficients will be zero, and they will not change the sum of squares.

5. Using the first set of regression results, test the hypothesis that all the constant terms in the fixed effects

model are equal to each other.

The R^2 in the pooled regression with one constant term is 0.9151612. The R^2 in the fixed effects regression is 0.9957743. So, the F statistic is F(49,605-50-6) = [(.9957743 - .9151612)/49]/[(1 - .9957743)/(605 – 50-6)] = 213.740 (This is given in the regression results) The critical F is 2.403, so the hypothesis is rejected.

6. The hypothesis of constant returns to scale is $\beta_2 + \beta_3 = 1$. Using the first regression, carry out a test of this hypothesis using the model that you chose in part 3.

We would use the fixed effects model. In the model shown, these are the coefficients on lnq2 and lnq3. The test statistic, using the Wald, or chi squared, would be (0.21431433 + 0.02548159-1)2/(0.000878247 + 0.0000346396 - 2×0.0000112294) = 649.03 This is much larger than the critical value of 3.84. Translating to a t statistic, the value would be -25.48.

7. In a cost function such as this, the assumption that the output variables are exogenous is sometimes justified by an appeal to the regulatory environment in which some regulatory body sets the prices for the firm and they must accept all demand that is forthcoming. The argument works for electricity or gas providers. It probably doesn't work for profit maximizing railroads. In general terms, how would you want to change your estimation strategy to deal with the possibility that these two variables are endogenous in the model.

We would need to find two instrumental variables. It's not clear what these might be. We could only speculate. Wherever they come from, call them z_1 and z_2, the next step would be 2 stage least squares. Nothing in the statement of the problem suggests that GMM provides any additional benefit.

8. The random effects model in the first results embodies an undesirable assumption of uncorrelatedness of α_i and the independent variables. The fixed effects model has many coefficients and is inefficient (possibly). The Mundlak approach represents a compromise of these two. Describe how to use Mundlak's estimator in this model.

The Mundlak is based on the proposition that we can project the effects on the means of the exogenous variables, that is,
α_i = means'$\gamma + w_i$. If we insert this in the fixed effects model, we come up with a random effects model in which the variables are the original time varying variables plus the group means of these variables. In this particular setting, it might make sense to think about ci also depending on the time invariant variables listed, which would put them back in the (now random effects) model.

9. After computing the fixed effects model in the first set of results below, I computed the 50 railroad specific intercept terms, $a_i = \overline{y}_i - \overline{\mathbf{x}}_i' b_{LSDV}$, $i = 1, ..., 50$. This gives me a sample of 50 observations. I then regressed this a_i on a constant and the railroad specific values of the four time invariant variables listed above. The results were as shown below. How (if at all) does this two step procedure relate to computing the fixed effects estimator and the random effects estimator in the second set of results below? Or, does this regression make no sense at all? What is your interpretation of this model? Is this two step procedure a valid estimator in the context of a particular model? Explain.

It has nothing to do with the fixed effects estimator, since the fixed effects embody all the time invariant information about each railroad. The regression suggests a sort of

Mundlak approach to the random effects model, however, based on 8 above. The model that seems to be suggested by the procedure is

$$\begin{aligned} y_{it} &= \alpha_i + x_{it}'\beta + \varepsilon_{it} \\ \alpha_i &= \alpha + z_i'\gamma + u_i \end{aligned}$$

We will have to assume that ui and (z_i, x_{it}) are uncorrelated. This does define the random effects model. However, note that estimation of the model in two steps is not the same as fitting the model by GLS. Inserting the second equation in the first produces the RE model, which we know consistently estimates α, β, γ. Doing this in two steps obtains a consistent estimator of β and an unbiased estimator of α_i. The implication is that

