

# panel-Copy1

May 23, 2023

## Tarea 2

### *Instrucciones*

Los resultados de los ejercicios propuestos se deben entregar como un notebook por correo electrónico a [juancaros@udec.cl](mailto:juancaros@udec.cl) el día 9/5 hasta las 21:00.

Es importante considerar que el código debe poder ejecutarse en cualquier computadora con la data original del repositorio. Recordar la convención para el nombre de archivo además de incluir en su documento títulos y encabezados por sección. La data a utilizar es **enia.csv**.

Las variables tienen la siguiente descripción:

- *ID*: firm unique identifier
- *year*: survey year
- *tamaño*: 1 large, 2 medium, 3 small, 4 micro (función de las ventas y el número de trabajadores)
- *sales*: sales (in log of 1,000 CLP)
- *age*: firm age at time of survey
- *foreign*: non-domestic firm (binary)
- *export*: production for export (binary)
- *workers*: log of number of workers
- *fomento*: firm receives public incentives (binary)
- *iyd*: firm does I+D (binary)
- *impuestos*: taxes (in million US)
- *utilidades*: firm revenue (in million US)

Para este análisis consideraremos tamaño como una variable continua, que identifica el tamaño de la empresa.

```
[1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm
import statsmodels.formula.api as smf
import sklearn
import scipy
import linearmodels.panel as lmp
import pytwoway as tw
import bipartitepandas as bpd
import seaborn as sns
from linearmodels.iv import IV2SLS

%matplotlib inline
```

Preguntas:

1. Cargar la base de datos *enia.csv* en el ambiente. Identifique los tipos de datos que se encuentran en la base, realice estadísticas descriptivas sobre las variables importantes (Hint: Revisar la distribuciones, datos faltantes, outliers, etc.) y limpie las variables cuando sea necesario. Para las preguntas 2-8 **EXCLUYA LA VARIABLE FOMENTO DE SU ANALISIS.**

**R:** El panel cuenta con 10,213 observaciones en el primer periodo, cayendo a 6,480. Sin embargo, existen 24,130 ID únicos de empresa, lo que implica que existe una cantidad importante de empresas que han entrado al panel en periodos subsiguientes para mantener el número total. Las empresas en el panel tienen importante variación en el total de ventas, edad y número de trabajadores. Un 8% de las empresas son extranjeras, 11% exporta y 22% hace alguna actividad de IyD.

```
[2]: enia=pd.read_csv('../data/enia.csv')
enia.dropna(inplace=True)
enia.export = enia.export.astype(int)
Xa = enia
bycount = enia['year'].groupby(enia['year']).count()
bIDcount = enia['ID'].groupby(enia['ID']).count()
enia['yr'] = enia['year'].astype(object)
enia = pd.get_dummies(enia)
enia = enia.set_index(["ID", "year"])
enia['utilidades']=np.log(enia['utilidades']-enia['utilidades'].min()+0.1)
enia = enia[enia["utilidades"] < 10000]
enia.describe()
```

C:\Users\juanc\AppData\Local\Temp\ipykernel\_4208\38379287.py:8: FutureWarning:  
In a future version, the Index constructor will not infer numeric dtypes when  
passed object-dtype sequences (matching Series behavior)  
enia = pd.get\_dummies(enia)

```
[2]:
```

|       | tamano       | sales        | age          | foreign      | export \     |
|-------|--------------|--------------|--------------|--------------|--------------|
| count | 39104.000000 | 39104.000000 | 39104.000000 | 39104.000000 | 39104.000000 |
| mean  | 2.248773     | 3.574172     | 15.305084    | 0.081859     | 0.111191     |

|     |          |           |            |          |          |
|-----|----------|-----------|------------|----------|----------|
| std | 1.153089 | 1.692742  | 12.488330  | 0.274153 | 0.314372 |
| min | 1.000000 | 0.000000  | 0.000000   | 0.000000 | 0.000000 |
| 25% | 1.000000 | 2.337643  | 7.000000   | 0.000000 | 0.000000 |
| 50% | 2.000000 | 3.553321  | 14.000000  | 0.000000 | 0.000000 |
| 75% | 3.000000 | 4.539098  | 20.000000  | 0.000000 | 0.000000 |
| max | 4.000000 | 10.309005 | 190.000000 | 1.000000 | 1.000000 |

