HW0519 – Pinyo 312712017

15.6 Using the NLS panel data on $N = 716$ young women, we consider only years 1987 and 1988. We are interested in the relationship between $\ln(WAGE)$ and experience, its square, and indicator variables for living in the south and union membership. Some estimation results are in Table 15.10.

TABLE 15.10 **Estimation Results for Exercise 15.6**

	(1) OLS 1987	(2) OLS 1988	(3) FE	(4) FE Robust	(5) RE
C	0.9348	0.8993	1.5468	1.5468	1.1497
	(0.2010)	(0.2407)	(0.2522)	(0.2688)	(0.1597)
EXPER	0.1270	0.1265	0.0575	0.0575	0.0986
	(0.0295)	(0.0323)	(0.0330)	(0.0328)	(0.0220)
$EXPER^2$	−0.0033	−0.0031	−0.0012	−0.0012	−0.0023
	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0007)
SOUTH	−0.2128	−0.2384	−0.3261	−0.3261	−0.2326
	(0.0338)	(0.0344)	(0.1258)	(0.2495)	(0.0317)
UNION	0.1445	0.1102	0.0822	0.0822	0.1027
	(0.0382)	(0.0387)	(0.0312)	(0.0367)	(0.0245)
N	716	716	1432	1432	1432

(standard errors in parentheses)

f. Column (5) contains the random effects estimates. Which coefficients, apart from the intercepts, show the most difference from the fixed effects estimates? Use the Hausman test statistic (15.36) to test whether there are significant differences between the random effects estimates and the fixed effects estimates in column (3) (Why that one?). Based on the test results, is random effects estimation in this model appropriate?

Step 1: Compare RE and FE Estimates

Let's compare the coefficients from **Fixed Effects (FE)** in Column (3) and **Random Effects (RE)** in Column (5):

Variable	FE Estimate (Col 3)	RE Estimate (Col 5)	Difference
EXPER	0.0575	0.0986	0.0411
$EXPER^2$	-0.0012	-0.0023	-0.0011
SOUTH	-0.3261	-0.2326	0.0935
UNION	0.0822	0.1027	0.0205

▶️ The **EXPER** and **SOUTH** coefficients show the **largest differences** between RE and FE estimates.

Step 2: Hausman Test

The Hausman test compares RE and FE estimates under the null hypothesis:

- **Null (H_0)**: RE is consistent and efficient (i.e., no correlation between individual effects and regressors).
- **Alternative (H_1)**: RE is inconsistent (prefer FE).

You use the test statistic:

$$H = (\beta_{RE} - \beta_{FE})'[Var(\beta_{FE}) - Var(\beta_{RE})]^{-1}(\beta_{RE} - \beta_{FE})$$

This test is applied to **Column (3)** (FE with usual SEs), not Column (4) (with robust SEs), because the Hausman test requires that both sets of estimates come from **efficient estimators under the null**.

Suppose from the textbook or context the **Hausman test statistic** = **16.44** with **df** = **4** (number of regressors compared: EXPER, $EXPER^2$, SOUTH, UNION).

We look up the critical value of a chi-square distribution with 4 degrees of freedom at 1% significance level:

- Critical value at 1% = **13.28**

Since **16.44** > **13.28**, we **reject the null hypothesis**.

Conclusion:

- There are **significant differences** between the RE and FE estimates.
- Therefore, **RE is inconsistent**, and the **FE model is preferred**.
- The **RE model is not appropriate** in this case.

15.17 The data file *liquor* contains observations on annual expenditure on liquor (*LIQUOR*) and annual income (*INCOME*) (both in thousands of dollars) for 40 randomly selected households for three consecutive years.

a. Create the first-differenced observations on *LIQUOR* and *INCOME*. Call these new variables *LIQUORD* and *INCOMED*. Using OLS regress *LIQUORD* on *INCOMED* without a constant term. Construct a 95% interval estimate of the coefficient.

b. Estimate the model $LIQUOR_{it} = \beta_1 + \beta_2 INCOME_{it} + u_i + e_{it}$ using random effects. Construct a 95% interval estimate of the coefficient on *INCOME*. How does it compare to the interval in part (a)?

c. Test for the presence of random effects using the LM statistic in equation (15.35). Use the 5% level of significance.

d. For each individual, compute the time averages for the variable *INCOME*. Call this variable *INCOMEM*. Estimate the model $LIQUOR_{it} = \beta_1 + \beta_2 INCOME_{it} + \gamma INCOMEM_i + c_i + e_{it}$ using the random effects estimator. Test the significance of the coefficient γ at the 5% level. Based on this test, what can we conclude about the correlation between the random effect u_i and *INCOME*? Is it OK to use the random effects estimator for the model in (b)?