$$a_i = \alpha_i + w_i$$

where $E[w_i|z_i] = 0$. It follows that α and γ are estimable by OLS. So, this is an alternative, less efficient way to estimate the parameters of the random effects model.

```
+-----------------------------------------------------+
| Ordinary    least squares regression                |
| LHS=AI       Mean                   =  -4.227825    |
|              Standard deviation     =   .5606726    |
| WTS=none     Number of observs.     =         50    |
| Model size   Parameters             =          6    |
|              Degrees of freedom     =         44    |
| Residuals    Sum of squares         =   10.04725    |
|              Standard error of e    =   .4778563    |
| Fit          R-squared              =   .3477224    |
|              Adjusted R-squared     =   .2735999    |
| Model test   F[  5,    44] (prob) =   4.69 (.0016) |
+-----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|   -4.44420462       .14204060   -31.288   .0000
 STOPS   |     .01119768       .00398614     2.809   .0074   21.1800000
 VIRAGE  |    -.23585127       .36712913     -.642   .5239     .70000000
 TUNNEL  |     .37900057       .19465879     1.947   .0579     .18000000
 NARROW_T|    -.02249261       .36654719     -.061   .9513     .66000000
 RACK    |     .41333733       .17440055     2.370   .0222     .22000000
```

FIRST SET OF RESULTS: TIME INVARIANT VARIABLES OMITTED FROM THE MODEL

```
+----------------------------------------------------+
| OLS Without Group Dummy Variables                  |
| Ordinary    least squares regression               |
| LHS=LNCT     Mean                  =   11.30622    |
|              Standard deviation    =   1.101691    |
| WTS=none     Number of observs.    =        605    |
| Model size   Parameters            =          7    |
|              Degrees of freedom    =        598    |
| Residuals    Sum of squares        =   62.19436    |
|              Standard error of e   =   .3224964    |
| Fit          R-squared             =   .9151612    |
|              Adjusted R-squared    =   .9143100    |
| Model test   F[  6,   598] (prob) =1075.11 (.0000) |
| Diagnostic   Log likelihood        =  -170.2812    |
|              Restricted(b=0)       =  -916.5494    |
|              Chi-sq [  6]  (prob) =1492.54 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =  -2.251823     |
|              Akaike Info. Criter. =  -2.251824     |
+----------------------------------------------------+
+----------------------------------------------------+
| Panel Data Analysis of LNCT        [ONE way]       |
|           Unconditional ANOVA (No regressors)      |
| Source      Variation   Deg. Free.     Mean Square |
| Between       720.242          49.     14.6988     |
| Residual      12.8465         555.     .231468E-01 |
| Total         733.089         604.     1.21372     |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 LNQ2    |     .58153570       .01463150    39.745   .0000   16.3175881
 LNQ3    |     .05791869       .00640043     9.049   .0000   12.4943868
 LNPK    |     .25475977       .03130808     8.137   .0000   10.1795011
 LNPL    |     .40014161       .08962528     4.465   .0000   13.2193536
 T       |     .00435867       .00372354     1.171   .2418   5.91570248
 STOPS   |     .00892884       .00096104     9.291   .0000   20.4760331
 Constant|   -6.99824742      1.16691897    -5.997   .0000
+----------------------------------------------------+
| Least Squares with Group Dummy Variables           |
| Ordinary    least squares regression               |
| LHS=LNCT     Mean                  =   11.30622    |
|              Standard deviation    =   1.101691    |
| WTS=none     Number of observs.    =        605    |
| Model size   Parameters            =         56    |
|              Degrees of freedom    =        549    |
```

```
| Residuals    Sum of squares       =    3.097795      |
|              Standard error of e  =    .7511733E-01 |
| Fit          R-squared            =    .9957743      |
|              Adjusted R-squared   =    .9953510      |
| Model test   F[ 55,    549] (prob) =2352.20 (.0000) |
+-----------------------------------------------------+
+-----------------------------------------------------+
| Panel:Groups   Empty        0,    Valid data      50 |
|                Smallest     1,    Largest         13 |
|                Average group size             12.10 |
+-----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 LNQ2    |     .21431433       .02963523     7.232   .0000   16.3175881
 LNQ3    |     .02548159       .00588554     4.330   .0000   12.4943868
 LNPK    |     .31551254       .01781963    17.706   .0000   10.1795011
 LNPL    |     .61669550       .03576588    17.243   .0000   13.2193536
 T       |     .00375773       .00112156     3.350   .0008   5.91570248
 STOPS   |     .01647699       .00248814     6.622   .0000   20.4760331
```

Matrix - Cov.Mat.

