

Datos de Panel
Problem Set 0
Repaso OLS, GLS & FGLS
Lectura y resumen de datos de panel en Stata

1. Considere el siguiente modelo de regresión:

$$ltotexp_i = \beta_0 + \beta_1 suppins_i + \beta_2 phylim_i + \beta_3 actlim_i + \beta_4 totchr_i + \beta_5 age_i + \beta_6 female_i + \beta_7 income_i + u_i, \quad i = 1, \dots, N$$

- a) Use la base de datos “mus03data.dta”, la cual contiene datos de corte transversal de gastos médicos, para estimar la ecuación por OLS usando comandos de matrices en Stata. Adicionalmente, reporte los errores estándar usuales de OLS y los estadísticos t asociados.
 - b) Utilice el comando `regress` para verificar los resultados obtenidos.
 - c) Implemente un test de significatividad individual para `totchr`.
 - d) Implemente un test de significatividad conjunta para todas las variables del modelo, excluyendo el intercepto.
2. En este ejercicio vamos a aprender cómo setear los datos como panel en Stata y cómo generar estadísticas descriptivas del panel. Adicionalmente, veremos cómo convertir los datos de *wide form* a *long form* y cómo generar un panel para simulaciones.
- a) Utilice la base `mus08psidextract.dta` y describa la base de datos de la manera usual y como un panel.
 - b) Utilice la base `pigweights.dta`. Los datos se encuentran en formato *wide*. Utilice el comando `reshape` para llevarlos a formato *long*. Luego, describa la base de la misma forma que en el inciso (a).
 - c) Genere un panel de 5000 observaciones con 10 períodos temporales y 500 unidades en el corte transversal. El panel debe estar en formato *long*. Genere observaciones de $x_{it} \sim \mathcal{N}(0, 1)$, $u_{it} \sim \mathcal{N}(0, 1)$ y además $y_{it} = 1 + x_{it} + u_{it}$. Estime por POLS.
3. Considere el siguiente modelo:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i, \quad i = 1, \dots, N$$
$$u_i = \sqrt{\exp(-1 + 0,2 \cdot x_{2i})} \cdot \varepsilon_i, \quad i = 1, \dots, N$$

con $\beta_1 = 1$, $\beta_2 = 1$, $\beta_3 = 1$, $x_2 \sim \mathcal{N}(0, 25)$, $x_3 \sim \mathcal{N}(0, 25)$ y $\varepsilon \sim \mathcal{N}(0, 25)$. Luego, el error u es heterocedástico con una varianza condicional igual a $25 \cdot \exp(-1 + 0,2 \cdot x_2)$.

- a) Genere 1000 muestras de $N=10$ observaciones a partir del modelo presentado. Para cada muestra estime por OLS, GLS y FGLS los parámetros del modelo y realice un test de hipótesis para contrastar que $H_0 : \beta_3 = 1$. Reporte tamaño del test al 1 %. Adicionalmente, reporte la media, mediana y desvío estándar de las estimaciones de β_1 , β_2 y β_3 .
- b) Repita el punto anterior con N igual a 20, 30, 100, 200 y 500.
- c) Describa detalladamente las propiedades de muestra finita de FGLS de acuerdo a lo que observó de los puntos anteriores.

Trabajo Práctico N° 0: **Repaso OLS, GLS y FGLS. Lectura y Resumen de Datos de Panel en Stata.**

Ejercicio 1.

Considerar el siguiente modelo de regresión:

$$ltotexp_i = \beta_0 + \beta_1 suppins_i + \beta_2 phylim_i + \beta_3 actlim_i + \beta_4 totchr_i + \beta_5 age_i + \beta_6 female_i + \beta_7 income_i + u_i, i = 1, \dots, N.$$

(a) Usar la base de datos “mus03data.dta”, la cual contiene datos de corte transversal de gastos médicos, para estimar la ecuación por OLS usando comandos de matrices en Stata. Adicionalmente, reportar los errores estándar usuales de OLS y los estadísticos t asociados.

	beta	se	t
suppins	.25564276	.04622641	5.5302312
phylim	.30205979	.05697091	5.3020003
actlim	.35600541	.06211178	5.7316894
totchr	.37582014	.01842273	20.399812
age	.00380163	.00365613	1.039797
female	-.08432753	.0455442	-1.8515536
income	.00254982	.0010194	2.5013046
_cons	6.7037374	.27675999	24.222206

(b) Utilizar el comando regress para verificar los resultados obtenidos.

Source	SS	df	MS	Number of obs	=	2,955
Model	1264.72124	7	180.674463	F(7, 2947)	=	124.98
Residual	4260.16814	2,947	1.44559489	Prob > F	=	0.0000
				R-squared	=	0.2289
				Adj R-squared	=	0.2271
Total	5524.88938	2,954	1.87030785	Root MSE	=	1.2023

ltotexp	Coefficient	Std. err.	t	P> t	[95% conf. interval]
suppins	.2556428	.0462264	5.53	0.000	.1650034 .3462821
phylim	.3020598	.0569709	5.30	0.000	.190353 .4137666
actlim	.3560054	.0621118	5.73	0.000	.2342185 .4777923
totchr	.3758201	.0184227	20.40	0.000	.3396974 .4119429
age	.0038016	.0036561	1.04	0.299	-.0033672 .0109705
female	-.0843275	.0455442	-1.85	0.064	-.1736292 .0049741
income	.0025498	.0010194	2.50	0.012	.000551 .0045486
_cons	6.703737	.27676	24.22	0.000	6.161075 7.2464

(c) Implementar un test de significatividad individual para totchr.

```
(1) totchr = 0
```

```
F( 1, 2947) = 416.15
Prob > F = 0.0000
```

Por lo tanto, se puede observar que, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que la variable *totchr* es estadísticamente significativa.

(d) *Implementar un test de significatividad conjunta para todas las variables del modelo, excluyendo el intercepto.*

```
(1) suppins = 0
(2) phylim = 0
(3) actlim = 0
(4) totchr = 0
(5) age = 0
(6) female = 0
(7) income = 0
```

```
F( 7, 2947) = 124.98
Prob > F = 0.0000
```

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que las variables del modelo, en conjunto, son estadísticamente significativas.

Ejercicio 2.

En este ejercicio, se va a aprender cómo setear los datos como panel en Stata y cómo generar estadísticas descriptivas del panel. Adicionalmente, se verá cómo convertir los datos de wide form a long form y cómo generar un panel para simulaciones.

(a) Utilizar la base “mus08psidextract.dta” y describir la base de datos de la manera usual y como un panel.

Stata.

(b) Utilizar la base “pigweights.dta”. Los datos se encuentran en formato wide. Utilizar el comando reshape para llevarlos a formato long. Luego, describir la base de la misma forma que en el inciso (a).

Stata.

(c) Generar un panel de 5000 observaciones con 10 períodos temporales y 500 unidades en el corte transversal. El panel debe estar en formato long. Generar observaciones de $x_{it} \sim \mathcal{N}(0, 1)$, $u_{it} \sim \mathcal{N}(0, 1)$ y, además, $y_{it} = 1 + x_{it} + u_{it}$. Estimar por POLS.

Source		SS	df	MS	Number of obs	=	5,000
Model		5006.33835	1	5006.33835	F(1, 4998)	=	4972.58
Residual		5031.92865	4,998	1.00678844	Prob > F	=	0.0000
Total		10038.267	4,999	2.00805501	R-squared	=	0.4987
					Adj R-squared	=	0.4986
					Root MSE	=	1.0034

y		Coefficient	Std. err.	t	P> t	[95% conf. interval]
x		1.01052	.0143303	70.52	0.000	.9824264 1.038614
_cons		1.02376	.0141901	72.15	0.000	.9959417 1.051579

Ejercicio 3.

Considerar el siguiente modelo:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i, i = 1, \dots, N.$$

$$u_i = \sqrt{e^{(-1+0.2x_{2i})}} \varepsilon_i, i = 1, \dots, N.$$

con $\beta_1 = 1, \beta_2 = 1, \beta_3 = 1, x_2 \sim \mathcal{N}(0, 25), x_3 \sim \mathcal{N}(0, 25)$ y $\varepsilon \sim \mathcal{N}(0, 25)$. Luego, el error u es heterocedástico con una varianza condicional igual a $25e^{(-1+0.2x_2)}$.

(a) Generar 1000 muestras de $N = 10$ observaciones a partir del modelo presentado. Para cada muestra, estimar por OLS, GLS y FGLS los parámetros del modelo y realizar un test de hipótesis para contrastar que $H_0: \beta_3 = 1$. Reportar tamaño del test al 1%. Adicionalmente, reportar la media, mediana y desvío estándar de las estimaciones de β_1, β_2 y β_3 .

(b) Repetir el inciso anterior con N igual a 20, 30, 100, 200 y 500.

	N_10	N_20	N_30	N_100	N_200	N_500
tam_test_1~s	.8	1.1	.8	1	.8	.8
media_b1_ols	1.0054159	1.0233711	.9913776	1.0073724	.99101667	1.0004401
mediana_b1~s	1.0432544	1.0316013	1.011929	1.000255	1.0050259	1.0013884
de_b1_ols	1.4418954	.90375325	.72169432	.39574087	.28065068	.18146336
media_b2_ols	.99360726	1.0100861	1.0053778	1.003707	.99668205	1.0001666
mediana_b2~s	1.001702	1.0022839	1.0072177	1.001808	.99718451	1.001999
de_b2_ols	.37958122	.24111343	.19670063	.11009015	.07716763	.05011949
media_b3_ols	1.0118828	.99105405	.98908491	1.0008243	.99812821	1.0005964
mediana_b3~s	1.0206553	1.0012873	.99137709	1.0031048	.99734056	1.0004818
de_b3_ols	.30943622	.18576126	.15294122	.07752204	.05484899	.03532914
tam_test_1~s	.8	1.6	.7	.8	.7	1.3
media_b1_gls	1.0280305	1.0098569	.99142743	1.0028527	.99508303	1.0002454
mediana_b1~s	1.0521936	1.0215993	.99141711	.97967514	.99639186	1.0007963
de_b1_gls	1.2535653	.78859652	.6326124	.34452127	.24216176	.15728003
media_b2_gls	1.0041153	.99966412	1.0056483	1.0009635	.99825912	1.0002415
mediana_b2~s	1.0016103	1.0011113	1.0057288	.99995628	.99866092	1.0011777
de_b2_gls	.27698551	.14838535	.10750797	.05216017	.03670627	.02233929
media_b3_gls	.99808488	.99177559	.99704105	1.0014807	.99794066	1.0011461
mediana_b3~s	1.0007759	.99349907	1.0007703	1.0030935	.99855053	1.0003986
de_b3_gls	.23968415	.13024611	.10009673	.05074781	.03363409	.02125601
tam_test_1~s	2.6	1.9	1.1	1.3	.8	1.1
media_b1_f~s	1.01992	1.01678	.9940748	1.0072039	.99260909	1.0000234
mediana_b1~s	1.0103357	1.023284	.98835871	.98473564	.99415511	.99863401
de_b1_fgls	1.3995429	.84752793	.66857393	.35431677	.24580091	.15830694
media_b2_f~s	1.0032334	1.0017987	1.0081194	1.0025778	.99772954	1.000189
mediana_b2~s	1.0093035	1.0049713	1.0057506	1.0009448	.99776992	1.0012414
de_b2_fgls	.33065581	.18431274	.13396221	.0558778	.0377256	.02264737
media_b3_f~s	1.006726	.99314471	.99556178	1.0007816	.99824046	1.0011263
mediana_b3~s	1.0098851	.99663675	.99534097	1.0025634	.99817607	1.0005233
de_b3_fgls	.29018109	.14444993	.11601262	.05232108	.03433107	.02134511

(c) Describir, detalladamente, las propiedades de muestra finita de FGLS de acuerdo a lo que se observó de los puntos anteriores.

Las propiedades de muestra finita de FGLS, de acuerdo a lo que se observó de los puntos anteriores, son:

- Sesgo: El estimador FGLS puede estar sesgado si el modelo subyacente no se especifica correctamente o si la estructura de correlación verdadera en los datos se impone de manera incorrecta. Sin embargo, a medida que el tamaño de muestra aumenta, este sesgo tiende a disminuir.
- Eficiencia relativa: La eficiencia relativa del estimador FGLS en comparación con otros estimadores (como el estimador OLS) puede variar dependiendo de la estructura de correlación verdadera en los datos y del modelo. Sin embargo, en algunos casos, el estimador FGLS puede proporcionar estimaciones más precisas que el estimador OLS, especialmente cuando la estructura de correlación de los errores es ignorada por las estimaciones por OLS.
- Varianza finita: La varianza del estimador FGLS depende del tamaño de muestra y de la estructura de correlación verdadera en los datos. A diferencia de las propiedades asintóticas, en muestras finitas, la varianza del estimador FGLS puede no converger a la varianza asintótica y puede ser mayor o menor dependiendo de las características específicas de los datos y del modelo.
- Robustez: El estimador FGLS puede ser más robusto que otros estimadores en presencia de violaciones de los supuestos de homocedasticidad y de correlación serial en los datos. Esto significa que el estimador FGLS puede proporcionar estimaciones más precisas, incluso cuando los supuestos clásicos no se cumplan completamente.

Datos de Panel
Problem Set 1
Modelo de Regresión Lineal

1. Utilice la base de datos provista “*cornwell.dta*”.

- a) A partir de los datos de los siete años, y utilizando los logaritmos de todas las variables, estime un modelo por POLS que relacione la tasa de crimen con *prbarr*, *prbconv*, *prbpris*, *avgsen* y *polpc* y que incluya un conjunto de *dummies* de año.
- b) Compute los errores estándar robustos a heteroscedasticidad arbitraria y a autocorrelación serial arbitraria.
- c) Implemente un contraste de Correlación Serial.
- d) Implemente un contraste de Heterocedasticidad.
- e) Asuma que se cumple el supuesto de exogeneidad estricta y que u_{it} sigue un proceso AR(1). Compute el estimador de FGLS siguiendo el enfoque de Prais-Winsten. Una descripción del procedimiento puede encontrarla en Wooldridge (2010), sección 7.8.6.

Observación: GLS necesita exogeneidad estricta para conseguir estimadores consistentes.

- f) Compute los errores estándar robustos a heteroscedasticidad arbitraria y a autocorrelación serial arbitraria para el modelo con las variables transformadas del inciso previo.

Sugerencia de Wooldridge. “..If we have any doubts about the homoskedasticity assumption, or whether the AR(1) assumption sufficiently captures the serial dependence, we can just apply the usual fully robust variance matrix and associated statistics to pooled OLS on the transformed variables. This allows us to probably obtain an estimator more efficient than POLS (on the original data) but also guards against the rather simple structure we imposed on Ω . Of course, failure of strict exogeneity generally causes the Prais-Winsten estimator of β to be inconsistent.”

2. En este ejercicio examinará un modelo para el costo total de producción en la industria aeronáutica a modo de ilustrar una aplicación de un modelo heteroscedástico por grupos. Considere la siguiente función de costos:

$$\ln \text{cost}_{jt} = \beta_1 + \beta_2 \ln \text{output}_{jt} + \beta_3 \text{load factor}_{jt} + \beta_4 \ln \text{fuel price}_{jt} \\ + \delta_2 \text{Firm}_2 + \delta_3 \text{Firm}_3 + \delta_4 \text{Firm}_4 + \delta_5 \text{Firm}_5 + \delta_6 \text{Firm}_6 + \varepsilon_{jt}$$

- a) Utilice la base de datos provista “*greene97.dta*”, la cual contiene datos para seis compañías aéreas observadas anualmente durante 15 años. Estime la ecuación por POLS.
- b) Ahora, asuma que dentro de cada compañía aérea se tiene que:

$$\text{Var}[\varepsilon_{jt} \mid \mathbf{x}_{jt}] = \sigma_j^2, \quad t = 1, \dots, T$$

Por lo tanto, si las varianzas fueran conocidas, el estimador de GLS sería:

$$\hat{\beta} = \left[\sum_{j=1}^N \left(\frac{1}{\sigma_j^2} \right) \mathbf{X}_j' \mathbf{X}_j \right]^{-1} \left[\sum_{j=1}^N \left(\frac{1}{\sigma_j^2} \right) \mathbf{X}_j' \mathbf{y}_j \right]$$

donde \mathbf{X}_j es una matriz $T \times K$. Sin embargo, en este caso práctico las varianzas son desconocidas. Luego, se le solicita computar el estimador de FGLS a través de los siguientes métodos:

- 1) Estime el modelo calculando el estimador necesario para la varianza específica de la compañía aérea a partir de los residuos de OLS, es decir, $\hat{\sigma}_j^2 = \frac{\mathbf{e}_j' \mathbf{e}_j}{n_j}$.
 - 2) Estimar el modelo tratándolo como una forma del modelo de heteroscedasticidad multiplicativa de Harvey (1976). Utilice el procedimiento en dos etapas.
 - c) Compare los resultados obtenidos en el inciso b).
3. Considere la siguiente ecuación de salarios:

$$y_{jt} = \beta_0 + \beta_1 x_{jt} + u_{jt}, \quad j = 1, 2, \dots, N; t = 1, 2 \quad (1)$$

donde $\beta_0 = \beta_1 = 1$, $u_j \sim N(0, \Omega)$, $\Omega = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$ y $x_j \sim U[1, 20]$

Genere 1000 muestras de $N = 5$ observaciones de corte transversal a partir del modelo (1). Para cada muestra estime por FGLS los parámetros del modelo y realice un test de hipótesis para contrastar que $H_0 : \beta_1 = 1$. Reporte tamaño del test al 1% y el poder del test cuando $\beta_1 = 0,8$. Luego, repita el procedimiento con $N = 500$. ¿Se aprecia algún cambio en el tamaño y/o en el poder del test ante el incremento de N ?

Trabajo Práctico N° 1: **Modelo de Regresión Lineal.**

Ejercicio 1.

Utilizar la base de datos provista “cornwell.dta”.

(a) A partir de los datos de los siete años, y utilizando los logaritmos de todas las variables, estimar un modelo por POLS que relacione la tasa de crimen con prbarr, prbconv, prbpris, avgscen y polpc y que incluya un conjunto de dummies de año.

