Advanced Econometrics,

Summary statistics:

The summary statistics of the panel data model is as follows:

Time	ID	LOGCAP
1958 : 26	3312 : 37	Min. : 2.821
1959 : 26	3313 : 37	1st Qu.: 6.500
1960 : 26	3315 : 37	Median : 7.275
1961 : 26	3316 : 37	Mean : 7.280
1962 : 26	3317 : 37	3rd Qu.: 7.825
1963 : 26	3321 : 37	Max. :11.194
(Other):806	(Other):740	
LOGLAB	LOGPROD	
Min. :-2.8869	Min. :-2.372	
1st Qu.: 0.5276	1st Qu.: 1.650	
Median : 1.1169	Median : 2.289	
Mean : 1.1327	Mean : 2.246	
3rd Qu.: 1.7241	3rd Qu.: 2.820	
Max. : 4.8415	Max. : 5.911	

Tests and analysis:

To estimate panel data, we need to make some assumptions regarding our modeling. The most common assumption is data homogeneity. To do so, we assume the model as $\alpha_{it} = \alpha, \beta_{it} = \beta, \gamma_{it} = \gamma$ and pool the regression model over i and t. Therefore, for the pooling model, the standard modeling is as follows:

$$\log(y_{it}) = \alpha_i + \beta\log(L_{it}) + \gamma\log(K_{it}) + \varepsilon_{it}$$

It does not allow for slope differences among individual components.

Pooling Model:

Residuals:

Min.	1st Qu.	Median	3rd Qu.	Max.
-3.1190050	-0.1561310	-0.0046955	0.1704016	1.2567695
Coefficients:				
	Estimate	Std. Error	t-value	Pr(>|t|)
(Intercept)	0.036817	0.094349	0.3902	0.6965
X1	0.184513	0.014828	12.4433	<2e-16 ***
X2	0.764332	0.015065	50.7363	<2e-16 ***
Total Sum of Squares: 1246.5				
Residual Sum of Squares: 81.509				
R-Squared: 0.93461				
Adj. R-Squared: 0.93447				
F-statistic: 6853.53 on 2 and 959 DF, p-value: < 2.22e-16				

Significance codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Y: LOGPROD   X1: LOGLAB   X2: LOGCAP

Additionally, for modeling the model heterogeneity, we treat the error component of the model as if it has three separate components. One is the individual effects which does not change overtime, also the time effect for the symmetric case shows the unobserved model effects:

$$\log(y_{it}) = \alpha + \beta \log(L_{it}) + \gamma \log(K_{it}) + \mu_i + u_{it}$$

The individual error component μ_i may be independent from the repressors or correlated to them. In the case of correlation, OLS is inconsistent. In this case we have a fixed effect. We assume μ_i as a set of n parameters and consider $\alpha_{it} = \alpha_i$ for all$t$. Unlike the pooling model, the individual effect accounts for individual differences, which leads to different intercepts. The within model in plm package in R, returns vectors of the values in deviation from the individual means.

One-way (individual) effect Within Model:

Residuals:
Min.   1st Qu.   Median   3rd Qu.   Max.
-2.8315776 -0.1076154 0.0046429 0.1110748 0.8995680
Coefficients:
Estimate Std. Error t-value Pr(>|t|)
X1 0.181205   0.017986 10.075 < 2.2e-16 ***
X2 0.734647   0.019415 37.839 < 2.2e-16 ***
Total Sum of Squares: 323.78
Residual Sum of Squares: 58.312
R-Squared: 0.8199
Adj. R-Squared: 0.81249
F-statistic: 2101.01 on 2 and 923 DF, p-value: < 2.22e-16

Testing for necessity of fixed effect requires comparison with pooled regression model:

Hypothesis:

$$\begin{cases} H_O: No\ fixed\ effects \\ H_A: Fixed\ effects \end{cases}$$

F-test for individual effects:

F = 10.199, df1 = 36, df2 = 923, p-value < 2.2e-16
Alternative hypothesis: significant effects

The alternative of no fixed effect has been rejected. And shows that this model is better suited for model compared to the pooled model.

Note that another method of removing the time-invariant component is first differencing. It works by lagging the model and removing the time-invariant component. This method is preferred if the errors are persistent over time, and $\Delta\varepsilon_{it}$ is serially uncorrelated.

