# Econometrics Project

Uysaler Kutay

# Table of contents

Introduction

Methodology

Results and Discussion

Model assumptions

Conclusion

# Introduction

The aim of this project was to estimate a production function, for which we used panel data techniques.

The dataset we used for analysis was ‘prodfn_data_1_5.dta’, which is made up of firm-level data with the following variables:

- individual identifier (ivar)
- time identifier (tvar)
- dependent variable (y = ln(output))
- explanatory variables (l = ln(labor) and k = ln(capital))

# Methodology

In order to estimate the production function we applied two commonly used panel data techniques; fixed effects (FE) and random effects (RE) models.

The FE model captures individual-specific effects by including individual fixed effects in the regression.

The RE model assumes that the individual-specific effects in the regression are uncorrelated with the explanatory variables.

# Results and Discussion

Data overview

Fixed Effects Model (FE)

Random Effects Model (RE)

# Data Overview

The dataset contained 3,020 observations with five previously mentioned variables: ivar, tvar, k, l, and y (ln(output)).

# Fixed Effects Model (FE)

The FE model was estimated using the plm package in R. The model includes ln(labor) and ln(capital) as explanatory variables.

The FE model yielded the following results:

- **Coefficients:** The estimated coefficients for ln(labor) and ln(capital) were 0.3832104 and 0.4701442, respectively, both statistically significant at a high level of confidence ($p < 0.001$).
- **R-squared:** The FE model explained approximately 21.3% of the variation in ln(output) (adjusted R-squared: 0.016052).
- **Specification Test:** The Hausman test was conducted to compare the FE and RE models. The test statistic of 0.19939 with a p-value of 0.9051 suggests that the null hypothesis of the FE model being consistent with the RE model cannot be rejected.

```
> # Examining model coefficients
> summary(fe_model)  # Print the summary of fixed effects model
Oneway (individual) effect Within Model

Call:
plm(formula = y ~ l + k, data = prodfn_data, model = "within",
    index = c("ivar", "tvar"))

Balanced Panel: n = 604, T = 5, N = 3020

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-1.9125650 -0.3000987 -0.0052542  0.2982728  1.4975965

Coefficients:
  Estimate Std. Error t-value  Pr(>|t|)
l 0.383210   0.034017  11.265 < 2.2e-16 ***
k 0.470144   0.020274  23.189 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    747.16
Residual Sum of Squares: 587.84
R-Squared:      0.21323
Adj. R-Squared: 0.016052
F-statistic: 327.126 on 2 and 2414 DF, p-value: < 2.22e-16
```

# Random Effects Model (RE)

The RE model was also estimated using the plm package. The model assumed that the individual-specific effects and the explanatory variables were uncorrelated.

The results of the RE model are as follows:

- **Coefficients:** The estimated coefficients for ln(labor) and ln(capital) were 0.3773880 and 0.4722017, respectively, both statistically significant ($p < 0.001$).
- **R-squared:** The RE model explained approximately 21.5% of the variation in ln(output) (adjusted R-squared: 0.21469).
- **Specification Test:** The p-value of the Hausman test (0.9051) indicates that the FE model is consistent with the RE model.

```
> summary(re_model)  # Print the summary of random effects model
Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)

Call:
plm(formula = y ~ l + k, data = prodfn_data, model = "random",
    index = c("ivar", "tvar"))

Balanced Panel: n = 604, T = 5, N = 3020

Effects:
                   var  std.dev share
idiosyncratic 0.243514 0.493471  0.99
individual    0.002483 0.049825  0.01
theta: 0.02455

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-1.7158602 -0.3366432 -0.0079159  0.3379671  1.4849595

Coefficients:
              Estimate Std. Error z-value Pr(>|z|)
(Intercept) -0.0046233  0.0092048 -0.5023   0.6155
l            0.3773880  0.0303618 12.4297   <2e-16 ***
k            0.4722017  0.0180555 26.1528   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    935.59
Residual Sum of Squares: 734.24
R-Squared:      0.21521
Adj. R-Squared: 0.21469
Chisq: 827.338 on 2 DF, p-value: < 2.22e-16
```

# Model assumptions

We performed several diagnostic tests in order to assess the validity of our model assumptions:

- Heteroscedasticity
- Serial correlation
- Normality of residuals
- Time Series Dependency

# Heteroscedasticity

We performed the Breusch-Pagan test for both the FE and RE models, yielding a chi-square test statistic of 0.52101 with a p-value of 0.4704.

