

Full name: Nguyen Nhut Vu Truong

Student ID: 413707008

Course: Financial Econometrics

HW0519

15.6

f. Random Effects vs. Fixed Effects

Comparing Random Effects vs. Fixed Effects Estimates

Looking at the coefficients in Table 15.10, the largest differences between RE (column 5) and FE (column 3) are:

- EXPER: 0.0986 (RE) vs 0.0575 (FE)
- EXPER²: -0.0023 (RE) vs -0.0012 (FE)

The Hausman test compares FE estimates from column (3) with RE estimates from column (5). We use column (3) rather than (4) because the Hausman test is based on the standard FE estimator without robust standard errors.

Conducting the Hausman Test

To perform the Hausman test using equation (15.36), let's calculate the t-statistic for each coefficient:

For EXPER:

- Difference in coefficients: $b_{FE} - b_{RE} = 0.0575 - 0.0986 = -0.0411$
- Standard errors: $se(b_{FE}) = 0.0330$, $se(b_{RE}) = 0.0220$
- Denominator: $\sqrt{[se(b_{FE})^2 - se(b_{RE})^2]} = \sqrt{[(0.0330)^2 - (0.0220)^2]} = \sqrt{[0.001089 - 0.000484]} = \sqrt{0.000605} = 0.0246$
- t-statistic: $-0.0411/0.0246 = -1.67$

For EXPER²:

- Difference in coefficients: $b_{FE} - b_{RE} = -0.0012 - (-0.0023) = 0.0011$
- Standard errors: $se(b_{FE}) = 0.0011$, $se(b_{RE}) = 0.0007$
- Denominator: $\sqrt{[se(b_{FE})^2 - se(b_{RE})^2]} = \sqrt{[(0.0011)^2 - (0.0007)^2]} = \sqrt{[0.00000121 - 0.00000049]} = \sqrt{0.00000072} = 0.00085$
- t-statistic: $0.0011/0.00085 = 1.29$

For SOUTH:

- Difference in coefficients: $b_{FE} - b_{RE} = -0.3261 - (-0.2326) = -0.0935$
- Standard errors: $se(b_{FE}) = 0.1258$, $se(b_{RE}) = 0.0317$

- Denominator: $\sqrt{[se(bFE)^2 - se(bRE)^2]} = \sqrt{[(0.1258)^2 - (0.0317)^2]} = \sqrt{[0.01583 - 0.00101]} = \sqrt{0.01482} = 0.1217$
- t-statistic: $-0.0935/0.1217 = -0.77$

For UNION:

- Difference in coefficients: $bFE - bRE = 0.0822 - 0.1027 = -0.0205$
- Standard errors: $se(bFE) = 0.0312$, $se(bRE) = 0.0245$
- Denominator: $\sqrt{[se(bFE)^2 - se(bRE)^2]} = \sqrt{[(0.0312)^2 - (0.0245)^2]} = \sqrt{[0.000973 - 0.000600]} = \sqrt{0.000373} = 0.0193$
- t-statistic: $-0.0205/0.0193 = -1.06$

Joint Test:

For the joint chi-square test with 4 degrees of freedom (corresponding to the 4 slope coefficients), we would calculate: $\chi^2 = \Sigma(t^2) = (-1.67)^2 + (1.29)^2 + (-0.77)^2 + (-1.06)^2 = 2.79 + 1.66 + 0.59 + 1.12 = 6.16$

The critical value for $\chi^2(4)$ at 5% significance level is 9.49.

Conclusion

Since our calculated $\chi^2 = 6.16 < 9.49$, we fail to reject the null hypothesis at the 5% level. This suggests that the random effects estimator is appropriate for this model, as there is not sufficient evidence of systematic differences between the fixed effects and random effects estimates.

This result indicates that the unobserved individual effects (u_i) are likely uncorrelated with the explanatory variables, making the random effects estimator valid and more efficient than the fixed effects estimator for this particular application.