|       | workers      | fomento      | iyd          | impuestos    | utilidades \ |
|-------|--------------|--------------|--------------|--------------|--------------|
| count | 39104.000000 | 39104.000000 | 39104.000000 | 39104.000000 | 39104.000000 |
| mean  | 1.757726     | 0.076105     | 0.224887     | 0.203856     | 5.500845     |
| std   | 1.186507     | 0.265169     | 0.417514     | 15.869466    | 0.060706     |
| min   | 0.000000     | 0.000000     | 0.000000     | -180.992528  | -2.302585    |
| 25%   | 0.778151     | 0.000000     | 0.000000     | 0.000000     | 5.499092     |
| 50%   | 1.785330     | 0.000000     | 0.000000     | 0.000007     | 5.499092     |
| 75%   | 2.661813     | 0.000000     | 0.000000     | 0.000167     | 5.499097     |
| max   | 5.845915     | 1.000000     | 1.000000     | 2981.494528  | 10.729529    |

|       | yr_2007      | yr_2009      | yr_2013      | yr_2015      | yr_2017      |
|-------|--------------|--------------|--------------|--------------|--------------|
| count | 39104.000000 | 39104.000000 | 39104.000000 | 39104.000000 | 39104.000000 |
| mean  | 0.261124     | 0.180595     | 0.185838     | 0.206731     | 0.165712     |
| std   | 0.439253     | 0.384687     | 0.388981     | 0.404966     | 0.371827     |
| min   | 0.000000     | 0.000000     | 0.000000     | 0.000000     | 0.000000     |
| 25%   | 0.000000     | 0.000000     | 0.000000     | 0.000000     | 0.000000     |
| 50%   | 0.000000     | 0.000000     | 0.000000     | 0.000000     | 0.000000     |
| 75%   | 1.000000     | 0.000000     | 0.000000     | 0.000000     | 0.000000     |
| max   | 1.000000     | 1.000000     | 1.000000     | 1.000000     | 1.000000     |

2. Ejecute un modelo Pooled OLS para explicar el numero de trabajadores. Seleccione las variables independientes a incluir en el modelo final e interprete su significado.

**R:** Segun el modelo inicial, todas las variables son significativas para explicar el numero de trabajadores excepto los impuestos, la cual es excluida en lo que continua. Los resultados muestran una correlacion positiva entre todos los factores excepto las utilidades. En particular, empresas que exportan, y aquellas que son extrajeras, tienen un numero de trabajadores promedio mayor.

```
[3]: y=enia['workers']
X=enia[['sales','age','foreign','export','iyd','utilidades','yr_2009','yr_2013','yr_2015','yr_2017']]
X=sm.add_constant(X)
model = lmp.PanelOLS(y, X)
mco = model.fit()
print(mco)
```

#### PanelOLS Estimation Summary

```
=====
Dep. Variable:          workers    R-squared:                0.6173
Estimator:              PanelOLS   R-squared (Between):      0.6323
No. Observations:       39104      R-squared (Within):       0.4109
Date:                   Tue, May 23 2023  R-squared (Overall):     0.6173
Time:                   14:30:52    Log-likelihood            -4.339e+04
```

|                 |            |                       |             |
|-----------------|------------|-----------------------|-------------|
| Cov. Estimator: | Unadjusted | F-statistic:          | 6306.1      |
| Entities:       | 24128      | P-value               | 0.0000      |
| Avg Obs:        | 1.6207     | Distribution:         | F(10,39093) |
| Min Obs:        | 1.0000     |                       |             |
| Max Obs:        | 5.0000     | F-statistic (robust): | 6306.1      |
|                 |            | P-value               | 0.0000      |
| Time periods:   | 5          | Distribution:         | F(10,39093) |
| Avg Obs:        | 7820.8     |                       |             |
| Min Obs:        | 6480.0     |                       |             |
| Max Obs:        | 1.021e+04  |                       |             |