One-way (individual) effect First-Difference Model:

Residuals:
Min. 1st Qu. Median Mean 3rd Qu. Max.
-1.23474 -0.06328 0.00310 0.00553 0.07099 2.61866
Coefficients:
Estimate Std. Error t-value Pr(>|t|)
X1 0.237957 0.031695 7.5077 1.417e-13 ***
X2 0.755561 0.033657 22.4492 < 2.2e-16 ***
Total Sum of Squares: 130.77
Residual Sum of Squares: 47.952
R-Squared: 0.63353
Adj. R-Squared: 0.63313
F-statistic: 1594.12 on 1 and 923 DF, p-value: < 2.22e-16

Comparing with the pooling model, the individual effects is selected over pooling:

F-test for individual effects:

F = 17.942, df1 = 36, df2 = 923, p-value < 2.2e-16
alternative hypothesis: significant effects

For estimating the long-run relation between the data, we can use the between model, which discards the information regarding intragroup variations. The between transformation returns a vector of individual means.

One-way (individual) effect Between Model:

Residuals:
Min. 1st Qu. Median 3rd Qu. Max.
-0.302195 -0.094842 -0.023554 0.124796 0.379229
Coefficients:
Estimate Std. Error t-value Pr(>|t|)
(Intercept) 0.049751 0.376517 0.1321 0.895657
X1 0.180493 0.059765 3.0201 0.004771 **
X2 0.778750 0.059389 13.1127 7.435e-15 ***
Total Sum of Squares: 35.49
Residual Sum of Squares: 0.87388
R-Squared: 0.97538
Adj. R-Squared: 0.97393
F-statistic: 673.405 on 2 and 34 DF, p-value: < 2.22e-16

F-test for individual effects:

F = 3.3916, df1 = 925, df2 = 34, p-value = 2.366e-05
alternative hypothesis: significant effects

As expected this analysis also reached the same conclusion that he individual effect model is selected over pooling model.

Unlike the fixed effect model, the random effect assumes that the coefficients vary randomly around a common average for all t. To test for random effects the hypothesis can be expressed as follows:

$$\begin{cases} H_O: No\ differences\ among\ individuals: \sigma_u^2 = 0 \\ H_A: Specific\ individual\ effects\ do\ not\ have\ zero\ variance: \sigma_u^2 > 0. \end{cases}$$

One-way (individual) effect Random Effect Model : (Swamy-Arora's transformation):

Effects:
var std.dev share
idiosyncratic 0.06318 0.25135 0.731
individual 0.02327 0.15255 0.269
theta: 0.6925
Residuals:
Min. 1st Qu. Median 3rd Qu. Max.
-2.93334541 -0.11285722 0.00061077 0.12433141 0.99621613
Coefficients:
Estimate Std. Error t-value Pr(>|t|)
(Intercept) 0.071742 0.113508 0.632 0.5275
X1 0.183081 0.017176 10.659 <2e-16 ***
X2 0.742706 0.018227 40.748 <2e-16 ***
Total Sum of Squares: 411.02
Residual Sum of Squares: 60.601
R-Squared: 0.85256
Adj. R-Squared: 0.85225
F-statistic: 2772.68 on 2 and 959 DF, p-value: < 2.22e-16

The model shows the null-hypothesis of zero-variance is rejected. Random effects is reliable under the assumption of exogenous individual characteristics, which means that whether they are independent with respect to the regressors in the random effect model. We can use the Hausman test for endogeneity. The hypothesis for this test is as follows:

$$\begin{cases} H_O: No\ correlation\ between\ independent\ variables\ and\ the\ error\ term\ \mu_i \\ H_A: There\ is\ a\ correlation\ between\ the\ independent\ variables\ and\ \mu_i \end{cases}$$

Hausman Test:

chisq = 2.2196, df = 2, p-value = 0.3296
alternative hypothesis: one model is inconsistent

The test results shows that we fail to reject the null hypothesis and both least squares estimator and instrumental variables estimator are consistent.