15.20

d.

```
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
```

```
> print(starReTest)
```

Lagrange Multiplier Test - (Honda)

```
data: readscore ~ small + aide + tchexper + boy + white_asian + freelunch
normal = 81.715, p-value < 2.2e-16
alternative hypothesis: significant effects
```

The random effects model reveals that school-level heterogeneity accounts for 17.1% of the total variance in reading scores, with the remaining 82.9% being idiosyncratic variance within schools. The coefficient estimates are very similar to the fixed effects model, with small classes increasing reading scores by 6.46 points, being white/Asian associated with a 7.35 point advantage, and free lunch eligibility reducing scores by 14.58 points. The Lagrange Multiplier test strongly rejects the null hypothesis of no random effects (test statistic = 81.715, $p < 2.2e-16$), providing compelling evidence that random effects are statistically necessary and that the simple pooled OLS model is inadequate. This confirms that school-specific unobserved heterogeneity significantly affects student reading performance and must be accounted for in the analysis.

Several variables are likely correlated with school effects, particularly **free lunch eligibility** and **race/ethnicity** due to residential segregation and socioeconomic sorting across schools, and potentially **teacher experience** if experienced teachers systematically select into certain types of schools. Only the experimental variables (small class size and teacher aide) should be uncorrelated with school effects since they were randomly assigned within schools.

Coefficient Estimates Comparison

Variable	Pooled OLS	Fixed Effects	Random Effects
small	5.82***	6.49***	6.46***
aide	0.82	1.00	0.99
tchexper	0.49***	0.29***	0.30***
boy	-6.16***	-5.46***	-5.51***
white_asian	3.91***	8.03***	7.35***
freelunch	-14.77***	-14.59***	-14.58***

e.

For the BOY variable, the random effects standard error is **larger** than the fixed effects standard error, making $se_{fe}^2 - se_{re}^2$ negative. This violates the theoretical

assumption that fixed effects should have larger standard errors than random effects.

The BOY coefficient cannot be tested using this individual t-test approach because the theoretical condition (FE variance > RE variance) is violated. This suggests the overall **joint Hausman test** is more appropriate.

Based on the individual coefficient Hausman tests at the 5% significance level, **no significant differences** were found between fixed effects and random effects estimates for any of the testable variables (SMALL, AIDE, TCHEXPER, WHITE_ASIAN, FREELUNCH), with TCHEXPER showing the largest difference but still marginally insignificant ($p = 0.053$).

This suggests that for these specific variables, the random effects assumption of zero correlation with school effects appears reasonable, indicating consistency between the two modeling approaches and supporting the validity of using either estimator for these coefficients.

```
> print(results_df)
```

	Variable	FE_Coeff	RE_Coeff	Difference	T_Statistic	P_Value	Significant_5pct
small	small	6.4902	6.4587	0.0315	1.1460	0.2518	FALSE
aide	aide	0.9961	0.9921	0.0039	0.1284	0.8978	FALSE
tchexper	tchexper	0.2856	0.3027	-0.0171	-1.9377	0.0527	FALSE
white_asian	white_asian	8.0280	7.3505	0.6775	1.2181	0.2232	FALSE
freelunch	freelunch	-14.5936	-14.5843	-0.0092	-0.0956	0.9239	FALSE
boy	boy	-5.4559	-5.5121	0.0561	NaN	NaN	NA

f.

```
> cat("Mundlak Test Results:\n")
Mundlak Test Results:
> cat("F-statistic:", round(mundlak_test$F[2], 2), "\n")
F-statistic: 21
> cat("P-value:", round(mundlak_test$Pr(>F)[2], 6), "\n")
P-value: 0
>
> # Interpretation
> if(mundlak_test$Pr(>F)[2] < 0.05) {
+   cat("\nCONCLUSION: Reject H0 - School averages are jointly significant.\n")
+   cat("This suggests correlation between unobserved school effects and regressors.\n")
+   cat("Fixed Effects estimation is preferred over Random Effects.\n")
+ } else {
+   cat("\nCONCLUSION: Fail to reject H0 - School averages not jointly significant.\n")
+   cat("Random Effects may be consistent.\n")
+ }
```