#### Parameter Estimates

|            | Parameter | Std. Err. | T-stat  | P-value | Lower CI | Upper CI |
|------------|-----------|-----------|---------|---------|----------|----------|
| const      | 0.7255    | 0.3375    | 2.1498  | 0.0316  | 0.0641   | 1.3870   |
| sales      | 0.4535    | 0.0032    | 143.91  | 0.0000  | 0.4474   | 0.4597   |
| age        | 0.0077    | 0.0003    | 24.809  | 0.0000  | 0.0071   | 0.0083   |
| foreign    | 0.0616    | 0.0144    | 4.2934  | 0.0000  | 0.0335   | 0.0898   |
| export     | 0.0896    | 0.0127    | 7.0463  | 0.0000  | 0.0647   | 0.1145   |
| iyd        | 0.2465    | 0.0094    | 26.109  | 0.0000  | 0.2280   | 0.2650   |
| utilidades | -0.2059   | 0.0614    | -3.3520 | 0.0008  | -0.3262  | -0.0855  |
| yr_2009    | -1.0986   | 0.0157    | -70.199 | 0.0000  | -1.1293  | -1.0679  |
| yr_2013    | 1.0897    | 0.0119    | 91.588  | 0.0000  | 1.0664   | 1.1130   |
| yr_2015    | 1.0346    | 0.0115    | 89.803  | 0.0000  | 1.0121   | 1.0572   |
| yr_2017    | 0.8317    | 0.0120    | 69.275  | 0.0000  | 0.8082   | 0.8552   |

3. Ejecute un modelo de efectos fijos para explicar el numero de trabajadores. Seleccione las variables independientes a incluir en el modelo final e interprete su significado.

**R:** El modelo de efectos fijos muestra que las ventas y la actividad de IyD tiene un efecto positivo en el numero de trabajadores (al igual que si la empresa es extranjera), mientras que las utilidades, antigüedad y capacidad de exportar no son significativas. Es importante recordar que el modelo FE considera la variación de las empresas en el tiempo, y por tanto remueve la heterogeneidad no observada que es fija y puede causar sesgo en OLS.

```
[4]: model=lmf.PanelOLS(y,X, entity_effects=True)
fe=model.fit(cov_type="robust")
print(fe)
```

#### PanelOLS Estimation Summary

|                   |          |                      |        |
|-------------------|----------|----------------------|--------|
| Dep. Variable:    | workers  | R-squared:           | 0.5559 |
| Estimator:        | PanelOLS | R-squared (Between): | 0.4194 |
| No. Observations: | 39104    | R-squared (Within):  | 0.5559 |

```

Date:                Tue, May 23 2023    R-squared (Overall):      0.4233
Time:                14:31:04           Log-likelihood            -3479.8
Cov. Estimator:      Robust

                               F-statistic:      1873.0
Entities:            24128              P-value                0.0000
Avg Obs:             1.6207            Distribution:          F(10,14966)
Min Obs:             1.0000
Max Obs:             5.0000            F-statistic (robust):    892.28
                               P-value                0.0000
Time periods:        5                Distribution:          F(10,14966)
Avg Obs:             7820.8
Min Obs:             6480.0
Max Obs:             1.021e+04

```

#### Parameter Estimates

|            | Parameter | Std. Err. | T-stat  | P-value | Lower CI | Upper CI |
|------------|-----------|-----------|---------|---------|----------|----------|
| const      | 0.6579    | 0.2709    | 2.4284  | 0.0152  | 0.1269   | 1.1890   |
| sales      | 0.1209    | 0.0135    | 8.9257  | 0.0000  | 0.0943   | 0.1474   |
| age        | -0.0011   | 0.0008    | -1.4406 | 0.1497  | -0.0026  | 0.0004   |
| foreign    | 0.0891    | 0.0357    | 2.4963  | 0.0126  | 0.0191   | 0.1591   |
| export     | 0.0295    | 0.0240    | 1.2254  | 0.2205  | -0.0177  | 0.0766   |
| iyd        | 0.0460    | 0.0124    | 3.7064  | 0.0002  | 0.0217   | 0.0704   |
| utilidades | 0.0305    | 0.0491    | 0.6201  | 0.5352  | -0.0658  | 0.1268   |
| yr_2009    | -0.3473   | 0.0404    | -8.5986 | 0.0000  | -0.4265  | -0.2681  |
| yr_2013    | 1.0278    | 0.0152    | 67.827  | 0.0000  | 0.9981   | 1.0575   |
| yr_2015    | 1.0003    | 0.0164    | 61.007  | 0.0000  | 0.9681   | 1.0324   |
| yr_2017    | 0.9704    | 0.0188    | 51.488  | 0.0000  | 0.9335   | 1.0074   |

F-test for Poolability: 4.1567

P-value: 0.0000

Distribution: F(24127,14966)

Included effects: Entity

4. Ejecute un modelo de efectos aleatorios para explicar el numero de trabajadores. Seleccione las variables independientes a incluir en el modelo final e interprete su significado.

