Thesis code knit

2024-01-05

```
library(pacman)
p_load(haven, tidyverse, prodest, estprod, plm, huxtable)
```

```
#load 2009-2013 Enterprise Survey data
data1 <- read_dta("C:\\Users\\Aayush\\Documents\\files prior to 1-7-2024\\Nepal_2009_2013.dta")
```

```
data2 <- data1 %>%
  # Select only manufacturing firms
  filter(a0 == 1) %>%
  #Select only rows valid for balanced panel
  group_by(id2009) %>%
  filter(all(c(2009, 2013) %in% year)) %>%
  ungroup()
```

```
data3 <- data2 %>%

  #select necessary columns for data analysis
  select(year, id2009, d2, n7a, n2a, n2e,e11,b7,k8, a6b,j30c,j30a,l1,b5,l4a,b7,d3c,e6,b2b,c30a,
         e1) %>%

  #filter rows with values greater than or equal to 0
  filter(if_all(c(d2,n7a,n2a,n2e,e11,b7,k8,a6b,j30c,j30a,l1,b5,l4a,b7,d3c,e6,b2b,c30a,e1), ~.>= 0)) %>%

  #adding no. of years of operation column to the data
  mutate(yofop = ifelse(year == 2009, 2009 - b5, ifelse(year == 2013, 2013 - b5, NA))) %>%

  #renaming columns
  rename(sales = d2, capital = n7a, labor = n2a, interm = n2e, ID = id2009, Informal="e11", Experience=
         Credit="k8",Size="a6b", Foreigntech="e6", Bussiness_permit="j30c", Tax_burden="j30a", local="e

  #take natural log of certain columns
  mutate(across(c(sales, capital,labor,interm), ~log(.))) %>%

  #Adjust for inflation for monetary values
  mutate(
    across(c(sales, capital, labor, interm),
           ~ifelse(year == 2013, (./142.52)*100, .))) %>%

  #Create dummy variables out of ordinal variables
  mutate(across(c(Informal, Credit, local, Foreigntech),
           ~case_when(. == 1 ~ 1,
                      TRUE ~ 0)),
    a6b = case_when(Size %in% c(1,2) ~ 1,
                    TRUE ~ 0))
```

```r
#levinsohn model
levinsohn_model <- levinsohn_petrin(data = data3, sales ~ labor | capital | interm,
                 id = "ID", time = "year", bootstrap = TRUE)
#olleypakes
olleypakes_model <- olley_pakes(data = data3, sales ~ labor | capital | interm,
                id = "ID", time = "year", bootstrap = TRUE)
```

```r
#filter again with coefficients
data4 <- data3%>%
  mutate(va=sales-interm) %>%
  mutate(logtfp=va-((levinsohn_model$t0[1])*labor)-((levinsohn_model$t0[2])*capital)) %>%
  mutate(avetfp=scale(logtfp))
```

```r
summary(data4)
```