The null hypothesis of no heteroscedasticity could not be rejected.

```
> print(bptest_fe)

	Lagrange Multiplier Test - (Breusch-Pagan)

data:  y ~ l + k
chisq = 0.52101, df = 1, p-value = 0.4704
alternative hypothesis: significant effects
```

```
> print(bptest_re)

	Lagrange Multiplier Test - (Breusch-Pagan)

data:  y ~ l + k
chisq = 0.52101, df = 1, p-value = 0.4704
alternative hypothesis: significant effects
```

# Serial Correlation

The Wooldridge test was employed to test for serial correlation in the FE and RE models.

The FE model exhibited evidence of serial correlation ($p < 0.001$), while the RE model did not indicate significant serial correlation ($p = 0.09367$)

```
> print(wooldridge_test_fe)

        Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  y ~ l + k
chisq = 94.076, df = 1, p-value < 2.2e-16
alternative hypothesis: serial correlation in idiosyncratic errors
```

```
> print(wooldridge_test_re)

        Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  y ~ l + k
chisq = 2.8102, df = 1, p-value = 0.09367
alternative hypothesis: serial correlation in idiosyncratic errors
```

# Normality of Residuals

The Shapiro-Wilk test was conducted to assess the normality assumption of the residuals.

Both the FE and RE models indicated a departure from normality ($p < 0.001$).

```
> print(shapiro_test_fe)

	Shapiro-Wilk normality test

data:  residuals(fe_model)
W = 0.99955, p-value = 0.748
```

```
> print(shapiro_test_re)

	Shapiro-Wilk normality test

data:  residuals(re_model)
W = 0.9994, p-value = 0.4727
```

# Time Series Dependency

The Durbin-Watson test was performed to examine time series dependency in the models.

The FE model indicated positive autocorrelation (DW statistic: 0.7168), while the RE model showed no significant autocorrelation (DW statistic: 1.9897).

```
> # Printing the Durbin-Watson test results
> print(dw_test_fe)

        Durbin-Watson test for serial correlation in panel models

data:  y ~ l + k
DW = 2.3522, p-value < 2.2e-16
alternative hypothesis: serial correlation in idiosyncratic errors

> print(dw_test_re)

        Durbin-Watson test for serial correlation in panel models

data:  y ~ l + k
DW = 1.9383, p-value = 0.09
alternative hypothesis: serial correlation in idiosyncratic errors
```

# Conclusion

Our estimation of the production function using panel data techniques provided valuable insights into the relationship between ln(output), ln(labor), and ln(capital).

Both the FE and RE models demonstrated that ln(labor) and ln(capital) had a significant positive impact on ln(output), while the Hausman test indicated that the random effects model is consistent and can be preferred over the fixed effects model.

```
> print(hausman_test)

        Hausman Test

data:  y ~ l + k
chisq = 0.19939, df = 2, p-value = 0.9051
alternative hypothesis: one model is inconsistent
```

# Conclusion

The diagnostic tests revealed some violations of model assumptions, including serial correlation, non-normality of residuals, and time series dependency in the FE model. Despite these violations, however, the RE model was more suitable due to its overall consistency and efficiency.

In summary, this project successfully estimated a production function using panel data techniques, and provided insights into the relationship between inputs (labor and capital) and output, contributing to the understanding of firm-level productivity dynamics.

Further investigation and potential remedies for the violated assumptions could potentially be explored in future research.