CONCLUSION: Reject H0 - School averages are jointly significant.
This suggests correlation between unobserved school effects and regressors.
Fixed Effects estimation is preferred over Random Effects.

```

Detailed OLS Mundlak Test:
> print(mundlak_test_ols)

Linear hypothesis test:
small_bar = 0
aide_bar = 0
tchexper_bar = 0
boy_bar = 0
white_asian_bar = 0
freelunch_bar = 0

Model 1: restricted model
Model 2: readscore ~ small + aide + tchexper + boy + white_asian + freelunch +
      small_bar + aide_bar + tchexper_bar + boy_bar + white_asian_bar +
      freelunch_bar

   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
1     5759 5247584
2     5753 5135095   6    112489 21.004 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> # Also show coefficients of school averages
> cat("\nCoefficients of school averages in Mundlak regression:\n")

Coefficients of school averages in Mundlak regression:
> mundlak_coefs <- coef(mundlak_ols)[c("small_bar", "aide_bar", "tchexper_bar",
+                                     "boy_bar", "white_asian_bar", "freelunch_bar")]
> print(round(mundlak_coefs, 4))
      small_bar      aide_bar      tchexper_bar      boy_bar white_asian_bar      freelunch_bar
      -23.1032       9.1872       0.8017      -50.2119      -7.2214      -4.1164

=== FINAL RECOMMENDATION ===
> if(fe_test$p.value < 0.05) {
+   cat("Fixed effects are significant (reject pooled OLS).\n")
+ }
Fixed effects are significant (reject pooled OLS).
> if(re_test$p.value < 0.05) {
+   cat("Random effects are significant (reject pooled OLS).\n")
+ }
Random effects are significant (reject pooled OLS).
> if(hausman_overall$p.value < 0.05) {
+   cat("Hausman test rejects RE in favor of FE.\n")
+   cat("RECOMMENDATION: Use Fixed Effects model.\n")
+ } else {
+   cat("Hausman test does not reject RE.\n")
+   if(mundlak_test$`Pr(>F)`[2] < 0.05) {
+     cat("But Mundlak test suggests FE is preferred.\n")
+     cat("RECOMMENDATION: Use Fixed Effects model.\n")
+   } else {
+     cat("Mundlak test also supports RE.\n")
+     cat("RECOMMENDATION: Random Effects model is appropriate.\n")
+   }
+ }
Hausman test rejects RE in favor of FE.
RECOMMENDATION: Use Fixed Effects model.

```

The Mundlak test was conducted to examine whether Random Effects or Fixed Effects estimation is more appropriate by testing for correlation between unobserved school effects and the regressors. The test involved augmenting the regression model with school-level averages of all time-varying variables and testing their joint significance.

The OLS Mundlak test yielded an F-statistic of 21.00 with a p-value $< 2.2e-16$, strongly rejecting the null hypothesis that the school averages are jointly equal to zero. This highly significant result indicates substantial correlation between unobserved school-

level heterogeneity and the explanatory variables, violating the key assumption required for Random Effects consistency. The coefficients on the school averages show meaningful variation (ranging from -50.21 for boy_bar to 9.19 for aide_bar), further confirming that school-level characteristics are systematically related to the regressors.

Therefore, the Mundlak test conclusively supports the use of Fixed Effects estimation over Random Effects, as the latter would produce inconsistent estimates due to the correlation between unobserved school effects and the regressors.

15.17

b.

```
> # Random Effects estimation
> model_re <- plm(liquor ~ income, data = liquor_panel, model = "random")
> summary(model_re)
Oneway (individual) effect Random Effect Model
(Swamy-Arora's transformation)

Call:
plm(formula = liquor ~ income, data = liquor_panel, model = "random")

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.
-2.263634 -0.697383  0.078697  0.552680  2.225798

Coefficients:
              Estimate Std. Error z-value Pr(>|z|)
(Intercept) 0.9690324   0.5210052   1.8599 0.0628957 .
income       0.0265755   0.0070126   3.7897 0.0001508 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> # 95% confidence interval for INCOME coefficient in RE model
> conf_int_re <- confint(model_re, level = 0.95)
> cat("\n95% Confidence Interval for INCOME coefficient (Random Effects):\n")

95% Confidence Interval for INCOME coefficient (Random Effects):
> print(conf_int_re)
              2.5 %      97.5 %
(Intercept) -0.05211904 1.99018381
income       0.01283111 0.04031983
```

```

=== COMPARISON OF CONFIDENCE INTERVALS ===
> cat("First Differences - INCOMED coefficient:\n")
First Differences - INCOMED coefficient:
> cat("Estimate:", round(coef(model_diff)[1], 4), "\n")
Estimate: 0.0297
> cat("95% CI: [", round(conf_int_diff[1], 4), ",", round(conf_int_diff[2], 4), "]\n\n")
95% CI: [ -0.0284 , 0.0879 ]

> cat("Random Effects - INCOME coefficient:\n")
Random Effects - INCOME coefficient:
> cat("Estimate:", round(coef(model_re)[2], 4), "\n")
Estimate: 0.0266
> cat("95% CI: [", round(conf_int_re[2,1], 4), ",", round(conf_int_re[2,2], 4), "]\n\n")
95% CI: [ 0.0128 , 0.0403 ]

