

tarea_2_nicolas_netz

May 11, 2023

1 Laboratorio de métodos aplicados: Tarea 1

Estudiante: Nicolás Netz

N°Mat: 2018458791

Correo: nnetz2018@udec.cl

Prof. Juan Carlos Caro

2 Importar los módulos necesarios y la data

En el presente notebook se utilizan distintas librerías para apoyar los cálculos y las visualizaciones. Además, se incluye una serie de funciones extra, para reutilizar código en distintas celdas.

Tarea 2

Instrucciones

Los resultados de los ejercicios propuestos se deben entregar como un notebook por correo electrónico a 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
- *tamano*: 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 analisis consideraremos tamaño como una variable continua, que identifica el tamaño de la empresa.

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**.
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.
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.
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.
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?
6. 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?
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?
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?
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).

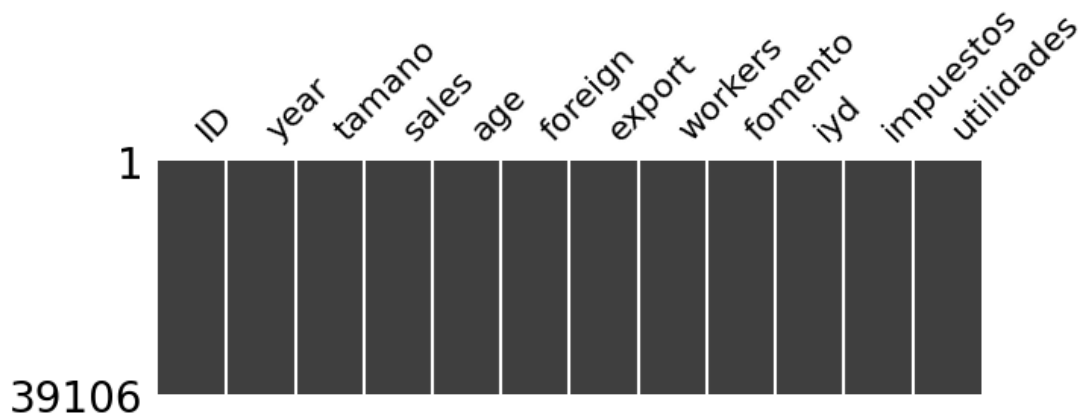
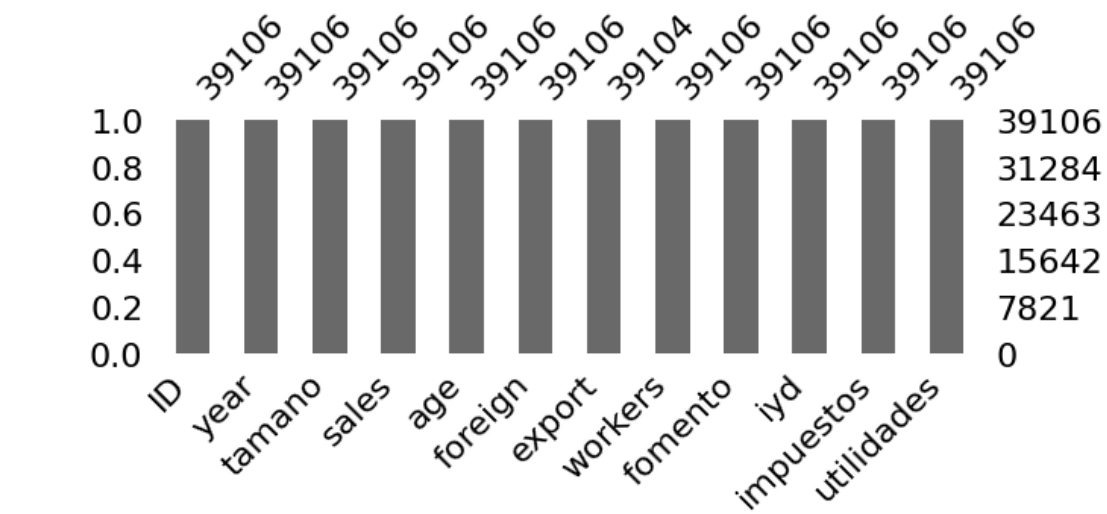
3 Pregunta 1: Cargar base de datos

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 distribución, datos faltantes, outliers, etc.) y limpie las variables cuando sea necesario. Para las preguntas 2-8 **EXCLUYA LA VARIABLE FOMENTO DE SU ANALISIS**.

