ECONOMETRIC MODELS II - CAT II

May 3, 2025

QUESTION ONE

Fixed Effects Model Setup

In general, the Fixed Effects (FE) model is given by;

$$Y_{it} = \alpha_i + \beta_1 X_{1,it} + \beta_2 X_{2,it} + \cdots + \beta_k X_{k,it} + U_{it} \tag{1}$$

for $i = 1, \ldots, n$ (individuals) and $t = 1, \ldots, T$ (time periods), which can be written as;

$$Y_{it} = \alpha_i + \mathbf{X}_{it}'\boldsymbol{\beta} + U_{it} \tag{2}$$

where;

(a) Y_{it} is the dependent variable

(b) α_i is the individual-specific fixed effect

(c) $\mathbf{X}_{it} = (X_{1,it}, X_{2,it}, \ldots, X_{k,it})'$ is a $k \times 1$ vector of regressors

(d) $\boldsymbol{\beta}$ is a $k \times 1$ vector of parameters

(e) U_{it} is the idiosyncratic error.

(i) Deriving the FE estimator, $\hat{\boldsymbol{\beta}}$ (4 Marks)

Step 1: Eliminating α_i using the "Within Transformation"

1. Averaging Equation (2) over time for each individual i, we have;

$$\bar{Y}_i = \alpha_i + \bar{\mathbf{X}}_i'\boldsymbol{\beta} + \bar{U}_i \tag{3}$$

2. Subtracting the time average from Equation (3) [the original equation], we have;

$$Y_{it} - \bar{Y}_i = (\mathbf{X}_{it} - \bar{\mathbf{X}}_i)'\boldsymbol{\beta} + (U_{it} - \bar{U}_i) \tag{4}$$

3. This gives the transformed FE model defined as;

$$\ddot{Y}_{it} = \ddot{\mathbf{X}}_{it}'\boldsymbol{\beta} + \ddot{U}_{it} \tag{5}$$

where;

(a) $\ddot{Y}_{it} = Y_{it} - \bar{Y}_i$
(b) $\ddot{\mathbf{X}}_{it} = \mathbf{X}_{it} - \bar{\mathbf{X}}_i$
(c) $\ddot{U}_{it} = U_{it} - \bar{U}_i$.

Step 2: Estimation
Stacking for all individuals and time periods, the FE estimator is obtained as follows;

1. From Equation (5), we have the estimate of $\hat{\boldsymbol{\beta}}$ defined as;

$$\hat{\boldsymbol{\beta}}_{\text{FE}} \ddot{\mathbf{X}}'_{it} = \ddot{Y}_{it} \tag{6}$$

2. Multiplying both sides of Equation (6) by $\ddot{\mathbf{X}}_{it}$, we have;

$$\hat{\boldsymbol{\beta}}_{\text{FE}} \ddot{\mathbf{X}}_{it} \ddot{\mathbf{X}}'_{it} = \ddot{\mathbf{X}}_{it} \ddot{Y}_{it} \tag{7}$$

3. Summing both sides of Equation (7) over all i and t, we have;

$$\hat{\boldsymbol{\beta}}_{\text{FE}} \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \ddot{\mathbf{X}}_{it} \ddot{\mathbf{X}}'_{it} \right) = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \ddot{\mathbf{X}}_{it} \ddot{Y}_{it} \right) \tag{8}$$

4. And this yields the required estimator for $\hat{\beta}$ as;

$$\hat{\boldsymbol{\beta}}_{\text{FE}} = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \ddot{\mathbf{X}}_{it} \ddot{\mathbf{X}}'_{it} \right)^{-1} \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \ddot{\mathbf{X}}_{it} \ddot{Y}_{it} \right) \tag{9}$$

(ii) Model with Dummy Variables (3 Marks)

Alternatively, dummy variables can be included for each individual i (excluding one or using one as a reference to avoid collinearity). The model (in Equation (5)) then becomes;

$$Y_{it} = \gamma_1 D_{i1} + \gamma_2 D_{i2} + \cdots + \gamma_{n-1} D_{i,n-1} + \mathbf{X}'_{it} \boldsymbol{\beta} + U_{it} \tag{10}$$

where;
(a) $D_{ij} = 1$ if observation i corresponds to individual j, and 0 otherwise

(b) γ_j estimates α_j relative to the base individual (excluded or referenced).

In matrix form, we have;

$$\mathbf{Y} = \mathbf{D}\boldsymbol{\gamma} + \mathbf{X}\boldsymbol{\beta} + \mathbf{U} \tag{11}$$

where;

(b) $\mathbf{D}$ is the $nT \times (n-1)$ dummy matrix

(b) $\boldsymbol{\gamma}$ is the vector of fixed effects

(c) $\mathbf{X}$ is the $nT \times k$ regressor matrix

(d) $\boldsymbol{\beta}$ is the slope coefficient vector.

as required

QUESTION TWO

(i) Copy and Complete Table 1 (2 Marks)

Given that;

(a) Model 1 has log-likelihood $= -11680$

(b) Model 2 has log-likelihood $= -9218$

(c) The p-value is $< 2.2 \times 10^{-16}$

The table 1 is completed by computing the following;

(1) **Degrees of freedom (df)**:

