

# ECONOMETRÍA DE DATOS DE PANEL

## Trabajo práctico N°2

Juan Guillermo Muñoz Delgado

### Exercise 1

(a) El ejercicio propone la siguiente ecuación para el logaritmo natural del salario:

$$\ln(\text{wage}_{it}) = \theta_t + \beta_1 \text{educ}_i + \beta_2 \text{black}_i + \beta_3 \text{hisp}_i + \beta_4 \text{exper}_{it} + \beta_5 \text{expersq}_{it} + \beta_6 \text{married}_{it} + \beta_7 \text{union}_{it} + c_i + u_{it}$$

Estimando los parámetros por mínimos cuadrados clásicos en el contexto de datos de panel se obtiene como resultado:

Cuadro 1: POLS

	Estimate	Std. Error	t-value	Pr(> t )
d81	0,058	0,030	1,926	0,054
d82	0,057	0,033	1,739	0,082
d83	0,051	0,035	1,443	0,149
d84	0,075	0,038	1,980	0,048
d85	0,090	0,040	2,239	0,025
d86	0,119	0,042	2,827	0,005
d87	0,147	0,044	3,349	0,001
educ	0,097	0,003	38,198	0
black	-0,139	0,024	-5,883	0
hisp	0,020	0,020	0,998	0,318
exper	0,076	0,012	6,424	0
expersq	-0,003	0,001	-3,397	0,001
married	0,107	0,016	6,822	0
union	0,183	0,017	10,640	0

Los errores estándar reportados anteriormente no son confiables dado que los errores compuestos del modelo están correlacionados, por construcción, a lo largo del tiempo para cada  $i$  ( $\text{Cov}(v_{1t}, v_{it-1}) \neq 0$ ). Esto es particularmente conflictivo a la hora de hacer inferencia estadística porque se está subestimando la desviación de los estimadores. Teniendo esto en cuenta, tenemos dos opciones: adoptar los coeficientes y corregir los errores estándar por los robustos ante la presencia de heterocedasticidad y/o correlación serial o estimar por mínimos cuadrados generalizados. Siguiendo el primer camino los resultados son los siguientes:

Cuadro 2: POLS ROBUSTO

	Estimate	Std. Error ROB
1981	0,150	0,028
1982	0,155	0,032
1983	0,154	0,038
1984	0,183	0,045
1985	0,201	0,051
1986	0,234	0,057
1987	0,266	0,0638
educ	0,091	0,003
black	-0,139	0,050
hisp	0,016	0,037
exper	0,067	0,015
expersq	-0,002	0,001
married	0,108	0,026
union	0,182	0,027

(b) Estimando la ecuación por efectos aleatorios se obtiene:

Cuadro 3: RE

	Estimate	Std. Error	z-value	Pr(> z )
d81	0,039	0,024	1,659	0,097
d82	0,029	0,029	0,995	0,320
d83	0,017	0,035	0,479	0,632
d84	0,038	0,041	0,931	0,352
d85	0,052	0,048	1,085	0,278
d86	0,085	0,054	1,556	0,120
d87	0,127	0,061	2,060	0,039
educ	0,093	0,003	31,197	0
black	-0,139	0,048	-2,906	0,004
hisp	0,023	0,042	0,546	0,585
exper	0,107	0,012	8,600	0
expersq	-0,005	0,001	-6,931	0
married	0,064	0,017	3,803	0,0001
union	0,106	0,018	5,935	0

Como se tienen variables que cambian a lo largo del tiempo para cada individuo  $i$ , los estimadores de POLS y RE no coinciden (contrario a lo que pasaba cuando solo se estimaba con las variables dicotómicas de tiempo, educ, black e hispan).

(c) La estimación por efectos fijos, teniendo en cuenta que todas aquellas variables que son constantes para cada individuo serán eliminadas por la transformación, tiene como resultado:

Cuadro 4: FE

	Estimate	Std. Error	t-value	Pr(> t )
d81	0,151	0,022	6,888	0
d82	0,253	0,024	10,360	0
d83	0,354	0,029	12,121	0
d84	0,490	0,036	13,529	0
d85	0,617	0,045	13,648	0
d86	0,765	0,056	13,638	0
d87	0,925	0,069	13,450	0
expersq	-0,005	0,001	-7,361	0
married	0,047	0,018	2,549	0,011
union	0,080	0,019	4,143	0,00004

Se puede notar que además de haber excluido a las variables de corte transversal, R también descartó la variable *exper* debido a su colinealidad perfecta a través del tiempo ( $exper_{it} = exper_{i1} + t \quad \forall t \neq 1$ ). Por otro lado, los coeficientes de *married* y *union* decrecieron, manteniendo el signo positivo, con respecto a la estimación por efectos aleatorios.