**R:** Los resultados del modelo de efectos aleatorios es comparable a OLS, sin embargo la magnitud es significativamente diferente. Es importante que, al igual que en FE, las diferencias con OLS tambien pueden estar asociadas a la existencia de muchas empresas con una sola observacion (puntos adicionales para realizar el analisis completo excluyendo las empresas que tienen solo una observacion).

```

[5]: model=lmp.RandomEffects(y,X)
      re=model.fit(cov_type="robust")

```

```
print(re)
```

#### RandomEffects Estimation Summary

```
=====
Dep. Variable:          workers    R-squared:                0.5343
Estimator:             RandomEffects    R-squared (Between):      0.6196
No. Observations:      39104    R-squared (Within):       0.4798
Date:                  Tue, May 23 2023    R-squared (Overall):      0.6036
Time:                  14:31:13    Log-likelihood            -2.422e+04
Cov. Estimator:        Robust

                               F-statistic:                4485.9
Entities:              24128    P-value                  0.0000
Avg Obs:               1.6207    Distribution:            F(10,39093)
Min Obs:               1.0000
Max Obs:               5.0000    F-statistic (robust):    4874.1
                               P-value                    0.0000
Time periods:          5    Distribution:            F(10,39093)
Avg Obs:               7820.8
Min Obs:               6480.0
Max Obs:               1.021e+04
```

#### Parameter Estimates

```
=====
               Parameter  Std. Err.    T-stat    P-value    Lower CI    Upper CI
-----
const          0.3126      0.2771     1.1281    0.2593     -0.2305     0.8557
sales          0.3677      0.0062    59.186    0.0000      0.3555     0.3799
age            0.0056      0.0003    16.187    0.0000      0.0049     0.0062
foreign        0.1446      0.0164     8.8244    0.0000      0.1125     0.1767
export         0.1333      0.0124    10.771    0.0000      0.1091     0.1576
iyd            0.1744      0.0076    23.020    0.0000      0.1596     0.1893
utilidades    -0.0856      0.0504    -1.7001    0.0891     -0.1843     0.0131
yr_2009        -0.9031      0.0246   -36.644    0.0000     -0.9514    -0.8548
yr_2013         1.0897      0.0095    114.57    0.0000      1.0710     1.1083
yr_2015         1.0348      0.0094    110.09    0.0000      1.0164     1.0532
yr_2017         0.9163      0.0104     88.059    0.0000      0.8959     0.9367
=====
```

5. Comente los resultados obtenidos en 2, 3 y 4. ¿Cuáles y por qué existen las diferencias entre los resultados?. En su opinión, ¿Cuál sería el más adecuado para responder la pregunta de investigación y por qué? ¿Qué variables resultaron ser robustas a la especificación?

**R:** Los resultados entre modelos son sustancialmente diferentes, y en base al resultado del test de Hausman, se favorece el modelo de efectos fijos. Dado aquello, las utilidades y capacidad de exportacion son irrelevantes, pero los otros aspectos de la empresa son importantes respecto del numero de trabajadores; estos aumentan con las ventas, IyD, y si es extranjera. En otras palabras, empresas que hacen investigacion y aquellas de gran tamaño (en ventas) son mas intensivas en numero de trabajadores, todo lo demas constante.

```
[8]: import numpy.linalg as la
from scipy import stats

def hausman(fe, re):
    diff = fe.params-re.params
    psi = fe.cov - re.cov
    dof = diff.size -1
    W = diff.dot(la.inv(psi)).dot(diff)
    pval = stats.chi2.sf(W, dof)
    return W, dof, pval

htest = hausman(fe, re)
print("Hausman Test: chi-2 = {0}, df = {1}, p-value = {2}".format(htest[0],
↪htest[1], htest[2]))
```