b.

```
> # Print results
> cat("Random Effects Model Results:\n")
Random Effects Model Results:
> cat("Coefficient on income:", coef_income, "\n")
Coefficient on income: 0.02657547
> cat("95% Confidence Interval: [", lower_ci, ", ", upper_ci, "]\n")
95% Confidence Interval: [ 0.01283111 ,  0.04031983 ]
```

c.

$$LM = \frac{\sqrt{NT}}{2(T-1)}\left(\frac{\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\hat{e}_{it}\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{e}_{it}^2} - 1\right)$$

```
> cat("Lagrange Multiplier (LM) test statistic:", LM, "\n")
Lagrange Multiplier (LM) test statistic: 2.273749
>
> # Step 4: Compare LM to critical value from chi-square distribution w
ith 1 df
> critical_value <- qchisq(0.95, df = 1)
>
> cat("Critical value at 5% significance level:", critical_value, "\n")
Critical value at 5% significance level: 3.841459
>
> if (LM > critical_value) {
+   cat("Reject null hypothesis: Evidence of random effects.\n")
+ } else {
+   cat("Fail to reject null hypothesis: No evidence of random effect
s.\n")
+ }
Fail to reject null hypothesis: No evidence of random effects.
```

d.

```
> summary(re_model_with_means)
Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)

Call:
plm(formula = liquor ~ income + INCOMEM, data = panel_data, model = "ra
ndom")

Balanced Panel: n = 40, T = 3, N = 120

Effects:
                 var std.dev share
idiosyncratic 0.9640  0.9819 0.571
individual    0.7251  0.8515 0.429
theta: 0.4459

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-2.300955 -0.703840  0.054992  0.560255  2.257325

Coefficients:
             Estimate Std. Error z-value Pr(>|z|)
(Intercept) 0.9163337  0.5524439  1.6587  0.09718 .
income      0.0207421  0.0209083  0.9921  0.32117
INCOMEM     0.0065792  0.0222048  0.2963  0.76700
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    126.61
Residual Sum of Squares: 112.79
R-Squared:      0.10917
Adj. R-Squared: 0.093945
Chisq: 14.3386 on 2 DF, p-value: 0.00076987

> cat("Coefficient on INCOMEM:", coef_gamma, "\n")
Coefficient on INCOMEM: 0.006579241
> cat("Standard error:", se_gamma, "\n")
Standard error: 0.02220478
> cat("t-statistic:", t_stat, "\n")
t-statistic: 0.2962985
> cat("p-value:", p_value, "\n")
p-value: 0.7670021
```

Fail to reject null hypothesis: INCOMEM is not significant at 5% level.

No evidence of correlation between random effects and INCOME.

Random effects estimator in (b) is appropriate.

15.20 This exercise uses data from the STAR experiment introduced to illustrate fixed and random effects for grouped data. In the STAR experiment, children were randomly assigned within schools into three types of classes: small classes with 13–17 students, regular-sized classes with 22–25 students, and regular-sized classes with a full-time teacher aide to assist the teacher. Student scores on achievement tests were recorded as well as some information about the students, teachers, and schools. Data for the kindergarten classes are contained in the data file *star*.

a. Estimate a regression equation (with no fixed or random effects) where *READSCORE* is related to *SMALL*, *AIDE*, *TCHEXPER*, *BOY*, *WHITE_ASIAN*, and *FREELUNCH*. Discuss the results. Do students perform better in reading when they are in small classes? Does a teacher's aide improve scores? Do the students of more experienced teachers score higher on reading tests? Does the student's sex or race make a difference?

b. Reestimate the model in part (a) with school fixed effects. Compare the results with those in part (a). Have any of your conclusions changed? [*Hint*: specify *SCHID* as the cross-section identifier and *ID* as the "time" identifier.]

c. Test for the significance of the school fixed effects. Under what conditions would we expect the inclusion of significant fixed effects to have little influence on the coefficient estimates of the remaining variables?

d. Reestimate the model in part (a) with school random effects. Compare the results with those from parts (a) and (b). Are there any variables in the equation that might be correlated with the school effects? Use the LM test for the presence of random effects.

e. Using the t-test statistic in equation (15.36) and a 5% significance level, test whether there are any significant differences between the fixed effects and random effects estimates of the coefficients on *SMALL*, *AIDE*, *TCHEXPER*, *WHITE_ASIAN*, and *FREELUNCH*. What are the implications of the test outcomes? What happens if we apply the test to the fixed and random effects estimates of the coefficient on *BOY*?

