**Maestría en Econometría - UTDT**

**Examen Final - Econometría de Datos de Panel**

**PRIMERA PARTE: Propiedades de Muestra Finita en Paneles Desbalanceados.**

*Considerar el siguiente modelo:*

*= + + , j= 1, 2, … , N; t= 1, 2, … , T, (1)*

*= + + + ,*

*= 1 si 0, = 0, en cualquier otro caso,*

*=*

*Las dos variables de la ecuación de selección (w y z) son independientes y normalmente distribuidas con media cero y varianza uno. La única variable de la ecuación de interés es = , por lo que se tiene una variable que está excluída de la ecuación de interés. El error idiosincrático de la ecuación de selección, , sigue una distribución normal con media cero y varianza uno. El error idiosincrático de la ecuación de interés (1) se define como: = 0,6 + 0,8, con siendo una variable independiente y normal estándar. Con esta especificación, la correlación entre los errores idiosincráticos de las ecuaciones de interés y de selección es 0,6. Además, definir tres variables aleatorias: , y . Estas variables son específicas de corte transversal. Son todas independientes y normales estándar. Con esta especificación general, definir tres modelos:*

*Modelo A: Este modelo asume que los dos términos de efectos no observables, y , están correlacionados. Más específicamente:*

*= + (2)*

*= + (3)*

*Modelo B: Este modelo asume que las variables observables y no observables están correlacionadas, pero que no hay correlación entre los efectos no observables entre ecuaciones.*

*= + (4)*

*= + (5)*

*Modelo C: Éste es el modelo más general porque permite ambos tipos de correlación entre los dos componentes no observables y entre estos componentes y las variables observables.*

*= + + (6)*

*= + + (7)*

*Todos los parámetros son iguales a uno (= = = 1). N= 20, 40, 100 y T= 2, 10.*

**Ejercicio 1.**

*Caso 1: Para cada modelo y combinación de N= 20, 40, 100 y T= 2, realizar un experimento de Monte Carlo de S= 1000 simulaciones y reportar el sesgo medio, el sesgo mediano, el error estándar, el RMSE y la desviación media absoluta de la estimación de Wooldridge para los tres parámetros del modelo (, y ). Explicar los resultados.*

En las tablas 1, 2 y 3, para cada modelo (A, B, C) y tamaño de muestra 20, 40 y 100, respectivamente, se presentan el sesgo medio, el sesgo mediano, el desvío estándar, el RMSE y la desviación media absoluta de la estimación de Wooldridge para los tres parámetros del modelo (, y ), considerando T= 2 y 1000 replicaciones.

Por un lado, se puede observar que, a medida que el tamaño de muestra aumenta, para cada modelo (A, B, C) y para los tres parámetros del modelo (, y ), mientras que el sesgo medio y el sesgo mediano aumentan, el desvío estándar, el RMSE y la AMD disminuyen. Por otro lado, se observa que, al comparar entre modelos, la estimación del modelo A se ajusta mejor al valor poblacional del parámetro (ecuación de interés), mientras que la estimación del modelo B se ajusta mejor a los valores poblacionales de los parámetros y (ecuación de selección).

Por último, cabe mencionar que, cuando se está en presencia de un modelo como el C, en donde existen ambos tipos de correlación (entre los dos componentes no observables y entre estos componentes y las variables observables), en comparación a cuando existe sólo uno de los dos tipos de correlación pero no los dos, este método de estimación (Wooldridge bajo el enfoque de Mundlak) no se ajusta bien al valor poblacional del parámetro de la ecuación de interés (RMSE y AMD resultan considerablemente mayores).