[6, 6] Cell:

	LNQ2	LNQ3	LNPK	LNPL	T	STOPS
LNQ2	0.000878247	-1.12294e-005	-7.17965e-005	-0.00018965	-1.0367e-005	-1.77992e-005
LNQ3	-1.12294e-005	3.46396e-005	-8.14176e-006	-1.33476e-006	3.12326e-006	-7.3875e-008
LNPK	-7.17965e-005	-8.14176e-006	0.000317539	-0.000192832	8.49508e-007	-1.81445e-006
LNPL	-0.00018965	-1.33476e-006	-0.000192832	0.0012792	-1.16043e-005	-4.58712e-006
T	-1.0367e-005	3.12326e-006	8.49508e-007	-1.16043e-005	1.2579e-006	-3.51269e-008
STOPS	-1.77992e-005	-7.3875e-008	-1.81445e-006	-4.58712e-006	-3.51269e-008	6.19086e-006

```
+--------------------------------------------------------------------+
|             Test Statistics for the Classical Model                 |
+--------------------------------------------------------------------+
|       Model            Log-Likelihood  Sum of Squares    R-squared |
|(1)  Constant term only     -916.54938  .7330886930D+03    .0000000 |
|(2)  Group effects only      306.82066  .1284645922D+02    .9824763 |
|(3)  X - variables only     -170.28114  .6219435608D+02    .9151612 |
|(4)  X and group effects     737.08990  .3097794979D+01    .9957743 |
+--------------------------------------------------------------------+
|                        Hypothesis Tests                            |
|         Likelihood Ratio Test            F Tests                   |
|         Chi-squared    d.f.  Prob.        F     num.  denom.  P value |
|(2) vs (1)  2446.740     49  .00000  635.027    49     555   .00000 |
|(3) vs (1)  1492.536      6  .00000 1075.110     6     598   .00000 |
|(4) vs (1)  3307.279     55  .00000 2352.201    55     549   .00000 |
|(4) vs (2)   860.538      6  .00000  287.948     6     549   .00000 |
|(4) vs (3)  1814.742     49  .00000  213.740    49     549   .00000 |
+--------------------------------------------------------------------+
+----------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)        |
| Estimates:  Var[e]              =    .564261D-02    |
|             Var[u]              =    .983613D-01    |
|             Corr[v(i,t),v(i,s)] =    .945746        |
| Lagrange Multiplier Test vs. Model (3) = 2941.42   |
| ( 1 df, prob value =  .000000)                      |
| (High values of LM favor FEM/REM over CR model.)   |
| Baltagi-Li form of LM Statistic =        1802.20   |
| Fixed vs. Random Effects (Hausman)      =    63.52 |
| ( 6 df, prob value =  .000000)                      |
| (High (low) values of H favor FEM (REM).)          |
|             Sum of Squares           .108770D+03    |
|             R-squared                .851628D+00    |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 LNQ2    |     .34260963       .02450705    13.980   .0000   16.3175881
 LNQ3    |     .03211634       .00562987     5.705   .0000   12.4943868
 LNPK    |     .30254035       .01753417    17.254   .0000   10.1795011
```

LNPL	.58213153	.03528976	16.496	.0000	13.2193536
T	.00278970	.00109781	2.541	.0110	5.91570248
STOPS	.01802224	.00187232	9.626	.0000	20.4760331
Constant	-5.84523658	.52101033	-11.219	.0000	

Matrix - Cov.Mat.