	ID	year	tamano	sales	age	foreign	export	workers	fomento	\
0	100003	2007	1	7.046558	22	1	1.0	3.486855	0	
1	100003	2009	1	7.875563	24	1	1.0	3.504607	0	
2	100003	2013	1	7.437399	23	1	1.0	4.691621	1	
3	100003	2015	1	7.356472	30	1	1.0	4.682614	1	
4	100003	2017	1	4.014772	32	1	1.0	4.611691	1	

	iyd	impuestos	utilidades
0	1	1.231345	7.113892
1	0	8.762233	33.976108
2	1	0.001886	3.137265
3	0	0.411670	1.298413
4	0	0.000721	0.001949

```
/Users/nicolas/.virtualenvs/lab-maa/lib/python3.8/site-  
packages/missingno/missingno.py:61: UserWarning: Plotting a sparkline on an  
existing axis is not currently supported. To remove this warning, set  
sparkline=False.  
warnings.warn(
```



3.1 Notas al leer la data

Se tienen 39106 entradas en 12 columnas, las cuales corresponden a:

- *ID*: firm unique identifier
- *year*: survey year
- *tamano*: 1 large, 2 medium, 3 small, 4 micro (funcion de las ventas y el numero 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)

A partir de los graficos anteriores se identificaron dos valores NAN en la columna export, ver primer gráfico. Estos valores se eliminarán. Además se hace un análisis exploratorio por columnas para validar que no hayan valores fuera de rango, o que los nulos estén codificados de otra manera, etc.

	ID	year	tamano	sales	age	foreign	export	workers	fomento	\
0	100003	2007	1	7.046558	22	1	1.0	3.486855	0	
1	100003	2009	1	7.875563	24	1	1.0	3.504607	0	
2	100003	2013	1	7.437399	23	1	1.0	4.691621	1	
3	100003	2015	1	7.356472	30	1	1.0	4.682614	1	
4	100003	2017	1	4.014772	32	1	1.0	4.611691	1	

	iyd	impuestos	utilidades
0	1	1.231345	7.113892
1	0	8.762233	33.976108
2	1	0.001886	3.137265
3	0	0.411670	1.298413
4	0	0.000721	0.001949

```
Index(['ID', 'year', 'tamano', 'sales', 'age', 'foreign', 'export', 'workers',
      'fomento', 'iyd', 'impuestos', 'utilidades'],
      dtype='object')
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 39106 entries, 0 to 39105
```

```
Data columns (total 12 columns):
```

#	Column	Non-Null Count	Dtype
0	ID	39106 non-null	int64
1	year	39106 non-null	int64
2	tamano	39106 non-null	int64
3	sales	39106 non-null	float64
4	age	39106 non-null	int64
5	foreign	39106 non-null	int64
6	export	39104 non-null	float64
7	workers	39106 non-null	float64
8	fomento	39106 non-null	int64
9	iyd	39106 non-null	int64
10	impuestos	39106 non-null	float64
11	utilidades	39106 non-null	float64

```
dtypes: float64(5), int64(7)
```

```
memory usage: 3.6 MB
```

	ID	year	tamano	sales	age \
count	39106.000000	39106.000000	39106.000000	39106.000000	39106.000000
mean	218083.624175	2011.786938	2.248734	3.574175	15.305605
std	128227.788409	3.781295	1.153078	1.692700	12.488596
min	100000.000000	2007.000000	1.000000	0.000000	0.000000
25%	105408.250000	2007.000000	1.000000	2.337709	7.000000
50%	200994.000000	2013.000000	2.000000	3.553321	14.000000
75%	302466.000000	2015.000000	3.000000	4.538989	20.000000
max	507526.000000	2017.000000	4.000000	10.309005	190.000000

	foreign	export	workers	fomento	iyd \
count	39106.000000	39104.000000	39106.000000	39106.000000	39106.000000
mean	0.081854	0.111191	1.757666	0.076101	0.224902
std	0.274146	0.314372	1.186514	0.265163	0.417523
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.778151	0.000000	0.000000
50%	0.000000	0.000000	1.785330	0.000000	0.000000
75%	0.000000	0.000000	2.661813	0.000000	0.000000
max	1.000000	1.000000	5.845915	1.000000	1.000000