POLS:

Source	SS	df	MS	Number of obs	=	630
Model	117.644669	11	10.6949699	F(11, 618)	=	74.49
Residual	88.735673	618	.143585231	Prob > F	=	0.0000
				R-squared	=	0.5700
				Adj R-squared	=	0.5624
Total	206.380342	629	.328108652	Root MSE	=	.37893

lcrmrte	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
lprbarr	-.7195033	.0367657	-19.57	0.000	-.7917042	-.6473024
lprbconv	-.5456589	.0263683	-20.69	0.000	-.5974413	-.4938765
lprbpris	.2475521	.0672268	3.68	0.000	.1155314	.3795728
lavgsen	-.0867575	.0579205	-1.50	0.135	-.2005023	.0269872
lpolpc	.3659886	.0300252	12.19	0.000	.3070248	.4249525
d82	.0051371	.057931	0.09	0.929	-.1086284	.1189026
d83	-.043503	.0576243	-0.75	0.451	-.1566662	.0696601
d84	-.1087542	.057923	-1.88	0.061	-.222504	.0049957
d85	-.0780454	.0583244	-1.34	0.181	-.1925835	.0364928
d86	-.0420791	.0578218	-0.73	0.467	-.15563	.0714719
d87	-.0270426	.056899	-0.48	0.635	-.1387815	.0846963
_cons	-2.082293	.2516253	-8.28	0.000	-2.576438	-1.588149

(b) Computar los errores estándar robustos a heteroscedasticidad arbitraria y a autocorrelación serial arbitraria.

POLS (con errores estándar robustos):

Linear regression	Number of obs	=	630
	F(11, 89)	=	37.19
	Prob > F	=	0.0000
	R-squared	=	0.5700
	Root MSE	=	.37893

(Std. err. adjusted for 90 clusters in county)

		Robust				
	lcrmrte	Coefficient	std. err.	t	P> t	[95% conf. interval]
	lprbarr	-.7195033	.1095979	-6.56	0.000	-.9372719 -.5017347
	lprbconv	-.5456589	.0704368	-7.75	0.000	-.6856152 -.4057025
	lprbpris	.2475521	.1088453	2.27	0.025	.0312787 .4638255
	lavgsen	-.0867575	.1130321	-0.77	0.445	-.3113499 .1378348
	lpolpc	.3659886	.121078	3.02	0.003	.1254092 .6065681
	d82	.0051371	.0367296	0.14	0.889	-.0678439 .0781181
	d83	-.043503	.033643	-1.29	0.199	-.1103509 .0233448
	d84	-.1087542	.0391758	-2.78	0.007	-.1865956 -.0309127
	d85	-.0780454	.0385625	-2.02	0.046	-.1546683 -.0014224
	d86	-.0420791	.0428788	-0.98	0.329	-.1272783 .0431201
	d87	-.0270426	.0381447	-0.71	0.480	-.1028353 .0487502
	_cons	-2.082293	.8647054	-2.41	0.018	-3.800445 -.3641423

(c) Implementar un contraste de Correlación Serial.

Stata.

Se rechaza la hipótesis nula de no correlación serial.

(d) Implementar un contraste de Heterocedasticidad.

Stata.

Se rechaza la hipótesis nula de homocedasticidad.

(e) Asumir que se cumple el supuesto de exogeneidad estricta y que u_{it} sigue un proceso AR(1). Computar el estimador de FGLS siguiendo el enfoque de Prais-Winsten. Una descripción del procedimiento se puede encontrar en Wooldridge (2010), sección 7.8.6. Observación: GLS necesita exogeneidad estricta para conseguir estimadores consistentes.

FGLS:

Source	SS	df	MS	Number of obs	=	630
Model	885.585523	12	73.7987936	F(12, 618)	=	2050.54
Residual	22.2417551	618	.035989895	Prob > F	=	0.0000
				R-squared	=	0.9755
				Adj R-squared	=	0.9750
Total	907.827278	630	1.44099568	Root MSE	=	.18971

tilde_lcrmrte	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
tilde_lprbarr	-.481208	.0333124	-14.45	0.000	-.5466271	-.4157888
tilde_lprbconv	-.3353095	.0209135	-16.03	0.000	-.3763796	-.2942395
tilde_lprbpris	-.1624321	.0339271	-4.79	0.000	-.2290585	-.0958058
tilde_lavgsen	-.0203981	.0289633	-0.70	0.482	-.0772766	.0364804
tilde_lpolpc	.3806954	.0298461	12.76	0.000	.3220834	.4393074
tilde_d82	.0120433	.0222954	0.54	0.589	-.0317405	.0558272
tilde_d83	-.0721363	.0288915	-2.50	0.013	-.1288737	-.0153989
tilde_d84	-.1092092	.0333946	-3.27	0.001	-.1747898	-.0436286
tilde_d85	-.1018016	.0364716	-2.79	0.005	-.1734249	-.0301784
tilde_d86	-.0775719	.0381852	-2.03	0.043	-.1525605	-.0025834
tilde_d87	-.0395482	.0394024	-1.00	0.316	-.1169271	.0378307
tilde_ones	-2.027131	.2099692	-9.65	0.000	-2.439471	-1.614792

(f) *Computar los errores estándar robustos a heteroscedasticidad arbitraria y a autocorrelación serial arbitraria para el modelo con las variables transformadas del inciso previo. Sugerencia de Wooldridge: “... If we have any doubts about the homoskedasticity assumption, or whether the AR(1) assumption sufficiently captures the serial dependence, we can just apply the usual fully robust variance matrix and associated statistics to pooled OLS on the transformed variables. This allows us to probably obtain an estimator more efficient than POLS (on the original data) but also guards against the rather simple structure we imposed on Ω . Of course, failure of strict exogeneity generally causes the Prais-Winsten estimator of β to be inconsistent.”*

FGLS (con errores estándar robustos):

Linear regression	Number of obs	=	630
	F(12, 89)	=	837.00
	Prob > F	=	0.0000
	R-squared	=	0.9755
	Root MSE	=	.18971

(Std. err. adjusted for 90 clusters in county)

tilde_lcrmrte	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
tilde_lprbarr	-.481208	.0718373	-6.70	0.000	-.6239471	-.3384689
tilde_lprbconv	-.3353095	.0440331	-7.61	0.000	-.4228023	-.2478168
tilde_lprbpris	-.1624321	.0500207	-3.25	0.002	-.2618222	-.0630421
tilde_lavgsen	-.0203981	.0258077	-0.79	0.431	-.0716774	.0308813
tilde_lpolpc	.3806954	.106039	3.59	0.001	.1699982	.5913925
tilde_d82	.0120433	.015186	0.79	0.430	-.0181309	.0422176
tilde_d83	-.0721363	.01852	-3.90	0.000	-.1089352	-.0353374
tilde_d84	-.1092092	.0215974	-5.06	0.000	-.1521229	-.0662956
tilde_d85	-.1018016	.0244324	-4.17	0.000	-.1503483	-.053255
tilde_d86	-.0775719	.0236838	-3.28	0.002	-.1246312	-.0305126
tilde_d87	-.0395482	.0252489	-1.57	0.121	-.0897173	.0106209
tilde_ones	-2.027131	.7465981	-2.72	0.008	-3.510606	-.5436569

Ejercicio 2.

En este ejercicio, se examinará un modelo para el costo total de producción en la industria aeronáutica a modo de ilustrar una aplicación de un modelo heterocedástico por grupos. Considerar la siguiente función de costos:

$$\ln cost_{jt} = \beta_1 + \beta_2 \ln output_{jt} + \beta_3 load\ factor_{jt} + \beta_4 \ln fuel\ price_{jt} + \delta_2 Firm_2 + \delta_3 Firm_3 + \delta_4 Firm_4 + \delta_5 Firm_5 + \delta_6 Firm_6 + \varepsilon_{jt}.$$

(a) Utilizar la base de datos provista “greene97.dta”, la cual contiene datos para seis compañías áreas observadas, anualmente, durante 15 años. Estimar la ecuación por POLS.

Source	SS	df	MS	Number of obs	=	90
Model	113.74827	8	14.2185338	F(8, 81)	=	3935.79
Residual	.292622888	81	.003612628	Prob > F	=	0.0000
				R-squared	=	0.9974
				Adj R-squared	=	0.9972
Total	114.040893	89	1.28135835	Root MSE	=	.06011

lc	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
lq	.9192845	.0298901	30.76	0.000	.8598126	.9787565
lf	-1.070396	.20169	-5.31	0.000	-1.471696	-.6690961
lpf	.4174918	.0151991	27.47	0.000	.3872503	.4477333
id						
2	-.0412359	.025184	-1.64	0.105	-.0913441	.0088722
3	-.2089211	.0427986	-4.88	0.000	-.294077	-.1237653
4	.1845557	.0607527	3.04	0.003	.0636769	.3054345
5	.0240547	.0799041	0.30	0.764	-.1349293	.1830387
6	.0870617	.0841995	1.03	0.304	-.080469	.2545924
_cons	9.705942	.193124	50.26	0.000	9.321686	10.0902

(b) Ahora, asumir que, dentro de cada compañía área, se tiene que:

$$\text{Var} [\varepsilon_{jt} | x_{jt}] = \sigma_j^2, t = 1, \dots, T.$$

Por lo tanto, si las varianzas fueran conocidas, el estimador GLS sería:

$$\hat{\beta} = \left[\sum_{j=1}^N \frac{1}{\sigma_j^2} X_j' X_j \right]^{-1} \sum_{j=1}^N \frac{1}{\sigma_j^2} X_j' y_j,$$

donde X_j es una matriz $T \times K$. Sin embargo, en este caso práctico, las varianzas son desconocidas. Luego, se solicita computar el estimador de FGLS a través de los siguientes métodos:

(i) Estimar el modelo calculando el estimador necesario para la varianza específica de la compañía áreas a partir de los residuos de OLS, es decir, $\hat{\sigma}_j^2 = \frac{e_j' e_j}{n_j}$.

Source	SS	df	MS	Number of obs	=	90
				F(8, 81)	=	5526.83
Model	118.222298	8	14.7777873	Prob > F	=	0.0000
Residual	.216579991	81	.002673827	R-squared	=	0.9982
				Adj R-squared	=	0.9980
Total	118.438878	89	1.33077391	Root MSE	=	.05171

lc	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
lq	.925765	.0267809	34.57	0.000	.8724795	.9790506
lf	-1.216307	.1855858	-6.55	0.000	-1.585565	-.8470495
lpf	.4056077	.0125488	32.32	0.000	.3806395	.4305758
id						
2	-.046026	.0237611	-1.94	0.056	-.0933031	.0012511
3	-.2020985	.0361494	-5.59	0.000	-.2740246	-.1301725
4	.1905462	.0551602	3.45	0.001	.0807946	.3002977
5	.0371723	.0704438	0.53	0.599	-.1029887	.1773334
6	.094588	.0743639	1.27	0.207	-.0533728	.2425488
_cons	9.942316	.1622899	61.26	0.000	9.61941	10.26522

(ii) Estimar el modelo tratándolo como una forma del modelo de heteroscedasticidad multiplicativa de Harvey (1976). Utilizar el procedimiento en dos etapas.

Heteroskedastic linear regression	Number of obs	=	90
Two-step GLS estimation	Wald chi2(8)	=	36250.22
	Prob > chi2	=	0.0000

	lc	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
lc							
	lq	.932333	.0295289	31.57	0.000	.8744574	.9902086
	lf	-1.115165	.1991174	-5.60	0.000	-1.505428	-.7249023
	lpf	.4086271	.0141468	28.88	0.000	.3808999	.4363543
	id						
	2	-.0387054	.0242462	-1.60	0.110	-.0862271	.0088163
	3	-.1929047	.0407071	-4.74	0.000	-.2726892	-.1131203
	4	.2082512	.0589261	3.53	0.000	.0927583	.3237442
	5	.0572051	.0780464	0.73	0.464	-.0957631	.2101733
	6	.1207862	.0825116	1.46	0.143	-.0409335	.2825059
	_cons	9.841375	.1768449	55.65	0.000	9.494765	10.18798
lnsigma2							
	id						
	2	.9333314	.8111556	1.15	0.250	-.6565043	2.523167
	3	.575379	.8111556	0.71	0.478	-1.014457	2.165215
	4	.639489	.8111556	0.79	0.430	-.9503466	2.229325
	5	.6042102	.8111556	0.74	0.456	-.9856255	2.194046
	6	.7988952	.8111556	0.98	0.325	-.7909405	2.388731
	_cons	-6.213752	.5735736	-10.83	0.000	-7.337935	-5.089568

Wald test of $\ln\sigma^2=0$: $\chi^2(5) = 1.56$

```
Prob > chi2 = 0.9063
```

(c) *Comparar los resultados obtenidos en el inciso (b).*

Los resultados obtenidos en el inciso (b) son semejantes en cuanto a valores estimados de los parámetros y a significatividad estadística.

Ejercicio 3.

Considerar la siguiente ecuación de salarios:

$$y_{jt} = \beta_0 + \beta_1 x_{jt} + u_{jt}, j = 1, 2, \dots, N; t = 1, 2 \quad (1)$$

donde $\beta_0 = \beta_1 = 1$, $u_j \sim \mathcal{N}(0, \Omega)$, $\Omega = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$ y $x_{jt} \sim U[1, 20]$.

Generar 1000 muestras de $N = 5$ observaciones de corte transversal a partir del modelo (1). Para cada muestra, estimar por FGLS los parámetros del modelo y realizar un test de hipótesis para contrastar que $H_0: \beta_1 = 1$. Reportar tamaño del test al 1% y el poder del test cuando $\beta_1 = 0,8$. Luego, repetir el procedimiento con $N = 500$. ¿Se aprecia algún cambio en el tamaño y/o en el poder del test ante el incremento de N ?

	N_5	N_500
tam_test_1	2.7	1
poder_tes~08	33.2	100

Por lo tanto, se puede observar que, ante el incremento de N , el tamaño del test tiende al nivel de significación del 1% y el poder del test tiende al 100%.

Datos de Panel
Problem Set 2
Modelos de Datos de Panel Lineales

1. Utilice nuevamente la base de datos “*cornwell.dta*” provista para el Problem Set 1. Considere el siguiente modelo de regresión:

$$\ln crmrte_{it} = \beta_0 + \beta_1 \ln prbarr_{it} + \beta_2 \ln prbconv_{it} + \beta_3 \ln prbpris_{it} + \beta_4 \ln avgse_{it} \\ + \beta_5 \ln polpc_{it} + \sum_{\tau=82}^{87} \beta_{\tau} \cdot I\{t = \tau\} + \mu_i + \varepsilon_{it}$$

- Utilizando el comando *egen* de STATA, construya las medias individuales de las variables del modelo.
 - Aplique la transformación *within* al modelo. Luego, estime el modelo transformado por POLS.
 - Comente sobre la validez de los errores estándar del inciso previo.
 - Utilice el comando *xtreg* para estimar nuevamente el modelo usando efectos fijos.
 - Estime el modelo usando diferencias finitas de primer orden.
2. Utilice la base de datos provista “*murder.dta*”. La base de datos es una muestra longitudinal de estados de EE.UU., para los años 1987, 1990 y 1993.

- Estime por OLS el efecto de las ejecuciones (x) sobre la tasa de homicidios (*murder rates*, m) controlando por desempleo (u) y año:

$$m_{i,t} = \alpha + \beta_x x_{i,t} + \beta_u u_{i,t} + \beta_{90} d_{90,t} + \beta_{93} d_{93,t} + \nu_{i,t}$$

Note que se omitió la dummy temporal para el año 1987. Interprete los resultados.

- ¿Por qué podría ser importante tener en consideración los efectos temporales agregados en el modelo?
- Ahora, considere la siguiente modificación en el modelo:

$$m_{i,t} = \alpha + \beta_x x_{i,t} + \beta_u u_{i,t} + \beta_{90} d_{90,t} + \beta_{93} d_{93,t} + c_i + e_{i,t}$$

donde c_i es un efecto individual por estado. Estime la ecuación usando efectos fijos.

- Repita la estimación del inciso previo usando diferencias finitas de primer orden.
- Brinde un ejemplo bajo el cual la variable de ejecuciones no sería estrictamente exógena (condicional en c_i). **Observación.** Para obtener estimaciones consistentes, el modelo de efectos fijos asume exogeneidad estricta de las variables explicativas condicionadas en c_i .
- Repita la estimación del inciso c) usando el estimador de GLS para diferencias finitas de primer orden. Compruebe que los coeficientes estimados son iguales a los obtenidos por FE.
- Reestimar el modelo del inciso c) usando efectos aleatorios. Implementar el test de Hausman. ¿Cuál es el mejor estimador?

3. Considere el siguiente modelo:

$$y_{it} = x_{it}\beta + \mu_i + \nu_{it}, \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T$$

donde $x_{it} \stackrel{iid}{\sim} \mathcal{N}(0, 1)$, $u_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_\mu^2)$, $\nu_{it} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_\nu^2)$ y $\mu_i \perp \nu_{it}$ para todo i, t . Suponga que $\beta = \sigma_\mu^2 = \sigma_\nu^2 = 1$ y $T = 10$. La idea es realizar experimentos de Monte Carlo para evaluar la eficiencia de distintos estimadores de β .

- a) Caso 1: $N = 5$. Realice un experimento de Monte Carlo con 1000 simulaciones. Reporte media, desvío estándar y RMSE de la estimación de β usando: POLS, RE y FE.
 - b) Repita el punto anterior con $N = 10, 30, 50, 100$ y 500.
 - c) Comente los resultados obtenidos y su conclusión de qué estimador debiera utilizarse en la práctica.
4. Basado en el Ejercicio 10.18 de Wooldridge (2010). Utilice la base de datos *wagepan.dta* para responder las preguntas a continuación.
- a) Utilizando *lwage* como variable dependiente, estimar un modelo que contenga un intercepto y las variables *dummy* de año *d81* a *d87*. Estime el modelo por POLS, RE, FE y FD. ¿Qué puede concluir acerca de los coeficientes de las variables *dummy*?
 - b) Añada las variables constantes en el tiempo *educ*, *black* e *hisp* al modelo, y estímelo por POLS y RE. ¿Cómo se comparan los coeficientes? ¿Qué ocurre si se estima la ecuación por FE?
 - c) ¿Son iguales los errores estándar de POLS y RE del inciso b)? ¿Cuáles son probablemente más fiables?
 - d) Obtenga los errores estándar robustos para POLS. ¿Prefiere estos o los errores estándar habituales de RE?
 - e) Obtenga los errores estándar robustos de RE. ¿Cómo se comparan con los errores estándar robustos de POLS, y por qué?

Trabajo Práctico N° 2: Modelos de Datos de Panel Lineales.

Ejercicio 1.

Utilizar, nuevamente, la base de datos “cornwell.dta” provista para el Problem Set 1. Considerar el siguiente modelo de regresión:

$$\ln crmrte_{it} = \beta_0 + \beta_1 \ln prbarr_{it} + \beta_2 \ln prbconv_{it} + \beta_3 \ln prbpris_{it} + \beta_4 \ln avgsen_{it} + \beta_5 \ln polpc_{it} + \sum_{\tau=1}^{87} \beta_{\tau} I\{t = \tau\} + \mu_i + \varepsilon_{it}.$$

(a) Utilizando el comando *egen* de STATA, construir las medias individuales de las variables del modelo.

Stata.

(b) Aplicar la transformación *within* al modelo. Luego, estimar el modelo transformado por POLS.

