**TIØ4317 Empirical and Quantitative Methods in Finance** **Exercise 7**
**16.03.2025** **Scribe: Adrian Fagerland**

## 1

Both models appear statistically significant. We see this by looking at “Prob > chi2 = 0.0000” for the first model (Random Effects) and “Prob > F = 0.0000” for the second model (Fixed Effects). These very low p-values indicate that, overall, the independent variables have a statistically significant relationship with the dependent variable (wage).

There are three R-squared reported. Overall R-squared is the regular $R^2$ one would get from a pooled OLS, and represents the percentage of variance in the dependent variable explained by the explanatory variables. The within R-squared refers to variation within individuals and is the r-squared from the regression of $y_{it} - \bar{y}_i$ against $x_{it} - \bar{x}_i$. The between R-squared refers to the variance between indivudals, and is the r-squared from the regression of $\bar{y}_i$ against $\bar{x}_i$.

## 2

Random Effects (RE) Model:

$$wage_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 w_{it} + \beta_3 n_{it} + u_i + \epsilon_{it}$$

Its assumptions are: * $E[\epsilon_{it}|X_i, u_i] = 0$ (Exogeneity) * $E[u_i|X_i] = 0$ (Random individual effects, uncorrelated with independent variables)

The estimator is GLS:

$$\hat{\beta}_{RE} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$$

Where, simplifying, $\Omega$ takes into account the variance of $u_i$ and $\epsilon_{it}$ to create a covariance matrix. This gives a $\hat{\beta}_{RE}$ that is different to regular OLS.

Fixed Effects (FE) Model:

$$wage_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 w_{it} + \beta_3 n_{it} + u_i + \epsilon_{it}$$

Its assumptions are: * $E[\epsilon_{it}|X_i, u_i] = 0$ * $u_i$ can be correlated with $X_i$.

It uses time-demeaned data (within transformation) to eliminate the individual-specific effects. The estimator is:

$$\hat{\beta}_{FE} = (\ddot{X}'\ddot{X})^{-1}\ddot{X}'\ddot{Y}$$

Where $\ddot{X}, \ddot{Y}$ is the within transformation, that is $\ddot{X} = X_{it} - \bar{X}_i$

The main difference is that RE assumes $u_i$ is uncorrelated with the explanatory variables, while FE allows for correlation. The estimator therefore changes to account for that.

## 3

I would prefer the Fixed Effects model. The Fixed Effects output reports corr(u_i, Xb) = -0.6300, indicating a substantial correlation between the individual effects and the regressors. This violates the Random Effects assumption of zero correlation, suggesting that Random Effects may be inconsistent. The F-test at the end of the Fixed Effects output (F(139, 888) = 32.74, Prob > F = 0.0000) strongly rejects the null hypothesis that all individual effects are zero, supporting the inclusion of individual-specific effects as in the Fixed Effects model. Fixed Effects is robust to unobserved heterogeneity correlated with the regressors, which is likely in wage data where individual traits (e.g., ability) may influence both wages and explanatory variables like w or k. While Random Effects might be more efficient if its assumptions hold, the evidence of correlation and significant individual effects favors Fixed Effects for consistency.

## 4

$\sigma_u$ is the standard deviation of the individual-specific random effects ($u_i$). Intuitively, it represent how much of the variance in the error can be attributed to time-invariant unobserved effects. $\sigma_e$ is the standard deviation of the idiosyncratic error term ($\epsilon_{it}$). Intuitively, it represents the overall variance in the model error. $\rho$ is the fraction of the total variance that is due to the individual-specific effects ($u_i$). Calculated as: $\rho = \frac{\sigma_u^2}{\sigma_u^2+\sigma_e^2}$.

The F test at the end of the second regression output tests that all $u_i = 0$. This tests the null hypothesis that all individual-specific effects are zero. The high F-statistic (32.74) and the very low p-value (0.0000) lead us to reject the null hypothesis. Therefore, there are significant individual effects.

## 5

The first model uses robust standard errors that adjust for clustering due to individuals (id). This is a good practice as errors for an individual are likely autocorrelated. The FE model does not use robust standard errors.

I would suggest using clustered standard errors in both models. In python, this could be added by adding vce(cluster id) in the regression.

## 6

We can use a Hausman test to compare.

$$H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})'(\hat{V}_{FE} - \hat{V}_{RE})^{-1}(\hat{\beta}_{FE} - \hat{\beta}_{RE})$$

Which under null is distributed as $\chi^2_K$, with K being the number of regressors.

Under the null hypothesis, both RE and FE are consistent, but RE is more efficient. Under the alternative hypothesis, RE is inconsistent, while FE is still consistent. The Hausman test checks if the coefficient estimates from the two models are significantly different. If they are, we reject the null and prefer FE.