	impuestos	utilidades
count	39106.000000	3.910600e+04
mean	0.203845	1.875158e+00

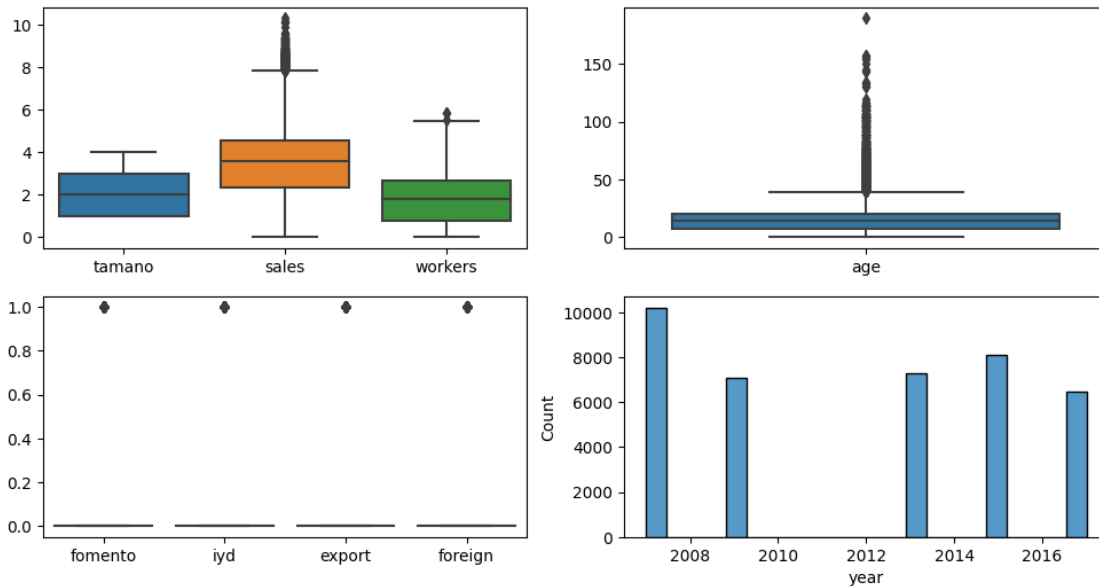
std	15.869061	2.306840e+02
min	-180.992528	-2.443698e+02
25%	0.000000	9.050000e-07
50%	0.000007	8.075000e-05
75%	0.000166	1.283282e-03
max	2981.494528	4.544069e+04