POLS:

Source	SS	df	MS	Number of obs	=	630
Model	7.81221835	11	.710201668	F(11, 619)	=	43.19
Residual	10.1785214	619	.016443492	Prob > F	=	0.0000
				R-squared	=	0.4342
				Adj R-squared	=	0.4242
Total	17.9907397	630	.02855673	Root MSE	=	.12823

within_lcrmrte	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
within_lprbarr	-.3597944	.0299699	-12.01	0.000	-.4186493	-.3009395
within_lprbconv	-.2858733	.0196143	-14.57	0.000	-.3243919	-.2473547
within_lprbpris	-.1827812	.0300086	-6.09	0.000	-.2417122	-.1238502
within_lavgsen	-.0044879	.0244449	-0.18	0.854	-.0525009	.043525
within_lpolpc	.4241142	.0243741	17.40	0.000	.3762483	.4719802
within_d82	.0125802	.0199141	0.63	0.528	-.0265271	.0516875
within_d83	-.0792813	.0197277	-4.02	0.000	-.1180225	-.04054
within_d84	-.1177281	.0199815	-5.89	0.000	-.1569678	-.0784884
within_d85	-.1119561	.0201954	-5.54	0.000	-.151616	-.0722962
within_d86	-.0818268	.0198078	-4.13	0.000	-.1207254	-.0429282
within_d87	-.0404704	.0194497	-2.08	0.038	-.0786657	-.0022751

(c) Comentar sobre la validez de los errores estándar del inciso previo.

Los errores estándar reportados tienden a ser pequeños comparados a los verdaderos. El problema se encuentra en que los grados de libertad de aplicar OLS al modelo transformado no coinciden con el denominador del estimador consistente para σ_{ε}^2 . Por consiguiente, excepto que T sea lo suficientemente grande, se necesita corregir este denominador.

(d) Utilizar el comando `xtreg` para estimar, nuevamente, el modelo usando efectos fijos.

FE:

```
Fixed-effects (within) regression      Number of obs   =      630
Group variable: county                Number of groups =      90

R-squared:                            Obs per group:
    Within = 0.4342                      min =          7
    Between = 0.4066                     avg =         7.0
    Overall = 0.4042                     max =          7

corr(u_i, Xb) = 0.2068                  F(11,89)        =     11.49
                                      Prob > F          =     0.0000
```

(Std. err. adjusted for 90 clusters in county)

		Robust				
	lcrmrte	Coefficient	std. err.	t	P> t	[95% conf. interval]
lprbarr		-.3597944	.0594678	-6.05	0.000	-.4779557 -.2416332
lprbconv		-.2858733	.051522	-5.55	0.000	-.3882464 -.1835001
lprbpris		-.1827812	.0452811	-4.04	0.000	-.2727538 -.0928085
lavgsen		-.0044879	.0333499	-0.13	0.893	-.0707535 .0617777
lpolpc		.4241142	.0849052	5.00	0.000	.2554095 .592819
d82		.0125802	.0160066	0.79	0.434	-.0192246 .044385
d83		-.0792813	.0195639	-4.05	0.000	-.1181544 -.0404081
d84		-.1177281	.0217118	-5.42	0.000	-.160869 -.0745872
d85		-.1119561	.0256583	-4.36	0.000	-.1629386 -.0609736
d86		-.0818268	.0236276	-3.46	0.001	-.1287745 -.0348792
d87		-.0404704	.0241765	-1.67	0.098	-.0885087 .0075678
_cons		-1.604135	.5102062	-3.14	0.002	-2.617904 -.5903664
sigma_u		.43487416				
sigma_e		.13871215				
rho		.90765322	(fraction of variance due to u_i)			

(e) Estimar el modelo usando diferencias finitas de primer orden.

FD:

Source	SS	df	MS	Number of obs	=	540
Model	9.60258308	11	.872962098	F(11, 529)	=	36.66
Residual	12.5963755	529	.023811674	Prob > F	=	0.0000
				R-squared	=	0.4326
				Adj R-squared	=	0.4208
Total	22.1989586	540	.041109183	Root MSE	=	.15431

D.lcrmrte	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
lprbarr						
D1.	-.3274942	.0299801	-10.92	0.000	-.3863889	-.2685995
lprbconv						
D1.	-.2381066	.0182341	-13.06	0.000	-.2739268	-.2022864
lprbpris						
D1.	-.1650462	.025969	-6.36	0.000	-.2160613	-.1140312
lavgsen						
D1.	-.0217607	.0220909	-0.99	0.325	-.0651574	.021636
lpolpc						
D1.	.3984264	.026882	14.82	0.000	.3456177	.451235
d82						
D1.	.0077134	.0170579	0.45	0.651	-.0257961	.0412229
d83						
D1.	-.0844391	.0234564	-3.60	0.000	-.1305182	-.03836
d84						
D1.	-.1246632	.0287464	-4.34	0.000	-.1811344	-.068192
d85						
D1.	-.121561	.03315	-3.67	0.000	-.1866827	-.0564392
d86						
D1.	-.0863333	.0366763	-2.35	0.019	-.1583823	-.0142842
d87						
D1.	-.0377932	.0399728	-0.95	0.345	-.116318	.0407316

Ejercicio 2.

Utilizar la base de datos provista “murder.dta”. La base de datos es una muestra longitudinal de estados de EE.UU., para los años 1987, 1990 y 1993.

(a) Estimar por OLS el efecto de las ejecuciones (x) sobre la tasa de homicidios (murder rates, m) controlando por desempleo (u) y año:

$$m_{i,t} = \alpha + \beta_x x_{i,t} + \beta_u u_{i,t} + \beta_{90} d_{90,t} + \beta_{93} d_{93,t} + v_{i,t},$$

Notar que se omitió la dummy temporal para el año 1987. Interpretar los resultados.

POLS:

Source	SS	df	MS	Number of obs	=	153
Model	977.390644	4	244.347661	F(4, 148)	=	3.05
Residual	11867.9475	148	80.1888343	Prob > F	=	0.0190
Total	12845.3381	152	84.5088034	R-squared	=	0.0761
				Adj R-squared	=	0.0511
				Root MSE	=	8.9548

mrdrtte	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
exec	.1627547	.1939295	0.84	0.403	-.2204738	.5459832
unem	1.390786	.4508653	3.08	0.002	.4998207	2.281751
d90	2.675335	1.816934	1.47	0.143	-.91515	6.26582
d93	1.607317	1.774768	0.91	0.367	-1.899842	5.114476
_cons	-1.864393	3.069517	-0.61	0.545	-7.930134	4.201349

(b) ¿Por qué podría ser importante tener en consideración los efectos temporales agregados en el modelo?

Tener en consideración los efectos temporales agregados en el modelo podría ser importante si la tasa de homicidios es afectada por factores macroeconómicos externos que afectan a todos los estados de EE.UU. de la misma manera. Por lo tanto, si no se incluyen estas variables, se debe suponer que cualquier cambio en la media de la tasa de homicidios en el tiempo se debe a las ejecuciones o a la tasa de desempleo y no a factores externos. Por otra parte, controlar por estas variables hace más factible que se cumpla el supuesto de ausencia de autocorrelación serial.

(c) Ahora, considerar la siguiente modificación en el modelo:

$$m_{i,t} = \alpha + \beta_x x_{i,t} + \beta_u u_{i,t} + \beta_{90} d_{90,t} + \beta_{93} d_{93,t} + c_i + e_{i,t},$$

donde c_i es un efecto individual por estado. Estimar la ecuación usando efectos fijos.

FE:

Fixed-effects (within) regression
 Group variable: id

Number of obs = 153
 Number of groups = 51

R-squared:
 Within = 0.0734
 Between = 0.0037
 Overall = 0.0108

Obs per group:
 min = 3
 avg = 3.0
 max = 3

corr(u_i, Xb) = 0.0010

F(4, 98) = 1.94
 Prob > F = 0.1098

mrdрте	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
exec	-.1383231	.1770059	-0.78	0.436	-.4895856	.2129395
unem	.2213158	.2963756	0.75	0.457	-.366832	.8094636
d90	1.556215	.7453273	2.09	0.039	.0771369	3.035293
d93	1.733242	.7004381	2.47	0.015	.3432454	3.123239
_cons	5.822104	1.915611	3.04	0.003	2.020636	9.623572
sigma_u	8.7527226					
sigma_e	3.5214244					
rho	.86068589	(fraction of variance due to u_i)				

F test that all u_i=0: F(50, 98) = 17.18 Prob > F = 0.0000

(d) Repetir la estimación del inciso previo usando diferencias finitas de primer orden.

FD:

Source	SS	df	MS	Number of obs	=	102
Model	119.103298	4	29.7758244	F(4, 98)	=	1.61
Residual	1812.28656	98	18.49272	Prob > F	=	0.1778
Total	1931.38986	102	18.9351947	R-squared	=	0.0617
				Adj R-squared	=	0.0234
				Root MSE	=	4.3003

cmrdрте	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
cexec	-.1150682	.1473871	-0.78	0.437	-.407553	.1774166
cunem	.1630854	.3079049	0.53	0.598	-.4479419	.7741126
cd90	1.51099	.6608967	2.29	0.024	.1994623	2.822518
cd93	1.725263	.8533453	2.02	0.046	.0318275	3.418699

(e) Brindar un ejemplo bajo el cual la variable de ejecuciones no sería, estrictamente, exógena (condicional en c_i). Observación: Para obtener estimaciones consistentes, el modelo de efectos fijos asume exogeneidad estricta de las variables explicativas condicionadas en c_i .

Un ejemplo bajo el cual la variable de ejecuciones ($x_{i,t}$) no sería estrictamente exógena (condicional en c_i) podría ser si los estados aumentan las ejecuciones futuras en respuesta a los shocks positivos actuales de la tasa de homicidios. Dado el tramo de tiempo relativamente corto de la base de datos, la retroalimentación de la tasa de homicidio a las

ejecuciones futuras puede no ser muy preocupante, ya que el proceso judicial en los casos de pena capital tiende a moverse lentamente. Por supuesto, si se acelerara debido a un aumento de la tasa de homicidios, eso podría violar la exogeneidad estricta. Con una serie temporal más larga, se podría añadir $x_{i,t+1}$ (e, incluso, valores de un futuro más lejano) y estimar la ecuación por FE, comprobando la significatividad estadística de la variable $x_{i,t+1}$. En el caso de que se encuentre que esta variable es estadísticamente significativa, se tendría evidencia a favor de que no se cumple el supuesto de exogeneidad estricta.

(f) Repetir la estimación del inciso (c) usando el estimador de GLS para diferencias finitas de primer orden. Comprobar que los coeficientes estimados son iguales a los obtenidos por FE.

```
bfdgls[4,1]
               mdrte
exec   - .13832306
unem    .22131582
d90     1.5562147
d93     1.7332421
```

(g) Reestimar el modelo del inciso (c) usando efectos aleatorios. Implementar el test de Hausman. ¿Cuál es el mejor estimador?

RE:

```
Random-effects GLS regression              Number of obs   =          153
Group variable: id                        Number of groups  =           51

R-squared:                                Obs per group:
    Within = 0.0680                               min =           3
    Between = 0.0731                              avg  =          3.0
    Overall = 0.0426                               max  =           3

corr(u_i, X) = 0 (assumed)                Wald chi2(4)      =           8.52
                                           Prob > chi2       =          0.0743
```

```
-----+-----
      mdrte | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-----+-----
      exec |   -.0543375   .1595008    -0.34   0.733    - .3669533   .2582784
      unem |    .3947507   .2848133     1.39   0.166    - .1634732   .9529745
      d90  |    1.732981   .7478556     2.32   0.020     .2672106   3.19875
      d93  |    1.699913   .7065606     2.41   0.016     .3150796   3.084746
      _cons |    4.635132   2.179451     2.13   0.033     .3634863   8.906778
-----+-----
      sigma_u |  8.2056677
      sigma_e |  3.5214244
      rho    |  .84447636   (fraction of variance due to u_i)
-----+-----
```

---- Coefficients ----				
	(b)	(B)	(b-B)	$\sqrt{\text{diag}(V_b - V_B)}$
	est_fe	est_re	Difference	Std. err.
exec	-.1383231	-.0543375	-.0839856	.0767503
unem	.2213158	.3947507	-.1734349	.0819749
d90	1.556215	1.732981	-.1767658	.
d93	1.733242	1.699913	.0333292	.

b = Consistent under H0 and Ha; obtained from xtreg.
 B = Inconsistent under Ha, efficient under H0; obtained from xtreg.

Test of H0: Difference in coefficients not systematic

$\chi^2(4) = (b-B)'[(V_b - V_B)^{-1}](b-B)$
 = 5.78
 Prob > χ^2 = 0.2165
 (V_b - V_B is not positive definite)

Por lo tanto, se puede observar que, considerando un nivel de significación del 10%, el mejor estimador es el de efectos fijos, ya que se rechaza la hipótesis nula de no correlación entre los regresores y los efectos fijos, por lo que el estimador de efectos aleatorios no es consistente.

Ejercicio 3.

Considerar el siguiente modelo:

$$y_{it} = x_{it}\beta + \mu_i + v_{it}, i = 1, 2, \dots, N; t = 1, 2, \dots, T,$$

donde $x_{it} \sim^{iid} \mathcal{N}(0, 1)$, $\mu_i \sim^{iid} \mathcal{N}(0, \sigma_\mu^2)$, $v_{it} \sim^{iid} \mathcal{N}(0, \sigma_v^2)$ y $\mu_i \perp v_{it}$ para todo i, t . Suponer $\beta = \sigma_\mu^2 = \sigma_v^2 = 1$ y $T = 10$. La idea es realizar experimentos de Monte Carlo para evaluar la eficiencia de distintos estimadores de β .

(a) Caso 1: $N = 5$. Realizar un experimento de Monte Carlo con 1000 simulaciones. Reportar media, desvío estándar y RMSE de la estimación de β usando: POLS, RE y FE.

(b) Repetir el punto anterior con $N = 10, 30, 50, 100$ y 500 .

	N_5	N_10	N_30	N_50	N_100	N_500
media_beta~s	.99804484	.99234775	1.0019568	1.0039147	1.0005582	.99963949
de_beta_pols	.19410442	.14170935	.07988537	.06148401	.04521915	.02023065
rmse_beta~s	.1940172	.14184504	.0798694	.06157782	.04519998	.02022375
media_beta~e	.99311721	.99036578	1.0021895	1.0027847	1.0005854	.99967618
de_beta_fe	.15215654	.10712643	.06006592	.04603343	.03388416	.01449802
rmse_beta_fe	.15223612	.10750542	.06007579	.0460946	.03387227	.01449439
media_beta~e	.99239395	.99075022	1.0022435	1.0029077	1.0005371	.99967278
de_beta_re	.15558624	.10656719	.05945567	.04579737	.0335924	.01453039
rmse_beta_re	.15569433	.10691477	.05946827	.04586672	.03357989	.01452681

(c) Comentar los resultados obtenidos y su conclusión de qué estimador debiera utilizarse en la práctica.

En primer lugar, es importante destacar que, dados los supuestos del modelo, los tres estimadores en consideración son consistentes. Por lo tanto, se debería esperar que, a medida que el tamaño muestral aumenta, la media de las estimaciones de β con los diferentes estimadores estén cerca del valor poblacional ($\beta = 1$). Ahora bien, para $N < 10$, el estimador FE es el que mejor funciona en términos de sesgo y de eficiencia. Luego, a partir de un tamaño de muestra de $N = 30$, ya se observa cómo el estimador RE es el más eficiente de todos, es decir, es el que presenta un menor desvío estándar, lo cual se vincula a que, dados los supuestos del modelo, es el estimador con la menor varianza asintótica. En resumen, si, en la práctica, se trabajara con un modelo donde se supone que se cumplen los supuestos del modelo del inciso, entonces, para N muy pequeño se optaría por utilizar el estimador FE, mientras que, a partir de $N = 30$, se optaría por el estimador RE.

Ejercicio 4.

Basado en el Ejercicio 10.18 de Wooldridge (2010). Utilizar la base de datos “wagepan.dta” para responder las preguntas a continuación.

(a) Utilizando *lwage* como variable dependiente, estimar un modelo que contenga un intercepto y las variables dummy de año *d81* a *d87*. Estimar el modelo por POLS, RE, FE y FD. ¿Qué se puede concluir acerca de los coeficientes de las variables dummy?

