HW2 - Econometrics

Himanshu, MDS202327

2025-02-25

Question 1

(a) Assuming the error follows a standard normal distribution (i.e. $\epsilon_i \sim N(0, 1)$ for i = 1, 2,. . ., n), find the probability of success $\Pr(y_i = 1)$? Derive the likelihood function for the ordinal probit model.

Consider the ordinal regression model,

$$z_i = x_i'\beta + \epsilon_i \quad \forall i = 1, \ldots, n$$
$$\gamma_{j-1} < z_i \leq \gamma_j \Rightarrow y_i = j \quad \forall i, j = 1, \ldots, J$$

The probability of success is given by,

$$\begin{aligned}
\Pr(y_i = j) &= \Pr(\gamma_{j-1} < z_i \leq \gamma_j) \\
&= \Pr(z_i \leq \gamma_j) - \Pr(z_i \leq \gamma_{j-1}) \\
&= \Pr(x_i'\beta + \epsilon_i \leq \gamma_j) - \Pr(x_i'\beta + \epsilon_i \leq \gamma_{j-1}) \\
&= \Pr(\epsilon_i \leq \gamma_j - x_i'\beta) - \Pr(\epsilon_i \leq \gamma_{j-1} - x_i'\beta) \\
&= \Phi(\gamma_j - x_i'\beta) - \Phi(\gamma_{j-1} - x_i'\beta)
\end{aligned}$$

The likelihood function is given by,

$$\begin{aligned}
l(\beta, y) &= \prod_{i=1}^{n}\prod_{j=1}^{J}[\Pr(y_i = j|\beta, \gamma)^{\mathrm{I}(y_i=j)}] \\
&= \prod_{i=1}^{n}\prod_{j=1}^{J}[\Phi(\gamma_j - x_i'\beta) - \Phi(\gamma_{j-1} - x_i'\beta)]^{\mathrm{I}(y_i=j)}
\end{aligned}$$

(b) Assuming the error follows a standard logistic distribution (i.e. $\epsilon_i \sim \mathcal{L}(0, 1)$ for i = 1, 2,. . ., n), find the probability of success $\Pr(y_i = 1)$? Derive the likelihood function for the ordinal logit model.

The probability of success is given by,

$$\begin{aligned}
\Pr(y_i = j) &= \Pr(\gamma_{j-1} < z_i \leq \gamma_j) \\
&= \Pr(z_i \leq \gamma_j) - \Pr(z_i \leq \gamma_{j-1}) \\
&= \Pr(x_i'\beta + \epsilon_i \leq \gamma_j) - \Pr(x_i'\beta + \epsilon_i \leq \gamma_{j-1}) \\
&= \Pr(\epsilon_i \leq \gamma_j - x_i'\beta) - \Pr(\epsilon_i \leq \gamma_{j-1} - x_i'\beta) \\
&= \Lambda(\gamma_j - x_i'\beta) - \Lambda(\gamma_{j-1} - x_i'\beta)
\end{aligned}$$

The likelihood function is given by,

$$
\begin{aligned}
l(\beta, y) &= \prod_{i=1}^{n}\prod_{j=1}^{J}[\Pr(y_i = j|\beta,\gamma)^{\mathrm{I}(y_i=j)}] \\
&= \prod_{i=1}^{n}\prod_{j=1}^{J}[\Lambda(\gamma_j - x_i'\beta) - \Lambda(\gamma_{j-1} - x_i'\beta)]^{\mathrm{I}(y_i=j)}
\end{aligned}
$$

(c) Consider the ordinal probit model. Show that adding a constant c to the cut-point γ_j and the mean $x_i'\beta$, does not change the outcome probability. How do we solve this first identification problem?