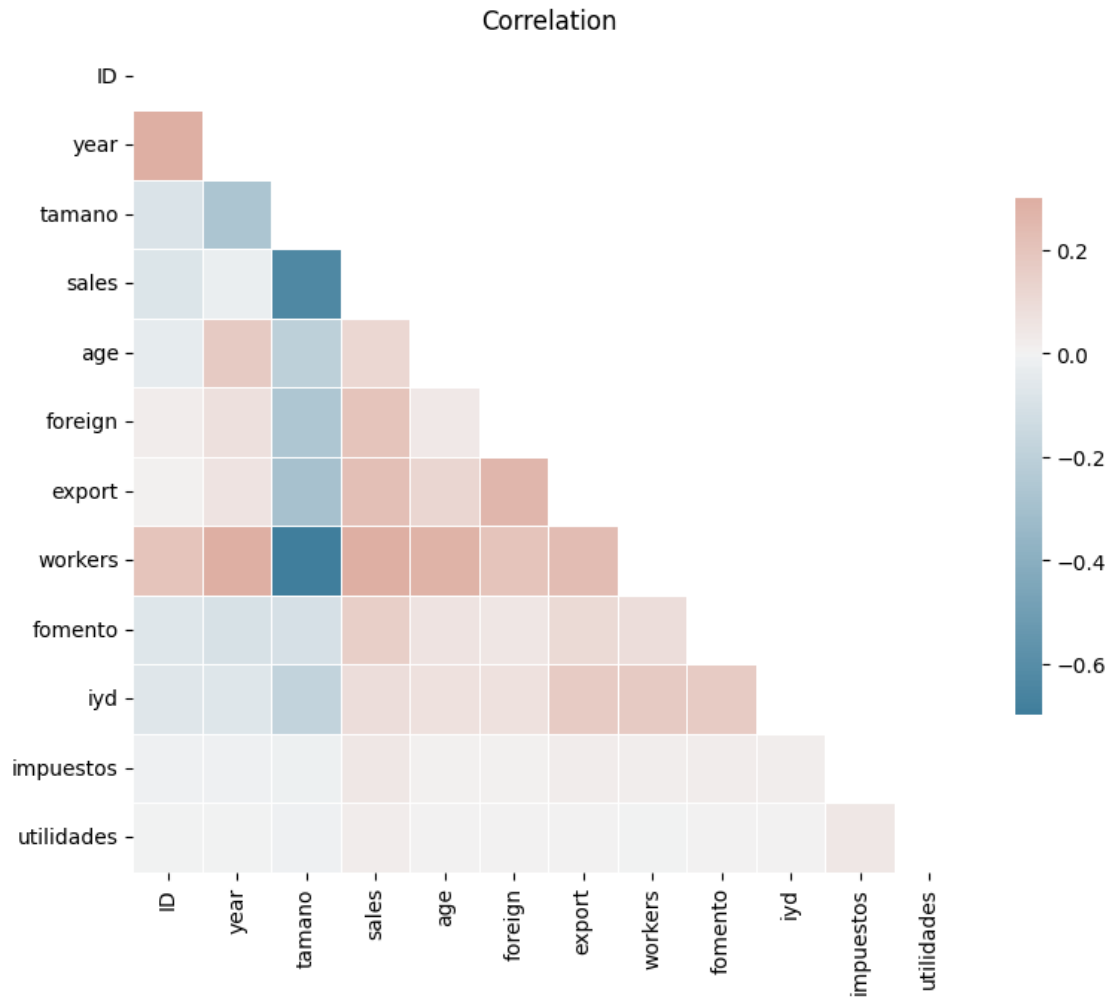
3.2 Notas sobre estadísticos descriptivos

Se revisó que los valores mínimos y máximos de las distintas variables estén en sus rangos.

Algunas cosas que notar: * Se encontraron impuestos con valor negativo * Se encontraron compañías que tenían 190 años de antigüedad. * Notar que los valores de tamaño de las ventas están en logaritmo de 1000 CLP

```
<Axes: xlabel='year', ylabel='Count'>
```





3.3 Notas sobre la matriz de correlación

Se identifica un alta correlación entre tamaño y workers, así como tamaño con sales. Debido a lo anterior, se excluyen estas variables como explicativas, pues pueden generar problemas de multicolinealidad

4 Notas sobre el set de datos

Se identifican algunos problemas de atrición, pues en distintos años se tienen distinta cantidad de compañías y no todas las compañías aparecen todos los años.

No obstante, se usan los datos de la forma en la que vienen.

5 Pregunta 2: Pooled OLS

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.

```

=====
                        PanelOLS Estimation Summary
=====
Dep. Variable:                workers    R-squared:                0.2576
Estimator:                    PanelOLS   R-squared (Between):      0.3038
No. Observations:             39104     R-squared (Within):      -0.5362
Date:                         Thu, May 11 2023   R-squared (Overall):      0.2576
Time:                         17:45:28    Log-likelihood            -5.635e+04
Cov. Estimator:               Unadjusted

                                F-statistic:                2260.4
Entities:                     24128     P-value                   0.0000
Avg Obs:                      1.6207    Distribution:             F(6,39097)
Min Obs:                      1.0000
Max Obs:                      5.0000     F-statistic (robust):    2260.4
                                P-value                   0.0000
Time periods:                 5         Distribution:             F(6,39097)
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	0.4574	0.0133	34.433	0.0000	0.4313	0.4834
sales	0.2376	0.0032	74.531	0.0000	0.2314	0.2439
age	0.0204	0.0004	48.617	0.0000	0.0196	0.0212
foreign	0.4079	0.0198	20.617	0.0000	0.3691	0.4466
export	0.3356	0.0176	19.069	0.0000	0.3011	0.3701
iyd	0.3036	0.0126	24.106	0.0000	0.2790	0.3283
utilidades	-7.313e-05	2.242e-05	-3.2617	0.0011	-0.0001	-2.918e-05