[7, 7] Cell:

	LNQ2	LNQ3	LNPK	LNPL	T	STOPS	ONE
LNQ2	0.000600596	-1.8565e-005	-4.58468e-005	-0.000125011	-7.65049e-006	-1.90247e-005	-0.0070151
LNQ3	-1.8565e-005	3.16955e-005	-7.1754e-006	3.2751e-006	2.96026e-006	-7.86842e-007	-6.42196e-005
LNPK	-4.58468e-005	-7.1754e-006	0.000307447	-0.000194747	5.30471e-007	-1.6234e-007	0.000282487
LNPL	-0.000125011	3.2751e-006	-0.000194747	0.00124537	-1.19631e-005	-9.09148e-007	-0.0123894
T	-7.65049e-006	2.96026e-006	5.30471e-007	-1.19631e-005	1.20518e-006	2.09577e-008	0.000233131
STOPS	-1.90247e-005	-7.86842e-007	-1.6234e-007	-9.09148e-007	2.09577e-008	3.5056e-006	0.000260404
ONE	-0.0070151	-6.42196e-005	0.000282487	-0.0123894	0.000233131	0.000260404	0.271452

SECOND SET OF RESULTS: TIME INVARIANT VARIABLES INCLUDED IN THE MODEL

```
+----------------------------------------------------+
| OLS Without Group Dummy Variables                  |
| Ordinary    least squares regression               |
| LHS=LNCT     Mean                 =   11.30622     |
|              Standard deviation   =   1.101691     |
| WTS=none     Number of observs.   =        605     |
| Model size   Parameters           =         11     |
|              Degrees of freedom   =        594     |
| Residuals    Sum of squares       =   33.26614     |
|              Standard error of e  =   .2366508     |
| Fit          R-squared            =   .9546219     |
|              Adjusted R-squared   =   .9538580     |
| Model test   F[ 10,    594] (prob) =1249.60 (.0000) |
+----------------------------------------------------+
+----------------------------------------------------+
| Panel Data Analysis of LNCT       [ONE way]        |
|          Unconditional ANOVA (No regressors)       |
| Source        Variation   Deg. Free.    Mean Square |
| Between        720.242          49.      14.6988   |
| Residual       12.8465         555.      .231468E-01 |
| Total          733.089         604.      1.21372   |
+----------------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
LNQ2	.60397404	.01291133	46.779	.0000	16.3175881
LNQ3	.05675679	.00662610	8.566	.0000	12.4943868
LNPK	.43028007	.02471100	17.412	.0000	10.1795011
LNPL	.48044792	.06629234	7.247	.0000	13.2193536
T	.00125984	.00277632	.454	.6500	5.91570248
STOPS	.01164985	.00077905	14.954	.0000	20.4760331
VIRAGE	-.05855252	.05349910	-1.094	.2738	.71570248
TUNNEL	-.17749327	.03217998	-5.516	.0000	.18842975
NARROW_T	-.18639735	.05662731	-3.292	.0010	.67603306
RACK	.58275984	.02598474	22.427	.0000	.23471074
Constant	-10.1709783	.87292761	-11.652	.0000	

```
+----------------------------------------------------+
| Least Squares with Group Dummy Variables           |
| Ordinary    least squares regression               |
| LHS=LNCT     Mean                 =   11.30622     |
|              Standard deviation   =   1.101691     |
| WTS=none     Number of observs.   =        605     |
| Model size   Parameters           =         60     |
|              Degrees of freedom   =        545     |
| Residuals    Sum of squares       =   3.097795     |
|              Standard error of e  =   .7539249E-01 |
| Fit          R-squared            =   .9957743     |
|              Adjusted R-squared   =   .9953169     |
| Model test   F[ 59,    545] (prob) =2176.75 (.0000) |
| Diagnostic   Log likelihood       =   737.0899     |
|              Restricted(b=0)      =  -916.5494     |
```

```
|               Chi-sq [ 59]  (prob) =3307.28 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =  -5.075537      |
|               Akaike Info. Criter. =  -5.076191      |
| Estd. Autocorrelation of e(i,t)     .663021         |
+-----------------------------------------------------+
```