**Tabla 1.** *Simulaciones mediante estimación de Wooldridge bajo el enfoque de Mundlak (con N= 20 y con 1000 replicaciones).*

|  |  |  |  |
| --- | --- | --- | --- |
| **Resultado** | **Modelo A** | **Modelo B** | **Modelo C** |
| Sesgo medio | 0,742 | 1,613 | 1,432 |
| Sesgo mediano | 0,477 | 1,222 | 1,035 |
| Desvío estándar | 3,182 | 3,358 | 5,683 |
| RMSE | 3,266 | 3,724 | 5,858 |
| AMD | 1,937 | 2,400 | 2,850 |
| Sesgo medio | 0,124 | 0,038 | 0,098 |
| Sesgo mediano | 0,002 | -0,032 | 0,004 |
| Desvío estándar | 0,886 | 0,534 | 0,731 |
| RMSE | 0,894 | 0,535 | 0,738 |
| AMD | 0,631 | 0,407 | 0,510 |
| Sesgo medio | 0,115 | 0,053 | 0,076 |
| Sesgo mediano | -0,047 | -0,027 | -0,055 |
| Desvío estándar | 1,003 | 0,564 | 0,741 |
| RMSE | 1,009 | 0,566 | 0,745 |
| AMD | 0,640 | 0,420 | 0,512 |

Fuente: Elaboración propia.

**Tabla 2.** *Simulaciones mediante estimación de Wooldridge bajo el enfoque de Mundlak (con N= 40 y con 1000 replicaciones).*

|  |  |  |  |
| --- | --- | --- | --- |
| **Resultado** | **Modelo A** | **Modelo B** | **Modelo C** |
| Sesgo medio | 0,946 | 1,854 | 1,797 |
| Sesgo mediano | 0,782 | 1,689 | 1,650 |
| Desvío estándar | 1,742 | 2,183 | 2,610 |
| RMSE | 1,982 | 2,863 | 3,168 |
| AMD | 1,413 | 2,164 | 2,295 |
| Sesgo medio | 0,239 | 0,177 | 0,217 |
| Sesgo mediano | 0,15 | 0,132 | 0,149 |
| Desvío estándar | 0,643 | 0,394 | 0,487 |
| RMSE | 0,685 | 0,431 | 0,533 |
| AMD | 0,496 | 0,317 | 0,39 |
| Sesgo medio | 0,261 | 0,184 | 0,232 |
| Sesgo mediano | 0,164 | 0,136 | 0,165 |
| Desvío estándar | 0,670 | 0,400 | 0,518 |
| RMSE | 0,719 | 0,440 | 0,568 |
| AMD | 0,515 | 0,327 | 0,409 |

Fuente: Elaboración propia.

**Tabla 3.** *Simulaciones mediante estimación de Wooldridge bajo el enfoque de Mundlak (con N= 100 y con 1000 replicaciones).*

|  |  |  |  |
| --- | --- | --- | --- |
| **Resultado** | **Modelo A** | **Modelo B** | **Modelo C** |
| Sesgo medio | 0,999 | 2,079 | 2,035 |
| Sesgo mediano | 0,932 | 2,013 | 1,958 |
| Desvío estándar | 0,971 | 1,077 | 1,361 |
| RMSE | 1,392 | 2,341 | 2,448 |
| AMD | 1,111 | 2,099 | 2,089 |
| Sesgo medio | 0,332 | 0,221 | 0,274 |
| Sesgo mediano | 0,297 | 0,200 | 0,242 |
| Desvío estándar | 0,409 | 0,245 | 0,320 |
| RMSE | 0,527 | 0,330 | 0,422 |
| AMD | 0,406 | 0,259 | 0,324 |
| Sesgo medio | 0,347 | 0,224 | 0,283 |
| Sesgo mediano | 0,296 | 0,216 | 0,261 |
| Desvío estándar | 0,422 | 0,258 | 0,328 |
| RMSE | 0,547 | 0,342 | 0,433 |
| AMD | 0,421 | 0,272 | 0,336 |

Fuente: Elaboración propia.