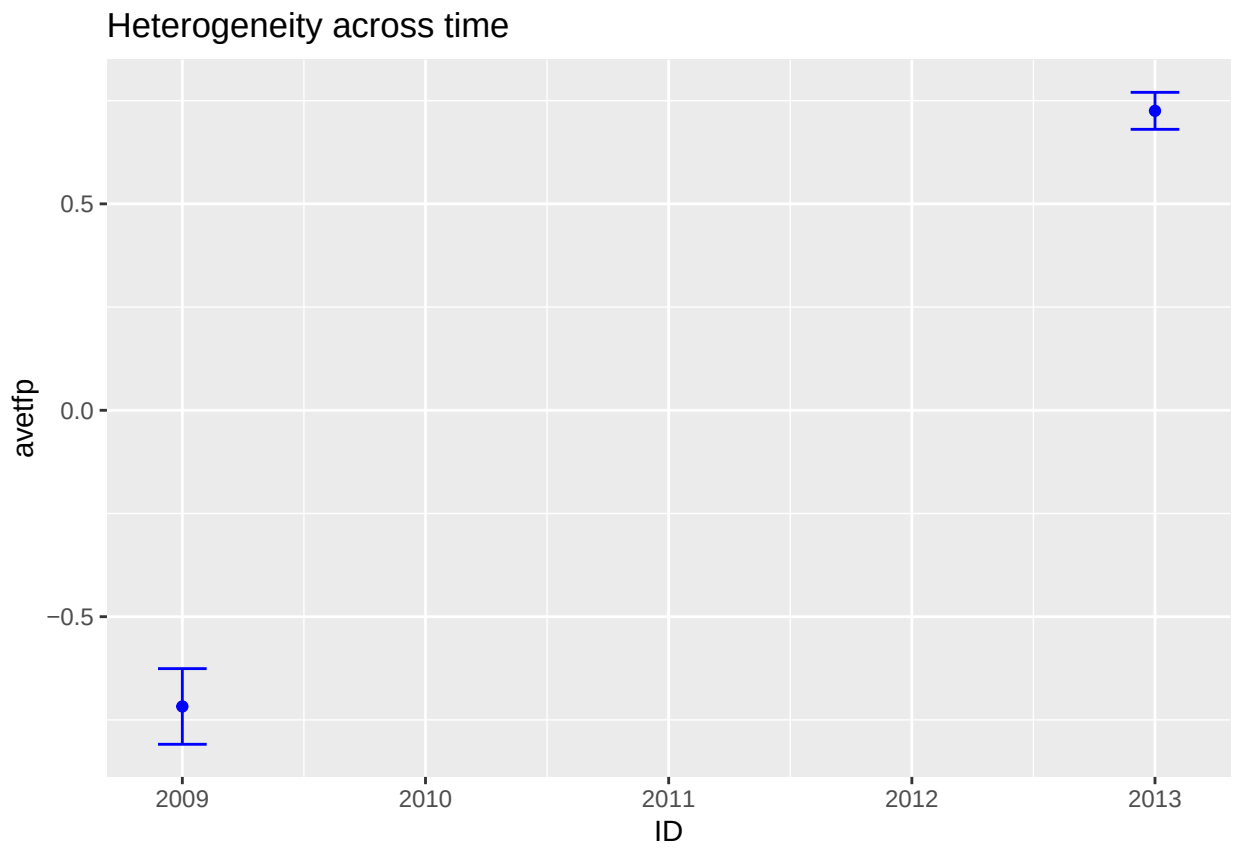
```
##       year            ID                sales           capital
##  Min.   :2009   Min.   :1.011e+09   Min.   : 9.133   Min.   : 8.078
##  1st Qu.:2009   1st Qu.:2.090e+09   1st Qu.:11.575   1st Qu.:11.005
##  Median :2009   Median :2.099e+09   Median :13.862   Median :12.439
##  Mean   :2011   Mean   :2.312e+09   Mean   :14.128   Mean   :13.200
##  3rd Qu.:2013   3rd Qu.:3.110e+09   3rd Qu.:16.208   3rd Qu.:15.607
##  Max.   :2013   Max.   :3.140e+09   Max.   :22.669   Max.   :19.807
##      labor           interm          Informal        Experience
##  Min.   : 8.363   Min.   : 8.112   Min.   :0.0000   Min.   : 1.0
##  1st Qu.:10.337   1st Qu.:11.069   1st Qu.:0.0000   1st Qu.:10.0
##  Median :12.439   Median :13.209   Median :0.0000   Median :15.0
##  Mean   :12.467   Mean   :13.533   Mean   :0.3443   Mean   :17.9
##  3rd Qu.:14.327   3rd Qu.:15.830   3rd Qu.:1.0000   3rd Qu.:25.0
##  Max.   :18.603   Max.   :22.515   Max.   :1.0000   Max.   :43.0
##      Credit            Size       Bussiness_permit   Tax_burden
##  Min.   :0.0000   Min.   :1.000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:1.500   1st Qu.:0.0000   1st Qu.:0.000
##  Median :1.0000   Median :2.000   Median :0.0000   Median :1.000
##  Mean   :0.5137   Mean   :1.902   Mean   :0.6557   Mean   :1.311
##  3rd Qu.:1.0000   3rd Qu.:2.000   3rd Qu.:1.0000   3rd Qu.:2.000
##  Max.   :1.0000   Max.   :3.000   Max.   :4.0000   Max.   :4.000
##       l1               b5            l4a              d3c
##  Min.   :  4.00   Min.   :1959   Min.   :  1.00   Min.   :  0.000
##  1st Qu.: 15.00   1st Qu.:1985   1st Qu.:  5.00   1st Qu.:  0.000
##  Median : 25.00   Median :1994   Median : 10.00   Median :  0.000
##  Mean   : 58.62   Mean   :1992   Mean   : 25.98   Mean   :  6.388
##  3rd Qu.: 48.50   3rd Qu.:2000   3rd Qu.: 20.00   3rd Qu.:  0.000
##  Max.   :900.00   Max.   :2007   Max.   :750.00   Max.   :100.000
##   Foreigntech           b2b              c30a          local
##  Min.   :0.00000   Min.   : 0.000   Min.   :0.00   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.: 0.000   1st Qu.:2.00   1st Qu.:0.0000
##  Median :0.00000   Median : 0.000   Median :3.00   Median :0.0000
##  Mean   :0.06557   Mean   : 1.574   Mean   :2.76   Mean   :0.2787
##  3rd Qu.:0.00000   3rd Qu.: 0.000   3rd Qu.:4.00   3rd Qu.:1.0000
##  Max.   :1.00000   Max.   :80.000   Max.   :4.00   Max.   :1.0000
##      yofop            a6b              va              logtfp
##  Min.   : 2.00   Min.   :0.000   Min.   :-2.3026   Min.   :-8.666
##  1st Qu.:12.00   1st Qu.:1.000   1st Qu.: 0.3248   1st Qu.:-5.563
```

```
##  Median :17.00   Median :1.000   Median : 0.4864   Median :-4.406
##  Mean   :18.92   Mean   :0.847   Mean   : 0.5955   Mean   :-4.724
##  3rd Qu.:25.50   3rd Qu.:1.000   3rd Qu.: 0.8473   3rd Qu.:-3.860
##  Max.   :54.00   Max.   :1.000   Max.   : 2.7726   Max.   :-2.547
##       avetfp.V1
##  Min.   :-3.443176
##  1st Qu.:-0.732491
##  Median : 0.277948
##  Mean   : 0.000000
##  3rd Qu.: 0.754572
##  Max.   : 1.901745
```