Replace γ_j^* by $\gamma_j + c$, γ_{j-1}^* by $\gamma_{j-1} + c$ and $(x_i'\beta)^*$ by $x_i'\beta + c$

The probability of success is given by,

$$
\begin{aligned}
\Pr(y_i = j) &= \Pr(\gamma_{j-1}^* < z_i \leq \gamma_j^*) \\
&= \Pr(z_i \leq \gamma_j^*) - \Pr(z_i \leq \gamma_{j-1}^*) \\
&= \Pr((x_i'\beta)^* + \epsilon_i \leq \gamma_j^*) - \Pr((x_i'\beta)^* + \epsilon_i \leq \gamma_{j-1}^*) \\
&= \Pr(x_i'\beta + c + \epsilon_i \leq \gamma_j + c) - \Pr(x_i'\beta + c + \epsilon_i \leq \gamma_{j-1} + c) \\
&= \Pr(x_i'\beta + \epsilon_i \leq \gamma_j) - \Pr(x_i'\beta + \epsilon_i \leq \gamma_{j-1}) \\
&= \Pr(\epsilon_i \leq \gamma_j - x_i'\beta) - \Pr(\epsilon_i \leq \gamma_{j-1} - x_i'\beta) \\
&= \Phi(\gamma_j - x_i'\beta) - \Phi(\gamma_{j-1} - x_i'\beta)
\end{aligned}
$$

Different combinations of (β, γ) can produce the same outcome probabilities giving rise to parameter identification problem. We need to anchor the location of the distribution to identify the model parameters by setting $\gamma_1 = 0$.

(d) Once again, consider the ordinal probit model. Show that rescaling the parameters (γ_j, β) and the scale of the distribution by some arbitrary constant d lead to same outcome probabilities. How do we solve the second identification problem?

Replace γ_j^* by $\gamma_j * d$, γ_{j-1}^* by $\gamma_{j-1} * d$ and $(x_i'\beta)^*$ by $x_i'\beta * d$

Note that scaling ϵ_i results in $\epsilon_i^* \sim N(0, d^2)$. Also $\frac{\epsilon_i^*}{d} \sim N(0,1)$

The probability of success is given by,

$$
\begin{aligned}
\Pr(y_i = j) &= \Pr(\gamma_{j-1}^* < z_i \leq \gamma_j^*) \\
&= \Pr(z_i \leq \gamma_j^*) - \Pr(z_i \leq \gamma_{j-1}^*) \\
&= \Pr((x_i'\beta)^* + \epsilon_i^* \leq \gamma_j^*) - \Pr((x_i'\beta)^* + \epsilon_i^* \leq \gamma_{j-1}^*) \\
&= \Pr((x_i'\beta)d + \epsilon_i^* \leq (\gamma_j)d) - \Pr((x_i'\beta)d + \epsilon_i^* \leq (\gamma_{j-1})d) \\
&= \Pr(\epsilon_i^* \leq (\gamma_j - x_i'\beta)d) - \Pr(\epsilon_i^* \leq (\gamma_{j-1} - x_i'\beta)d) \\
&= \Pr(\frac{\epsilon_i^*}{d} \leq \gamma_j - x_i'\beta) - \Pr(\frac{\epsilon_i^*}{d} \leq \gamma_{j-1} - x_i'\beta) \\
&= \Phi(\gamma_j - x_i'\beta) - \Phi(\gamma_{j-1} - x_i'\beta)
\end{aligned}
$$

Different combinations of (β, γ) can produce the same outcome probabilities giving rise to parameter identification problem. We need to anchor the scale of the distribution to identify the model parameters by assuming $\text{var}(\epsilon_i) = 1$.

(e) Consider the data present in the file `Feb14Data.xlsx`. This file contains 1,492 observations from the February 2014 Political Survey conducted during February 14 - 23, 2014, by the Princeton Survey Research Associates and sponsored by the Pew Research Center for the People and the Press. Based on this data, do the following.