**Ejercicio 2.**

*Caso 2: Para cada modelo y combinación de N= 20, 40, 100 y T= 10, considerar lo siguiente:*

**(a)** *Tomar la primera generación de los datos (S= 1) como si fuera una muestra obtenida de la realidad.*

**(b)** *Estimar el modelo por Wooldridge y guardar los coeficientes estimados (llamarlos , y ), como así también los residuos de la ecuación de selección, .*

**(c)** *Realizar un procedimiento de bootstrapping para construir intervalos de 95% de confiabilidad para los tres estimadores de la siguiente manera: (i) Utilizar un procedimiento de muestreo aleatorio simple con reemplazo para obtener una nueva muestra de errores de la ecuación de selección. (Nota: Observar que se debe generar para cada j una nueva muestra de residuos de dimensión T). Con estos nuevos residuos, con y y con los valores originales de w, z (es decir, con los valores generados para S= 1) y , construir una nueva variable dependiente del modelo de selección. (ii) Usando esta nueva variable dependiente del modelo de selección, los valores de y y , construir una nueva variable dependiente de la ecuación de interés. (iii) Con las nuevas variables dependientes de ambas ecuaciones construidas, estimar por Wooldridge los tres parámetros del modelo. (iv) Volver al paso (i) y repetir el procedimiento. Repetir (iv) para construir B= 1000 muestras de bootstrapping y, en cada caso, guardar los coeficientes estimados por Wooldridge de , y . Reportar, para B= 1000, intervalos de bootstrapping de 95% de confiabilidad. Comparar estos resultados con intervalos de 95% de confiabilidad construidos a partir de la matriz de varianzas y covarianzas asintótica para el modelo estimado con los datos de S= 1. ¿Qué conclusiones se puede sacar?*

En las tablas 4 y 5, para cada combinación de modelo (A, B, C) y tamaño de muestra (20, 40, 100), se presentan intervalos de 95% de confiabilidad construidos a partir de *bootstrapping* (para B= 1000 muestras) y a partir de la matriz de varianzas y covarianzas, respectivamente, para los tres parámetros del modelo (, y ), considerando T= 10.

**Tabla 4.** *Intervalos de 95% de confiabilidad construidos a partir de bootstrapping (con B= 1000).*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tamaño muestra** | **Parámetro** | **Modelo A** | **Modelo B** | **Modelo C** |
| N= 20 |  | [1,654; 5,301] | [0,218; 2,058] | [0,816; 1,377] |
|  | [1,064; 2,705] | [0,877; 1,604] | [2,242; 3,762] |
|  | [1,035; 2,396] | [0,941; 1,605] | [1,865; 2,956] |
| N= 40 |  | [6,032; 7,982] | [3,346; 5,001] | [2,522; 6,037] |
|  | [1,610; 2,411] | [1,576; 2,299] | [1,203; 1,969] |
|  | [1,968; 2,806] | [2,014; 2,729] | [1,678; 2,549] |
| N= 100 |  | [5,650; 6,336] | [5,335; 6,007] | [6,290; 7,070] |
|  | [1,657; 2,042] | [1,701; 2,093] | [1,830; 2,260] |
|  | [1,699; 2,079] | [1,609; 1,982] | [1,643; 2,030] |

Fuente: Elaboración propia.

**Tabla 5.** *Intervalos de 95% de confiabilidad construidos a partir de la matriz de varianzas y covarianzas asintótica.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tamaño muestra** | **Parámetro** | **Modelo A** | **Modelo B** | **Modelo C** |
| N= 20 |  | [-1,381; 4,474] | [-1,145; 2,928] | [-1,965; 3,385] |
|  | [0,411; 1,615] | [0,530; 1,544] | [0,803; 2,764] |
|  | [0,373; 1,849] | [0,607; 1,617] | [0,826; 2,277] |
| N= 40 |  | [0,866; 4,322] | [0,052; 3,579] | [0,143; 4,056] |
|  | [0,711; 1,872] | [0,867; 1,467] | [0,619; 1,642] |
|  | [0,771; 2,108] | [0,993; 1,805] | [0,698; 1,986] |
| N= 100 |  | [1,561; 3,182] | [1,599; 2,962] | [1,563; 3,549] |
|  | [0,953; 1,611] | [0,994; 1,696] | [1,046; 1,808] |
|  | [1,005; 1,732] | [0,976; 1,640] | [1,015; 1,747] |

Fuente: Elaboración propia.