Hausman Test: chi-2 = 816.8930431782032, df = 10, p-value = 4.8141832456423486e-169

```
[6]: print(lmp.compare({"FE": fe, "RE": re, "Pooled": mco}))
```

| Model Comparison    |                      |                    |                    |
|---------------------|----------------------|--------------------|--------------------|
|                     | FE                   | RE                 | Pooled             |
| Dep. Variable       | workers              | workers            | workers            |
| Estimator           | PanelOLS             | RandomEffects      | PanelOLS           |
| No. Observations    | 39104                | 39104              | 39104              |
| Cov. Est.           | Robust               | Robust             | Unadjusted         |
| R-squared           | 0.5559               | 0.5343             | 0.6173             |
| R-Squared (Within)  | 0.5559               | 0.4798             | 0.4109             |
| R-Squared (Between) | 0.4194               | 0.6196             | 0.6323             |
| R-Squared (Overall) | 0.4233               | 0.6036             | 0.6173             |
| F-statistic         | 1873.0               | 4485.9             | 6306.1             |
| P-value (F-stat)    | 0.0000               | 0.0000             | 0.0000             |
| const               | 0.6579<br>(2.4284)   | 0.3126<br>(1.1281) | 0.7255<br>(2.1498) |
| sales               | 0.1209<br>(8.9257)   | 0.3677<br>(59.186) | 0.4535<br>(143.91) |
| age                 | -0.0011<br>(-1.4406) | 0.0056<br>(16.187) | 0.0077<br>(24.809) |
| foreign             | 0.0891<br>(2.4963)   | 0.1446<br>(8.8244) | 0.0616<br>(4.2934) |
| export              | 0.0295<br>(1.2254)   | 0.1333<br>(10.771) | 0.0896<br>(7.0463) |
| iyd                 | 0.0460<br>(3.7064)   | 0.1744<br>(23.020) | 0.2465<br>(26.109) |
| utilidades          | 0.0305               | -0.0856            | -0.2059            |

|         |           |           |           |
|---------|-----------|-----------|-----------|
|         | (0.6201)  | (-1.7001) | (-3.3520) |
| yr_2009 | -0.3473   | -0.9031   | -1.0986   |
|         | (-8.5986) | (-36.644) | (-70.199) |
| yr_2013 | 1.0278    | 1.0897    | 1.0897    |
|         | (67.827)  | (114.57)  | (91.588)  |
| yr_2015 | 1.0003    | 1.0348    | 1.0346    |
|         | (61.007)  | (110.09)  | (89.803)  |
| yr_2017 | 0.9704    | 0.9163    | 0.8317    |
|         | (51.488)  | (88.059)  | (69.275)  |
| =====   |           |           |           |
| Effects | Entity    |           |           |
| -----   |           |           |           |

T-stats reported in parentheses

- Ejecute un modelo de efectos aleatorios correlacionados (CRE) para explicar el numero de trabajadores. Seleccione las variables independientes a incluir en el modelo final e interprete su significado. Es este modelo adecuado, dada la data disponible, para modelar el componente no observado?

**R:** El modelo CRE entrega resultados similares a FE, sin embargo la antigüedad de la empresa ahora es significativa, y el impacto de las ventas es mayor. Todas las variables incluidas para modelar el componente de heterogeneidad no observada son significativos tambien, lo cual es beneficioso para el ajuste del modelo.

```
[7]: X = enia.
      ↪drop(['workers', 'tamano', 'fomento', 'yr_2009', 'yr_2013', 'yr_2015', 'yr_2017'],
      ↪axis=1, inplace=False)
Xm = X.groupby(by = 'ID').transform('mean')
Xm.columns = ['m{}'.format(column) for column in Xm.columns]
Xc = pd.merge(enia,Xm, left_index=True, right_index=True)
Y = Xc['workers']
X =
      ↪Xc[['sales', 'age', 'foreign', 'export', 'iyd', 'utilidades', 'msales', 'mage', 'mexport', 'miyd', 'm
X=sm.add_constant(X)

model=lmf.RandomEffects(y,X)
cre=model.fit(cov_type="robust")
print(cre)
```

| RandomEffects Estimation Summary |                  |                      |            |
|----------------------------------|------------------|----------------------|------------|
| =====                            |                  |                      |            |
| Dep. Variable:                   | workers          | R-squared:           | 0.5622     |
| Estimator:                       | RandomEffects    | R-squared (Between): | 0.6252     |
| No. Observations:                | 39104            | R-squared (Within):  | 0.5420     |
| Date:                            | Tue, May 23 2023 | R-squared (Overall): | 0.6229     |
| Time:                            | 14:31:32         | Log-likelihood       | -2.301e+04 |
| Cov. Estimator:                  | Robust           |                      |            |
|                                  |                  | F-statistic:         | 3345.9     |