(i) Present a descriptive summary of the data as in Table 1 of the lecture slides.

```r
# Import necessary libraries
library(dplyr)
library(knitr)
library(MASS)
library(plm)

# Read the data
feb14data = read.csv("Feb14Data.csv", header = TRUE)


# List of tolerant states
tolerant = c('Alaska', 'Arizona', 'California', 'Colorado', 'Connecticut',
             'Delaware', 'Hawaii', 'Illinois', 'Maine', 'Maryland',
             'Massachusetts', 'Michigan', 'Montana', 'Nevada',
             'New Hampshire', 'New Jersey', 'New Mexico', 'Oregon',
             'Rhode Island', 'Vermont', 'Washington', 'Washington DC',
             'District of Columbia')


# Mappings to group categories
public_opinion_map = c("medicinal"="Legal only for medicinal use",
                       "notlegal"="Oppose Legalization",
                       "personal"="Legal for personal use")


religion_keys = c(paste0("Protestant (Baptist, Methodist, Non-denominational, ",
                  "Lutheran, Presbyterian, Pentecostal, Episcopalian, ",
                  "Reformed, etc.)"),
                  "Nothing in particular",
                  "Roman Catholic (Catholic)",
                  "Agnostic (not sure if there is a God)",
                  "Christian (VOL.)",
                  paste0("Mormon (Church of Jesus Christ of Latter-day ",
                  "Saints/LDS)"),
                  "Jewish (Judaism)",
                  "Hindu",
                  "Buddhist",
                  "Atheist (do not believe in God)",
                  paste0("Orthodox (Greek, Russian, or some other ",
                  "orthodox church)"),
                  "Unitarian (Universalist) (VOL.)",
                  "Muslim (Islam)")
```

```
# Mappings to group categories
religion_values = c("Protestant", "Liberal", "Roman Catholic",
                    "Liberal", "Christian", "Conservative",
                    "Conservative", "Conservative", "Conservative",
                    "Liberal", "Conservative", "Liberal", "Conservative")


religion_map = setNames(religion_values, religion_keys)


education_map = c("HS"="High School & Below",
                  "Some College"="Below Bachelors",
                  "Postgraduate Degree"="Bachelors & Above",
                  "Associate Degree"="Below Bachelors",
                  "HS Incomplete"="High School & Below",
                  "Less than HS"="High School & Below",
                  "Four year college"="Bachelors & Above",
                  "Some Postgraduate"="Bachelors & Above")


race_map = c("White"="White",
             "Black or African-American"="African American",
             "Hispanic or Latino"="Other Races",
             "Asian or Asian-American"="Other Races",
             "Native American"="Other Races")


income_map = c("Less than $10,000"=log(5000),
               "10 to under $20,000"=log(15000),
               "20 to under $30,000"=log(25000),
               "30 to under $40,000"=log(35000),
               "40 to under $50,000"=log(45000),
               "50 to under $75,000"=log(62500),
               "75 to under $100,000"=log(87500),
               "100 to under $150,000"=log(125000),
               "$150,000 or more"=log(150000))


party_map = c("No preference (VOL.)"="Independent & Others",
              "Independent"="Independent & Others",
              "Democrat"="Democrat",
              "Republican"="Republican",
              "Other party (VOL.)"="Independent & Others")


map = list("Public.Opinion"=public_opinion_map,
        "Religion"=religion_map,
        "Education"=education_map,
        "Race"=race_map,
        "Income"=income_map,
        "Party.Affiliation"=party_map)
```

```r
# Transform state
state_list = c()
for (state in feb14data$State){
    if (state %in% tolerant){
        state_list = append(state_list, 'Tolerant')
    }
    else{
        state_list = append(state_list, 'Not Tolerant')
    }
}


feb14data$State = state_list


# Log to the age
feb14data$Age = log(feb14data$Age)


# Transform Public Opinion, Religion, Education, Race, Income, Party Affiliation
for (col in names(map)) {
  feb14data[[col]] <- map[[col]][as.character(feb14data[[col]]) ]
}


# Return data frame with descriptive statistics accordingly
for (col in colnames(feb14data)){
    df = data.frame()
    vec = feb14data[,col]

    if (typeof(vec) == "character"){
        for (i in unique(vec)){
            out = c(i, sum(vec==i), round(sum(vec==i)*100/length(vec),2))
            df = rbind(df, out)
        }
    colnames(df) = c(col, 'Counts', 'Percentage')
    print(df)
    }

    if (typeof(vec)=="integer" | typeof(vec) == "double"){
        out = c(col, round(mean(vec),2), round(sd(vec),2))
        df = rbind(df, out)
    colnames(df) = c('Column', 'Mean', 'Std. Deviation')
    print(df)
    }
}
```

Variable	Category	Counts	Percentage
Public Opinion	Legal only for medicinal use	640	42.9
	Oppose Legalization	218	14.61
	Legal for personal use	634	42.49
Past Use	Yes	719	48.19
Religion	Protestant	550	36.86
	Liberal	348	23.32
	Roman Catholic	290	19.44
	Christian	182	12.20
	Conservative	122	8.18
Sex	Male	791	53.02
Education	High School & Below	507	33.98
	Below Bachelors	434	29.09
	Bachelors & Above	551	36.93
Race	White	1149	77.01
	African American	202	13.54
	Other Races	141	9.45
Party Affiliation	Independent & Others	648	43.43
	Democrat	511	34.25
	Republican	333	22.32
State	Tolerant	556	37.27
Eventually Legalized	Yes, it will	1154	77.35

Table 1: Descriptive summary of ordinal variables.