**SEGUNDA PARTE: Propiedades de Muestra Finita en Paneles Dinámicos.**

*Recientemente, ha habido un renovado interés de la literatura macroeconómica en examinar el crecimiento en el largo plazo utilizando modelos dinámicos con datos de panel para países (ver Mankiw, Romer y Weil, 1992; Levine y Renelt, 1992, entre otros). También la literatura de economía laboral ha utilizado los modelos dinámicos, por ejemplo, para el análisis de la relación existente entre salarios y desempleo, la denominada “wage curve” por Blanchflower y Oswald (1994) (ver Blanchard y Katz, 1997; Galiani, 1999).*

*En general, la dimensión de los macropaneles para analizar este tipo de problemas está caracterizada por un valor de T entre 10 y 30 y un valor de N entre 40 y 50.*

*En términos de los modelos tradicionales de datos de panel, se sabe que LSDV es sesgado e inconsistente para T fijo y N tendiendo a infinito; mientras que los estimadores que utilizan procedimientos de IV o GMM son consistentes cuando N tiende a infinito.*

*En este trabajo, se analizará cómo estimar y hacer inferencia en paneles que no tienen una dimensión muy pequeña en T y no tienen una dimensión muy grande en N. Por ejemplo, ¿un valor de T= 30 es lo suficientemente grande como para ignorar el sesgo de LSDV? o ¿un valor de N= 50 es lo suficientemente grande para que los métodos de IV o GMM den estimadores consistentes?*

*Para responder este tipo de interrogantes, considerar el siguiente modelo:*

*= + + + , j= 1, 2, … , N; t= 1, 2, … , T + 10, (8)*

*con (0, 1), (0, 1), = 0 y las primeras 10 observaciones de series temporales se descartan, es decir, que el tamaño muestral es NT. El regresor adicional, , es una variable estrictamente exógena generada de la siguiente manera: = 0,8 + , donde (0, 0,9).*

**Ejercicio 1.**

*Caso 1: = 0, T= 10, N= 30, = 0,5. Realizar un experimento de Monte Carlo con 1000 simulaciones. Reportar media, desvío estándar y RMSE de la estimación de usando: LSDV, AB-GMM1, AB-GMM2, BB-GMM1, AH y Kiviet. Reportar el tamaño del test para = 0,5 en todos los casos. Comentar los resultados obtenidos y la conclusión de qué estimador debiera utilizarse en la práctica. (Nota: AH es el estimador de Anderson-Hsiao que utiliza los niveles rezagados como instrumento para la primera diferencia; Kiviet es el estimador de LSDV corregido por el sesgo asintótico. En este último caso, utilizar AH como estimación para construir la matriz C).*

En la tabla 6, para cada método de estimación (LSDV, AB-GMM1, AB-GMM2, BB-GMM1, AH, Kiviet), se presentan la media, el desvío estándar, el RMSE y el tamaño del test (con un nivel de significatividad del 1%) de la estimación de del modelo (8), considerando los parámetros = 0,5, = 0, N= 30 y T= 10 y considerando 1000 replicaciones.

**Tabla 6.** *Simulaciones mediante estimaciones por LSDV, AB-GMM1, AB-GMM2, BB-GMM1, AH y Kiviet (con = 0,5, = 0, N= 30 y T= 10 y con 1000 replicaciones).*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Resultado** | **LSDV** | **AB-GMM1** | **AB-GMM2** | **BB-GMM1** | **AH** | **Kiviet** |
| Media | 0,333 | 0,400 | 0,397 | 0,687 | 0,504 | 0,492 |
| Desvío estándar | 0,056 | 0,100 | 0,104 | 0,091 | 0,135 | 0,078 |
| RMSE | 0,177 | 0,142 | 0,146 | 0,208 | 0,135 | 0,079 |
| Tamaño del test | 62,8 | 6,7 | 4,0 | 54,3 | 0,0 | 2,5 |

Fuente: Elaboración propia.

Con esta configuración de parámetros, por un lado, se puede observar que, mientras que, en las estimaciones LSDV, AB-GMM1, AB-GMM2 y Kiviet, la media de es menor al valor poblacional (= 0,5), en las estimaciones BB-GMM1 y AH, es mayor. Por otro lado, se observa que, en la estimación BB-GMM1, el RMSE empeora (es mayor). Por último, en las estimaciones LSDV y BB-GMM1, el tamaño del test empeora (se aleja mucho del 1%).