```
+----------------------------------------------------+
| Panel:Groups   Empty       0,   Valid data      50 |
|                Smallest    1,   Largest         13 |
|                Average group size           12.10 |
| There are  4 vars. with no within group variation. |
| VIRAGE   TUNNEL   NARROW_T RACK                    |
| Look for huge standard errors and fixed parameters.|
| F.E. results are based on a generalized inverse.   |
| They will be highly erratic. (Problematic model.)  |
| Unable to compute std.errors for dummy var. coeffs.|
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 LNQ2    |     .21431433       .02974378     7.205   .0000   16.3175881
 LNQ3    |     .02548159       .00590710     4.314   .0000   12.4943868
 LNPK    |     .31551254       .01788490    17.641   .0000   10.1795011
 LNPL    |     .61669550       .03589689    17.180   .0000   13.2193536
 T       |     .00375773       .00112567     3.338   .0008   5.91570248
 STOPS   |     .01647699       .00249726     6.598   .0000   20.4760331
 VIRAGE  |       .000000    ......(Fixed Parameter).......
 TUNNEL  |       .000000    ......(Fixed Parameter).......
 NARROW_T|       .000000    ......(Fixed Parameter).......
 RACK    |       .000000    ......(Fixed Parameter).......
+--------------------------------------------------------------------+
|             Test Statistics for the Classical Model                |
+--------------------------------------------------------------------+
|       Model            Log-Likelihood  Sum of Squares    R-squared |
|(1)  Constant term only     -916.54938  .7330886930D+03    .0000000 |
|(2)  Group effects only      306.82066  .1284645922D+02    .9824763 |
|(3)  X - variables only       19.00043  .3326613925D+02    .9546219 |
|(4)  X and group effects     737.08990  .3097794979D+01    .9957743 |
+--------------------------------------------------------------------+
|                        Hypothesis Tests                            |
|         Likelihood Ratio Test           F Tests                    |
|         Chi-squared   d.f.  Prob.       F    num. denom.   P value |
|(2) vs (1)  2446.740     49  .00000  635.027    49     555   .00000 |
|(3) vs (1)  1871.100     10  .00000 1249.603    10     594   .00000 |
|(4) vs (1)  3307.279     59  .00000 2176.754    59     545   .00000 |
|(4) vs (2)   860.538     10  .00000  171.510    10     545   .00000 |
|(4) vs (3)  1436.179     49  .00000  108.318    49     545   .00000 |
+--------------------------------------------------------------------+
  Error   425: REGR;PANEL. Could not invert VC matrix for Hausman test.
+--------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)     |
| Estimates:  Var[e]              =   .568403D-02  |
|             Var[u]              =   .503196D-01  |
|             Corr[v(i,t),v(i,s)] =   .898506      |
| Lagrange Multiplier Test vs. Model (3) = 2330.30 |
| ( 1 df, prob value =  .000000)                   |
| (High values of LM favor FEM/REM over CR model.) |
| Baltagi-Li form of LM Statistic =        1427.77 |
| Fixed vs. Random Effects (Hausman)     =     .00 |
| (10 df, prob value = 1.000000)                   |
| (High (low) values of H favor FEM (REM).)        |
|             Sum of Squares          .664209D+02  |
|             R-squared               .909412D+00  |
+--------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 LNQ2    |     .38937211       .02314079    16.826   .0000   16.3175881
 LNQ3    |     .03403754       .00564514     6.030   .0000   12.4943868
 LNPK    |     .30471515       .01743458    17.478   .0000   10.1795011
 LNPL    |     .56876577       .03520706    16.155   .0000   13.2193536
 T       |     .00238707       .00110142     2.167   .0302   5.91570248
 STOPS   |     .01779681       .00164611    10.811   .0000   20.4760331
 VIRAGE  |    -.17791631       .17329440    -1.027   .3046    .71570248
 TUNNEL  |     .20298377       .09501443     2.136   .0327    .18842975
 NARROW_T|    -.04830923       .17275122     -.280   .7797    .67603306
 RACK    |     .43134224       .08205919     5.256   .0000    .23471074
 Constant|   -6.44689245       .50866766   -12.674   .0000
```