|               |           |                       |             |
|---------------|-----------|-----------------------|-------------|
| Entities:     | 24128     | P-value               | 0.0000      |
| Avg Obs:      | 1.6207    | Distribution:         | F(15,39088) |
| Min Obs:      | 1.0000    |                       |             |
| Max Obs:      | 5.0000    | F-statistic (robust): | 4340.1      |
|               |           | P-value               | 0.0000      |
| Time periods: | 5         | Distribution:         | F(15,39088) |
| Avg Obs:      | 7820.8    |                       |             |
| Min Obs:      | 6480.0    |                       |             |
| Max Obs:      | 1.021e+04 |                       |             |

#### Parameter Estimates

|             | Parameter | Std. Err. | T-stat  | P-value | Lower CI | Upper CI |
|-------------|-----------|-----------|---------|---------|----------|----------|
| const       | 2.0545    | 0.8735    | 2.3520  | 0.0187  | 0.3424   | 3.7666   |
| sales       | 0.2164    | 0.0080    | 26.914  | 0.0000  | 0.2006   | 0.2322   |
| age         | -0.0017   | 0.0005    | -3.3925 | 0.0007  | -0.0027  | -0.0007  |
| foreign     | 0.0955    | 0.0159    | 6.0233  | 0.0000  | 0.0645   | 0.1266   |
| export      | 0.0216    | 0.0177    | 1.2228  | 0.2214  | -0.0130  | 0.0562   |
| iyd         | 0.0323    | 0.0094    | 3.4233  | 0.0006  | 0.0138   | 0.0507   |
| utilidades  | 0.0312    | 0.0555    | 0.5611  | 0.5747  | -0.0777  | 0.1400   |
| msales      | 0.1950    | 0.0058    | 33.699  | 0.0000  | 0.1837   | 0.2064   |
| mage        | 0.0106    | 0.0006    | 16.781  | 0.0000  | 0.0094   | 0.0118   |
| mexport     | 0.1316    | 0.0243    | 5.4124  | 0.0000  | 0.0840   | 0.1793   |
| miyd        | 0.3076    | 0.0145    | 21.166  | 0.0000  | 0.2791   | 0.3361   |
| mutilidades | -0.4798   | 0.1763    | -2.7205 | 0.0065  | -0.8254  | -0.1341  |
| yr_2009     | -0.6831   | 0.0240    | -28.453 | 0.0000  | -0.7301  | -0.6360  |
| yr_2013     | 1.0381    | 0.0090    | 115.00  | 0.0000  | 1.0204   | 1.0557   |
| yr_2015     | 1.0154    | 0.0090    | 112.23  | 0.0000  | 0.9976   | 1.0331   |
| yr_2017     | 0.9139    | 0.0101    | 90.440  | 0.0000  | 0.8940   | 0.9337   |

7. Usando el modelo CRE, prediga la distribucion del componente no observado. Que puede inferir respecto de la heterogeneidad fija en el tiempo y su impacto en el numero de trabajadores?

**R:** La heterogeneidad no observada que se predice del modelo CRE tiene un comportamiento normal, lo cual en principio puede ser inapropiado dado que genera valores negativos en el log-workers, lo cual no deberia ocurrir. Sin embargo, en promedio captura una parte importante de la variacion de las empresas en cuanto a su numero de trabajadores.

```
[9]: Xpred = X
Xpred['sales']=0
Xpred['age']=0
Xpred['export']=0
Xpred['iyd']=0
Xpred['utilidades']=0
Xpred['foreign']=0
```