```

=====

```

5.1 Respuesta

Se decidió que las variables a utilizar fueron:

- sales: float
- age: int
- foreign : binaria
- export: binaria
- iyd: binaria
- utilidades: float

Notar que se excluyeron otras variables: Tamaño se excluye pues está calculada a partir de sales y workers, lo que implicaría problemas de multicolinealidad. Impuestos igual se excluye, pues en teoría deberían ser una fracción de las utilidades (aunque la correlación en este caso no resultó perfecta.)

A partir de la tabla de resultados anterior, donde se usó PooledOLS para estimar la cantidad de trabajadores en función de las variables anteriormente mencionadas, se obtiene la siguiente interpretación de los resultados del modelo:

1. Los resultados de este modelo corresponden a los efectos promedio de las variables, puesto que la data es de panel. Así, el modelo ignora que las observaciones en realidad corresponden a los mismos individuos en distintos momentos del tiempo.
2. El modelo en general explica el 25% de la varianza, según el R^2 . Notamos además que el $R^2_{between}$ corresponde a un 30% de la varianza, lo que quiere decir que las diferencias entre individuos (empresas) explican ese porcentaje de la varianza.
3. El estadístico F para la prueba de hipótesis:

$$H_0 : \beta_i = 0 \quad \forall i \quad H_1 : \beta_i \neq 0 \quad \exists i$$

Rechaza la hipótesis nula con un 99.9% de confianza, indicando que alguno de los estimadores β es significativo.

4. Al revisar las pruebas de T-student, podemos ver que muchos rechazan la hipótesis nula de:

$$H_0 : \beta_i = 0 \quad H_1 : \beta_i \neq 0$$

a un 99% de confianza. Esto quiere decir que los coeficientes son significativos, y se puede interpretar que:

- frente a aumentos en las ventas, se observan en promedio aumentos en la cantidad de trabajadores.
- frente a aumentos en la edad de las empresas, se observa en promedio un pequeño aumento en la cantidad de trabajadores.
- si la empresa es extranjera, entonces tiene en promedio más trabajadores.
- si la empresa exporta, entonces tiene en promedio más trabajadores.
- si la empresa invierte en i+D, entonces tiene en promedio más trabajadores.
- si la empresa disminuye sus utilidades, entonces tiene en promedio menos trabajadores.

6 Pregunta 3: Efectos Fijos

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

PanelOLS Estimation Summary			
=====			
Dep. Variable:	workers	R-squared:	0.1734
Estimator:	PanelOLS	R-squared (Between):	-0.3116
No. Observations:	39104	R-squared (Within):	0.1734
Date:	Thu, May 11 2023	R-squared (Overall):	-0.1992

```

Time:                  17:45:28   Log-likelihood              -1.563e+04
Cov. Estimator:       Robust

                        F-statistic:              523.30
Entities:             24128   P-value              0.0000
Avg Obs:              1.6207   Distribution:      F(6,14970)
Min Obs:              1.0000
Max Obs:              5.0000   F-statistic (robust): 196.14
                                P-value              0.0000
Time periods:         5       Distribution:      F(6,14970)
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	2.1909	0.0361	60.641	0.0000	2.1201	2.2617
sales	-0.1688	0.0060	-27.995	0.0000	-0.1806	-0.1570
age	0.0127	0.0014	9.0612	0.0000	0.0099	0.0154
foreign	0.1056	0.0493	2.1414	0.0323	0.0089	0.2023
export	0.0588	0.0336	1.7466	0.0807	-0.0072	0.1247
iyd	-0.1733	0.0180	-9.6435	0.0000	-0.2086	-0.1381
utilidades	4.289e-05	2.848e-06	15.060	0.0000	3.731e-05	4.847e-05

F-test for Poolability: 4.3604

P-value: 0.0000

Distribution: F(24127,14970)

Included effects: Entity

6.1 Respuesta

Se decidió que las variables a utilizar fueron las del modelo anterior.

Así, usando la tabla de resultados anterior donde se usó Efectos Fijos, se interpretan los siguientes resultados

1. Se utiliza la transformación de efectos fijos para eliminar el efecto inobservable antes de la estimación, esto es, quitar el promedio de las variables en el tiempo.
2. El modelo explica el 17% de la varianza, según el R^2 within.
3. El estadístico F para la prueba de hipótesis:

$$H_0 : \beta_i = 0 \quad \forall i \quad H_1 : \beta_i \neq 0 \quad \exists i$$

Rechaza la hipótesis nula con un 99.9% de confianza, indicando que alguno de los estimadores β es significativo.