- (d) En la siguiente tabla se muestran la estimación por diferencias finitas y se comparan restando los coeficientes para ver que tan cerca del cero se encuentran:

Cuadro 5:

	Estimate [FD]	[FD]-[FE]
d81	0,156	0,004
d82	0,264	0,010
d83	0,372	0,017
d84	0,516	0,025
d85	0,651	0,033
d86	0,808	0,042
d87	0,981	0,056
expersq	-0,006	0
married	0,038	0,008
union	0,041	0,038

- (e) Multiplicando la variable educacional por la variable dicotómica temporal y estimando por efectos fijos se obtiene:

Cuadro 6: FE  $\theta_t educ_{it}$ 

	Estimate	Std. Error	t-value	Pr(> t )
d81	0,098	0,146	0,674	0,500
d82	0,247	0,149	1,655	0,098
d83	0,409	0,156	2,625	0,009
d84	0,640	0,165	3,873	0,0001
d85	0,773	0,178	4,343	0,00001
d86	0,970	0,194	4,995	0,00000
d87	1,189	0,214	5,566	0,00000
expersq	-0,006	0,001	-7,001	0
married	0,047	0,018	2,588	0,010
union	0,079	0,019	4,085	0,00004
d81*educ	0,005	0,012	0,408	0,683
d82*educ	0,002	0,012	0,134	0,893
d83*educ	-0,003	0,013	-0,213	0,831
d84*educ	-0,010	0,013	-0,770	0,441
d85*educ	-0,009	0,013	-0,705	0,481
d86*educ	-0,012	0,013	-0,903	0,367
d87*educ	-0,016	0,014	-1,139	0,255

Los resultados de los estimadores parecen indicar que los retornos a la educación en términos de salario decrecieron con el tiempo, sin embargo, también se puede apreciar como todos los estimadores asociados a las nuevas variables tienen un pobre nivel de significancia estadística.

(f) El ejercicio propone la siguiente ecuación:

$$\ln(\text{wage}_{it}) = \theta_t + \beta_1 \text{educ}_i + \beta_2 \text{black}_i + \beta_3 \text{hisp}_i + \beta_4 \text{exper}_{it} + \beta_5 \text{expersq}_{it} + \beta_6 \text{married}_{it} + \beta_7 \text{union}_{it} + \text{union}_{it+1} + c_i + u_{it}$$

Estimando por efectos fijos se llega a:

Cuadro 7: FE con  $\text{union}_{it+1}$ 

	Estimate	Std. Error	t-value	Pr(> t )
d81	0,153	0,023	6,747	0
d82	0,258	0,026	9,830	0
d83	0,362	0,033	11,007	0
d84	0,502	0,042	11,901	0
d85	0,634	0,054	11,753	0
d86	0,784	0,068	11,548	0
expersq	-0,005	0,001	-6,207	0
married	0,045	0,021	2,149	0,032
union	0,076	0,022	3,508	0,0005
union <sub>it+1</sub>	0,050	0,022	2,224	0,026

Concentrándonos en la última fila podemos ver como el estimador es positivo y posee un nivel de significatividad individual aceptado bajo un p-valor de 0.05. El supuesto de exogeneidad estricta nos pide:

$$E[u_{it}|x_{i1}, x_{i2}, \dots x_{iT}] = 0 \quad t = 1, \dots, T$$

Lo cual no se cumple en la medida que cada una de las ecuaciones parece estar explicada con una variables del periodo siguiente.

- (g) Para replicar los resultados del *paper* se agrega una constante al modelo que hasta ahora no se había tenido en cuenta. Aquellas columnas que corresponden a estimaciones por mínimos cuadrados clásicos tendrán un control adicional por variables de tiempo. La ecuación a estimar está compuesta por nueve variables explicativas principales: *union*, *educ*, *exper*, *expersq*, *hisp*, *black*, *hur*, *married* y *poorhlth*. Adicionalmente, se controla por sector industrial y ubicación geográfica. El ejercicio consiste en calcular por POLS y FE en los casos en los que se incluye las variables ocupacionales (denotadas en el data frame como occ) y en los que no. Las tablas que se muestran a continuación contiene las estimaciones propias para estos experimentos, donde solo voy a mostrar las estimaciones de los parámetros enseñados por el *paper* para su fácil comparación<sup>1</sup>.

Cuadro 8: Relevancia de OCC para OLS

	$\beta_{OLS}[1]$	Std. Error[1]	$\beta_{OLS}[2]$	Std. Error[2]
(Intercept)	0,320	0,163	0,388	0,165
union	0,148	0,026	0,177	0,025
educ	0,084	0,010	0,073	0,010
exper	0,059	0,017	0,057	0,017
expersq	-0,002	0,0	-0,002	0,000
hisp	-0,059	0,040	-0,047	0,039
black	-0,150	0,046	-0,126	0,044
hur	-0,129	0,031	-0,114	0,031
married	0,110	0,024	0,102	0,023
poorhlth	-0,055	0,067	-0,032	0,069

Cuadro 9: Relevancia de OCC para FE

	$\beta_{FE}[3]$	Std. Error[3]	$\beta_{FE}[4]$	Std. Error[4]
union	0,080	0,020	0,079	0,019
exper	0,111	0,009	0,112	0,008
expersq	-0,004	0,001	-0,004	0,001
hur	0,048	0,029	0,050	0,029
married	0,038	0,018	0,040	0,018
poorhlth	-0,010	0,047	-0,017	0,047

<sup>1</sup>Si se quiere ver el total de los resultados se puede visitar el siguiente enlace