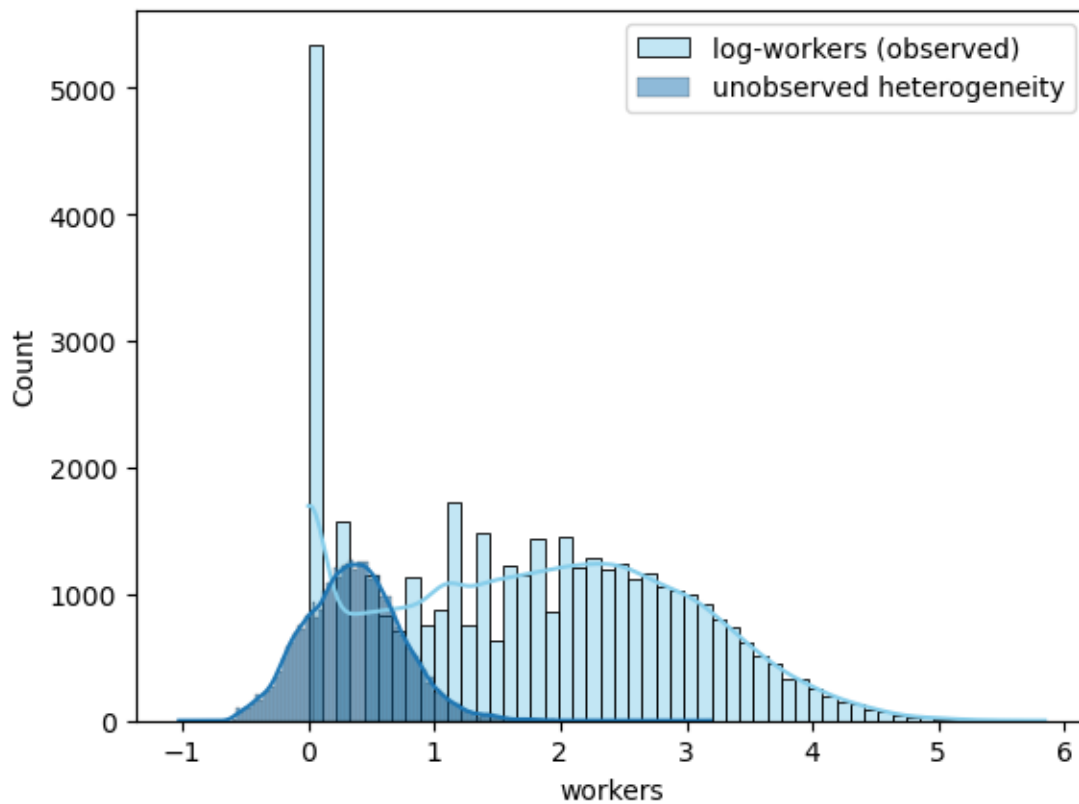
```

Xpred['yr_2009']=0
Xpred['yr_2013']=0
Xpred['yr_2015']=0
Xpred['yr_2017']=0
yhat = cre.predict(Xpred)

sns.histplot(data=y, color="skyblue", label="log-workers (observed)", kde=True)
sns.histplot(data=yhat, color="red", label="unobserved heterogeneity", kde=True)

plt.legend()
plt.show()

```



8. Usando sus respuestas anteriores, que modelo prefiere? que se puede inferir en general respecto del efecto de las variables explicativas sobre el numero de trabajadores?

**R:** En vista de los resultados, el ajuste y los coeficientes entregados, se favorece el modelo CRE, siendo este mas preciso que FE y RE, ademas de entregar diferencias importantes en coeficientes que podrian haber estado potencialmente sesgados (como ventas).

```
[11]: print(lmp.compare({"FE": fe, "RE": re, "CRE": cre}))
```

Model Comparison

|                     | FE                   | RE                   | CRE                  |
|---------------------|----------------------|----------------------|----------------------|
| Dep. Variable       | workers              | workers              | workers              |
| Estimator           | PanelOLS             | RandomEffects        | RandomEffects        |
| No. Observations    | 39104                | 39104                | 39104                |
| Cov. Est.           | Robust               | Robust               | Robust               |
| R-squared           | 0.5559               | 0.5343               | 0.5622               |
| R-Squared (Within)  | 0.5559               | 0.4798               | 0.5420               |
| R-Squared (Between) | 0.4194               | 0.6196               | 0.6252               |
| R-Squared (Overall) | 0.4233               | 0.6036               | 0.6229               |
| F-statistic         | 1873.0               | 4485.9               | 3345.9               |
| P-value (F-stat)    | 0.0000               | 0.0000               | 0.0000               |
| const               | 0.6579<br>(2.4284)   | 0.3126<br>(1.1281)   | 2.0545<br>(2.3520)   |
| sales               | 0.1209<br>(8.9257)   | 0.3677<br>(59.186)   | 0.2164<br>(26.914)   |
| age                 | -0.0011<br>(-1.4406) | 0.0056<br>(16.187)   | -0.0017<br>(-3.3925) |
| foreign             | 0.0891<br>(2.4963)   | 0.1446<br>(8.8244)   | 0.0955<br>(6.0233)   |
| export              | 0.0295<br>(1.2254)   | 0.1333<br>(10.771)   | 0.0216<br>(1.2228)   |
| iyd                 | 0.0460<br>(3.7064)   | 0.1744<br>(23.020)   | 0.0323<br>(3.4233)   |
| utilidades          | 0.0305<br>(0.6201)   | -0.0856<br>(-1.7001) | 0.0312<br>(0.5611)   |
| yr_2009             | -0.3473<br>(-8.5986) | -0.9031<br>(-36.644) | -0.6831<br>(-28.453) |
| yr_2013             | 1.0278<br>(67.827)   | 1.0897<br>(114.57)   | 1.0381<br>(115.00)   |
| yr_2015             | 1.0003<br>(61.007)   | 1.0348<br>(110.09)   | 1.0154<br>(112.23)   |
| yr_2017             | 0.9704<br>(51.488)   | 0.9163<br>(88.059)   | 0.9139<br>(90.440)   |
| msales              |                      |                      | 0.1950<br>(33.699)   |
| mage                |                      |                      | 0.0106<br>(16.781)   |
| mexport             |                      |                      | 0.1316<br>(5.4124)   |
| miyd                |                      |                      | 0.3076<br>(21.166)   |
| mutilidades         |                      |                      | -0.4798<br>(-2.7205) |
| Effects             | Entity               |                      |                      |