POLS:

Source	SS	df	MS	Number of obs	=	4,360
Model	92.9668229	7	13.2809747	F(7, 4352)	=	50.54
Residual	1143.56282	4,352	.262767192	Prob > F	=	0.0000
				R-squared	=	0.0752
				Adj R-squared	=	0.0737
Total	1236.52964	4,359	.283672779	Root MSE	=	.51261

lwage	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
d81	.1193902	.0310529	3.84	0.000	.0585107	.1802697
d82	.1781901	.0310529	5.74	0.000	.1173106	.2390696
d83	.2257865	.0310529	7.27	0.000	.1649069	.2866666
d84	.2968181	.0310529	9.56	0.000	.2359386	.3576976
d85	.3459333	.0310529	11.14	0.000	.2850538	.4068128
d86	.4062418	.0310529	13.08	0.000	.3453623	.4671213
d87	.4730023	.0310529	15.23	0.000	.4121228	.5338818
_cons	1.393477	.0219577	63.46	0.000	1.350429	1.436525

RE:

Random-effects GLS regression	Number of obs	=	4,360
Group variable: nr	Number of groups	=	545
R-squared:	Obs per group:		
Within = 0.0000	min =		8
Between = 0.0000	avg =		8.0
Overall = 0.0752	max =		8
corr(u_i, X) = 0 (assumed)	Wald chi2(7)	=	738.94
	Prob > chi2	=	0.0000

lwage	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
d81	.1193902	.021487	5.56	0.000	.0772765	.1615039
d82	.1781901	.021487	8.29	0.000	.1360764	.2203038
d83	.2257865	.021487	10.51	0.000	.1836728	.2679001
d84	.2968181	.021487	13.81	0.000	.2547044	.3389318
d85	.3459333	.021487	16.10	0.000	.3038196	.388047
d86	.4062418	.021487	18.91	0.000	.3641281	.4483555
d87	.4730023	.021487	22.01	0.000	.4308886	.515116
_cons	1.393477	.0219577	63.46	0.000	1.350441	1.436513
sigma_u	.37007665					
sigma_e	.35469771					
rho	.52120938	(fraction of variance due to u_i)				

FE:

```
Fixed-effects (within) regression          Number of obs   =      4,360
Group variable: nr                        Number of groups =       545

R-squared:                                Obs per group:
    Within = 0.1625                        min =          8
    Between = .                            avg =         8.0
    Overall = 0.0752                       max =          8

corr(u_i, Xb) = 0.0000                    F(7,3808)       =     105.56
                                           Prob > F        =      0.0000
```

lwage	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
d81	.1193902	.021487	5.56	0.000	.0772631	.1615173
d82	.1781901	.021487	8.29	0.000	.136063	.2203172
d83	.2257865	.021487	10.51	0.000	.1836594	.2679135
d84	.2968181	.021487	13.81	0.000	.254691	.3389452
d85	.3459333	.021487	16.10	0.000	.3038063	.3880604
d86	.4062418	.021487	18.91	0.000	.3641147	.4483688
d87	.4730023	.021487	22.01	0.000	.4308753	.5151294
_cons	1.393477	.0151936	91.71	0.000	1.363689	1.423265
sigma_u	.39074676					
sigma_e	.35469771					
rho	.54824631	(fraction of variance due to u_i)				

F test that all u_i=0: F(544, 3808) = 9.71 Prob > F = 0.0000

FD:

Source	SS	df	MS	Number of obs	=	3,815
Model	19.3631642	7	2.76616631	F(7, 3808)	=	14.06
Residual	749.249837	3,808	.196756785	Prob > F	=	0.0000
Total	768.613001	3,815	.201471298	R-squared	=	0.0252
				Adj R-squared	=	0.0234
				Root MSE	=	.44357

D.lwage	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
d81						
D1.	.1193902	.0190006	6.28	0.000	.0821379	.1566425
d82						
D1.	.1781901	.0268709	6.63	0.000	.1255074	.2308728
d83						
D1.	.2257865	.03291	6.86	0.000	.1612636	.2903093
d84						
D1.	.2968181	.0380011	7.81	0.000	.2223136	.3713226
d85						
D1.	.3459333	.0424866	8.14	0.000	.2626347	.4292319
d86						
D1.	.4062418	.0465417	8.73	0.000	.3149927	.4974908
d87						
D1.	.4730023	.0502708	9.41	0.000	.3744421	.5715626

Tabla comparativa:

	(1) POLS	(2) RE	(3) FE	(4) FD
d81	0.119*** (0.0311)	0.119*** (0.0215)	0.119*** (0.0215)	0.119*** (0.0190)
d82	0.178*** (0.0311)	0.178*** (0.0215)	0.178*** (0.0215)	0.178*** (0.0269)
d83	0.226*** (0.0311)	0.226*** (0.0215)	0.226*** (0.0215)	0.226*** (0.0329)
d84	0.297*** (0.0311)	0.297*** (0.0215)	0.297*** (0.0215)	0.297*** (0.0380)
d85	0.346*** (0.0311)	0.346*** (0.0215)	0.346*** (0.0215)	0.346*** (0.0425)
d86	0.406*** (0.0311)	0.406*** (0.0215)	0.406*** (0.0215)	0.406*** (0.0465)
d87	0.473*** (0.0311)	0.473*** (0.0215)	0.473*** (0.0215)	0.473*** (0.0503)
_cons	1.393*** (0.0220)	1.393*** (0.0220)	1.393*** (0.0152)	
N	4360	4360	4360	3815
r2	0.0752		0.163	0.0252

Standard errors in parentheses
 * p<0.10, ** p<0.05, *** p<0.01

Por lo tanto, lo que se puede concluir acerca de los coeficientes de las variables *dummy* es que son numéricamente idénticos.

(b) Añadir las variables constantes en el tiempo *educ*, *black* e *hisp* al modelo, y estimar por *POLS* y *RE*. ¿Cómo se comparan los coeficientes? ¿Qué ocurre si se estima la ecuación por *FE*?

POLS:

Source	SS	df	MS	Number of obs	=	4,360
				F(10, 4349)	=	73.66
Model	179.091659	10	17.9091659	Prob > F	=	0.0000
Residual	1057.43798	4,349	.243145087	R-squared	=	0.1448
				Adj R-squared	=	0.1429
Total	1236.52964	4,359	.283672779	Root MSE	=	.4931

lwage	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
d81	.1193902	.029871	4.00	0.000	.0608279	.1779526
d82	.1781901	.029871	5.97	0.000	.1196277	.2367524
d83	.2257865	.029871	7.56	0.000	.1672241	.2843488
d84	.2968181	.029871	9.94	0.000	.2382557	.3553804
d85	.3459333	.029871	11.58	0.000	.287371	.4044957
d86	.4062418	.029871	13.60	0.000	.3476794	.4648041
d87	.4730023	.029871	15.83	0.000	.41444	.5315647
educ	.0770943	.0043766	17.62	0.000	.0685139	.0856747
black	-.1225637	.0237021	-5.17	0.000	-.1690319	-.0760955
hisp	.024623	.0213056	1.16	0.248	-.0171468	.0663928
_cons	.4966384	.0566686	8.76	0.000	.3855391	.6077377

RE:

Random-effects GLS regression	Number of obs	=	4,360
Group variable: nr	Number of groups	=	545
R-squared:	Obs per group:		
Within = 0.1625	min =		8
Between = 0.1296	avg =		8.0
Overall = 0.1448	max =		8
corr(u_i, X) = 0 (assumed)	Wald chi2(10)	=	819.51
	Prob > chi2	=	0.0000

lwage	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
d81	.1193902	.021487	5.56	0.000	.0772765	.1615039
d82	.1781901	.021487	8.29	0.000	.1360764	.2203038
d83	.2257865	.021487	10.51	0.000	.1836728	.2679001
d84	.2968181	.021487	13.81	0.000	.2547044	.3389318
d85	.3459333	.021487	16.10	0.000	.3038196	.388047
d86	.4062418	.021487	18.91	0.000	.3641281	.4483555
d87	.4730023	.021487	22.01	0.000	.4308886	.515116
educ	.0770943	.009177	8.40	0.000	.0591076	.0950809
black	-.1225637	.0496994	-2.47	0.014	-.2199728	-.0251546
hisp	.024623	.0446744	0.55	0.582	-.0629371	.1121831
_cons	.4966384	.1122718	4.42	0.000	.2765897	.7166871
sigma_u	.34337144					
sigma_e	.35469771					
rho	.48377912	(fraction of variance due to u_i)				

FE:

```
Fixed-effects (within) regression      Number of obs   =      4,360
Group variable: nr                    Number of groups =       545

R-squared:                            Obs per group:
    Within = 0.1625                      min =          8
    Between = .                          avg =         8.0
    Overall = 0.0752                     max =          8

corr(u_i, Xb) = 0.0000                F(7,3808)       =      105.56
                                      Prob > F         =       0.0000
```

```
-----+-----
      lwage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-----+-----
      d81 |   .1193902   .021487     5.56   0.000   .0772631   .1615173
      d82 |   .1781901   .021487     8.29   0.000   .136063   .2203172
      d83 |   .2257865   .021487    10.51   0.000   .1836594   .2679135
      d84 |   .2968181   .021487    13.81   0.000   .254691   .3389452
      d85 |   .3459333   .021487    16.10   0.000   .3038063   .3880604
      d86 |   .4062418   .021487    18.91   0.000   .3641147   .4483688
      d87 |   .4730023   .021487    22.01   0.000   .4308753   .5151294
      educ |           0   (omitted)
      black |           0   (omitted)
      hisp |           0   (omitted)
      _cons |  1.393477   .0151936   91.71   0.000   1.363689   1.423265
-----+-----
      sigma_u | .39074676
      sigma_e | .35469771
      rho    | .54824631   (fraction of variance due to u_i)
-----+-----
F test that all u_i=0: F(544, 3808) = 9.71                Prob > F = 0.0000
```

Tabla comparativa:

	(1) POLS	(2) RE	(3) FE
d81	0.119*** (0.0299)	0.119*** (0.0215)	0.119*** (0.0215)
d82	0.178*** (0.0299)	0.178*** (0.0215)	0.178*** (0.0215)
d83	0.226*** (0.0299)	0.226*** (0.0215)	0.226*** (0.0215)
d84	0.297*** (0.0299)	0.297*** (0.0215)	0.297*** (0.0215)
d85	0.346*** (0.0299)	0.346*** (0.0215)	0.346*** (0.0215)
d86	0.406*** (0.0299)	0.406*** (0.0215)	0.406*** (0.0215)
d87	0.473*** (0.0299)	0.473*** (0.0215)	0.473*** (0.0215)
educ	0.0771*** (0.00438)	0.0771*** (0.00918)	0 (.)
black	-0.123*** (0.0237)	-0.123** (0.0497)	0 (.)
hisp	0.0246 (0.0213)	0.0246 (0.0447)	0 (.)
_cons	0.497*** (0.0567)	0.497*** (0.112)	1.393*** (0.0152)
N	4360	4360	4360
r2	0.145		0.163

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Por un lado, se puede observar que las estimaciones de POLS y RE son numéricamente idénticas, ya que, si el modelo incluye sólo efectos temporales agregados y covariables específicas del individuo que no tienen variación temporal, entonces, los coeficientes de POLS son iguales a los de RE.

Por otra parte, lo que ocurre si se estima la ecuación por FE es que los coeficientes asociados a las variables constantes en el tiempo no se pueden estimar y, en consecuencia, cuando se incluyen variables constantes en el tiempo, la estimación de la constante en FE no es igual a la estimación de la constante en POLS/RE.

(c) ¿Son iguales los errores estándar de POLS y RE del inciso (b)? ¿Cuáles son, probablemente, más fiables?

Los errores estándar de POLS y RE del inciso (b) no son iguales. Los errores estándar de POLS suponen, además de homocedasticidad, que no hay correlación serial en el error compuesto, es decir, que no considera la posible presencia de heterogeneidad individual no observable. Los errores estándar de RE, al menos, en su estructura estándar, permiten la presencia de correlación serial (en particular, la cual es igual para todos los pares de períodos (t, s)). Esto puede ser demasiado restrictivo, pero es menos restrictivo que los habituales errores estándar de POLS.

(d) Obtener los errores estándar robustos para POLS. ¿Son preferibles estos o los errores estándar habituales de RE?

POLS (con errores estándar robustos):

Linear regression	Number of obs	=	4,360
	F(10, 544)	=	49.41
	Prob > F	=	0.0000
	R-squared	=	0.1448
	Root MSE	=	.4931

(Std. err. adjusted for 545 clusters in nr)

		Robust				
lwage	Coefficient	std. err.	t	P> t	[95% conf. interval]	
d81	.1193902	.0244086	4.89	0.000	.0714435	.1673369
d82	.1781901	.0241987	7.36	0.000	.1306558	.2257243
d83	.2257865	.0243796	9.26	0.000	.1778968	.2736761
d84	.2968181	.0271485	10.93	0.000	.2434894	.3501468
d85	.3459333	.0263181	13.14	0.000	.2942358	.3976309
d86	.4062418	.0273064	14.88	0.000	.3526029	.4598807
d87	.4730023	.025996	18.20	0.000	.4219374	.5240672
educ	.0770943	.0090198	8.55	0.000	.0593763	.0948122
black	-.1225637	.0532662	-2.30	0.022	-.2271964	-.017931
hisp	.024623	.0411235	0.60	0.550	-.0561573	.1054033
_cons	.4966384	.1097474	4.53	0.000	.2810579	.7122189

Estos errores estándar robustos son preferibles a los errores estándar habituales de RE, ya que estos errores estándar robustos permiten cualquier tipo de correlación serial y de heterocedasticidad de los disturbios que varían en el tiempo.

(e) Obtener los errores estándar robustos de RE. ¿Cómo se comparan con los errores estándar robustos de POLS y por qué?

RE (con errores estándar robustos):

```
Random-effects GLS regression              Number of obs   =       4,360
Group variable: nr                        Number of groups =        545

R-squared:                                Obs per group:
    Within = 0.1625                          min =           8
    Between = 0.1296                         avg =          8.0
    Overall = 0.1448                         max =           8

corr(u_i, X) = 0 (assumed)                  Wald chi2(10)    =       494.13
                                           Prob > chi2      =       0.0000
```

(Std. err. adjusted for 545 clusters in nr)

		Robust				
	lwage	Coefficient	std. err.	z	P> z	[95% conf. interval]
d81		.1193902	.0244086	4.89	0.000	.0715502 .1672302
d82		.1781901	.0241987	7.36	0.000	.1307616 .2256186
d83		.2257865	.0243796	9.26	0.000	.1780033 .2735696
d84		.2968181	.0271485	10.93	0.000	.2436081 .3500281
d85		.3459333	.0263181	13.14	0.000	.2943508 .3975159
d86		.4062418	.0273064	14.88	0.000	.3527222 .4597613
d87		.4730023	.025996	18.20	0.000	.422051 .5239536
educ		.0770943	.0090198	8.55	0.000	.0594157 .0947728
black		-.1225637	.0532662	-2.30	0.021	-.2269636 -.0181638
hisp		.024623	.0411235	0.60	0.549	-.0559775 .1052236
_cons		.4966384	.1097474	4.53	0.000	.2815375 .7117392
sigma_u		.34337144				
sigma_e		.35469771				
rho		.48377912	(fraction of variance due to u_i)			

Tabla comparativa:

	(1)	(2)
	POLS (robust)	RE (robust)
d81	0.119*** (0.0244)	0.119*** (0.0244)
d82	0.178*** (0.0242)	0.178*** (0.0242)
d83	0.226*** (0.0244)	0.226*** (0.0244)
d84	0.297*** (0.0271)	0.297*** (0.0271)
d85	0.346*** (0.0263)	0.346*** (0.0263)
d86	0.406*** (0.0273)	0.406*** (0.0273)
d87	0.473*** (0.0260)	0.473*** (0.0260)
educ	0.0771*** (0.00902)	0.0771*** (0.00902)
black	-0.123** (0.0533)	-0.123** (0.0533)
hisp	0.0246 (0.0411)	0.0246 (0.0411)
_cons	0.497*** (0.110)	0.497*** (0.110)
N	4360	4360
r2	0.145	

Standard errors in parentheses
 * p<0.10, ** p<0.05, *** p<0.01

Por lo tanto, se puede observar que estos errores estándar son numéricamente idénticos a los errores estándar robustos de POLS porque se tiene un solo estimador y, entonces, hay una sola varianza robusta.

- f) Obtenga la estimación de GMM de y_{it} utilizando todos los instrumentos posibles en niveles para el modelo en primeras diferencias y y_{it-1} como instrumento para el modelo en niveles.
- g) Repita las estimaciones de los incisos e) y f) incluyendo efectos fijos de tiempo.

2. En este ejercicio se ilustra el hecho de que los estimadores de Arellano-Bond y de Blundell-Bond pueden extenderse en forma directa a modelos que incluyan regresores estrictamente exógenos y regresores secuencialmente exógenos.

En su paper original, Arellano y Bond modelaron el empleo de las empresas (n_{it}) utilizando un modelo de ajuste parcial para reflejar los costos de contratación y despido, incluyendo dos rezagos de la variable empleo. Otras variables incluidas fueron el nivel salarial actual y el rezagado (w_{it}), el stock de capital actual, rezagado una y dos veces (k_{it}) y la producción agregada actual, rezagada una y dos veces en el sector de la empresa (y_{it}). Todas las variables se expresan en logaritmos. También se incluye un conjunto de variables dummy de tiempo.

- a) Estime el modelo por OLS. Compute los errores estándar robustos a heterocedasticidad y correlación serial.
- b) Estime el modelo por FE. Compute los errores estándar robustos a heterocedasticidad y correlación serial.
- c) Implemente el estimador de Anderson-Hsiao usando n_{it-2} como instrumento.
- d) Estime la ecuación de empleo usando el estimador de Arellano-Bond. Asuma que la única endogeneidad presente es en el rezago de la variable dependiente.
- e) Ahora, considere como hicieron Blundell y Bond (1998) que los salarios y el stock de capital no deben tomarse como estrictamente exógenos en este contexto (como se hizo en los modelos anteriores). Reestime el modelo usando el estimador de A-B y considerando a los salarios y el stock de capital como regresores secuencialmente exógenos.
- f) Adicionalmente, Blundell y Bond (1998) eliminan de su modelo los rezagos más largos (de dos periodos) del empleo y el capital, y prescinden del nivel de producto agregado sectorial. Considerando esta cuestión, compute el estimador de Blundell-Bond.

3. Cuando hay muchos instrumentos, surgen dos problemas principales:

Trabajo Práctico N° 1: Modelo de Regresión Lineal

Ejercicio 1.

8 WLOL]DU OD EDVH GH GDWRV SURYLVWD ³FRUQZHOO G\

(a) A partir de los datos de los siete años, y utilizando los logaritmos de todas las variables, estimar un modelo PLS que relacione la tasa de crimen con prbarr, prbconv, prbpris, avgsen y polpc y que incluya un conjunto de dummies de año.

PLS:

Source	SS	df	MS	Number of obs	=	630		
Model	117.644669	11	10.6949699	Prob > F	=	0.0000	F(11, 618) =	74.49
Residual	88.735673	618	.143585231	R	=		- squared	= 0.5700
Total	206.380342	629	.328108652	Root MSE	=	.37893	Adj R - squared	= 0.5624

lcrmte	Coefficient	Std. err.	t	P> t	[95% conf. interval]
lprbarr	-.7195033	.0367657	-19.57	0.000	-.7917042 - .6473024
lprbconv	-.5456589	.0263683	-20.69	0.000	-.5974413 - .4938765
lprbpris	.2475521	.0672268	3.68	0.000	.1155314 .3795728
lavgse	-.0867575	.0579205	-1.50	0.135	-.2005023 .0269872
lpolpc	.3659886	.0300252	12.19	0.000	.3070248 .4249525
d82	.0051371	.057931	0.09	0.929	-.1086284 .1189026
d83	-.043503	.0576243	-0.75	0.451	-.1566662 .0696601
d84	-.1087542	.057923	-1.88	0.061	-.222504 .0049957
d85	-.0780454	.0583244	-1.34	0.181	-.1925835 .0364928
d86	-.0420791	.0578218	-0.73	0.467	-.15563 .0714719
d87	-.0270426	.056899	-0.48	0.635	-.1387815 .0846963
_cons	-2.082293	.2516253	-8.28	0.000	-2.576438 - 1.588149

(b) Computar los errores estándar robustos a heteroscedasticidad arbitraria y a autocorrelación serial arbitraria.

Trabajo Práctico N° 3: **Modelos de Paneles Dinámicos.**

Ejercicio 1.

Considerar la base de datos “mod_abdata.dta” que fue utilizada por Arellano y Bond en su famoso paper de 1991. Se trata de un panel de 140 empresas británicas encuestadas, anualmente, entre 1976 y 1984. El panel original no es balanceado, pero la versión para este ejercicio se trata de un panel balanceado de empresas con observaciones para, exactamente, 6 años entre 1977 y 1982. La variable que identifica la empresa es *id* y la variable que identifica el tiempo es *year*. La variable *n* es el empleo de la empresa. Luego, considerar un modelo muy simplificado del siguiente tipo:

$$\ln n_{it} = \rho \ln n_{it-1} + \varepsilon_{it},$$

$$\varepsilon_{it} = c_i + v_{it},$$

$$E(c_i) = E(v_{it}) = E(c_i v_{it}) = 0,$$

donde n_{it} es el empleo de la empresa *i* en el año *t*.