(a) Model 1 having 2 predictors $\rightarrow$ 2 df + 1 intercept = 3 parameters

(b) Model 2 having 4 predictors $\rightarrow$ 4 df + 1 intercept = 5 parameters

(c) Therefore, the Df (difference) $= 5 - 3 = 2$

(2) **Likelihood Ratio Test Statistic (Chi-sq)**:

$$\text{Chisq} = 2\times(\text{LogLik}_{\text{Model 2}} - \text{LogLik}_{\text{Model 1}}) = 2\times(-9218-(-11680)) = 2\times2462 = 4924$$

Therefore, the completed table is given as follows;

Table 1: Likelihood Ratio Test

Model	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	3	-11680			
2	5	-9218	2	4924	$< 2.2 \times 10^{-16}$

(ii) Which Model Fits the Data Better? Explain (2 Marks)

Model 2 fits the data better because the Log-likelihood is higher (-9218 against -11680).

Also, the Likelihood Ratio test yields a very large chi-square value (4924) with 2 df, implying a better fit.

Finally, since the p-value is $< 2.2 \times 10^{-16}$, is highly significant, we conclude that Model 1 is sufficiently, a smaller model. Therefore, Model 2 significantly improves the fit by including SLOPE and ET.

(iii) R Program

```
# (iii-1) Read in the dataset                                    (1 Mark)
data <- read.csv("data.csv")

# (iii-2) Fit the two logistic regression models              (4 Marks)
model1<-glm(STATUS ~ TASP + ELEVATION, data = data, family=binomial)
model2<-glm(STATUS~TASP + SLOPE+ELEVATION+ET, data=data, family= binomial)

# (iii-3) Run the likelihood ratio test                        (2 Marks)
anova(model1, model2, test="Chisq")
```

QUESTION THREE

(i) Model Fit Comparison (2 Marks)

To assess model fit, we compare R-squared, SEs, F-statistics, as in Table 1;

Table 2: Model Fit Statistics

Criterion	**OLS Model**	**Fixed Effects Model**
Residual Std. Error	0.2834	0.283
Multiple R-squared	0.0152	0.9986
Adjusted R-squared	0.0150	0.9986
F-statistic	82.22	3.54×10^{-5}
P-value	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$

From Table 2, the FE model fits the data much better because;

(a) R^2 increased from 0.015 to 0.9986. Thus, the model accounts for almost all the variability in ln(hr).

(b) There is much higher F-statistic for Fixed Effects.

(c) There is very slight improvement in residual standard error.

Therefore, the Fixed Effects account for the unobserved year-specific effects, which are significant. Hence, it is a better fit to the data.

(ii) Fixed Effects Model Equation (3 Marks)

Given the following information;

(a) Dependent variable: ln(hr)

(b) Independent variable: ln(wg)

(c) Fixed Effects: Dummy variables for year

The fitted Fixed Effects model is defined as;

$$\ln(\text{hr})_{it} = \beta_1 \cdot \ln(\text{wg})_{it} + \sum_{y=1979}^{1988} \alpha_y \cdot D_y \tag{12}$$

where;

(a) D_y are year dummies (one omitted to avoid multicollinearity)

(b) $\beta_1 = 0.0825$ is the coefficient of ln(wg)

(c) α_y are the fixed effects for each year

Alternatively, the fitted Fixed Effects model is defined as;

$$\begin{aligned}\ln(\mathrm{hr})_{it} = 0.082531 \cdot \ln(\mathrm{wg})_{it} & \\ + 7.454946 \cdot factor(year)1979 + 7.445318 \cdot factor(year)1980 & \\ + 7.452165 \cdot factor(year)1981 + 7.430298 \cdot factor(year)1982 & \\ + 7.397614 \cdot factor(year)1983 + 7.421926 \cdot factor(year)1984 & \\ + 7.452550 \cdot factor(year)1985 + 7.444488 \cdot factor(year)1986 & \\ + 7.458266 \cdot factor(year)1987 + 7.463137 \cdot factor(year)1988 &\end{aligned} \tag{13}$$

(iii) R Scatterplot Output Sketch (2 Marks)

```
scatterplot(yhat ~ lnwg | LaborSupply$year, boxplots = FALSE,
            xlab = "Hourly wage", ylab = "yhat", smooth = FALSE)
```

Interpretation and Expectations

(a) There should be multiple small scatter plots, specifically, one for each year, showing ln(wg) on the horizontal or x-axis against the fitted values $\hat{y}$, on the vertical or y-axis.

(b) Each panel should show the linear relationship for that year, indicating pure scatter of predicted values.

(c) Expectation: An upward linear trend is expected in each panel, reflecting the positive relationship between wage and hours worked [as indicated by the ln(wg) coefficient].

(iv) R Program

(iv-1) Read Dataset (1 Mark)

```
# Load necessary packages
library(car)      # For scatterplot
library(plm)      # For fixed effects
library(dplyr)    # Optional for data manipulation

# Read the dataset
LaborSupply <- read.csv("LaborSupply.csv")
```

(iv-2) Fit Models (4 Marks)

```
# OLS Model
ols_model <- lm(lnhr ~ lnwg, data = LaborSupply)
summary(ols_model)

# Fixed Effects Model using plm
fe_model <- plm(lnhr ~ lnwg+factor(year), data=LaborSupply, index=c("id", "year"),
            model = "within")
summary(fe_model)
```