ECONOMETRÍA DE DATOS DE PANEL

Trabajo práctico N°4

Juan Guillermo Muñoz Delgado

Exercise 1

(a) Para este ejercicio se realizará un experimento de Montecarlo con 1000 simulaciones (S) con el objetivo de determinar el comportamiento de la corrección de Wooldridge en paneles desbalanceados teniendo en cuenta distintos modelos. Para esto se definen a continuación las siguientes métricas de desempeño:

$\ddot{\theta} \equiv$ Mediana del parámetro

$\bar{\theta} \equiv$ Media del parámetro

$\hat{\theta} \equiv$ Parámetro estimado

$\bar{\hat{\theta}} - \theta \equiv$ Sesgo medio (SM)

$\ddot{\hat{\theta}} - \theta \equiv$ Sesgo mediano (SMN)

$\sqrt{\sum_{s=1}^{S} \frac{(\hat{\theta}_s - \bar{\hat{\theta}})^2}{S}} \equiv$ Desvió estándar (DE)

$\sqrt{\sum_{s=1}^{S} \frac{(\hat{\theta}_s - \theta)^2}{S}} \equiv$ RMSE

$\frac{\sum_{s=1}^{S} |\hat{\theta}_s - \bar{\hat{\theta}}|}{S} \equiv$ Desvío medio absoluto (DMA)

Los resultados de las simulaciones se muestran a continuación:

Cuadro 1: N = 20, T = 2

	A			B			C		
	β	γ_1	γ_2	β	γ_1	γ_2	β	γ_1	γ_2
SM	0.024	0.304	-0.295	0.014	0.023	0.077	-0.021	-0.241	-0.178
SMN	0.011	-0.356	-0.339	0.012	-0.206	-0.183	-0.021	-0.369	-0.341
DE	1e-15	1e-15	9e-16	3e-16	8e-17	4e-16	6e-16	9e-16	9e-16
RMSE	7e-04	0.009	0.009	4e-04	7e-3	0.002	6e-04	0.007	0.005
DMA	3e-17	5e-17	3e-17	9e-18	2e-18	1e-17	2e-17	3e-17	2e-17

Cuadro 2: N = 40, T = 2

	A			B			C		
	β	γ_1	γ_2	β	γ_1	γ_2	β	γ_1	γ_2
SM	7e-3	-0.374	-0.373	0.004	-0.211	-0.216	-7e-3	-0.368	-0.365
SMN	0.010	-0.394	-0.389	-0.004	0.247	-0.250	0.010	-0.392	-0.383
DE	1e-15	1e-15	7e-16	1e-15	1e-15	5e-16	6e-16	1e-18	8e-16
RMSE	2e-5	0.011	0.011	1e-4	0.006	0.006	2e-05	0.011	0.011
DMA	5e-17	5e-17	2e-17	5e-17	1e-17	1e-17	1e-17	5e-20	2e-17

Cuadro 3: N = 100, T = 2

	A			B			C		
	β	γ_1	γ_2	β	γ_1	γ_2	β	γ_1	γ_2
SM	0.001	-0.405	-0.403	0.006	-0.264	-0.267	-0.007	-0.405	-0.407
SMN	-0.005	-0.409	-0.408	0.013	-0.277	-0.268	-0.002	-0.411	-0.406
DE	2e-15	2e-16	3e-16	3e-15	4e-17	5e-17	1e-15	1e-16	8e-17
RMSE	4e-5	0.012	0.012	1e-04	0.008	0.008	2e-4	0.012	0.012
DMA	6e-17	7e-18	1e-17	1e-16	1e-18	1e-18	4e-17	4e-18	2e-18

Las medidas de la estimación de β siguen el comportamiento usual de mejorar al aumentar el tamaño del corte transversal, sin embargo, esto no se repite para los demás estimadores. El modelo B tienen las mejores métricas cuando el N es pequeño pero esa ventaja deja de ser tan clara al aumentar el tamaño de la muestra. En diversas ocasiones el signo del sesgo medio y mediano son distintos, lo que sugiere que la muestra es muy variable y posee valores atípicos.

(b) Si bien los estimadores del inciso anterior serán consistentes y asintóticamente normales, los desvío estándar que se reportan en el ultimo paso de la corrección de Wooldridge no son confiables dado que no se está teniendo en cuenta la introducción de la estimación del cociente de Mills. En esta ocasión se usó el proceso de bootstrapping detallado en el enunciado con el fin de encontrar un intervalo de confianza al 95 % para cada uno de los estimadores. A continuación se muestran los resultados para cada uno de los modelos y configuraciones propuestas:

Cuadro 4: N = 20, T = 10

	A			B			C		
	β	γ_1	γ_2	β	γ_1	γ_2	β	γ_1	γ_2
Lím. Inferior	0.942	0.386	0.382	0.983	0.421	0.414	0.974	0.326	0.329
Lím. Superior	0.974	0.403	0.399	1.006	0.439	0.432	1.002	0.344	0.346

Cuadro 5: N = 40, T = 10

	A			B			C		
	β	γ_1	γ_2	β	γ_1	γ_2	β	γ_1	γ_2
Lím. Inferior	0.962	0.359	0.359	0.987	0.411	0.413	0.975	0.322	0.321
Lím. Superior	0.983	0.371	0.372	1.003	0.424	0.425	0.994	0.334	0.332