(a) Estimar el modelo por OLS. ¿Qué sesgo se esperaría encontrar y por qué?

POLS:

Source	SS	df	MS	Number of obs	=	770
Model	1396.09073	1	1396.09073	F(1, 768)	=	76542.09
Residual	14.0079495	768	.018239518	Prob > F	=	0.0000
				R-squared	=	0.9901
				Adj R-squared	=	0.9901
Total	1410.09868	769	1.83367839	Root MSE	=	.13505

n	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
nL1	.9967362	.0036027	276.66	0.000	.9896639	1.003809
_cons	-.0379493	.0063799	-5.95	0.000	-.0504734	-.0254252

El sesgo que se esperaría encontrar es el sesgo de paneles dinámicos, el cual se desprende que $\ln n_{it-1}$ está correlacionado con los efectos fijos, c_i , que se encuentran en el término de error. En general, bajo muchos supuestos, OLS sobrestima el valor real del parámetro ρ .

(b) Estimar el modelo usando efectos fijos (FE). ¿Permite la transformación within eliminar el sesgo de paneles dinámicos?

FE:

```
Fixed-effects (within) regression      Number of obs   =       770
Group variable: id                    Number of groups =       138

R-squared:                             Obs per group:
    Within = 0.4926                      min =          5
    Between = 0.9979                     avg =         5.6
    Overall = 0.9901                     max =          6

corr(u_i, Xb) = 0.9382                  F(1,631)        =      612.49
                                         Prob > F         =      0.0000
```

```
-----+-----
            n | Coefficient  Std. err.      t    P>|t|    [95% conf. interval]
-----+-----
      nL1 |    .869605   .0351375    24.75   0.000   .8006043   .9386056
    _cons |    .1076112  .0405095     2.66   0.008   .0280614   .1871609
-----+-----
sigma_u |    .18358137
sigma_e |    .1315487
rho     |    .66073284   (fraction of variance due to u_i)
-----+-----
F test that all u_i=0: F(137, 631) = 1.30                Prob > F = 0.0194
```

La transformación *within* no permite eliminar el sesgo de paneles dinámicos, ya que, ahora, el problema se encuentra en que $\ln \widetilde{n_{it-1}} = \ln n_{it-1} - \overline{\ln n_{i-1}}$ está correlacionado con $\widetilde{v_{it}} = v_{it} - \bar{v}_i$, aun cuando v_{it} no tiene correlación serial. En particular, el término $\ln n_{it-1}$ correlaciona negativamente con $\frac{-1}{T-1} v_{it-1}$ que se encuentra dentro de \bar{v}_i , mientras que, simétricamente, $\frac{-1}{T-1} \ln n_{it}$ y v_{it} también se encuentran correlacionados negativamente. Adicionalmente, hay otros pares de términos que correlacionan, pero su impacto es de segundo orden. Por último, cabe mencionar que Nickell mostró que, si $\rho > 0$, este sesgo es siempre negativo.

(c) Considerar una transformación de diferencias finitas de primer orden del modelo. ¿Continúa siendo la variable dependiente rezagada potencialmente endógena?

Considerando una transformación de diferencias finitas de primer orden del modelo, la variable dependiente rezagada continúa siendo potencialmente endógena, ya que el término $\Delta \ln n_{it-1} = \ln n_{it-1} - \ln n_{it-2}$ está correlacionado con v_{it-1} en $\Delta v_{it} = v_{it} - v_{it-1}$.

(d) Implementar el estimador de Anderson-Hsiao a partir del comando `ivregress` en Stata.

IV (Anderson-Hsiao):

```
Instrumental variables 2SLS regression      Number of obs   =      632
                                           Wald chi2(1)    =      .
                                           Prob > chi2     =      .
                                           R-squared       =      .
                                           Root MSE       =      .25024
```

D.n	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
nL1						
D1.	2.015601	.4911953	4.10	0.000	1.052876	2.978327

```
Instrumented: D.nL1
Instruments: nL2
```

(e) Ahora, obtener la estimación GMM de ρ utilizando todos los instrumentos posibles en niveles para el modelo en primeras diferencias. Para ello, utilizar el comando `xtabond2`.

GMM One-Step (Arellano-Bond):

Dynamic panel-data estimation, one-step difference GMM

```
Group variable: id      Number of obs   =      632
Time variable : year   Number of groups =      138
Number of instruments = 10  Obs per group: min =      4
Wald chi2(0) = .          avg =      4.58
Prob > chi2 = .          max =      5
```

n	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
nL1	1.146045	.0865907	13.24	0.000	.9763309	1.31576

```
Instruments for first differences equation
GMM-type (missing=0, separate instruments for each period unless collapsed)
L(1/5).L.n
```

```
Arellano-Bond test for AR(1) in first differences: z = -4.21 Pr > z = 0.000
Arellano-Bond test for AR(2) in first differences: z = -2.35 Pr > z = 0.019
```

```
Sargan test of overid. restrictions: chi2(9) = 122.88 Prob > chi2 = 0.000
(Not robust, but not weakened by many instruments.)
```


GMM Two-Step (Arellano-Bond):

Dynamic panel-data estimation, two-step difference GMM

```
-----
Group variable: id                Number of obs   =      632
Time variable : year              Number of groups =      138
Number of instruments = 10         Obs per group: min =       4
Wald chi2(0)   =                  .               avg   =      4.58
Prob > chi2    =                  .               max   =       5
-----
```

```
-----
      n | Coefficient  Std. err.      z    P>|z|    [95% conf. interval]
-----+-----
      nL1 |   1.176208   .0771686    15.24   0.000    1.024961    1.327456
-----
```

Warning: Uncorrected two-step standard errors are unreliable.

Instruments for first differences equation

GMM-type (missing=0, separate instruments for each period unless collapsed)
L(1/5).L.n

```
-----
Arellano-Bond test for AR(1) in first differences: z =  -2.97   Pr > z =  0.003
Arellano-Bond test for AR(2) in first differences: z =  -1.71   Pr > z =  0.087
-----
```

Sargan test of overid. restrictions: chi2(9) = 122.88 Prob > chi2 = 0.000
(Not robust, but not weakened by many instruments.)

Hansen test of overid. restrictions: chi2(9) = 48.86 Prob > chi2 = 0.000
(Robust, but weakened by many instruments.)

(f) *Obtener la estimación de GMM de ρ utilizando todos los instrumentos posibles en niveles para el modelo en primeras diferencias e Δy_{it-1} como instrumento para el modelo en niveles.*

SGMM One-Step (Blundell-Bond):

Dynamic panel-data estimation, one-step system GMM

Group variable: id	Number of obs	=	770
Time variable : year	Number of groups	=	138
Number of instruments = 15	Obs per group: min	=	5
Wald chi2(1) = 75831.06	avg	=	5.58
Prob > chi2 = 0.000	max	=	6

	n	Coefficient	Std. err.	z	P> z	[95% conf. interval]
nL1		1.107192	.0200965	55.09	0.000	1.067803 1.14658
_cons		-.1644167	.0233703	-7.04	0.000	-.2102216 -.1186118

Instruments for first differences equation

GMM-type (missing=0, separate instruments for each period unless collapsed)
L(1/5).L.n

Instruments for levels equation

Standard
_cons
GMM-type (missing=0, separate instruments for each period unless collapsed)
D.L.n

Arellano-Bond test for AR(1) in first differences: z = -5.47 Pr > z = 0.000
Arellano-Bond test for AR(2) in first differences: z = -2.27 Pr > z = 0.023

Sargan test of overid. restrictions: chi2(13) = 168.40 Prob > chi2 = 0.000
(Not robust, but not weakened by many instruments.)

Difference-in-Sargan tests of exogeneity of instrument subsets:

GMM instruments for levels
Sargan test excluding group: chi2(9) = 127.56 Prob > chi2 = 0.000
Difference (null H = exogenous): chi2(4) = 40.84 Prob > chi2 = 0.000

SGMM Two-Step (Blundell-Bond):

Dynamic panel-data estimation, two-step system GMM

```
-----
Group variable: id                      Number of obs   =      770
Time variable : year                   Number of groups =      138
Number of instruments = 15              Obs per group: min =       5
Wald chi2(1)   = 23169.14                avg =      5.58
Prob > chi2    =      0.000                max =       6
-----
```

	n	Coefficient	Std. err.	z	P> z	[95% conf. interval]
nL1		1.149856	.0316392	36.34	0.000	1.087845 1.211868
_cons		-.1738829	.0370401	-4.69	0.000	-.2464801 -.1012857

Warning: Uncorrected two-step standard errors are unreliable.

Instruments for first differences equation

GMM-type (missing=0, separate instruments for each period unless collapsed)
L(1/5).L.n

Instruments for levels equation

Standard

_cons

GMM-type (missing=0, separate instruments for each period unless collapsed)
D.L.n

```
-----
Arellano-Bond test for AR(1) in first differences: z = -3.18 Pr > z = 0.001
Arellano-Bond test for AR(2) in first differences: z = -1.71 Pr > z = 0.087
-----
```

Sargan test of overid. restrictions: chi2(13) = 168.40 Prob > chi2 = 0.000
(Not robust, but not weakened by many instruments.)

Hansen test of overid. restrictions: chi2(13) = 56.19 Prob > chi2 = 0.000
(Robust, but weakened by many instruments.)

Difference-in-Hansen tests of exogeneity of instrument subsets:

GMM instruments for levels

Hansen test excluding group: chi2(9) = 49.15 Prob > chi2 = 0.000

Difference (null H = exogenous): chi2(4) = 7.04 Prob > chi2 = 0.134

(g) Repetir las estimaciones de los incisos (e) y (f) incluyendo efectos fijos de tiempo.

GMM One-Step (Arellano-Bond):

Dynamic panel-data estimation, one-step difference GMM

```
-----
Group variable: id                      Number of obs   =      632
Time variable : year                    Number of groups =      138
Number of instruments = 15              Obs per group: min =       4
Wald chi2(0)   =                      .                avg   =      4.58
Prob > chi2    =                      .                max   =       5
-----
```

	n	Coefficient	Std. err.	z	P> z	[95% conf. interval]
nL1		.2667934	.1114731	2.39	0.017	.0483102 .4852767
yr1977		.0072053	.0133804	0.54	0.590	-.0190198 .0334304
yr1978		-.000092	.0101049	-0.01	0.993	-.0198972 .0197132
yr1980		-.0390927	.010015	-3.90	0.000	-.0587218 -.0194636
yr1981		-.1440582	.0107307	-13.42	0.000	-.1650901 -.1230264
yr1982		-.214458	.0194763	-11.01	0.000	-.2526308 -.1762851

Instruments for first differences equation

Standard

D.(yr1977 yr1978 yr1979 yr1980 yr1981 yr1982)

GMM-type (missing=0, separate instruments for each period unless collapsed)

L(1/5).L.n

```
-----
Arellano-Bond test for AR(1) in first differences: z = 0.30 Pr > z = 0.764
Arellano-Bond test for AR(2) in first differences: z = -0.93 Pr > z = 0.351
-----
```

```
Sargan test of overid. restrictions: chi2(9) = 36.01 Prob > chi2 = 0.000
(Not robust, but not weakened by many instruments.)
```

GMM Two-Step (Arellano-Bond):

Dynamic panel-data estimation, two-step difference GMM

```
-----
Group variable: id                      Number of obs   =      632
Time variable : year                    Number of groups =      138
Number of instruments = 15              Obs per group: min =       4
Wald chi2(0)   =                      .                avg   =      4.58
Prob > chi2    =                      .                max   =       5
-----
```

	n	Coefficient	Std. err.	z	P> z	[95% conf. interval]
nL1		.4092502	.0923988	4.43	0.000	.2281518 .5903486
yr1977		.0090912	.0095385	0.95	0.341	-.009604 .0277864
yr1978		.0020047	.0051366	0.39	0.696	-.0080629 .0120723
yr1980		-.0313017	.0073163	-4.28	0.000	-.0456414 -.0169621
yr1981		-.1342723	.0163176	-8.23	0.000	-.1662543 -.1022903
yr1982		-.2001323	.0235447	-8.50	0.000	-.2462791 -.1539855

Warning: Uncorrected two-step standard errors are unreliable.

Instruments for first differences equation

Standard

D.(yr1977 yr1978 yr1979 yr1980 yr1981 yr1982)

GMM-type (missing=0, separate instruments for each period unless collapsed)

L(1/5).L.n

```
-----
Arellano-Bond test for AR(1) in first differences: z = -0.90 Pr > z = 0.370
Arellano-Bond test for AR(2) in first differences: z = -1.22 Pr > z = 0.222
-----
```

```
Sargan test of overid. restrictions: chi2(9) = 36.01 Prob > chi2 = 0.000
(Not robust, but not weakened by many instruments.)
```

```
Hansen test of overid. restrictions: chi2(9) = 11.00 Prob > chi2 = 0.276
(Robust, but weakened by many instruments.)
```

SGMM One-Step (Blundell-Bond):

Dynamic panel-data estimation, one-step system GMM

```
-----
Group variable: id                      Number of obs   =      770
Time variable : year                    Number of groups =      138
Number of instruments = 20              Obs per group: min =       5
Wald chi2(6)   =   84467.08              avg   =      5.58
Prob > chi2    =       0.000              max   =       6
-----
```

	n	Coefficient	Std. err.	z	P> z	[95% conf. interval]
nL1		1.070702	.0198077	54.05	0.000	1.031879 1.109524
yr1977		.0438478	.0151864	2.89	0.004	.0140831 .0736126
yr1978		.0518471	.0129701	4.00	0.000	.0264261 .0772681
yr1979		.0418566	.0129664	3.23	0.001	.0164428 .0672704
yr1981		-.0742759	.0129883	-5.72	0.000	-.0997325 -.0488193
yr1982		-.0521052	.0133172	-3.91	0.000	-.0782065 -.026004
_cons		-.1213363	.0250374	-4.85	0.000	-.1704086 -.072264

Instruments for first differences equation

GMM-type (missing=0, separate instruments for each period unless collapsed)
L(1/5).L.n

Instruments for levels equation

Standard
yr1977 yr1978 yr1979 yr1980 yr1981 yr1982
_cons

GMM-type (missing=0, separate instruments for each period unless collapsed)
D.L.n

```
-----
Arellano-Bond test for AR(1) in first differences: z =  -6.30  Pr > z =  0.000
Arellano-Bond test for AR(2) in first differences: z =  -2.36  Pr > z =  0.018
-----
```

Sargan test of overid. restrictions: chi2(13) = 79.17 Prob > chi2 = 0.000
(Not robust, but not weakened by many instruments.)

Difference-in-Sargan tests of exogeneity of instrument subsets:

GMM instruments for levels

Sargan test excluding group: chi2(9) = 21.45 Prob > chi2 = 0.011
Difference (null H = exogenous): chi2(4) = 57.72 Prob > chi2 = 0.000

iv(yr1977 yr1978 yr1979 yr1980 yr1981 yr1982, eq(level))

Sargan test excluding group: chi2(8) = 56.13 Prob > chi2 = 0.000
Difference (null H = exogenous): chi2(5) = 23.05 Prob > chi2 = 0.000

SGMM Two-Step (Blundell-Bond):

Dynamic panel-data estimation, two-step system GMM

```
-----
Group variable: id                Number of obs   =       770
Time variable : year              Number of groups =       138
Number of instruments = 20         Obs per group: min =        5
Wald chi2(6)  = 19554.89          avg          =    5.58
Prob > chi2   =      0.000        max          =        6
-----
```

	n	Coefficient	Std. err.	z	P> z	[95% conf. interval]
nL1		1.115279	.0252032	44.25	0.000	1.065881 1.164676
yr1977		.042046	.0187107	2.25	0.025	.0053737 .0787184
yr1978		.0475315	.0087333	5.44	0.000	.0304146 .0646484
yr1979		.0347642	.0075048	4.63	0.000	.0200551 .0494734
yr1981		-.0766901	.0109072	-7.03	0.000	-.0980679 -.0553123
yr1982		-.0535137	.0128686	-4.16	0.000	-.0787358 -.0282917
_cons		-.1778026	.0308186	-5.77	0.000	-.238206 -.1173993

Warning: Uncorrected two-step standard errors are unreliable.

Instruments for first differences equation

GMM-type (missing=0, separate instruments for each period unless collapsed)
L(1/5).L.n

Instruments for levels equation

Standard
yr1977 yr1978 yr1979 yr1980 yr1981 yr1982
_cons

GMM-type (missing=0, separate instruments for each period unless collapsed)
D.L.n

```
-----
Arellano-Bond test for AR(1) in first differences: z = -3.22 Pr > z = 0.001
Arellano-Bond test for AR(2) in first differences: z = -1.56 Pr > z = 0.118
-----
```

Sargan test of overid. restrictions: chi2(13) = 79.17 Prob > chi2 = 0.000
(Not robust, but not weakened by many instruments.)

Hansen test of overid. restrictions: chi2(13) = 28.91 Prob > chi2 = 0.007
(Robust, but weakened by many instruments.)

Difference-in-Hansen tests of exogeneity of instrument subsets:

GMM instruments for levels

Hansen test excluding group: chi2(9) = 6.67 Prob > chi2 = 0.671
Difference (null H = exogenous): chi2(4) = 22.23 Prob > chi2 = 0.000

iv(yr1977 yr1978 yr1979 yr1980 yr1981 yr1982, eq(level))

Hansen test excluding group: chi2(8) = 22.21 Prob > chi2 = 0.005
Difference (null H = exogenous): chi2(5) = 6.69 Prob > chi2 = 0.244

Ejercicio 2.

En este ejercicio, se ilustrará el hecho de que los estimadores de Arellano-Bond y de Blundell-Bond pueden extenderse, en forma directa, a modelos que incluyan regresores estrictamente exógenos y regresores secuencialmente exógenos.

En su paper original, Arellano y Bond modelaron el empleo de las empresas (n) utilizando un modelo de ajuste parcial para reflejar los costos de contratación y despido, incluyendo dos rezagos de la variable empleo. Otras variables incluidas fueron el nivel salarial actual y el rezagado (w), el stock de capital actual, rezagado una y dos veces (k), y la producción agregada actual, rezagada una y dos veces en el sector de la empresa (ys). Todas las variables se expresan en logaritmos. También se incluye un conjunto de variables dummy de tiempo.