4. Al revisar las pruebas de T-student, podemos ver que muchos rechazan la hipótesis nula de:

$$H_0 : \beta_i = 0 \quad H_1 : \beta_i \neq 0$$

a un 99% de confianza. Esto quiere decir que los coeficientes son significativos, y se puede interpretar que:

- frente a aumentos en las ventas, se observa una disminución en la cantidad de trabajadores. (contrario a lo observado en pooled)
- frente a aumentos en la edad de las empresas, se observa un pequeño aumento en la cantidad de trabajadores.
- si la empresa es extranjera, se observa un aumento en trabajadores
- si la empresa es exportadora, se observa un aumento en trabajadores
- si la empresa invierte en i+D, entonces tiene en promedio menos trabajadores. (contrario a lo observado en pooled)
- si la empresa aumenta sus utilidades, entonces tiene en promedio más trabajadores. (Lo mismo que en pooled pero al revés)

7 Pregunta 4: Efectos aleatorios

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

RandomEffects Estimation Summary			
=====			
Dep. Variable:	workers	R-squared:	0.0302
Estimator:	RandomEffects	R-squared (Between):	0.2134
No. Observations:	39104	R-squared (Within):	-0.1009
Date:	Thu, May 11 2023	R-squared (Overall):	0.1898
Time:	17:45:29	Log-likelihood	-3.866e+04
Cov. Estimator:	Robust		
		F-statistic:	202.89
Entities:	24128	P-value	0.0000
Avg Obs:	1.6207	Distribution:	F(6,39097)
Min Obs:	1.0000		
Max Obs:	5.0000	F-statistic (robust):	604.61
		P-value	0.0000
Time periods:	5	Distribution:	F(6,39097)
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	0.9063	0.0144	62.718	0.0000	0.8779	0.9346
sales	0.0763	0.0029	26.277	0.0000	0.0706	0.0820

age	0.0217	0.0006	33.492	0.0000	0.0205	0.0230
foreign	0.5012	0.0229	21.890	0.0000	0.4564	0.5461
export	0.3883	0.0180	21.524	0.0000	0.3530	0.4237
iyd	0.1318	0.0111	11.843	0.0000	0.1100	0.1536
utilidades	8.193e-07	1.532e-05	0.0535	0.9574	-2.922e-05	3.085e-05

=====

```
Effects                0.578154
Residual               0.340098
Percent due to Effects 0.629624
Name: Variance Decomposition, dtype: float64
```

7.1 Respuesta

Se decidió que las variables a utilizar fueron las del modelo PooledOLS, pues el modelo de efectos aleatorios sí permite considerar variables que son constantes en el tiempo.

Así, usando la tabla de resultados anterior donde se usó Efectos Aleatorios, se interpretan los siguientes resultados

1. Se utiliza la transformación de efectos aleatorios pues se supone que la heterogeneidad no observada no está correlacionada con ninguna de las variables explicativas.
2. El modelo explica el 17% de la varianza, según el R^2 within.
3. El estadístico F para la prueba de hipótesis:

$$H_0 : \beta_i = 0 \quad \forall i \quad H_1 : \beta_i \neq 0 \quad \exists i$$

Rechaza la hipótesis nula con un 99.9% de confianza, indicando que alguno de los estimadores β es significativo.

4. Al revisar las pruebas de T-student, podemos ver que muchos rechazan la hipótesis nula de:

$$H_0 : \beta_i = 0 \quad H_1 : \beta_i \neq 0$$

a un 99% de confianza. Esto quiere decir que los coeficientes son significativos, menos el **utilidades**, y se puede interpretar que:

- frente a aumentos en las ventas, se observan aumentos en la cantidad de trabajadores.
 - frente a aumentos en la edad de las empresas, se observa un pequeño aumento en la cantidad de trabajadores.
 - si la empresa es extranjera, entonces tiene más trabajadores.
 - si la empresa exporta, entonces tiene más trabajadores.
 - si la empresa invierte en i+D, entonces tiene más trabajadores
5. Finalmente, al descomponer la varianza en los efectos y el residuo, se puede observar que la ésta se explica principalmente por los efectos, en un 62%.

8 Pregunta 5: Comparación de resultados

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?