Variable	Mean	Std. Dev
log(Age)	3.72	0.44
log(Income)	10.63	0.98
Household Size	2.74	1.42

Table 2: Descriptive summary of continuous variables.

(ii) Please use the data to analyze public opinion on extent marijuana legalization in the US i.e., estimate Model 8 and replicate the results presented in Table 2 in lecture slides.

```
# Order the vectors accordingly
feb14data$Public.Opinion = factor(feb14data$Public.Opinion,
      levels = c('Oppose Legalization',
              'Legal only for medicinal use',
              'Legal for personal use'),
      ordered = TRUE)


feb14data$Education = factor(feb14data$Education,
                      levels = c('High School & Below',
                                 'Below Bachelors',
                                 'Bachelors & Above'))
edu_contr = contr.treatment(n = 3, base = 1)
colnames(edu_contr) = c(".Below Bachelors",
                            ".Bachelors & Above")
contrasts(feb14data$Education) = edu_contr
```

```r
# Order the vectors accordingly
feb14data$Race = factor(feb14data$Race,
                        levels = c('White', 'African American',
                                   'Other Races'))

race_contr = contr.treatment(n = 3, base = 1)
colnames(race_contr) = c('.African American', '.Other Races')
contrasts(feb14data$Race) = race_contr

feb14data$Party.Affiliation = factor(feb14data$Party.Affiliation,
                        levels = c('Republican', 'Democrat',
                                   'Independent & Others'))

party_contr = contr.treatment(n = 3, base = 1)
colnames(party_contr) = c('.Democrat', '.Independent & Others')
contrasts(feb14data$Party.Affiliation) = party_contr

feb14data$Religion = factor(feb14data$Religion,
                        levels = c('Protestant', 'Roman Catholic',
                                   'Christian', 'Conservative', 'Liberal'))

rel_contr = contr.treatment(n = 5, base = 1)
colnames(rel_contr) = c('.Roman Catholic', '.Christian',
                        '.Conservative', '.Liberal')
contrasts(feb14data$Religion) = rel_contr

# Ordinal Probit Model
model8 = polr(Public.Opinion ~ Age + Income + Past.Use + Sex + Education +
              Household.Size + State + Eventually.Legalized + Race +
              Party.Affiliation + Religion, data = feb14data,
              method = 'probit', Hess=TRUE)

# Estimates from summary
ctable = coef(summary(model8))
p = pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2
ctable = cbind(ctable, "p value" = p)

# Intercept Estimate
int = -model8$zeta[1]
names(int) = 'Intercept'

# Cut point estimate
cp = model8$zeta[2] - model8$zeta[1]
names(cp) = 'Cut-point'

# Intercept Std. Error
se_int = ctable[,'Std. Error'][18]

# Cut point Std. Error
cov12 = vcov(model8)[18,19]
se1 = ctable[,'Std. Error'][18]
se2 = ctable[,'Std. Error'][19]
se_cp = sqrt(se1^2 + se2^2 - 2 * cov12)
```

```r
# Intercept t-value and p-value
t_int = int / se_int
p_int = 2 * pnorm(abs(t_int), lower.tail = FALSE)

# Cut point t-value and p-value
t_cp = cp / se_cp
p_cp = 2 * pnorm(abs(t_cp), lower.tail = FALSE)

# LR Statistic
model_red = polr(factor(Public.Opinion)~1, data = feb14data,
                     Hess = TRUE, method = 'probit')
lr = -2*(logLik(model_red) - logLik(model8))

# McFadden R^2
Rm = 1 - logLik(model8)/logLik(model_red)

# Hit Rate
pred = predict(model8,type = "class")
unique_vals = unique(pred)
num_map = seq_along(unique_vals)
hit_rate = 100*mean(match(pred, unique_vals) ==
            match(feb14data$Public.Opinion, unique_vals))

summ2df = function(coefs) {
    est = coefs[, "Value"]
    est = append(int, est)
    est = append(est, cp)

    se = coefs[, "Std. Error"]
    se = append(se_int, se)
    se = append(se, se_cp)

    t = coefs[, "t value"]
    t = append(t_int, t)
    t = append(t, t_cp)

    p = coefs[,"p value"]
    p = append(p_int, p)
    p = append(p, p_cp)

    df = data.frame(est = est, se = se, t = t, p = p)
    colnames(df) = c('Estimate', 'Std. Error', 't-value', 'p-value')
    return(df)
}

df = summ2df(ctable)
df = df %>% mutate_if(is.numeric, round, digits=2)
df
```