(a) *Estimar el modelo por OLS. Computar los errores estándar robustos a heterocedasticidad y correlación serial.*

POLS:

```
Linear regression                                Number of obs   =          632
                                                F(14, 137)      =       15042.46
                                                Prob > F        =          0.0000
                                                R-squared       =          0.9948
                                                Root MSE      =          .09885
```

(Std. err. adjusted for 138 clusters in id)

	n	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
nL1		1.083681	.0479746	22.59	0.000	.9888145	1.178548
nL2		-.1204015	.0432502	-2.78	0.006	-.2059257	-.0348772
w		-.4314672	.1861579	-2.32	0.022	-.7995817	-.0633527
wL1		.3933175	.1806983	2.18	0.031	.035999	.7506359
k		.3214569	.0546692	5.88	0.000	.2133524	.4295614
kL1		-.2087172	.0674584	-3.09	0.002	-.3421117	-.0753228
kL2		-.0811552	.030786	-2.64	0.009	-.1420324	-.020278
ys		.5156912	.1862924	2.77	0.006	.1473108	.8840716
ysL1		-.7065917	.2745098	-2.57	0.011	-1.249416	-.1637674
ysL2		.2489473	.1450994	1.72	0.088	-.0379767	.5358714
yr1977		0	(omitted)				
yr1978		0	(omitted)				
yr1979		.0161153	.0087992	1.83	0.069	-.0012845	.0335151
yr1980		.0267825	.0153105	1.75	0.082	-.003493	.057058
yr1981		-.0111743	.0255106	-0.44	0.662	-.0616197	.0392712
yr1982		-.0017447	.0217911	-0.08	0.936	-.044835	.0413456
_cons		-.1238146	.2952534	-0.42	0.676	-.7076579	.4600287

(b) *Estimar el modelo por FE. Computar los errores estándar robustos a heterocedasticidad y correlación serial.*

FE:

```
Fixed-effects (within) regression          Number of obs   =          632
Group variable: id                        Number of groups =          138

R-squared:                                Obs per group:
    Within = 0.7708                        min =              4
    Between = 0.9706                       avg =             4.6
    Overall = 0.9674                       max =              5

corr(u_i, Xb) = 0.6273                     F(14,137)        =       128.03
                                           Prob > F          =        0.0000
```

(Std. err. adjusted for 138 clusters in id)

	n	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
nL1		.712259	.0546499	13.03	0.000	.6041925	.8203255
nL2		-.2216269	.0557228	-3.98	0.000	-.3318149	-.1114389
w		-.504334	.1902322	-2.65	0.009	-.8805051	-.1281629
wL1		.1750077	.1374862	1.27	0.205	-.0968619	.4468772
k		.3667223	.0660277	5.55	0.000	.2361571	.4972875
kL1		-.0648159	.052492	-1.23	0.219	-.1686152	.0389834
kL2		.0287852	.0412597	0.70	0.487	-.0528031	.1103735
ys		.5252203	.1803162	2.91	0.004	.1686574	.8817831
ysL1		-.5622163	.2111979	-2.66	0.009	-.9798456	-.1445871
ysL2		.1339081	.1695363	0.79	0.431	-.2013383	.4691544
yr1977		0	(omitted)				
yr1978		.0218097	.0273806	0.80	0.427	-.0323336	.0759531
yr1979		.0241949	.0257897	0.94	0.350	-.0268024	.0751922
yr1980		.0319888	.0201233	1.59	0.114	-.0078037	.0717813
yr1981		-.0005961	.0168409	-0.04	0.972	-.0338978	.0327056
yr1982		0	(omitted)				
_cons		1.248446	.8176095	1.53	0.129	-.3683202	2.865213
sigma_u		.29806935					
sigma_e		.09040774					
rho		.91575291	(fraction of variance due to u_i)				

(c) Implementar el estimador de Anderson-Hsiao usando n_{it-2} como instrumento.

IV (Anderson-Hsiao):

Instrumental variables 2SLS regression	Number of obs	=	494
	Wald chi2(13)	=	14.10
	Prob > chi2	=	0.3669
	R-squared	=	.
	Root MSE	=	.53695

D.n	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
nL1						
D1.	5.823993	14.07042	0.41	0.679	-21.75353	33.40152
nL2						
D1.	-.9846567	2.216574	-0.44	0.657	-5.329062	3.359749
w						
D1.	-.3323156	.4888956	-0.68	0.497	-1.290533	.6259022
wL1						
D1.	1.433551	3.563217	0.40	0.687	-5.550226	8.417328
k						
D1.	.0698672	.7653782	0.09	0.927	-1.430246	1.569981
kL1						
D1.	-1.708326	4.482217	-0.38	0.703	-10.49331	7.076659
kL2						
D1.	-.5052786	1.434586	-0.35	0.725	-3.317016	2.306459
ys						
D1.	1.344045	1.935232	0.69	0.487	-2.448939	5.13703
ysL1						
D1.	-2.722142	6.196766	-0.44	0.660	-14.86758	9.423296
ysL2						
D1.	-.1505585	.837905	-0.18	0.857	-1.792822	1.491705
yr1977						
D1.	0	(omitted)				
yr1978						
D1.	.0677291	.381059	0.18	0.859	-.6791328	.814591
yr1979						
D1.	.1150868	.3885855	0.30	0.767	-.6465268	.8767003
yr1980						
D1.	.1093503	.2766837	0.40	0.693	-.4329398	.6516404
yr1981						
D1.	0	(omitted)				
yr1982						
D1.	0	(omitted)				
_cons	.1190119	.350733	0.34	0.734	-.5684122	.806436

Instrumented: D.nL1

Instruments: D.nL2 D.w D.wL1 D.k D.kL1 D.kL2 D.ys D.ysL1 D.ysL2 D.yr1978
D.yr1979 D.yr1980 nL2

(d) Estimar la ecuación de empleo usando el estimador de Arellano-Bond. Asumir que la única endogeneidad presente es en el rezago de la variable dependiente.

GMM One-Step (Arellano-Bond):

Dynamic panel-data estimation, one-step difference GMM

```
-----
Group variable: id                      Number of obs   =      414
Time variable : year                   Number of groups =      138
Number of instruments = 20              Obs per group: min =       3
Wald chi2(0)   =                      .                avg   =      3.00
Prob > chi2    =                      .                max   =       3
-----
```

	n	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	

n							
L1.		.5325962	.4060438	1.31	0.190	-.2632351	1.328428
L2.		-.1678165	.108895	-1.54	0.123	-.3812468	.0456137
w							
--.		-.5435347	.1878885	-2.89	0.004	-.9117894	-.1752799
L1.		.0465042	.2149028	0.22	0.829	-.3746976	.467706
k							
--.		.3597198	.0798932	4.50	0.000	.203132	.5163076
L1.		-.0203542	.1486021	-0.14	0.891	-.311609	.2709006
L2.		.0531949	.0564035	0.94	0.346	-.0573539	.1637438
ys							
--.		.6720783	.1618321	4.15	0.000	.3548932	.9892634
L1.		-.3962257	.2155005	-1.84	0.066	-.818599	.0261475
L2.		-.061621	.1883471	-0.33	0.744	-.4307746	.3075325
yr1979		-.0019098	.0289454	-0.07	0.947	-.0586417	.0548221
yr1980		.0153498	.0194	0.79	0.429	-.0226734	.0533731
yr1982		-.0047222	.0199879	-0.24	0.813	-.0438977	.0344534

Instruments for first differences equation

Standard

D.(w L.w k L.k L2.k ys L.ys L2.ys yr1977 yr1978 yr1979 yr1980 yr1981 yr1982)

GMM-type (missing=0, separate instruments for each period unless collapsed)
L(1/5).L.n

```
-----
Arellano-Bond test for AR(1) in first differences: z = -0.93 Pr > z = 0.351
Arellano-Bond test for AR(2) in first differences: z = -1.57 Pr > z = 0.117
-----
```

Sargan test of overid. restrictions: chi2(7) = 18.99 Prob > chi2 = 0.008
(Not robust, but not weakened by many instruments.)

Hansen test of overid. restrictions: chi2(7) = 12.33 Prob > chi2 = 0.090
(Robust, but weakened by many instruments.)

GMM Two-Step (Arellano-Bond):

Dynamic panel-data estimation, two-step difference GMM

```

-----
Group variable: id                      Number of obs   =      414
Time variable : year                   Number of groups =      138
Number of instruments = 20              Obs per group:  min =       3
Wald chi2(0)   =                      .                avg  =     3.00
Prob > chi2    =                      .                max  =       3
-----

```

	n	Coefficient	Corrected std. err.	z	P> z	[95% conf. interval]	

	n						
	L1.	.6080291	.6739661	0.90	0.367	-.7129203	1.928978
	L2.	-.137942	.1688863	-0.82	0.414	-.468953	.1930691

	w						
	--.	-.4912924	.2335444	-2.10	0.035	-.949031	-.0335539
	L1.	.1263267	.2293468	0.55	0.582	-.3231847	.5758381

	k						
	--.	.2912765	.0889614	3.27	0.001	.1169154	.4656376
	L1.	-.0170639	.219071	-0.08	0.938	-.4464351	.4123073
	L2.	.0263268	.0736175	0.36	0.721	-.1179608	.1706145

	ys						
	--.	.5452599	.1895989	2.88	0.004	.1736528	.916867
	L1.	-.307353	.2411337	-1.27	0.202	-.7799662	.1652603
	L2.	-.1253987	.2036039	-0.62	0.538	-.524455	.2736577

	yr1979	.008789	.0360821	0.24	0.808	-.0619306	.0795087
	yr1980	.0244669	.0224526	1.09	0.276	-.0195394	.0684731
	yr1982	-.0189478	.0283143	-0.67	0.503	-.0744428	.0365472

Instruments for first differences equation

Standard

D. (w L.w k L.k L2.k ys L.ys L2.ys yr1977 yr1978 yr1979 yr1980 yr1981
yr1982)

GMM-type (missing=0, separate instruments for each period unless collapsed)
L(1/5).L.n

```

-----
Arellano-Bond test for AR(1) in first differences: z =  -0.64  Pr > z =  0.523
Arellano-Bond test for AR(2) in first differences: z =  -1.41  Pr > z =  0.159
-----

```

Sargan test of overid. restrictions: chi2(7) = 18.99 Prob > chi2 = 0.008
(Not robust, but not weakened by many instruments.)

Hansen test of overid. restrictions: chi2(7) = 12.33 Prob > chi2 = 0.090
(Robust, but weakened by many instruments.)

(e) Ahora, considerar, como hicieron Blundell y Bond (1998), que los salarios y el stock de capital no deben tomarse como estrictamente exógenos en este contexto (como se hizo en los modelos anteriores). Reestimar el modelo usando el estimador de A-B y considerando a los salarios y al stock de capital como regresores secuencialmente exógenos.

GMM One-Step (Arellano-Bond):

Dynamic panel-data estimation, one-step difference GMM

```
-----
Group variable: id                      Number of obs   =      414
Time variable : year                    Number of groups =      138
Number of instruments = 33              Obs per group: min =       3
Wald chi2(0)   =                       avg          =     3.00
Prob > chi2    =                       max          =       3
-----
```

	n	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	

	n						
	L1.	.8328315	.1219725	6.83	0.000	.5937698	1.071893
	L2.	-.1640736	.0770492	-2.13	0.033	-.3150873	-.0130599
	w						
	--.	-.5056556	.3148758	-1.61	0.108	-1.122801	.1114896
	L1.	.2750362	.3347117	0.82	0.411	-.3809867	.931059
	k						
	--.	.3384097	.1810356	1.87	0.062	-.0164135	.6932329
	L1.	-.2157422	.1197218	-1.80	0.072	-.4503925	.0189082
	L2.	-.0006972	.0481602	-0.01	0.988	-.0950895	.0936951
	ys						
	--.	.7245478	.2528721	2.87	0.004	.2289276	1.220168
	L1.	-.5540143	.4752081	-1.17	0.244	-1.485405	.3773765
	L2.	.0545983	.2887007	0.19	0.850	-.5112447	.6204413
	yr1979	-.0061002	.0303218	-0.20	0.841	-.0655299	.0533295
	yr1980	.0188299	.0214161	0.88	0.379	-.0231449	.0608046
	yr1982	-.0067336	.0217373	-0.31	0.757	-.049338	.0358708

Instruments for first differences equation

Standard

D.(ys L.ys L2.ys yr1977 yr1978 yr1979 yr1980 yr1981 yr1982)

GMM-type (missing=0, separate instruments for each period unless collapsed)

L(1/5).(L.n L.w L.k)

```
-----
Arellano-Bond test for AR(1) in first differences: z =  -3.83  Pr > z =  0.000
Arellano-Bond test for AR(2) in first differences: z =  -2.12  Pr > z =  0.034
-----
```

Sargan test of overid. restrictions: chi2(20) = 38.36 Prob > chi2 = 0.008
(Not robust, but not weakened by many instruments.)

Hansen test of overid. restrictions: chi2(20) = 22.98 Prob > chi2 = 0.290
(Robust, but weakened by many instruments.)

Difference-in-Hansen tests of exogeneity of instrument subsets:

iv(ys L.ys L2.ys yr1977 yr1978 yr1979 yr1980 yr1981 yr1982)

Hansen test excluding group: chi2(14) = 14.84 Prob > chi2 = 0.389

Difference (null H = exogenous): chi2(6) = 8.14 Prob > chi2 = 0.228

GMM Two-Step (Arellano-Bond):

Dynamic panel-data estimation, two-step difference GMM

```
-----
Group variable: id                      Number of obs   =    414
Time variable : year                    Number of groups =    138
Number of instruments = 33              Obs per group: min =     3
Wald chi2(0)   =                      avg         =    3.00
Prob > chi2    =                      max         =     3
-----
```

	n	Coefficient	Corrected std. err.	z	P> z	[95% conf. interval]	

	n						
	L1.	.9403457	.1894156	4.96	0.000	.5690979	1.311593
	L2.	-.1526986	.0928117	-1.65	0.100	-.3346062	.029209
	w						
	--.	-.4490707	.3993756	-1.12	0.261	-1.231833	.3336911
	L1.	.3385343	.3136958	1.08	0.281	-.2762982	.9533668
	k						
	--.	.1701686	.2126183	0.80	0.424	-.2465556	.5868928
	L1.	-.2258957	.1315513	-1.72	0.086	-.4837316	.0319401
	L2.	-.0419358	.0564683	-0.74	0.458	-.1526117	.0687401
	ys						
	--.	.689388	.2434718	2.83	0.005	.2121921	1.166584
	L1.	-.5276126	.4812107	-1.10	0.273	-1.470768	.4155429
	L2.	.1320996	.2419835	0.55	0.585	-.3421794	.6063786
	yr1979	.0193174	.0342941	0.56	0.573	-.0478977	.0865326
	yr1980	.0379698	.0228282	1.66	0.096	-.0067727	.0827124
	yr1982	-.0206929	.0242533	-0.85	0.394	-.0682285	.0268427

Instruments for first differences equation

Standard

D.(ys L.ys L2.ys yr1977 yr1978 yr1979 yr1980 yr1981 yr1982)

GMM-type (missing=0, separate instruments for each period unless collapsed)

L(1/5).(L.n L.w L.k)

```
-----
Arellano-Bond test for AR(1) in first differences: z = -2.47 Pr > z = 0.014
Arellano-Bond test for AR(2) in first differences: z = -1.74 Pr > z = 0.081
-----
```

Sargan test of overid. restrictions: chi2(20) = 38.36 Prob > chi2 = 0.008
(Not robust, but not weakened by many instruments.)

Hansen test of overid. restrictions: chi2(20) = 22.98 Prob > chi2 = 0.290
(Robust, but weakened by many instruments.)

Difference-in-Hansen tests of exogeneity of instrument subsets:

iv(ys L.ys L2.ys yr1977 yr1978 yr1979 yr1980 yr1981 yr1982)

Hansen test excluding group: chi2(14) = 14.84 Prob > chi2 = 0.389

Difference (null H = exogenous): chi2(6) = 8.14 Prob > chi2 = 0.228

(f) Adicionalmente, Blundell y Bond (1998) eliminan de su modelo los rezagos más largos (de dos períodos) del empleo y el capital y prescinden del nivel de producto agregado sectorial. Considerando esta cuestión, computar el estimador de Blundell-Bond.

SGMM One-Step (Blundell-Bond):

Dynamic panel-data estimation, one-step system GMM

```
-----
Group variable: id                      Number of obs   =      690
Time variable : year                    Number of groups =      138
Number of instruments = 47              Obs per group: min =       5
Wald chi2(9)   =   37427.45              avg   =      5.00
Prob > chi2    =       0.000              max   =       5
-----
```

	n	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
	n						
L1.		.9025721	.0412484	21.88	0.000	.8217267	.9834175
	w						
--.		-.5397871	.1984403	-2.72	0.007	-.928723	-.1508513
L1.		.3047706	.1800251	1.69	0.090	-.0480722	.6576134
	k						
--.		.4734141	.0900715	5.26	0.000	.2968772	.6499511
L1.		-.3942878	.086912	-4.54	0.000	-.5646322	-.2239433
	yr1978						
	yr1979						
	yr1980						
	yr1982						
	_cons						
		.0347845	.0217776	1.60	0.110	-.0078988	.0774678
		.044848	.0173516	2.58	0.010	.0108395	.0788565
		.0291248	.0171671	1.70	0.090	-.004522	.0627716
		.0333513	.0144621	2.31	0.021	.0050062	.0616964
		.8194633	.3304982	2.48	0.013	.1716988	1.467228

Instruments for first differences equation

GMM-type (missing=0, separate instruments for each period unless collapsed)
L(1/5).(L.n L.w L.k)

Instruments for levels equation

Standard

yr1977 yr1978 yr1979 yr1980 yr1981 yr1982

_cons

GMM-type (missing=0, separate instruments for each period unless collapsed)
D.(L.n L.w L.k)

Arellano-Bond test for AR(1) in first differences: z = -4.61 Pr > z = 0.000

Arellano-Bond test for AR(2) in first differences: z = -1.18 Pr > z = 0.238

Sargan test of overid. restrictions: chi2(37) = 86.05 Prob > chi2 = 0.000
(Not robust, but not weakened by many instruments.)

Hansen test of overid. restrictions: chi2(37) = 52.82 Prob > chi2 = 0.044
(Robust, but weakened by many instruments.)