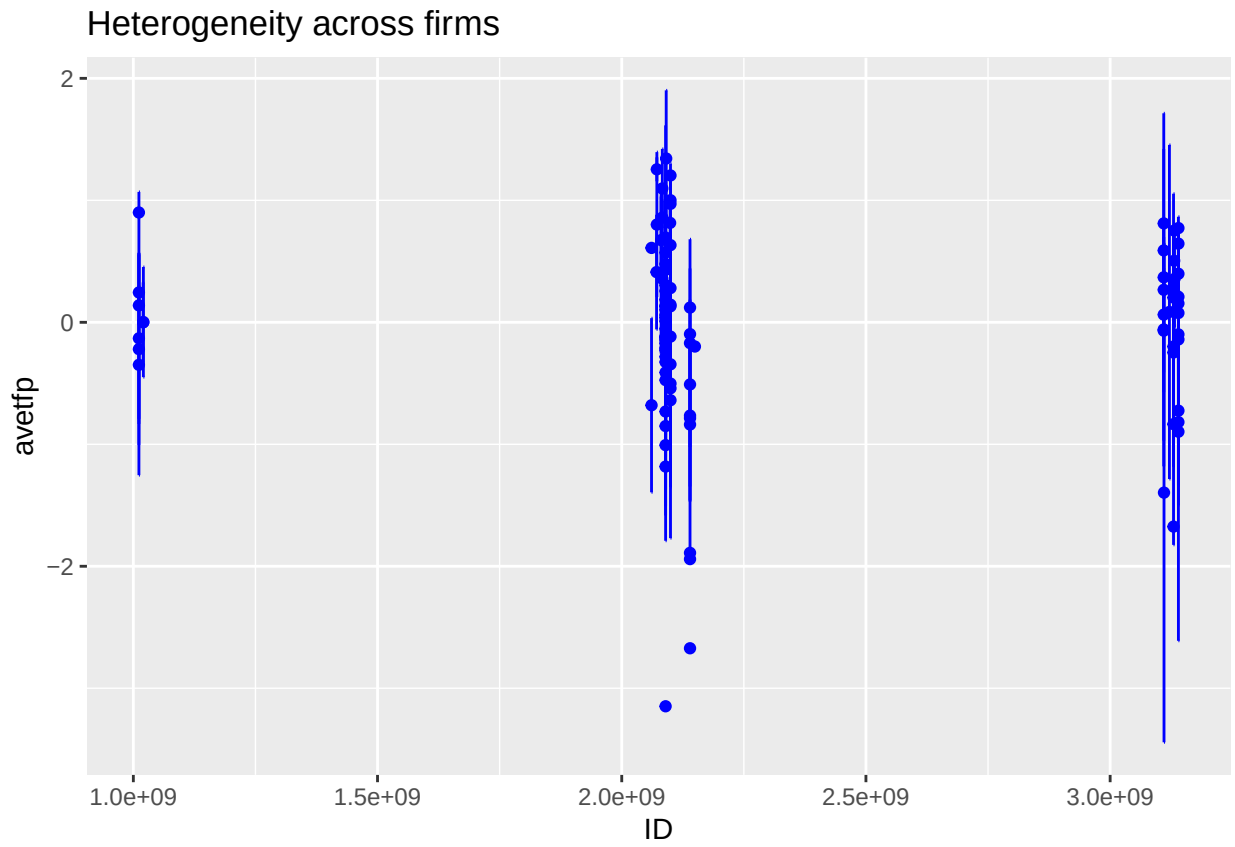
```
cor(data4[, c("avetfp", "Informal", "Experience", "Credit", "Size", "Foreigntech", "Tax_burden", "Bussi
```

```
##                      avetfp     Informal  Experience       Credit        Size
## avetfp            1.00000000  0.182034592  0.08187739 -0.264366027 -0.23966981
## Informal          0.18203459  1.000000000  0.04246105 -0.008298968 -0.05131846
## Experience        0.08187739  0.042461051  1.00000000 -0.090844342 -0.07282489
## Credit           -0.26436603 -0.008298968 -0.09084434  1.000000000  0.17830334
## Size             -0.23966981 -0.051318457 -0.07282489  0.178303336  1.00000000
## Foreigntech      -0.18335942 -0.052555814 -0.01283394 -0.139743287  0.18202741
## Tax_burden        0.08399380  0.003293553 -0.07928627 -0.018921413  0.24843240
## Bussiness_permit  0.19274325  0.218072874 -0.11619905  0.056442775  0.17426023
## local             0.26878840  0.062657718 -0.09076028 -0.102336424 -0.27127849
##                  Foreigntech   Tax_burden Bussiness_permit       local
## avetfp           -0.18335942  0.083993803       0.19274325  0.26878840
## Informal         -0.05255581  0.003293553       0.21807287  0.06265772
## Experience       -0.01283394 -0.079286265      -0.11619905 -0.09076028
## Credit           -0.13974329 -0.018921413       0.05644278 -0.10233642
## Size              0.18202741  0.248432403       0.17426023 -0.27127849
## Foreigntech       1.00000000  0.121762273       0.06656780 -0.16466098
## Tax_burden        0.12176227  1.000000000       0.24630221 -0.10076129
## Bussiness_permit  0.06656780  0.246302210       1.00000000 -0.08735813
## local            -0.16466098 -0.100761292      -0.08735813  1.00000000
```

```
# Plotting to observe heterogeneity for time
ggplot(data4, aes(x = year, y = avetfp)) +
  stat_summary(fun = mean, geom = "point", color = "blue") +
  geom_errorbar(stat = "summary", fun.data = "mean_se", color = "blue", width = 0.2) +
  labs(title = "Heterogeneity across time", x = "ID", y = "avetfp")
```