Cuadro 6: N = 100, T = 10

	A			B			C		
	β	γ_1	γ_2	β	γ_1	γ_2	β	γ_1	γ_2
Lím. Inferior	0.963	0.354	0.353	0.988	0.402	0.404	0.986	0.313	0.316
Lím. Superior	0.977	0.361	0.360	0.998	0.410	0.412	0.998	0.321	0.324

Como esperariamos, el intervalo se va achicando en la medida que aumenta N. No obstante, solo el modelo B para N=20 y N=40 contuvo el valor verdadero de β, pero no debería preocupar tanto teniendo en cuenta que siempre se está muy cerca del valor real. Por otro lado, los intervalos de γ_1 y γ_2 nunca contienen el valor verdadero del parámetro y están muy alejadas del mismo. Podríamos concluir que las estimaciones por bootstrapping son muy buenas para describir el comportamiento del regresor de la ecuación de interés pero se debe tener cuidado con los resultados para la variable de selección.

Exercise 2

(a) Se realiza la estimación por Pooled Probit dando como resultados:

Cuadro 7: Pooled Probit

	Estimador	Std. Error	t-valor	Pr(>|t|)
$Intercepto$	$-0{,}5400$	0,02795	$-19{,}32$	$< 2e-16$
$employ_{it-1}$	01,3871	0,04340	31,96	$< 2e-16$

En el modelo más restrictivo, para que los desvíos estándar y estadísticos t sean asintóticamente válidos deberíamos pedir que (1) a través del corte tranversal el efecto fijo no observado condicionado en las variables explicativas sean normales con media cero y varianza definida (2) Que para dato de corte transversal la variable dependiente sea independiente condicionada en las variables explicativas y el efecto fijo no observado (3) exogeneidad estricta. Si quisiéramos relajar los dos primeros supuestos que son ampliamente restrictivos podríamos usar un estimador sandwich trabajando los supuestos comunes para el Hessiano junto a los del estimador de Berndt–Hall–Hall–Hausman.

(b) Para los estimadores antes encontrados la probabilidad de tener empleo en un periodo dado que lo tuvo en el periodo anterior (1) es aproximadamente 0,8015, mientras que la misma probabilidad pero esta vez condicionada a haber estado desempleado el periodo anterior (2) es de aproximadamente 0,2945. Para calcular los desvíos estándar de las estimaciones antes planteadas podríamos seguir una vía teórica usando el teorema central del límite y estimar las varianzas por el método delta. También se podría seguir una vía más práctica y estimar usando algún proceso de bootstraping como en el ejercicio 1.

(c) Para el nuevo modelo planteado por el inciso los resultados son:

Cuadro 8: Pooled Probit + temp dummys

	Estimador	Std. Error	t-valor	Pr(>|t|)
$Intercepto$	−0,8827	0,0547	−16,109	$< 2e-16$
$employ_{it-1}$	1,3181	00,0450	29,272	$< 2e-16$
$y83$	0,3421	0,0740	4,621	$3{,}81e-06$
$y84$	0,4517	0,0747	6,046	$1{,}49e-09$
$y85$	0,5244	00,0758	6,911	$4{,}82e-12$
$y86$	0,3907	0,0767	5,090	$3{,}57e-07$
$y87$	0,5286	0,0767	6,889	$5{,}61e-126$

Usando esta estructura la probabilidad (1) pasa a ser aproximadamente 0,9155, mientras que la (2) se mueve a 0,4954. Las probabilidades difieren ampliamente de las encontradas anteriormente, dada la significatividad de las variables temporales podemos asegurar que hay una mejora sustancial en la especificidad del modelo al estimarlo por dos vías.

(d) Los resultados para el modelo dinámico se muestran a continuación:

Cuadro 9: RE Probit + temp dummys

	Estimador	Std. Error	t-valor	Pr(>|t|)
$Intercepto$	−1,00344	0,0652	−15,376	$< 2e-16$
$employ_{it-1}$	0,8914	0,0672	13,26	$< 2e-16$
$employ_{81}$	0,5718	0,0881	6,490	$8{,}60e-11$
$y83$	0,434	0,0794	5,464	$4{,}67e-08$
$y84$	0,6507	0,0832	7,815	$5{,}49e-15$
$y85$	0,7982	0,0878	9,088	$< 2e-16$
$y86$	0,6896	0,0894	7,708	$1{,}27e-14$
$y87$	0,8401	0,0904	9,287	$5{,}61e-126$
$sigma$	0,5585	0,0549	10,173	$< 2e-16$

(e) Viendo los resultados de sigma, que corresponde a las medidas para la heterogeneidad no observada, podemos asegurar que existe evidencia para la dependencia condicional del empleo actual con respecto al empleo en el periodo anterior.

(f) Transformando las variables usando las estimaciones de sigma para que funcionen en una normal estándar y promediando como lo estipula el ejercicio, se llega a que la probabilidad (1) es de 0.7676 y la probabilidad dos es de 0.4871. Comparándolas con las de (c) vemos que (2) prácticamente no cambió, pero (1) disminuyó ampliamente.