Difference-in-Hansen tests of exogeneity of instrument subsets:

GMM instruments for levels

Hansen test excluding group: chi2(25) = 35.67 Prob > chi2 = 0.077

Difference (null H = exogenous): chi2(12) = 17.15 Prob > chi2 = 0.144

iv(yr1977 yr1978 yr1979 yr1980 yr1981 yr1982, eq(level))

Hansen test excluding group: chi2(33) = 44.68 Prob > chi2 = 0.084

Difference (null H = exogenous): chi2(4) = 8.14 Prob > chi2 = 0.087

SGMM Two-Step (Blundell-Bond):

Dynamic panel-data estimation, two-step system GMM

```
-----
Group variable: id                      Number of obs   =      690
Time variable : year                   Number of groups =      138
Number of instruments = 47              Obs per group: min =       5
Wald chi2(9)   =   30749.14              avg          =    5.00
Prob > chi2    =       0.000              max          =       5
-----
```

	n	Coefficient	Corrected std. err.	z	P> z	[95% conf. interval]	

	n						
	L1.	.892057	.0524149	17.02	0.000	.7893258	.9947883
	w						
	--.	-.4138471	.2405795	-1.72	0.085	-.8853744	.0576801
	L1.	.23221	.1663529	1.40	0.163	-.0938358	.5582558
	k						
	--.	.4563271	.1222931	3.73	0.000	.2166371	.6960171
	L1.	-.3582417	.1210988	-2.96	0.003	-.595591	-.1208924
yr1978		.0405244	.0265643	1.53	0.127	-.0115406	.0925895
yr1979		.0513874	.0198615	2.59	0.010	.0124596	.0903151
yr1980		.0296077	.0200915	1.47	0.141	-.0097708	.0689863
yr1982		.0290802	.014089	2.06	0.039	.0014662	.0566942
cons		.6712912	.3744471	1.79	0.073	-.0626117	1.405194

Instruments for first differences equation

GMM-type (missing=0, separate instruments for each period unless collapsed)
L(1/5).(L.n L.w L.k)

Instruments for levels equation

Standard

yr1977 yr1978 yr1979 yr1980 yr1981 yr1982

_cons

GMM-type (missing=0, separate instruments for each period unless collapsed)
D.(L.n L.w L.k)

Arellano-Bond test for AR(1) in first differences: z = -4.46 Pr > z = 0.000

Arellano-Bond test for AR(2) in first differences: z = -1.27 Pr > z = 0.203

Sargan test of overid. restrictions: chi2(37) = 86.05 Prob > chi2 = 0.000
(Not robust, but not weakened by many instruments.)

Hansen test of overid. restrictions: chi2(37) = 52.82 Prob > chi2 = 0.044
(Robust, but weakened by many instruments.)

Difference-in-Hansen tests of exogeneity of instrument subsets:

GMM instruments for levels

Hansen test excluding group: chi2(25) = 35.67 Prob > chi2 = 0.077

Difference (null H = exogenous): chi2(12) = 17.15 Prob > chi2 = 0.144

iv(yr1977 yr1978 yr1979 yr1980 yr1981 yr1982, eq(level))

Hansen test excluding group: chi2(33) = 44.68 Prob > chi2 = 0.084

Difference (null H = exogenous): chi2(4) = 8.14 Prob > chi2 = 0.087

Ejercicio 3.

Cuando hay muchos instrumentos, surgen dos problemas principales:

- Sobreestimación (overfitting) de la variable endógena.
- Mala estimación de la matriz de pesos W .

En estos casos, se proponen las siguientes soluciones:

- Probar diferentes especificaciones de IV recortando el número de rezagos en la matriz de instrumentos Z .
- Colapsar/combinar instrumentos. Se modifica la matriz de instrumentos para el individuo i :

$$Z_i = \begin{bmatrix} y_{i1} & 0 & 0 & \dots \\ y_{i1} & y_{i2} & 0 & \dots \\ y_{i3} & y_{i2} & y_{i1} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Si el modelo funciona, debería dar resultados similares con distintos instrumentos. Retomar el Ejercicio 2.e para ver una aplicación de esta cuestión. Estimar el modelo de empleo restringiendo el máximo rezago a 3 y 4 períodos. Por último, estimar el modelo colapsando instrumentos. Analizar si los resultados obtenidos son robustos.

Tabla comparativa:

	(1) GMM (OS) 2	(2) GMM (TS) 2	(3) GMM (OS) 3	(4) GMM (TS) 3	(5) GMM (OS) 4	(6) GMM (TS) 4	(7) GMM (OS) 5	(8) GMM (TS) 5
nL1	0.833*** (0.122)	0.940*** (0.189)	0.949*** (0.182)	0.900*** (0.226)	0.847*** (0.121)	0.930*** (0.174)	1.108*** (0.387)	0.877* (0.455)
N	414	414	414	414	414	414	414	414

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Ejercicio 4.

Considerar, nuevamente, el modelo del primer ejercicio. Obtener el estimador LSDVC propuesto por Kiviet (1995) a partir del comando `xtlsdvc`. Luego, estimar la matriz de varianzas y covarianzas de los coeficientes de Kiviet siguiendo el procedimiento explicado en clase.

LSDVC (Kiviet):

LSDVC dynamic regression
(SE not computed)

n	Coefficient	Std. err.	z	P> z	[95% conf. interval]
n					
L1.	.9890308

LSDVC (Kiviet):

LSDVC dynamic regression
(SE not computed)

n	Coefficient	Std. err.	z	P> z	[95% conf. interval]
n					
L1.	.7863675
yr1978	.1430045
yr1979	.1365801
yr1980	.0957011
yr1981	.0105706

UNIVERSIDAD TORCUATO DI TELLA
MAESTRÍA EN ECONOMÍA Y ECONOMETRÍA
2022

Datos de Panel
Problem Set 4
Modelos Lineales en Paneles Desbalanceados

1. Utilice la base de datos “*keane.dta*” la cual contiene el historial de empleo y escolaridad de una muestra de hombres para los años 1981 a 1987. Luego, considere la siguiente ecuación de salarios:

$$\ln(wage_{it}) = \beta_0 + \beta_1 exper_{it} + \beta_2 educ_{it} + c_i + u_{it}, \quad t = 1, 2, \dots, T \quad (1)$$

donde $\ln(wage_{it})$ es el logaritmo del salario por hora, $exper_{it}$ son los años de experiencia en el mercado laboral y $educ_{it}$ son los años de escolaridad. Responda las siguientes preguntas:

- a) Estime la ecuación usando efectos fijos. ¿Cuál es el sesgo potencial en este contexto?
 - b) Implemente el contraste de sesgo de selección propuesto por Wooldridge (1995) bajo el enfoque de Mundlak (1978).
 - c) Implemente el contraste de sesgo de selección propuesto por Wooldridge (1995) bajo el enfoque de Chamberlain (1980).
2. Considerando nuevamente la ecuación de salarios del ejercicio previo, realice los siguientes procedimientos:
 - a) Estime el modelo por Wooldridge (1995) bajo el enfoque de Chamberlain (1980).
 - b) Estime el modelo por Wooldridge (1995) bajo el enfoque de Mundlak (1978).
 - c) Comente sobre los errores estándar de las estimaciones anteriores.
 - d) Estime los errores estándar vía bootstrapping.
 - e) Estime los errores estándar analíticos (varianza asintótica).

Trabajo Práctico N° 4: **Modelos Lineales en Paneles Desbalanceados.**

Ejercicio 1.

Utilizar la base de datos “keane.dta”, la cual contiene el historial de empleo y escolaridad de una muestra de hombres para los años 1981 a 1987. Luego, considerar la siguiente ecuación de salarios:

$$\ln(\text{wage}_{it}) = \beta_0 + \beta_1 \text{exper}_{it} + \beta_2 \text{educ}_{it} + c_i + u_{it}, t = 1, 2, \dots, T, \quad (1)$$

donde $\ln(\text{wage}_{it})$ es el logaritmo del salario por hora, exper_{it} son los años de experiencia en el mercado laboral y educ_{it} son los años de escolaridad. Responder las siguientes preguntas:

(a) Estimar la ecuación usando efectos fijos. ¿Cuál es el sesgo potencial en este contexto?

FE:

```
Fixed-effects (within) regression      Number of obs   =      5,837
Group variable: id                   Number of groups =      1,531

R-squared:                            Obs per group:
    Within = 0.2373                      min =          1
    Between = 0.1857                     avg =         3.8
    Overall = 0.1767                     max =          7

corr(u_i, Xb) = -0.3197                F(2,4304)       =      669.73
                                      Prob > F          =      0.0000
```

```
-----+-----
      lwage | Coefficient  Std. err.      t    P>|t|    [95% conf. interval]
-----+-----
      exper |   .0964067   .0027721    34.78  0.000   .0909721   .1018414
       educ |   .1697764   .0243797     6.96  0.000   .1219795   .2175732
       _cons |   7.270616   .295975    24.56  0.000   6.690352   7.850879
-----+-----
      sigma_u |   .45083563
      sigma_e |   .31573611
       rho    |   .67092951   (fraction of variance due to u_i)
-----+-----
F test that all u_i=0: F(1530, 4304) = 5.46                Prob > F = 0.0000
```

En este contexto de efectos fijos, la selección muestral por truncamiento incidental es un problema si la selección está relacionada con los errores idiosincráticos de la ecuación de interés. Por lo tanto, si se piensa que, efectivamente, lo anterior se cumple y que se están observando los salarios “más altos” (los mejores salarios que se ofrecieron), entonces, el truncamiento tendría como consecuencia una sobreestimación de los retornos a la educación.

(b) Implementar el contraste de sesgo de selección propuesto por Wooldridge (1995) bajo el enfoque de Mundlak (1978).

Se rechaza la hipótesis nula, por lo que existe evidencia suficiente de que hay sesgo de selección.

(c) Implementar el contraste de sesgo de selección propuesto por Wooldridge (1995) bajo el enfoque de Chamberlain (1980).

Se rechaza la hipótesis nula, por lo que existe evidencia suficiente de que hay sesgo de selección.

Ejercicio 2.

Considerando, nuevamente, la ecuación de salarios del ejercicio previo, realizar los siguientes procedimientos:

(a) Estimar el modelo por Wooldridge (1995) bajo el enfoque de Chamberlain (1980).

POLS (Chamberlain):

Source	SS	df	MS	Number of obs	=	5,837
Model	357.875358	15	23.8583572	F(15, 5821)	=	111.85
Residual	1241.62941	5,821	.213301737	Prob > F	=	0.0000
Total	1599.50477	5,836	.274075526	R-squared	=	0.2237
				Adj R-squared	=	0.2217
				Root MSE	=	.46185

lwage	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
exper	.0566567	.0130844	4.33	0.000	.0310064	.0823071
educ	.101945	.0187839	5.43	0.000	.0651215	.1387684
exper81	.1106423	.0330149	3.35	0.001	.0459208	.1753637
educ81	.0986618	.0204867	4.82	0.000	.0585003	.1388233
exper82	-.0102639	.0284707	-0.36	0.718	-.0660771	.0455492
educ82	-.2174353	.0366788	-5.93	0.000	-.2893393	-.1455313
exper83	-.0704014	.0284495	-2.47	0.013	-.1261731	-.0146298
educ83	.0856036	.0363551	2.35	0.019	.0143341	.1568731
year#c.lambdac_chamberlain						
81	-.2683315	.0873923	-3.07	0.002	-.4396528	-.0970101
82	-.3211014	.0890068	-3.61	0.000	-.4955878	-.1466151
83	-.3500805	.0833289	-4.20	0.000	-.5134361	-.1867249
84	-.3390861	.0853207	-3.97	0.000	-.5063464	-.1718258
85	-.3585597	.0887959	-4.04	0.000	-.5326328	-.1844867
86	-.3195615	.09277	-3.44	0.001	-.5014252	-.1376977
87	-.3590845	.0978012	-3.67	0.000	-.5508112	-.1673579
_cons	8.87783	.1514153	58.63	0.000	8.580999	9.17466

(b) Estimar el modelo por Wooldridge (1995) bajo el enfoque de Mundlak (1978).

POLS (Mundlak):

Source	SS	df	MS	Number of obs	=	5,837
Model	354.236597	11	32.203327	F(11, 5825)	=	150.64
Residual	1245.26817	5,825	.213779943	Prob > F	=	0.0000
				R-squared	=	0.2215
				Adj R-squared	=	0.2200
Total	1599.50477	5,836	.274075526	Root MSE	=	.46236

lwage	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
exper	.0678465	.007649	8.87	0.000	.0528517	.0828413
educ	.0881093	.0296264	2.97	0.003	.0300305	.146188
mean_exper	-.0275146	.0156551	-1.76	0.079	-.0582044	.0031753
mean_educ	-.0167755	.0291599	-0.58	0.565	-.0739398	.0403887
year#c.lambda_mundlak						
81	-.2756566	.0778676	-3.54	0.000	-.4283059	-.1230073
82	-.3239156	.0765798	-4.23	0.000	-.4740404	-.1737908
83	-.3637717	.0737034	-4.94	0.000	-.5082579	-.2192856
84	-.3154769	.0727725	-4.34	0.000	-.4581381	-.1728158
85	-.3101603	.073666	-4.21	0.000	-.454573	-.1657477
86	-.2515261	.074834	-3.36	0.001	-.3982286	-.1048236
87	-.2525642	.0770484	-3.28	0.001	-.4036077	-.1015207
_cons	8.781247	.1284501	68.36	0.000	8.529437	9.033056

(c) *Comentar sobre los errores estándar de las estimaciones anteriores.*

La varianza asintótica de los estimadores de la segunda etapa necesita ser corregida por heterocedasticidad y correlación serial arbitraria, así como, además, por la estimación de la primera etapa.

(d) *Estimar los errores estándar vía bootstrapping.*

Stata.

(e) *Estimar los errores estándar analíticos (varianza asintótica).*

Stata.

Datos de Panel
Problem Set 5
Modelos de Variable Dependiente Discreta

1. El archivo *wagepan.dta* contiene los datos utilizados por Vella y Verbeek (1998). Estos datos contienen información para 545 hombres que trabajaron cada año de 1980 a 1987. Utilice los datos para analizar el impacto de la escolaridad ($educ_{it}$) en la probabilidad de estar afiliado a un sindicato ($union_{it}$). Las variables se describen en el conjunto de datos. Observe que la educación no cambia con el tiempo.

- a) Use Pooled OLS para estimar el modelo:

$$P(union_{it} = 1 | educ_{it}) = \beta_0 + \beta_1 educ_{it} \quad (1)$$

De acuerdo a los resultados obtenidos, ¿tiene impacto un año más de escolaridad sobre la probabilidad de estar afiliado a un sindicato?

- b) Use Pooled Probit para estimar el modelo:

$$P(union_{it} = 1 | educ_{it}) = \Phi(\beta_0 + \beta_1 educ_{it}) \quad (2)$$

Comente sobre el impacto de un año más de educación en la probabilidad de estar afiliado a un sindicato.

- c) Use Pooled Logit para estimar el modelo:

$$P(union_{it} = 1 | educ_{it}) = \Lambda(\beta_0 + \beta_1 educ_{it}) = \frac{e^{\beta_0 + \beta_1 educ_{it}}}{1 + e^{\beta_0 + \beta_1 educ_{it}}} \quad (3)$$

Comente sobre el impacto de un año más de educación en la probabilidad de estar afiliado a un sindicato. Compute el error estándar para esta estimación.

- d) Estime la siguiente extensión del modelo (2):

$$P(union_{it} = 1 | educ_{it}, c_i) = \Phi(\beta_0 + \beta_1 educ_{it} + c_i) \quad (4)$$

donde c_i son efectos no observables individuales. Use el modelo Probit de efectos aleatorios. ¿Cuál es el problema que surge al momento de estimar el efecto parcial de interés?

- e) Estime la siguiente extensión del modelo (3):

$$P(union_{it} = 1 | educ_{it}, c_i) = \Lambda(\beta_0 + \beta_1 educ_{it} + c_i) \quad (5)$$

donde c_i son efectos no observables individuales. Use el modelo Logit de efectos aleatorios. ¿Surge el mismo problema que en el inciso anterior al momento de estimar el efecto parcial de interés?

- f) Compute el denominado estimador Logit de efectos fijos para el modelo (5). ¿Se puede computar el efecto de un año más de educación sobre la probabilidad de estar afiliado a un sindicato? Explique.
- g) Considere la siguiente extensión del modelo (4):

$$P(\text{union}_{it} = 1 \mid \text{educ}_{it}, \text{black}_{it}, \text{married}_{it}, c_i) = \Phi(\beta_0 + \beta_1 \text{educ}_{it} + \beta_2 \text{black}_{it} + \beta_3 \text{married}_{it} + c_i) \quad (6)$$

donde black_{it} es una variable binaria que toma valor 1 si la persona es afroamericana y married_{it} es una variable binaria que toma valor 1 si la persona es casada. Asuma la siguiente versión de Mundlak (1978) del supuesto de Chamberlain (1980):

$$c_i \mid X_i \sim \text{Normal}(\psi + \xi \cdot \overline{\text{married}_i}, \sigma_a^2) \quad (7)$$

El modelo dado por (6) y (7) es un caso de lo que en la literatura se denomina modelo Probit de efectos aleatorios de Chamberlain. Al asumir solamente (6) y (7) se tiene que:

$$\begin{aligned} P(\text{union}_{it} = 1 \mid \text{educ}_{it}, \text{black}_{it}, \text{married}_{it}) &= \\ \Phi \left[(\beta_0 + \beta_1 \text{educ}_{it} + \beta_2 \text{black}_{it} + \beta_3 \text{married}_{it} + \psi + \xi \overline{\text{married}_i}) (1 + \sigma_a^2)^{-1/2} \right] &\equiv \\ \Phi [\beta_{0,a} + \beta_{1,a} \text{educ}_{it} + \beta_{2,a} \text{black}_{it} + \beta_{3,a} \text{married}_{it} + \xi_a \overline{\text{married}_i}] & \end{aligned}$$

Use Pooled Probit para estimar el modelo. Estime el efecto de la escolaridad sobre la probabilidad de estar sindicalizado para una persona afroamericana casada.

2. Considere los datos del ejercicio previo para analizar la probabilidad de estar afiliado a un sindicato según la situación de afiliación sindical del año previo.

- a) Use Pooled Probit para estimar el modelo:

$$P(\text{union}_{it} = 1 \mid \text{union}_{it-1}) = \Phi(\psi + \rho \cdot \text{union}_{it-1}) \quad (8)$$

A continuación, obtenga una estimación para

$$P(\text{union}_{it} = 1 \mid \text{union}_{it-1} = 1)$$

y para

$$P(\text{union}_{it} = 1 \mid \text{union}_{it-1} = 0).$$

. Comente sobre el efecto marginal de estar afiliado a un sindicato en el año $t - 1$ en la probabilidad de estar afiliado a un sindicato en el año t .

- b) Adicione al modelo el conjunto completo de variables binarias temporales. Vuelva a estimar las probabilidades solicitadas para cada año de la muestra.
- c) Estime un modelo de efectos no observables dinámico. Use el modelo Probit de efectos aleatorios incluyendo $\text{union}_{i,80}$ como una variable explicativa adicional. Luego, promedie las probabilidades estimadas a lo largo de $\text{union}_{i,80}$ para obtener la probabilidad promedio de estar afiliado a un sindicato en el año 1987 dado que estaba afiliado en el período anterior.

Trabajo Práctico N° 5: **Modelos de Variable Dependiente Discreta.**

Ejercicio 1.