-----

T-stats reported in parentheses

9. Considere que la variable *fomento* es una politica publica donde aleatoriamente se selecciono un grupo de empresas para recibir recursos financieros dedicados a incentivar I+D. Utilizando fomento como instrumento, estime un modelo en dos etapas para entender el impacto causal de la inversion en I+D sobre el numero de trabajadores, y compare versus el modelo MCO (puntos adicionales para hacerlo en un contexto de panel).

**R:** Al comparar los modelos se observa que la relacion entre IyD y numero de trabajadores esta sobre-estimada al usar MCO, mientras que al utilizar el instrumento, la relacion entre las variables baja sustancialmente. Lo que se puede inferir es que aquellas empresas que realizan IyD producto de la politica publica, contratan 12% mas trabajadores respecto de aquellas que no lo hacen, y no 30% como sugiere los resultados de MCO (potencial endogeneidad ya que pueden existir otros factores que inducen un incremento en el potencial de IyD y el numero de trabajadores, como el rubro de la empresa, por ejemplo).

```
[11]: from statsmodels.api import add_constant
#enia = sm.add_constant(enia)
res_ols = IV2SLS(enia.workers,
    ↪enia[['sales', 'foreign', 'age', 'export', 'utilidades', 'iyd', 'const']], None,
    ↪None).fit(
    cov_type="unadjusted")
res_IV = IV2SLS(enia.workers,
    ↪enia[['sales', 'foreign', 'age', 'export', 'utilidades', 'const']], enia['iyd'],
    ↪enia['fomento']).fit(
    cov_type="unadjusted")

from linearmodels.iv import compare
print(compare({"OLS": res_ols, "2SLS": res_IV}))
```

Model Comparison

|                  | OLS                | 2SLS               |
|------------------|--------------------|--------------------|
| Dep. Variable    | workers            | workers            |
| Estimator        | OLS                | IV-2SLS            |
| No. Observations | 39104              | 39104              |
| Cov. Est.        | unadjusted         | unadjusted         |
| R-squared        | 0.2498             | 0.2458             |
| Adj. R-squared   | 0.2497             | 0.2457             |
| F-statistic      | 1.302e+04          | 1.236e+04          |
| P-value (F-stat) | 0.0000             | 0.0000             |
| sales            | 0.2485<br>(78.177) | 0.2505<br>(75.422) |
| age              | 0.0203<br>(48.170) | 0.0206<br>(46.281) |

|             |           |           |
|-------------|-----------|-----------|
| export      | 0.4179    | 0.4549    |
|             | (24.237)  | (18.788)  |
| utilidades  | -0.4655   | -0.4442   |
|             | (-5.4171) | (-5.1224) |
| iyd         | 0.3102    | 0.1257    |
|             | (24.505)  | (1.4715)  |
| const       | 3.0031    | 2.9113    |
|             | (6.3630)  | (6.1276)  |
| =====       |           |           |
| Instruments |           | fomento   |
| -----       |           |           |

T-stats reported in parentheses