The general FGLS test is unrestricted and more general compared to the random effects model and is robust against heteroscedasticity and serial correlation. The results for this test can be seen as follows:

Residuals:
Min. 1st Qu. Median Mean 3rd Qu. Max.
-3.07455 -0.14372 0.01957 0.02416 0.19674 1.30146
Coefficients:
Estimate Std. Error z-value Pr(>|z|)
(Intercept) -0.0150676 0.0515049 -0.2925 0.7699
X1 0.1866622 0.0077683 24.0287 <2e-16 ***
X2 0.7749961 0.0084248 91.9901 <2e-16 ***
Total Sum of Squares: 1246.5
Residual Sum of Squares: 82.284
Multiple R-squared: 0.93399

Breusch-Pagan test for heteroscedasticity:

BP = 236.09, df = 2, p value < 2.2e16

The null hypothesis for the test is homoscedasticity, that we fail to reject the null and the error terms are heteroskedastic.

To find robust coefficients under random effects, we can estimate a robust covariance matrix as follows:

Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.071742 0.262799 0.2730 0.7849
X1 0.183081 0.041527 4.4087 1.157e-05 ***
X2 0.742706 0.040262 18.4468 < 2.2e-16 ***

Test for normality:

There are different methods to test for normality, I used the test for skewness and kurtosis with the null hypothesis of normality. As we can see in the following, the test for the data set shows that we fail to reject the null hypothesis and the non-normality assumption is rejected.

Skewness test for normality:

data: Y X1 X2
p-value
0.011, <2.2e-16, 0.001

Kurtosis test for normality:

data: Y X1 X2
p-value

<2.2e-16 , <2.2e-16 , <2.2e-16

Test for serial correlation:

As we've seen in the model, we have individual effects in the model, which can influence our analysis. The serial correlation due to these individual effects do not vanish overtime. In this part we test for serial correlation in idiosyncratic error terms.

Breusch-Godfrey/Wooldridge test for serial correlation:

chisq = 210.48, df = 2, p-value < 2.2e-16
alternative hypothesis: serial correlation in idiosyncratic errors

We fail to reject the null hypothesis of serial correlation in the data.

Wooldridge's first-difference test for serial correlation:

F = 3.4372, df1 = 1, df2 = 886, p-value = 0.06408
alternative hypothesis: serial correlation in differenced errors

The model shows that there exists serial correlation in the first differenced errors. Therefore, we cannot conclude that the error term is a random walk.

Pesaran CD test for cross-sectional dependence:

z = 1.5912, p value = 0.1116
alternative hypothesis: cross-sectional dependence

The results shows cross-sectional dependencies in the data.

Test for unobserved effects:

Wooldridge's test for unobserved individual effects:

z = 3.7088, p-value = 0.0002083
alternative hypothesis: unobserved effect

The null hypothesis in Wooldridge test is the covariance matrix of the residuals for each individual is diagonal (no unobserved effect), which based on this test we fail to reject the null hypothesis. This test does not rely on homoscedasticity and is independent of the model distributions. Fail to rejection of the null hypothesis supports the use of pooled OLS model.

The following model provides the robust model using pooled OLS:

Coefficients:
(Intercept)   X1   X2
0.036817   0.184513   0.764332
Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.036817   0.226107 0.1628   0.8707
X1   0.184513   0.034362 5.3697 9.896e-08 ***
X2   0.764332   0.032442 23.5601 < 2.2e-16 ***

Conclusion:

In this empirical study, the influence of labor and capital on the production has been analyzed. The model is a panel model. In the first section, a summary statistics has been provided. In the next part first I estimated the model using pooled regression, then I conducted a fixed effect model and comparing the fixed effects and pooled model supported the existence of individual effects in the model. In continuation, examining the data supported the use of random effects.

Before conducting any test for heteroscedasticity and non-normality, I used general FGLS model on random model and reported the results. The reason for such a choice is that this model does not depend on the assumption of homoscedasticity and normal distribution.

The test Breusch-Pegan test supported heteroscedasticity of error terms and the non-normality has not been supported by normality tests. Additionally, in the overall panel data, we do not see serial correlation, but we have cross-sectional dependence and serial correlation in the lagged model.

Implementing the Wooldridge's test for unobserved individual effects supported the use of pooled OLS, therefore, after correcting for robustness, I concluded the use of pooled OLS and reported the results.

References:

Croissant, Yves, and Giovanni Millo. "Panel data econometrics in R: The plm package." Journal of Statistical Software 27.2 (2008): 1-43.

https://bookdown.org/ccolonescu/RPoE4/panel-data-models.html