El archivo “wagepan.dta” contiene los datos utilizados por Vella y Verbeek (1998). Estos datos contienen información para 545 hombres que trabajaron cada año de 1980 a 1987. Utilizar los datos para analizar el impacto de la escolaridad ($educ_{it}$) en la probabilidad de estar afiliado a un sindicato ($union_{it}$). Las variables se describen en el conjunto de datos. Observar que la educación no cambia con el tiempo.

(a) Usar Pooled OLS para estimar el modelo:

$$P(union_{it} = 1 | educ_{it}) = \beta_0 + \beta_1 educ_{it} \quad (1)$$

De acuerdo a los resultados obtenidos, ¿tiene impacto un año más de escolaridad sobre la probabilidad de estar afiliado a un sindicato?

POLS:

Source	SS	df	MS	Number of obs	=	4,360
Model	.030271965	1	.030271965	F(1, 4358)	=	0.16
Residual	804.314682	4,358	.184560505	Prob > F	=	0.6855
				R-squared	=	0.0000
				Adj R-squared	=	-0.0002
Total	804.344954	4,359	.18452511	Root MSE	=	.42961

union	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
educ	-.0015092	.0037264	-0.40	0.686	-.0088148	.0057964
_cons	.261795	.0443282	5.91	0.000	.1748892	.3487009

Por lo tanto, se puede observar que un año más de escolaridad no tiene impacto sobre la probabilidad de estar afiliado a un sindicato, ya que la variable *educ* no es estadísticamente significativa.

(b) Usar Pooled Probit para estimar el modelo:

$$P(union_{it} = 1 | educ_{it}) = \Phi(\beta_0 + \beta_1 educ_{it}) \quad (2)$$

Comentar sobre el impacto de un año más de educación en la probabilidad de estar afiliado a un sindicato.

Pooled Probit:

Probit regression
Log likelihood = -2422.7142

Number of obs = 4,360
LR chi2(1) = 0.17
Prob > chi2 = 0.6758
Pseudo R2 = 0.0000

	union	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
	educ	-.0051242	.0122467	-0.42	0.676	-.0291273	.0188789
	_cons	-.6331107	.1454783	-4.35	0.000	-.9182429	-.3479784

Efectos marginales (promedio) en Pooled Probit:

Average marginal effects
Model VCE: OIM

Number of obs = 4,360

Expression: Pr(union), predict()
dy/dx wrt: educ

		Delta-method				[95% conf. interval]	
		dy/dx	std. err.	z	P> z		
	educ	-.0016074	.0038414	-0.42	0.676	-.0091364	.0059216

Efectos marginales (condicionales) en Pooled Probit:

Conditional marginal effects
Model VCE: OIM

Number of obs = 4,360

Expression: Pr(union), predict()
dy/dx wrt: educ
At: educ = 11.76697 (mean)

		Delta-method				[95% conf. interval]	
		dy/dx	std. err.	z	P> z		
	educ	-.0016074	.0038416	-0.42	0.676	-.0091369	.005922

Por lo tanto, se puede observar que un año más de escolaridad no tiene impacto sobre la probabilidad de estar afiliado a un sindicato, ya que la variable *educ* no es estadísticamente significativa.

(c) Usar Pooled Logit para estimar el modelo:

$$P(\text{union}_{it} = 1 \mid \text{educ}_{it}) = \Lambda(\beta_0 + \beta_1 \text{educ}_{it}) = \frac{e^{\beta_0 + \beta_1 \text{educ}_{it}}}{1 + e^{\beta_0 + \beta_1 \text{educ}_{it}}} \quad (3)$$

Comentar sobre el impacto de un año más de educación en la probabilidad de estar afiliado a un sindicato. Compute el error estándar para esta estimación.

Pooled Logit:

Logistic regression
Log likelihood = -2422.7197
Number of obs = 4,360
LR chi2(1) = 0.16
Prob > chi2 = 0.6857
Pseudo R2 = 0.0000

	union	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
	educ	-.0081586	.0201411	-0.41	0.685	-.0476344	.0313172
	_cons	-1.034725	.2393549	-4.32	0.000	-1.503852	-.5655975

Efectos marginales (promedio) en Pooled Logit:

Average marginal effects
Model VCE: OIM
Number of obs = 4,360

Expression: Pr(union), predict()
dy/dx wrt: educ

		Delta-method				[95% conf. interval]	
		dy/dx	std. err.	z	P> z		
	educ	-.0015051	.0037154	-0.41	0.685	-.008787	.0057769

Efectos marginales (condicionales) en Pooled Logit:

Conditional marginal effects
Model VCE: OIM
Number of obs = 4,360

Expression: Pr(union), predict()
dy/dx wrt: educ
At: educ = 11.76697 (mean)

		Delta-method				[95% conf. interval]	
		dy/dx	std. err.	z	P> z		
	educ	-.0015051	.0037155	-0.41	0.685	-.0087873	.0057771

Por lo tanto, se puede observar que un año más de escolaridad no tiene impacto sobre la probabilidad de estar afiliado a un sindicato, ya que la variable *educ* no es estadísticamente significativa.

(d) Estimar la siguiente extensión del modelo (2):

$$P(\text{union}_{it} = 1 \mid \text{educ}_{it}, c_i) = \Phi(\beta_0 + \beta_1 \text{educ}_{it} + c_i), \quad (4)$$

donde c_i son efectos no observables individuales. Usar el modelo Probit de efectos aleatorios. ¿Cuál es el problema que surge al momento de estimar el efecto parcial de interés?

RE Probit:

```
Random-effects probit regression          Number of obs   = 4,360
Group variable: nr                      Number of groups = 545

Random effects u_i ~ Gaussian           Obs per group:
                                         min = 8
                                         avg = 8.0
                                         max = 8

Integration method: mvaghermite         Integration pts. = 12

Log likelihood = -1672.7504              Wald chi2(1)    = 1.04
                                         Prob > chi2     = 0.3080
```

	union	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
	educ	-.0502723	.0493148	-1.02	0.308	-.1469276	.0463829
	_cons	-.8013208	.5835448	-1.37	0.170	-1.945048	.342406
	/lnsig2u	1.099022	.1137102			.8761544	1.32189
	sigma_u	1.732406	.0984961			1.549725	1.936622
	rho	.7500768	.0213163			.7060247	.789496

LR test of rho=0: chibar2(01) = 1499.93 Prob >= chibar2 = 0.000

Efectos marginales (promedio) en RE Probit:

```
Average marginal effects          Number of obs = 4,360
Model VCE: OIM

Expression: Pr(union=1 | u_i=0), predict(pu0)
dy/dx wrt: educ
```

		Delta-method				[95% conf. interval]	
		dy/dx	std. err.	z	P> z		
	educ	-.0076285	.0075519	-1.01	0.312	-.022243	.007173

Efectos marginales (condicionales) en RE Probit:

```
Conditional marginal effects          Number of obs = 4,360
Model VCE: OIM

Expression: Pr(union=1 | u_i=0), predict(pu0)
dy/dx wrt: educ
At: educ = 11.76697 (mean)
```

		Delta-method				[95% conf. interval]	
		dy/dx	std. err.	z	P> z		
	educ	-.0076024	.0074781	-1.02	0.309	-.0222592	.0070543

El problema que surge al momento de estimar el efecto parcial de interés es que éste depende de c_i , el cual no es estimado, por lo que no es posible estimar la magnitud del efecto parcial, a menos que se imponga el valor de c_i . Este valor impuesto puede ser $c_i =$

0, lo cual tiene sentido ya que se está asumiendo que la distribución de c_i es $c_i | X_i \sim \text{Normal}(0, \sigma_c^2)$.

(e) Estimar la siguiente extensión del modelo (3):

$$P(\text{union}_{it} = 1 | \text{educ}_{it}, c_i) = \Lambda(\beta_0 + \beta_1 \text{educ}_{it} + c_i), \quad (5)$$

donde c_i son efectos no observables individuales. Usar el modelo Logit de efectos aleatorios. ¿Surge el mismo problema que en el inciso anterior al momento de estimar el efecto parcial de interés?

RE Logit:

```
Random-effects logistic regression          Number of obs   =   4,360
Group variable: nr                        Number of groups =    545

Random effects u_i ~ Gaussian              Obs per group:
                                         min =         8
                                         avg =        8.0
                                         max =         8

Integration method: mvaghermite           Integration pts. =    12

Log likelihood = -1670.7204                Wald chi2(1)     =     0.98
                                         Prob > chi2      =    0.3227
```

	union	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
	educ	-.0831898	.0841158	-0.99	0.323	-.2480538	.0816742
	_cons	-1.497913	.9974077	-1.50	0.133	-3.452797	.4569698
	/lnsig2u	2.248281	.1156237			2.021663	2.4749
	sigma_u	3.077571	.17792			2.747885	3.446812
	rho	.7421998	.0221233			.6965274	.7831387

LR test of rho=0: chibar2(01) = 1504.00 Prob >= chibar2 = 0.000

Efectos marginales (promedio) en RE Logit:

```
Average marginal effects                  Number of obs = 4,360
Model VCE: OIM

Expression: Pr(union=1 | u_i=0), predict(pu0)
dy/dx wrt:  educ
```

		Delta-method				[95% conf. interval]	
		dy/dx	std. err.	z	P> z		
	educ	-.0059833	.0061477	-0.97	0.330	-.0180325	.0060659

Efectos marginales (condicionales) en RE Logit:

Conditional marginal effects
Model VCE: OIM

Number of obs = 4,360

Expression: Pr(union=1 | u_i=0), predict(pu0)
dy/dx wrt: educ
At: educ = 11.76697 (mean)

		Delta-method				
	dy/dx	std. err.	z	P> z	[95% conf. interval]	
educ	-.0059476	.0060407	-0.98	0.325	-.0177872	.005892

Sí, surge el mismo problema que en el inciso anterior a la hora de estimar el efecto parcial de interés.

(f) Computar el denominado estimador Logit de efectos fijos para el modelo (5). ¿Se puede computar el efecto de un año más de educación sobre la probabilidad de estar afiliado a un sindicato? Explicar.

El efecto de un año más de educación sobre la probabilidad de estar afiliado a un sindicato no se puede computar, ya que, en este método no se identifican los coeficientes de los regresores que no varían en el tiempo, como es el caso de la variable *educ* en este ejercicio. Sin embargo, incluso si el coeficiente estuviera identificado (en el caso de que la variable varíe en el tiempo), sucedería lo de los incisos anteriores de que se debería imponer el valor de c_i sumado a que, en este caso, es difícil saber qué valor imponer, ya que no se está asumiendo que se conoce la distribución de c_i .

(g) Considerar la siguiente extensión del modelo (4):

$$P(\text{union}_{it} = 1 \mid \text{educ}_{it}, \text{black}_{it}, \text{married}_{it}, c_i) = \Phi(\beta_0 + \beta_1 \text{educ}_{it} + \beta_2 \text{black}_{it} + \beta_3 \text{married}_{it} + c_i), \quad (6)$$

donde black_{it} es una variable binaria que toma valor 1 si la persona es afroamericana y married_{it} es una variable binaria que toma valor 1 si la persona es casada. Asumir la siguiente versión de Mundlak (1978) del supuesto de Chamberlain (1980):

$$c_i \mid X_i \sim \text{Normal}(\psi + \xi \overline{\text{married}_i}, \sigma_a^2) \quad (7)$$

El modelo dado por (6) y (7) es un caso de lo que, en la literatura, se denomina modelo Probit de efectos aleatorios de Chamberlain. Al asumir sólo (6) y (7), se tiene que:

$$\begin{aligned} P(\text{union}_{it} = 1 \mid \text{educ}_{it}, \text{black}_{it}, \text{married}_{it}, c_i) &= \\ \Phi[(\beta_0 + \beta_1 \text{educ}_{it} + \beta_2 \text{black}_{it} + \beta_3 \text{married}_{it} + \psi + \xi \overline{\text{married}_i}) (1 + \sigma_a^2)^{-\frac{1}{2}}] &= \\ \Phi(\beta_{0,a} + \beta_{1,a} \text{educ}_{it} + \beta_{2,a} \text{black}_{it} + \beta_{3,a} \text{married}_{it} + \xi_a \overline{\text{married}_i}). \end{aligned}$$

Usar Pooled Probit para estimar el modelo. Estimar el efecto de la escolaridad sobre la probabilidad de estar sindicalizado para una persona afroamericana casada.

Pooled Probit:

```
Probit regression                                Number of obs = 4,360
                                                Wald chi2(4) = 15.55
                                                Prob > chi2 = 0.0037
Log pseudolikelihood = -2389.9119              Pseudo R2 = 0.0136
```

(Std. err. adjusted for 545 clusters in nr)

			Robust			
	union	Coefficient	std. err.	z	P> z	[95% conf. interval]
	educ	-.0031662	.0219937	-0.14	0.886	-.0462731 .0399408
	black	.4707704	.1297179	3.63	0.000	.216528 .7250127
	married	.0425222	.0532229	0.80	0.424	-.0617927 .1468371
mean_married		.2052462	.1316841	1.56	0.119	-.0528499 .4633423
_cons		-.8279919	.2671794	-3.10	0.002	-1.351654 -.3043298

Efectos marginales (condicionales) en Pooled Probit:

```
Conditional marginal effects                    Number of obs = 4,360
Model VCE: Robust
```

```
Expression: Pr(union), predict()
dy/dx wrt: educ
At: educ = 11.47692
      black = 1
      married = 1
      mean_married = .6923077
```

			Delta-method			
		dy/dx	std. err.	z	P> z	[95% conf. interval]
	educ	-.0012358	.0085842	-0.14	0.886	-.0180605 .0155888

Ejercicio 2.

Considerar los datos del ejercicio previo para analizar la probabilidad de estar afiliado a un sindicato según la situación de afiliación sindical del año previo.

(a) Usar Pooled Probit para estimar el modelo:

$$P(\text{union}_{it} = 1 \mid \text{union}_{it-1}) = \Phi(\psi + \rho \text{union}_{it-1}) \quad (8)$$

A continuación, obtener una estimación para

$$P(\text{union}_{it} = 1 \mid \text{union}_{it-1} = 1)$$

y para

$$P(\text{union}_{it} = 1 \mid \text{union}_{it-1} = 0).$$

Comentar sobre el efecto marginal de estar afiliado a un sindicato en el año t-1 en la probabilidad de estar afiliado a un sindicato en el año t.

Pooled Probit:

Probit regression	Number of obs =	3,815
	LR chi2(1) =	1416.84
	Prob > chi2 =	0.0000
Log likelihood = -1406.9979	Pseudo R2 =	0.3349

union	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
union						
L1.	1.953334	.055067	35.47	0.000	1.845404	2.061263
_cons	-1.348154	.0328877	-40.99	0.000	-1.412613	-1.283695

El efecto marginal de estar afiliado a un sindicato en el año t-1 en la probabilidad de estar afiliado a un sindicato en el año t es 0,639.

(b) Adicionar al modelo el conjunto completo de variables binarias temporales. Volver a estimar las probabilidades solicitadas para cada año de la muestra.

Pooled Probit (con variables binarias temporales):

Probit regression	Number of obs = 3,815
	LR chi2(7) = 1432.42
	Prob > chi2 = 0.0000
Log likelihood = -1399.2079	Pseudo R2 = 0.3386

	union	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
union							
l1.		1.967931	.0554571	35.49	0.000	1.859237	2.076625
d82		.0412915	.0960685	0.43	0.667	-.1469994	.2295824
d83		-.0513702	.0981709	-0.52	0.601	-.2437816	.1410412
d84		.0046824	.0980729	0.05	0.962	-.1875369	.1969016
d85		-.1584597	.1003127	-1.58	0.114	-.3550689	.0381495
d86		-.1384916	.1007772	-1.37	0.169	-.3360114	.0590282
d87		.1744346	.0957908	1.82	0.069	-.013312	.3621812
_cons		-1.339921	.0712489	-18.81	0.000	-1.479566	-1.200275

- El efecto marginal de estar afiliado a un sindicato en el año 1981 en la probabilidad de estar afiliado a un sindicato en el año 1982 es 0,651.
- El efecto marginal de estar afiliado a un sindicato en el año 1982 en la probabilidad de estar afiliado a un sindicato en el año 1983 es 0,636.
- El efecto marginal de estar afiliado a un sindicato en el año 1983 en la probabilidad de estar afiliado a un sindicato en el año 1984 es 0,646.
- El efecto marginal de estar afiliado a un sindicato en el año 1984 en la probabilidad de estar afiliado a un sindicato en el año 1985 es 0,614.
- El efecto marginal de estar afiliado a un sindicato en el año 1985 en la probabilidad de estar afiliado a un sindicato en el año 1986 es 0,618.
- El efecto marginal de estar afiliado a un sindicato en el año 1986 en la probabilidad de estar afiliado a un sindicato en el año 1987 es 0,667.

(c) *Estimar un modelo de efectos no observables dinámico. Usar el modelo Probit de efectos aleatorios incluyendo $union_{i,80}$ como una variable explicativa adicional. Luego, promediar las probabilidades estimadas a lo largo de $union_{i,80}$ para obtener la probabilidad promedio de estar afiliado a un sindicato en el año 1987 dado que estaba afiliado en el período anterior.*

RE Probit:

```
Random-effects probit regression          Number of obs   = 3,815
Group variable: nr                      Number of groups = 545

Random effects u_i ~ Gaussian           Obs per group:
                                         min = 7
                                         avg = 7.0
                                         max = 7

Integration method: mvaghermite         Integration pts. = 12

Log likelihood = -1293.5235              Wald chi2(8)     = 335.28
                                         Prob > chi2      = 0.0000
```

	union	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
union	L1.	.8886017	.0923747	9.62	0.000	.7075507	1.069653
	d82	.0398224	.1134088	0.35	0.725	-.1824547	.2620995
	d83	-.0636919	.1163085	-0.55	0.584	-.2916524	.1642686
	d84	-.0119455	.116519	-0.10	0.918	-.2403185	.2164276
	d85	-.224562	.1192335	-1.88	0.060	-.4582554	.0091315
	d86	-.2661777	.1199779	-2.22	0.027	-.5013301	-.0310253
	d87	.1277544	.1136075	1.12	0.261	-.0949122	.3504211
	union80	1.479094	.1659071	8.92	0.000	1.153922	1.804266
	_cons	-1.791908	.1172376	-15.28	0.000	-2.02169	-1.562127
/lnsig2u		.2114328	.1648497			-.1116665	.5345322
sigma_u		1.111507	.0916157			.9456968	1.306388
rho		.5526622	.0407552			.4721123	.6305396
LR test of rho=0: chibar2(01) = 160.37							
						Prob >= chibar2 = 0.000	

La probabilidad promedio de estar afiliado a un sindicato en el año 1987 dado que estaba afiliado en el período anterior es 0,397.