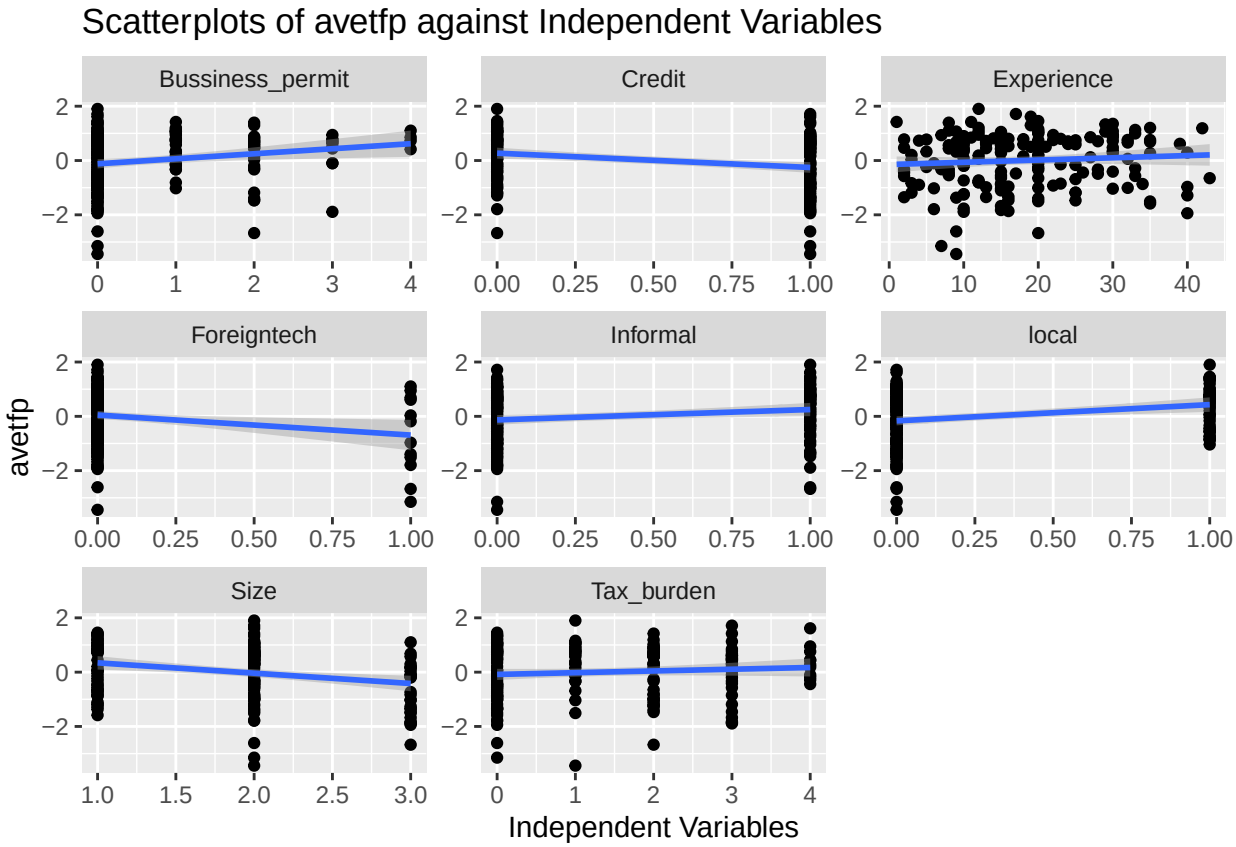

```
# Plotting to observe heterogeneity for firms
ggplot(data4, aes(x = ID, y = avetfp)) +
  stat_summary(fun = mean, geom = "point", color = "blue") +
  geom_errorbar(stat = "summary", fun.data = "mean_se", color = "blue", width = 0.2) +
  labs(title = "Heterogeneity across firms", x = "ID", y = "avetfp")
```