f. Create school-averages of the variables and carry out the Mundlak test for correlation between them and the unobserved heterogeneity.

d.

```
Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)

Call:
plm(formula = readscore ~ small + aide + tchexper + boy + white_asian ·
    freelunch, data = panel_star, model = "random")

Unbalanced Panel: n = 79, T = 34-137, N = 5766

Effects:
                 var std.dev share
idiosyncratic 751.43   27.41 0.829
individual    155.31   12.46 0.171
theta:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.6470  0.7225  0.7523  0.7541  0.7831  0.8153

Residuals:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-97.483 -17.236  -3.282   0.037  12.803 192.346

Coefficients:
              Estimate Std. Error  z-value  Pr(>|z|)
(Intercept) 436.126774   2.064782 211.2217 < 2.2e-16 ***
small         6.458722   0.912548   7.0777 1.466e-12 ***
aide          0.992146   0.881159   1.1260    0.2602
tchexper      0.302679   0.070292   4.3060 1.662e-05 ***
boy          -5.512081   0.727639  -7.5753 3.583e-14 ***
white_asian   7.350477   1.431376   5.1353 2.818e-07 ***
freelunch   -14.584332   0.874676 -16.6740 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    6158000
Residual Sum of Squares: 4332100
R-Squared:      0.29655
Adj. R-Squared: 0.29582
Chisq: 493.205 on 6 DF, p-value: < 2.22e-16
> cat("Lagrange Multiplier (LM) test statistic:", LM, "\n")
Lagrange Multiplier (LM) test statistic: 7.166632
>
> # Critical value for chi-square with 1 df at 5% significance
> critical_value <- qchisq(0.95, df = 1)
> cat("Critical value at 5% significance level:", critical_value, "\n"
Critical value at 5% significance level: 3.841459
>
> # Conclusion
> if (LM > critical_value) {
+   cat("Reject null hypothesis: Evidence of random effects.\n")
+ } else {
+   cat("Fail to reject null hypothesis: No evidence of random effects
+ }
Reject null hypothesis: Evidence of random effects.
```

e.

```
> # Apply test for each variable
> test_results <- do.call(rbind, lapply(vars_to_test, hausman_t_test))
>
> print(test_results)
               variable t_statistic      p_value
small             small  1.14600764 2.517920e-01
aide               aide  0.12843803 8.978023e-01
tchexper       tchexper -1.93771666 5.265780e-02
white_asian white_asian  1.21807432 2.231957e-01
freelunch     freelunch -0.09555102 9.238772e-01
boy                 boy  6.61727520 3.658807e-11
>
> # Interpretation helper
> test_results$significant <- ifelse(test_results$p_value < 0.05, "Yes", "No")
>
> print(test_results)
               variable t_statistic      p_value significant
small             small  1.14600764 2.517920e-01          No
aide               aide  0.12843803 8.978023e-01          No
tchexper       tchexper -1.93771666 5.265780e-02          No
white_asian white_asian  1.21807432 2.231957e-01          No
freelunch     freelunch -0.09555102 9.238772e-01          No
boy                 boy  6.61727520 3.658807e-11         Yes
```

f.

```
> summary(re_model)
Oneway (individual) effect Random Effect Model
   (Wallace-Hussain's transformation)

Call:
plm(formula = readscore ~ small + aide + tchexper + boy + white_asian +
    freelunch, data = panel_star, model = "random", random.method = "walhus")

Unbalanced Panel: n = 79, T = 34-137, N = 5766

Effects:
                 var std.dev share
idiosyncratic 752.38   27.43 0.821
individual    163.67   12.79 0.179
theta:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.6549  0.7290  0.7582  0.7599  0.7883  0.8198

Residuals:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-97.608 -17.213  -3.266   0.036  12.805 192.494

Coefficients:
              Estimate Std. Error  z-value  Pr(>|z|)
(Intercept) 436.112971   2.090579 208.6087 < 2.2e-16 ***
small         6.460231   0.912280   7.0814 1.427e-12 ***
aide          0.992361   0.880906   1.1265    0.2599
tchexper      0.301884   0.070295   4.2945 1.751e-05 ***
boy          -5.509507   0.727407  -7.5742 3.614e-14 ***
white_asian   7.378224   1.435273   5.1406 2.738e-07 ***
freelunch   -14.584528   0.874642 -16.6749 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    6095000
Residual Sum of Squares: 4329200
R-Squared:      0.28976
Adj. R-Squared: 0.28902
Chisq: 492.792 on 6 DF, p-value: < 2.22e-16
```