```

Based on the confidence intervals, the **Random Effects model provides more precise and statistically significant results** than the First Differences model. While the First Differences estimate (0.0297) includes zero in its confidence interval [-0.0284, 0.0879] suggesting no significant effect, the Random Effects model shows a significant positive relationship (0.0266) with a confidence interval [0.0128, 0.0403] that doesn't include zero.

c. CONCLUSION

Reject H_0 at 5% level. Random effects are present - use RE instead of pooled OLS.

Honda (default) - matches equation (15.35)

```
lm_test <- plmtest(model_pooled, effect = "individual")
```

Lagrange Multiplier Test for Random Effects:

```

> cat("Manual LM statistic:", round(LM_manual, 4), "\n")
Manual LM statistic: 4.5475
> cat("Built-in LM statistic:", round(lm_test$statistic, 4), "\n")
Built-in LM statistic: 4.5475
> cat("P-value:", round(lm_test$p.value, 6), "\n")
P-value: 3e-06

```

d.

CONCLUSION: Fail to reject $H_0: \gamma = 0$ at 5% level.

The coefficient on INCOMEM is not statistically significant.

No evidence of correlation between the random effect u_i and INCOME.

It is OK to use the random effects estimator for the model in part (b).

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

Call:
plm(formula = liquor ~ income + incomem, data = liquor_mundlak_panel,
     model = "random")

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
```

```
> print(comparison_table)
              Model Income_Coefficient Standard_Error CI_Lower CI_Upper
1      First Differences              0.0297          0.0292 -0.0284  0.0879
2      Random Effects                0.0266          0.0070  0.0128  0.0403
3 RE with Time Averages              0.0207          0.0209 -0.0202  0.0617
```

Formal hypothesis test for $\gamma = 0$:

```
> mundlak_test <- linearHypothesis(model_mundlak, "incomem = 0")
> print(mundlak_test)
```

Linear hypothesis test:
incomem = 0

Model 1: restricted model
Model 2: liquor ~ income + incomem

	Res.Df	Df	Chisq	Pr(>Chisq)
1	118			
2	117	1	0.0878	0.767

Summary:

```
=== FINAL SUMMARY ===
> cat("1. First Differences vs Random Effects:\n")
1. First Differences vs Random Effects:
> cat("  - FD coefficient:", round(coef(model_diff)[1], 4), "\n")
  - FD coefficient: 0.0297
> cat("  - RE coefficient:", round(coef(model_re)[2], 4), "\n")
  - RE coefficient: 0.0266
> cat("  - Difference suggests presence of individual effects\n\n")
  - Difference suggests presence of individual effects

>
> cat("2. LM Test for Random Effects:\n")
2. LM Test for Random Effects:
> cat("  - LM statistic:", round(lm_test$statistic, 4), "\n")
  - LM statistic: 4.5475
> cat("  - P-value:", round(lm_test$p.value, 6), "\n")
  - P-value: 3e-06
> cat("  - Conclusion:", ifelse(lm_test$p.value < 0.05, "Random effects present", "No random effects"), "\n\n")
  - Conclusion: Random effects present

>
> cat("3. Mundlak Test (gamma = 0):\n")
3. Mundlak Test (gamma = 0):
> cat("  - gamma coefficient:", round(gamma_coef, 4), "\n")
  - gamma coefficient: 0.0066
> cat("  - P-value:", round(gamma_p, 6), "\n")
  - P-value: 0.767527
> cat("  - Conclusion:", ifelse(gamma_p < 0.05, "RE not consistent", "RE is OK"), "\n")
  - Conclusion: RE is OK
```