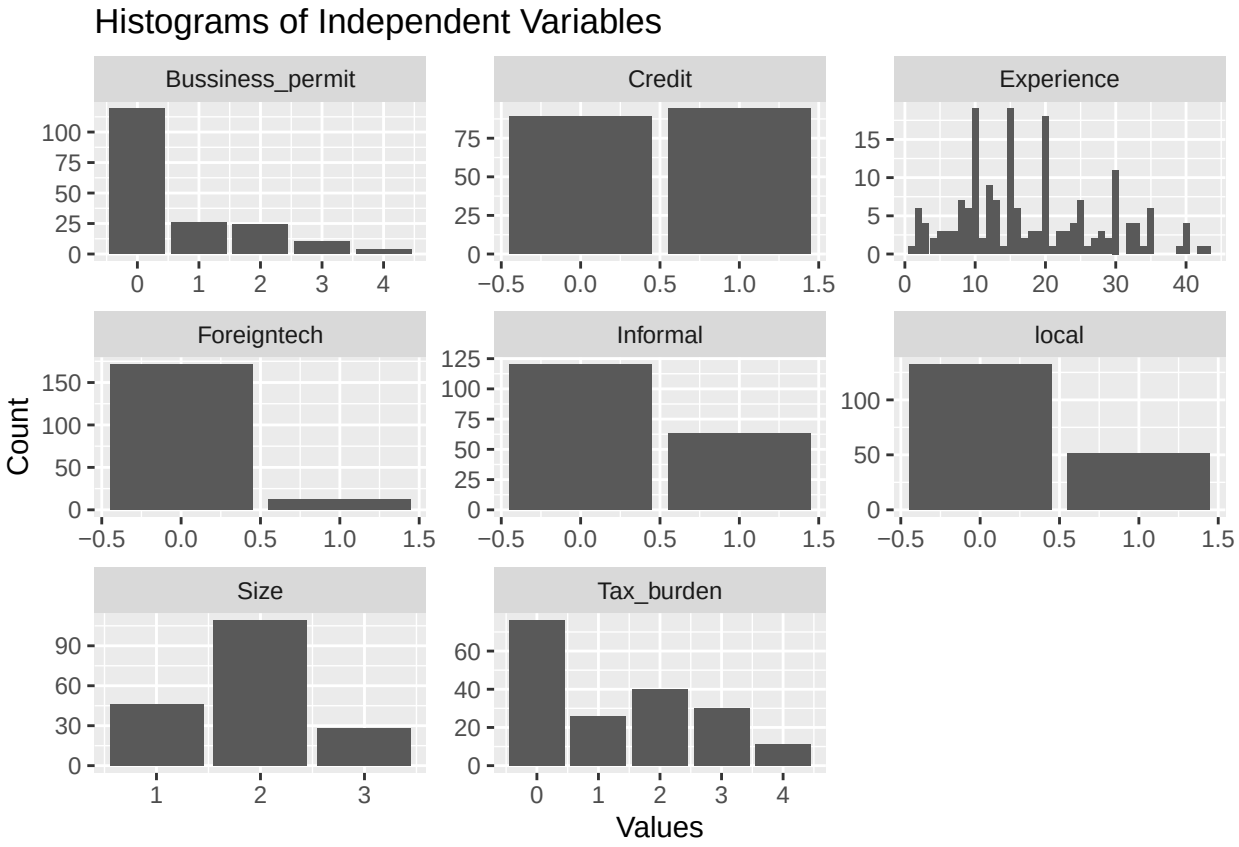

```
# Make dataframe long for plot
data4_long <- tidyr::pivot_longer(data4, cols = c("Informal", "Experience", "Credit", "Size", "Foreignt

# Scatterplot with wrap
ggplot(data4_long, aes(x = value, y = avetfp)) +
  geom_point() +
  geom_smooth(method = lm) +
  facet_wrap(~name, scales = "free") +
  labs(title = "Scatterplots of avetfp against Independent Variables", x = "Independent Variables", y =
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
# Histogram with wrap
ggplot(data4_long, aes(x = value)) +
  geom_bar() +
  facet_wrap(~name, scales = "free") +
  labs(title = "Histograms of Independent Variables", x = "Values", y = "Count")
```




```
# OLS models
ols_model1 <- lm(avetfp ~ Informal, data = data4)
ols_model2 <- lm(avetfp ~ Informal + Experience , data = data4)
ols_model3 <- lm(avetfp ~ Informal + Experience + Credit, data = data4)
ols_model4 <- lm(avetfp ~ Informal + Experience + Credit + Size, data = data4)
ols_model5 <- lm(avetfp ~ Informal + Experience + Credit + Size + Foreigntech, data = data4)
ols_model6 <- lm(avetfp ~ Informal + Experience + Credit + Size + Foreigntech + Tax_burden, data = data4
ols_model7 <- lm(avetfp ~ Informal + Experience + Credit + Size + Foreigntech + Tax_burden + Bussiness_p
ols_model8 <- lm(avetfp ~ Informal + Experience + Credit + Size + Foreigntech + Tax_burden + Bussiness_p

# Store OLS models in a list
ols_models <- list(
  ols_model1, ols_model2, ols_model3, ols_model4,
  ols_model5, ols_model6, ols_model7, ols_model8
)

# Generate stargazer table for OLS regression
huxreg(ols_models) %>%
  set_caption("Ols Regression Models") %>%
  set_number_format(2) %>%
  set_width(0.95) %>%
  set_height(0.95) %>%
  set_position("center")
```