Variable	Estimate	Std. Dev	t-value	p-value
(Intercept)	0.34	0.48	0.72	0.472
Log(Age)	-0.35**	0.08	-4.51	0.000
Log(Income)	0.09**	0.04	2.47	0.014
Past Use	0.69**	0.06	10.67	0.000
Male	0.06	0.06	1.00	0.319
Below Bachelors	0.05	0.08	0.62	0.538
Bachelors & Above	0.24**	0.08	3.01	0.003
Household Size	-0.02	0.02	-0.87	0.382
Tolerant States	0.07	0.07	1.04	0.298
Eventually Legal	0.57**	0.07	7.76	0.000
African American	0.03	0.10	0.26	0.795
Other Races	-0.28**	0.11	-2.56	0.010
Democrat	0.44**	0.09	5.03	0.000
Independent & Others	0.36**	0.08	4.56	0.000
Roman Catholic	0.10	0.09	1.19	0.234
Christian	0.16	0.10	1.60	0.110
Conservative	0.09	0.12	0.76	0.447
Liberal	0.39**	0.09	4.39	0.000
Oppose|Medicinal	-0.35	0.48	-0.72	0.472
Medicinal|Personal	1.12**	0.48	2.33	0.020
Cut Point	1.46**	0.05	29.58	0.000

Measure	Value
LR χ^2 Statistics	377.001 (df=2)
McFadden's R^2	0.125 (df=19)
Hit Rate	58.914

Table 3: Result summary of model 8.

(iii) Compute the covariate effects for variables presented in Table 3.

```
compute_ce = function(cov, num = FALSE, delta=1){
    X = model.matrix(model8, temp)
    X = X[,2:dim(X)[2]]
    X0 = X
    X1 = X
    if (num == FALSE){
        X1[,cov] = 1
        X0[,cov] = 0
    }
    else {
        X1[,cov] = log(exp(X[,cov]) + delta)
    }
    pred_x1 = X1 %*% beta_hat
    pred_x0 = X0 %*% beta_hat

    prob_x1 = pnorm(outer(gamma_hat, pred_x1, "-"))
```

```r
    prob_x0 = pnorm(outer(gamma_hat, pred_x0, "-"))

    prob_x1 = t(apply(prob_x1, 2, function(p) c(p[1], diff(p))))
    prob_x0 = t(apply(prob_x0, 2, function(p) c(p[1], diff(p))))

    ce = prob_x1 - prob_x0
    avg_ce = colMeans(ce)
    return(round(avg_ce,3))
}

# Convert categorical variables to factors
temp = feb14data %>% mutate(across(c(Public.Opinion, Past.Use,Religion, Sex,
        Education, Race, Party.Affiliation, Eventually.Legalized), as.factor))

beta_hat = coef(model8)
gamma_hat = c(model8$zeta[1],model8$zeta[2], Inf)

df = data.frame()
out = compute_ce('Age', num = TRUE, delta = 10)
df = rbind(df, append('Age', out))
out = compute_ce('Income', num = TRUE, delta = 10000)
df = rbind(df, append('Income', out))

cov_list = c('Past.UseYes','Education.Bachelors & Above',
             'Eventually.LegalizedYes, it will', 'Race.Other Races',
             'Party.Affiliation.Democrat',
             'Party.Affiliation.Independent & Others', 'Religion.Liberal')

for (cov in cov_list){
    out = compute_ce(cov)
    df = rbind(df, append(cov, out))
}

colnames(df) = c('Covariate', 'P(not legal)', 'P(medicinal use)', 'P(personal use')
print(df)
```