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	0.1734	0.0302	0.2576
R-Squared (Within)	0.1734	-0.1009	-0.5362
R-Squared (Between)	-0.3116	0.2134	0.3038
R-Squared (Overall)	-0.1992	0.1898	0.2576
F-statistic	523.30	202.89	2260.4
P-value (F-stat)	0.0000	0.0000	0.0000
const	2.1909*** (60.641)	0.9063*** (62.718)	0.4574*** (34.433)
sales	-0.1688*** (-27.995)	0.0763*** (26.277)	0.2376*** (74.531)
age	0.0127*** (9.0612)	0.0217*** (33.492)	0.0204*** (48.617)
foreign	0.1056** (2.1414)	0.5012*** (21.890)	0.4079*** (20.617)
export	0.0588* (1.7466)	0.3883*** (21.524)	0.3356*** (19.069)
iyd	-0.1733*** (-9.6435)	0.1318*** (11.843)	0.3036*** (24.106)
utilidades	4.289e-05*** (15.060)	8.193e-07 (0.0535)	-7.313e-05*** (-3.2617)
Effects	Entity		

T-stats reported in parentheses

Hausman Test: $\chi^2 = 2215.270908122931$, $df = 6$, $p\text{-value} = 0.0$

8.1 Respuesta

Primero, podemos ver que los tres modelos explican algún porcentaje de la varianza, al ver los R-squared. Además, notamos que todos los modelos presentan algún coeficiente significativo y que la mayoría de estos rechazan las hipótesis nulas en las pruebas de t student para los coeficientes.

Estas diferencias se deben a la construcción de estos modelos. PooledOLS no considera los efectos

en el tiempo y trabaja entregando el efecto promedio, Efectos Fijos remueve la heterogeneidad no observada quitando el promedio en el tiempo a las variables y efectos aleatorios considera la heterogeneidad no observada como un componente aleatorio.

En general, casi todas las variables resultaron ser robustas a la especificación en el sentido de pasar las pruebas t. No obstante, algunas resultan tener efectos inversos sobre la cantidad de trabajadores.


Para determinar cuál sería el modelo más adecuado, primero se descarta el modelo PooledOLS, pues este no considera los efectos temporales de la data.

Segundo, para elegir entre el modelo de Efectos fijos y el modelo de Efectos aleatorios, se utiliza el test de Hausmann para validar el supuesto de que la covarianza entre la heterogeneidad no observada y las variables explicativas es igual a 0. Recordando que el modelo de efectos aleatorios supone que $Cov(\alpha_i, X_{it}) = 0$. Esto implica que efectos aleatorios y efectos fijos son consistentes, pero si es distinto de 0, entonces sólo efectos fijos es consistente. Considerando esto, se lleva a cabo la prueba de hipótesis de Hausmann, calculada en la celda anterior, tal que:

$$H_0 : Cov(\alpha_i, X_{it}) = 0 \quad H_1 : Cov(\alpha_i, X_{it}) \neq 0$$

Con el estadístico:

$$W = \frac{(\beta_{fe} - \beta_{re})^2}{\sigma_{\beta_{fe}}^2 - \sigma_{\beta_{re}}^2} \sim \chi_1$$

Se obtuvo que el valor $p \approx 0$, por lo que se rechaza la hipótesis nula, lo que implica que efectos fijos tiene estimadores consistentes. Debido a lo anterior y a que  se modelo considera los efectos del tiempo, sería el más adecuado para la pregunta de investigación.

9 Pregunta 6: Efectos aleatorios correlacionados

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?

RandomEffects Estimation Summary			
=====			
Dep. Variable:	workers	R-squared:	0.2182
Estimator:	RandomEffects	R-squared (Between):	0.3294
No. Observations:	39104	R-squared (Within):	0.1734
Date:	Thu, May 11 2023	R-squared (Overall):	0.3532
Time:	17:45:29	Log-likelihood	-3.445e+04
Cov. Estimator:	Robust		
		F-statistic:	909.06
Entities:	24128	P-value	0.0000
Avg Obs:	1.6207	Distribution:	F(12,39091)
Min Obs:	1.0000		
Max Obs:	5.0000	F-statistic (robust):	1312.0

Time periods: 5 P-value 0.0000
 Distribution: F(12,39091)
 Avg Obs: 7820.8
 Min Obs: 6480.0
 Max Obs: 1.021e+04