**Ejercicio 2.**

*Repetir el ejercicio anterior con N= 50. Comparar los resultados en ambos ejercicios.*

En la tabla 7, para cada método de estimación (LSDV, AB-GMM1, AB-GMM2, BB-GMM1, AH, Kiviet), se presentan la media, el desvío estándar, el RMSE y el tamaño del test (con un nivel de significatividad del 1%) de la estimación de del modelo (8), considerando los parámetros = 0,5, = 0, N= 50 y T= 10 y considerando 1000 replicaciones.

**Tabla 7.** *Simulaciones mediante estimaciones por LSDV, AB-GMM1, AB-GMM2, BB-GMM1, AH y Kiviet (con = 0,5, = 0, N= 50 y T= 10 y con 1000 replicaciones).*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Resultado** | **LSDV** | **AB-GMM1** | **AB-GMM2** | **BB-GMM1** | **AH** | **Kiviet** |
| Media | 0,337 | 0,433 | 0,430 | 0,641 | 0,503 | 0,494 |
| Desvío estándar | 0,043 | 0,083 | 0,091 | 0,077 | 0,102 | 0,061 |
| RMSE | 0,169 | 0,107 | 0,114 | 0,161 | 0,102 | 0,061 |
| Tamaño del test | 87,3 | 3,7 | 3,1 | 41,4 | 0,0 | 2,8 |

Fuente: Elaboración propia.

Comparando con el ejercicio anterior, con esta configuración de parámetros (un N mayor y todo lo demás igual), por un lado, se puede observar que, para todas las estimaciones, la media de se acerca más al valor poblacional (= 0,5), el desvío estándar es menor y el RMSE mejora (es menor). Por otro lado, se observa que, mientras que, para las estimaciones LSDV y Kiviet, el tamaño del test empeora (se aleja más del 1%), para el resto de las estimaciones, mejora (se acerca más al 1%).

**Ejercicio 3.**

*Ahora, suponer que = 0, T= 20, N= 30, = 0,8. Repetir el ejercicio anterior.*

En la tabla 8, para cada método de estimación (LSDV, AB-GMM1, AB-GMM2, BB-GMM1, AH, Kiviet), se presentan la media, el desvío estándar, el RMSE y el tamaño del test (con un nivel de significatividad del 1%) de la estimación de del modelo (8), considerando los parámetros = 0,8, = 0, N= 30 y T= 20 y considerando 1000 replicaciones.

**Tabla 8.** *Simulaciones mediante estimaciones por LSDV, AB-GMM1, AB-GMM2, BB-GMM1, AH y Kiviet (con = 0,8, = 0, N= 30 y T= 20 y con 1000 replicaciones).*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Resultado** | **LSDV** | **AB-GMM1** | **AB-GMM2** | **BB-GMM1** | **AH** | **Kiviet** |
| Media | 0,691 | 0,666 | 0,630 | 0,961 | 0,873 | 0,779 |
| Desvío estándar | 0,034 | 0,060 | 0,088 | 0,016 | 1,569 | 0,044 |
| RMSE | 0,114 | 0,146 | 0,191 | 0,162 | 1,570 | 0,049 |
| Tamaño del test | 84,4 | 41,8 | 9,5 | 100,0 | 0,0 | 7,7 |

Fuente: Elaboración propia.

Comparando con el ejercicio 1, con esta configuración de parámetros (un mayor, un T mayor y todo lo demás igual), por un lado, se puede observar que, para la estimación LSDV, la media de se acerca más al parámetro poblacional (= 0,8), el desvío estándar es menor, el RMSE mejora (es menor) y el tamaño del test empeora (se aleja más del 1%). Por otro lado, se observa que, para el resto de las estimaciones, el RMSE empeora (es mayor) y el tamaño del test también (se aleja más del 1%); en particular, para la estimación BB-GMM1, el tamaño del test es igual a 100%.