Covariate	**ΔP(not legal)**	**ΔP(medicinal use)**	**ΔP(personal use)**
Log(Age)	0.015	0.012	-0.028
Log(Income)	-0.005	-0.003	0.008
Past Use	-0.129	-0.113	0.243
Bachelors & Above	-0.045	-0.035	0.08
Eventually Legal	-0.126	-0.06	0.186
Other Races	0.059	0.031	-0.089
Democrat	-0.08	-0.066	0.147
Independent & Others	-0.07	-0.051	0.121
Liberal	-0.068	-0.066	0.134

Table 4: Average covariate effects from Model 8.

Question 2

```r
firm = read.csv('Grunfeld220obs.csv')

# Remove American Steel
pos_idx = firm$firm!="American Steel"
firm = firm[pos_idx,]

# Checking the data is balanced or unbalanced
firm %>%is.pbalanced()
```

(a) Estimate a pooled effects model on the covariates: `value` and `capital`. Summarize the results and interpret the coefficients.

```r
# Pooled Effects Model
pe_model = plm(inv ~ value + capital, data = firm, index = c("firm", "year"),
               effect = "individual", model = "pooling")
summary(pe_model)
```

```
## Pooling Model
##
## Call:
## plm(formula = inv ~ value + capital, data = firm, effect = "individual",
##     model = "pooling", index = c("firm", "year"))
##
## Balanced Panel: n = 10, T = 20, N = 200
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -291.6757  -30.0137    5.3033   34.8293  369.4464
##
## Coefficients:
##                Estimate  Std. Error t-value  Pr(>|t|)
## (Intercept) -42.7143694   9.5116760 -4.4907 1.207e-05 ***
## value         0.1155622   0.0058357 19.8026 < 2.2e-16 ***
## capital       0.2306785   0.0254758  9.0548 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    9359900
## Residual Sum of Squares: 1755900
## R-Squared:      0.81241
## Adj. R-Squared: 0.8105
## F-statistic: 426.576 on 2 and 197 DF, p-value: < 2.22e-16
```

Summary of Results

The pooled effects model assumes that all firms are homogeneous and does not account for individual firm-specific effects and hence coefficient estimates can be interpreted the same way as OLS estimates.

$\mathbf{R^2}$ **(`0.812`)** indicates **81.2%** of the variation in investment is explained by the model.

F-statistic (`426.576`) with significant **p-value**, confirms that the model as whole is statistically significant.

Intercept (`-42.71`): This suggests that when `value` and `capital` are both zero, the predicted level of investment is -42.71. While this has no practical interpretation (since investment cannot be negative), it serves as the baseline for the regression model.

Value (`0.1156`): For every one-unit increase in `value`, investment (`inv`) is expected to increase by 0.1156 units, holding `capital` constant. The coefficient is *highly significant*, suggesting that higher `value` levels are strongly associated with higher investment.

Capital (`0.2307`): For every one-unit increase in `capital`, investment increases by 0.2307 units, holding `value` constant. This coefficient is also *highly significant*, suggesting that higher `capital` levels are strongly associated with higher investment.

(b) Now, consider the panel structure of the data. Estimate a fixed-effects model, using the `plm` function, by regressing `inv` on `value` and `capital`. Do not ignore the indexing of data by `firm` and `year`. Summarize the results and interpret the coefficients.

```r
# Fixed Effects Model
fe_model = plm(inv ~ value + capital, data = firm, index = c("firm", "year"),
                effect = "individual", model = "within")
summary(fe_model)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = inv ~ value + capital, data = firm, effect = "individual",
##     model = "within", index = c("firm", "year"))
##
## Balanced Panel: n = 10, T = 20, N = 200
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -184.00857  -17.64316    0.56337   19.19222  250.70974
##
## Coefficients:
##         Estimate Std. Error t-value  Pr(>|t|)
## value   0.110124   0.011857  9.2879 < 2.2e-16 ***
## capital 0.310065   0.017355 17.8666 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    2244400
## Residual Sum of Squares: 523480
## R-Squared:      0.76676
## Adj. R-Squared: 0.75311
## F-statistic: 309.014 on 2 and 188 DF, p-value: < 2.22e-16
```

Summary of Results

The fixed effects model estimates the relationship between the dependent variable (**`inv`**) and the independent variables (**`value and capital`**) accounting for firm-specific unobserved heterogeneity and hence the interpretation changes as the coefficients capture the within-group effect.