Parameter Estimates

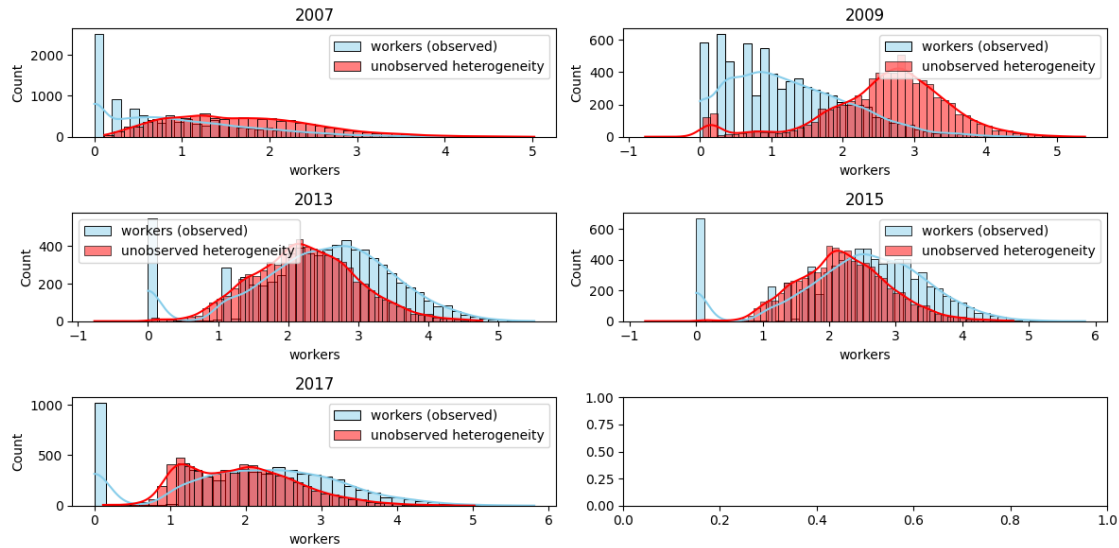
	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
const	0.1072	0.0133	8.0397	0.0000	0.0810	0.1334
sales	-0.1688	0.0040	-42.461	0.0000	-0.1766	-0.1610
age	0.0127	0.0009	14.294	0.0000	0.0109	0.0144
foreign	0.1056	0.0344	3.0740	0.0021	0.0383	0.1729
export	0.0588	0.0237	2.4802	0.0131	0.0123	0.1052
iyd	-0.1733	0.0133	-13.030	0.0000	-0.1994	-0.1473
utilidades	4.289e-05	2.26e-06	18.977	0.0000	3.846e-05	4.732e-05
msales	0.4929	0.0060	82.372	0.0000	0.4812	0.5047
mage	0.0040	0.0010	3.7858	0.0002	0.0019	0.0060
mforeign	0.2727	0.0443	6.1574	0.0000	0.1859	0.3594
mexport	0.3032	0.0336	9.0196	0.0000	0.2373	0.3691
miyd	0.5401	0.0202	26.707	0.0000	0.5004	0.5797
mutilidades	-0.0002	8.019e-06	-28.935	0.0000	-0.0002	-0.0002

9.1 Respuesta

Se eligen las mismas variables que para los modelos anteriores. Se obtiene un buen resultado para las pruebas F y para los parámetros β . Notar que los resultados para las variables que no son en promedio son iguales a los de efectos fijos. Cabe tambien destacar, que se tiene un mayor r-squared en el resultado, en comparación a los otros modelos. Notar que para este modelo se supone que $E(U_{it}|X_{it}) = 0$.

10 Pregunta 7: CRE - Componente no observado

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?



10.1 Respuesta

Se calculó la distribución del componente no observado usando el modelo CRE. Notar que para cada año, según se puede ver en la figura anterior, se tiene una distribución distinta, por lo que se presenta un sesgo que está más marcado en los años 2009 y 2013.

11 Pregunta 8: Preferencia de modelo

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?

11.1 Respuesta

En base a lo anteriormente analizado, prefiero el modelo de Efectos aleatorios correlacionados, pues tiene una mayor capacidad al incluir tanto la variación de cada individuo en el tiempo como la heterogeneidad no observada, a pesar de tener una mayor cantidad de supuestos fuertes.

12 Pregunta 9: Experimento

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).

No responde