Table 1: Ols Regression Models

	(1.00)	(2.00)	(3.00)	(4.00)	(5.00)	(6.00)	(7.00)	(8.00)
(Intercept)	-0.13	-0.26	0.04	0.60 *	0.58 *	0.52	0.51	0.17
	(0.09)	(0.16)	(0.17)	(0.27)	(0.27)	(0.27)	(0.26)	(0.28)
Informal	0.38 *	0.38 *	0.37 *	0.35 *	0.34 *	0.33 *	0.23	0.20
	(0.15)	(0.15)	(0.15)	(0.15)	(0.14)	(0.14)	(0.14)	(0.14)
Experience		0.01	0.01	0.00	0.00	0.00	0.01	0.01
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Credit			-0.52 ***	-0.45 **	-0.52 ***	-0.50 ***	-0.51 ***	-0.48 ***
			(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.13)
Size				-0.30 **	-0.24 *	-0.30 *	-0.33 **	-0.26 *
				(0.11)	(0.11)	(0.12)	(0.11)	(0.11)
Foreigntech					-0.74 *	-0.78 **	-0.81 **	-0.71 *
					(0.29)	(0.28)	(0.28)	(0.27)
Tax_burden						0.12 *	0.08	0.09
						(0.05)	(0.05)	(0.05)
Bussiness_permit							0.21 **	0.22 **
							(0.07)	(0.07)
local								0.45 **
								(0.15)
N	183.00	183.00	183.00	183.00	183.00	183.00	183.00	183.00
R2.00	0.03	0.04	0.10	0.14	0.17	0.19	0.23	0.27
logLik	-256.08	-255.56	-249.03	-245.50	-242.10	-239.70	-235.05	-230.53
AIC	518.16	519.12	508.07	502.99	498.19	495.39	488.10	481.05