R^2 (`0.7667`) indicates **76.67%** of the variation in investment is explained by the model after accounting for firm-specific effects.

F-statistic (`309.014`) with significant **p-value**, confirms that the model as whole is statistically significant.

Value (`0.1101`): For every one-unit increase in `value`, investment (`inv`) is expected to increase by 0.1101 units, holding `capital` constant and accounting for time-invariant firm-specific factors. The coefficient is *highly significant*, suggesting that higher `value` levels are strongly associated with higher investment.

Capital (`0.3101`): For every one-unit increase in `capital`, investment increases by 0.3101 units, holding `value` constant and accounting for time-invariant firm-specific factors. This coefficient is also *highly significant*, suggesting that higher `capital` levels are strongly associated with higher investment.

The FE model removes time-invariant firm-specific characteristics, even though `value` and `capital` are not time-invariant, FE only exploits within-firm variations over time, ignoring between-firm variation. This means that the interpretation is still within a firm rather than across firms. The results suggest that investment is positively influenced by both firm value and capital stock.

(c) Once again, consider the panel structure of the data. Estimate a random-effects model, using the `plm` function, by regressing `inv` on `value` and `capital`. Do not ignore the indexing of data by `firm` and `year`. Summarize the results and interpret the coefficients.

```r
# Random Effects Model
re_model = plm(inv ~ value + capital -1, data = firm, index = c("firm", "year"),
                effect = "individual", model = "random")
summary(re_model)
```

```
## Oneway (individual) effect Random Effect Model
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = inv ~ value + capital - 1, data = firm, effect = "individual",
##     model = "random", index = c("firm", "year"))
##
## Balanced Panel: n = 10, T = 20, N = 200
##
## Effects:
##                    var std.dev share
## idiosyncratic 2784.46   52.77 0.281
## individual    7123.06   84.40 0.719
## theta: 0.8615
##
## Residuals:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -177.63  -27.14   -2.65   -6.94   11.60  258.93
##
## Coefficients:
##         Estimate Std. Error z-value  Pr(>|z|)
## value   0.102822   0.009976  10.307 < 2.2e-16 ***
## capital 0.307445   0.017305  17.766 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    2380800
## Residual Sum of Squares: 559900
## R-Squared:      0.76921
## Adj. R-Squared: 0.76804
## Chisq: 672.801 on 2 DF, p-value: < 2.22e-16
```

Summary of Results

The random effects model assumes that firm-specific effects are uncorrelated with `value` and `capital`, which is not necessarily true in real-world data. Unlike fixed effects model, random effects model uses both within-firm and between-firm variation to estimate coefficients. Since the independent variables `value` and `capital` are not time-invariant, random effects model's estimates will be similar to fixed effects model only if the assumption of no correlation between firm effects and explanatory variables holds (which can be tested using the Hausman test).

The **theta coefficient (`0.8615`)** suggests that a substantial portion of the variation is due to individual (firm-level) effects rather than purely time-specific fluctuations.

The **idiosyncratic error variance** (within-firm variation) is `2784.46` with a standard deviation of `52.77`.

The **individual (firm-specific) variance** is `7123.06` with a standard deviation of `84.40`.

R^2 (`0.769`) indicates **76.9%** of the variation in investment is explained by the model.

χ^2 statistic (`672.801`) with significant **p-value**, confirms that the model as whole is statistically significant.

Value (`0.1028`): For every one-unit increase in `value`, investment (`inv`) is expected to increase by 0.1028 units, holding `capital` constant, while accounting for both within-firm and between-firm variations. The coefficient is *highly significant*, suggesting that higher `value` levels are strongly associated with higher investment.

Capital (`0.3074`): For every one-unit increase in `capital`, investment increases by 0.3074 units, holding `value` constant, while accounting for both within-firm and between-firm variations. This coefficient is also *highly significant*, suggesting that higher `capital` levels are strongly associated with higher investment.