*** $p < 0.00$; ** $p < 0.01$; * $p < 0.05$.

```r
# Create a panel data object
panel_data <- pdata.frame(data4, index = c("ID", "year"))


# Run fixed effects models
fixed_model1 <- plm(avetfp ~ Informal, data = panel_data, model = "within")
fixed_model2 <- plm(avetfp ~ Informal + Experience , data = panel_data, model = "within")
fixed_model3 <- plm(avetfp ~ Informal + Experience + Credit, data = panel_data, model = "within")
fixed_model4 <- plm(avetfp ~ Informal + Experience + Credit + Size, data = panel_data, model = "within"
fixed_model5 <- plm(avetfp ~ Informal + Experience + Credit + Size + Foreigntech, data = panel_data, mo
fixed_model6 <- plm(avetfp ~ Informal + Experience + Credit + Size + Foreigntech + Tax_burden, data = pa
fixed_model7 <- plm(avetfp ~ Informal + Experience + Credit + Size + Foreigntech + Tax_burden + Bussines
fixed_model8 <- plm(avetfp ~ Informal + Experience + Credit + Size + Foreigntech + Tax_burden + Bussines


# Store fixed effects models in a list
fixed_models <- list(
  fixed_model1, fixed_model2, fixed_model3, fixed_model4,
  fixed_model5, fixed_model6, fixed_model7, fixed_model8
)


# Generate stargazer table for panel regression
huxreg(fixed_models) %>%
  set_caption("Panel Regression Models") %>%
  set_number_format(2) %>%
  set_width(0.95) %>%
  set_height(0.95)
```

```r
fixef(fixed_model8) #Cross section wise intercept
```

```
## 1011401033 1011401036 1011401037 1011401038 1011401040 1011401041 1020101072
##  -1.529198  -1.067006  -1.259925  -0.015284  -0.781112  -0.443181  -1.527514
## 1020501011 2061001003 2061001004 2071301016 2071901026 2071901034 2080601056
##  -1.141162  -0.514561  -1.975747  -0.988862  -1.001311  -0.313746  -0.924983
## 2080610011 2083106002 2083106022 2083203006 2089801001 2089801007 2089801009
##  -0.885472  -0.333412  -2.077866  -1.685133  -1.863357  -2.328304  -0.851913
## 2089801010 2089801011 2089801013 2089801014 2089801016 2089801019 2089801021
##  -0.936903  -1.358474  -1.047459  -2.169920  -0.436623  -0.629388  -1.083857
## 2089801022 2089801023 2089801024 2089801025 2089801027 2089801028 2089801035
##  -2.073850  -3.081298  -1.609145  -2.074753  -0.836309  -1.270333  -2.512413
## 2089801036 2089801037 2089801040 2089801041 2089801042 2089801049 2089801050
##  -1.239632  -1.290040  -1.572498  -1.409431  -1.213708  -2.267724  -2.021215
## 2089801056 2089801057 2089801062 2089801067 2089801074 2089801076 2091001039
##  -2.017630  -1.506841  -0.712774  -0.686728  -1.121300  -1.672087  -1.560432
## 2091001085 2099001050 2099801018 2099801025 2099801030 2099801036 2099801047
##  -0.589541  -1.220805  -1.510795  -1.243373  -1.519385  -0.598263  -0.249291
## 2099801049 2099801051 2099801054 2099801058 2099801062 2099801065 2099801079
##  -1.230581  -1.345404  -1.080777  -0.258300  -0.765593   0.335769  -1.193270
## 2139801064 2139801066 2139801068 2139801069 2139801071 2139801072 2139801073
##  -3.406701  -1.623351  -2.375679  -1.923034  -1.703365  -2.798345  -2.646284
## 2139801078 2139801080 2139801081 2149801105 3109801003 3109801010 3109801011
```

Table 2: Panel Regression Models

	(1.00)	(2.00)	(3.00)	(4.00)	(5.00)	(6.00)	(7.00)	(8.00)
Informal	0.64 *	0.64 *	0.61 *	0.60 *	0.60 *	0.51	0.36	0.47
	(0.29)	(0.29)	(0.28)	(0.28)	(0.28)	(0.27)	(0.26)	(0.26)
Experience		0.02	0.02	0.01	0.02	0.02	0.02	0.03
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Credit			-0.47	-0.44	-0.49	-0.56 *	-0.50	-0.47
			(0.27)	(0.27)	(0.28)	(0.27)	(0.26)	(0.25)
Size				0.39	0.46	0.24	0.26	0.22
				(0.34)	(0.35)	(0.34)	(0.32)	(0.32)
Foreigntech					-0.57	-0.92	-0.78	-0.78
					(0.62)	(0.60)	(0.57)	(0.56)
Tax_burden						0.31 **	0.22 *	0.22 *
						(0.11)	(0.10)	(0.10)
Bussiness_permit							0.33 **	0.34 **
							(0.11)	(0.11)
local								0.63 *
								(0.29)
N	183.00	183.00	183.00	183.00	183.00	183.00	183.00	183.00
R2.00	0.06	0.07	0.11	0.12	0.13	0.23	0.31	0.35

*** p < 0.00; ** p < 0.01; * p < 0.05.

```
##  -4.487150  -4.654956  -3.154938  -2.925842  -0.873195  -2.119222  -0.896221
## 3109801013 3109801016 3109801019 3109801023 3110301056 3121401031 3129801001
##  -1.299438  -0.758849  -1.148912  -1.918962  -2.722841  -1.691507  -2.909746
## 3129801002 3129801003 3129801004 3129801006 3129801012 3131101002 3131101049
##  -1.729234  -2.227449  -1.375058  -1.078447  -0.739252  -2.597273  -1.841639
## 3131101088 3131701066 3139801001 3139801004 3139801014 3139801035 3139801036
##  -1.512694  -0.821137  -0.722667  -1.652350  -1.903011  -2.063841  -1.249990
## 3139801037 3139801040 3139801041 3139801046 3139801049 3139801050
##  -1.019741  -1.367764  -1.626264  -1.419694  -1.726737  -1.114807
```

```
pFtest(fixed_model8,ols_model8)  #compare FE and Ols
```

```
##
##  F test for individual effects
```

```
##
## data:  avetfp ~ Informal + Experience + Credit + Size + Foreigntech +  ...
## F = 0.69723, df1 = 103, df2 = 71, p-value = 0.9531
## alternative hypothesis: significant effects
```

```
plmtest(fixed_model8, c("time"), type="bp")
```

```
##
##  Lagrange Multiplier Test - time effects (Breusch-Pagan)
##
## data:  avetfp ~ Informal + Experience + Credit + Size + Foreigntech +  ...
## chisq = 1202.8, df = 1, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

```
# Create a panel data object
panel_data <- pdata.frame(data4, index = c("ID", "year"))


# Run random effects models
random_model1 <- plm(avetfp ~ Informal, data = panel_data, model = "random")
random_model2 <- plm(avetfp ~ Informal + Experience , data = panel_data, model = "random")
random_model3 <- plm(avetfp ~ Informal + Experience + Credit, data = panel_data, model = "random")
random_model4 <- plm(avetfp ~ Informal + Experience + Credit + Size, data = panel_data, model = "random
random_model5 <- plm(avetfp ~ Informal + Experience + Credit + Size + Foreigntech, data = panel_data, m
random_model6 <- plm(avetfp ~ Informal + Experience + Credit + Size + Foreigntech + Tax_burden, data = p
random_model7 <- plm(avetfp ~ Informal + Experience + Credit + Size + Foreigntech + Tax_burden + Bussin
random_model8 <- plm(avetfp ~ Informal + Experience + Credit + Size + Foreigntech + Tax_burden + Bussin

# Store random effects models in a list
random_models <- list(
  random_model1, random_model2, random_model3, random_model4,
  random_model5, random_model6, random_model7, random_model8
)


# Generate stargazer table for panel regression
huxreg(random_models) %>%
  set_caption("Panel Regression Models") %>%
  set_number_format(2) %>%
  set_width(0.95) %>%
  set_height(0.95)
```

```
## Warning in huxreg(random_models): Unrecognized statistics: logLik, AIC
## Try setting ‘statistics‘ explicitly in the call to ‘huxreg()‘
```

```
#hausman test

phtest(fixed_model8, random_model8)
```

```
##
##  Hausman Test
```

```
##
## data:  avetfp ~ Informal + Experience + Credit + Size + Foreigntech +  ...
## chisq = 17.041, df = 8, p-value = 0.02968
## alternative hypothesis: one model is inconsistent
```

Table 3: Panel Regression Models

	(1.00)	(2.00)	(3.00)	(4.00)	(5.00)	(6.00)	(7.00)	(8.00)
(Intercept)	-0.13	-0.26	0.04	0.60 *	0.58 *	0.52	0.51	0.17
	(0.09)	(0.16)	(0.17)	(0.27)	(0.27)	(0.27)	(0.26)	(0.28)
Informal	0.38 *	0.38 *	0.37 *	0.35 *	0.34 *	0.33 *	0.23	0.20
	(0.15)	(0.15)	(0.15)	(0.15)	(0.14)	(0.14)	(0.14)	(0.14)
Experience		0.01	0.01	0.00	0.00	0.00	0.01	0.01
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Credit			-0.52 ***	-0.45 **	-0.52 ***	-0.50 ***	-0.51 ***	-0.48 ***
			(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.13)
Size				-0.30 **	-0.24 *	-0.30 *	-0.33 **	-0.26 *
				(0.11)	(0.11)	(0.12)	(0.11)	(0.11)
Foreigntech					-0.74 **	-0.78 **	-0.81 **	-0.71 **
					(0.29)	(0.28)	(0.28)	(0.27)
Tax_burden						0.12 *	0.08	0.09
						(0.05)	(0.05)	(0.05)
Bussiness_permit							0.21 **	0.22 **
							(0.07)	(0.07)
local								0.45 **
								(0.15)
N	183.00	183.00	183.00	183.00	183.00	183.00	183.00	183.00
R2.00	0.03	0.04	0.10	0.14	0.17	0.19	0.23	0.27

*** p < 0.00; ** p < 0.01; * p < 0.05.