**Ejercicio 4.**

*Ahora, suponer que = 0, T= 20, N= 30, = 0,92. Repetir el ejercicio anterior prestando particular atención al comentario de BB acerca de los instrumentos débiles. ¿Qué se puede comentar al respecto?*

En la tabla 9, para cada método de estimación (LSDV, AB-GMM1, AB-GMM2, BB-GMM1, AH, Kiviet), se presentan la media, el desvío estándar, el RMSE y el tamaño del test (con un nivel de significatividad del 1%) de la estimación de del modelo (8), considerando los parámetros = 0,92, = 0, N= 30 y T= 20 y considerando 1000 replicaciones.

**Tabla 9.** *Simulaciones mediante estimaciones por LSDV, AB-GMM1, AB-GMM2, BB-GMM1, AH y Kiviet (con = 0,92, = 0, N= 30 y T= 20 y con 1000 replicaciones).*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Resultado** | **LSDV** | **AB-GMM1** | **AB-GMM2** | **BB-GMM1** | **AH** | **Kiviet** |
| Media | 0,844 | 0,841 | 0,842 | 1,011 | 0,920 | 0,905 |
| Desvío estándar | 0,024 | 0,045 | 0,053 | 0,005 | 0,087 | 0,030 |
| RMSE | 0,080 | 0,091 | 0,094 | 0,091 | 0,087 | 0,033 |
| Tamaño del test | 87,7 | 29,9 | 10,0 | 100,0 | 0,0 | 6,5 |

Fuente: Elaboración propia.

Comparando con el ejercicio 1, con esta configuración de parámetros (un mayor, un T mayor y todo lo demás igual), por un lado, se puede observar que, para todas las estimaciones, la media de se acerca más al parámetro poblacional (= 0,92), el desvío estándar es menor, el RMSE mejora (es menor) y el tamaño del test empeora (se aleja más del 1%); en particular, para la estimación BB-GMM1, el tamaño del test es igual a 100%.

**Ejercicio 5.**

*Ahora, suponer que = 0, T= 30, N= 50, = 0,5. Repetir el ejercicio anterior prestando particular atención al sesgo del estimador LSDV. ¿Es T= 30 suficiente como para ignorar el sesgo de LSDV?*

En la tabla 10, para cada método de estimación (LSDV, AB-GMM1, AB-GMM2, BB-GMM1, AH, Kiviet), se presentan la media, el desvío estándar, el RMSE y el tamaño del test (con un nivel de significatividad del 1%) de la estimación de del modelo (8), considerando los parámetros = 0,5, = 0, N= 50 y T= 30 y considerando 1000 replicaciones.

**Tabla 10.** *Simulaciones mediante estimaciones por LSDV, AB-GMM1, AB-GMM2, BB-GMM1, AH y Kiviet (con = 0,5, = 0, N= 50 y T= 30 y con 1000 replicaciones).*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Resultado** | **LSDV** | **AB-GMM1** | **AB-GMM2** | **BB-GMM1** | **AH** | **Kiviet** |
| Media | 0,446 | 0,460 | 0,423 | 0,764 | 0,502 | 0,499 |
| Desvío estándar | 0,024 | 0,029 | 0,059 | 0,040 | 0,049 | 0,026 |
| RMSE | 0,059 | 0,049 | 0,097 | 0,267 | 0,049 | 0,026 |
| Tamaño del test | 38,6 | 11,1 | 1,1 | 100,0 | 0,0 | 0,5 |

Fuente: Elaboración propia.

Comparando con el ejercicio 1, con esta configuración de parámetros (un T mayor, un N mayor y todo lo demás igual), por un lado, se puede observar que, para todas las estimaciones, la media de se acerca más al parámetro poblacional (= 0,5) (excepto para la estimación BB-GMM1), el desvío estándar es menor y el RMSE mejora (es menor) (excepto para la estimación BB-GMM1). Por otro lado, se observa que, mientras que, para las estimaciones LSDV, AB-GMM2 y Kiviet, el tamaño del test mejora (se acerca más al 1%), para el resto de las estimaciones, empeora (se aleja más del 1%); en particular, para la estimación BB-GMM1, el tamaño del test es igual a 100%.

Por otra parte, a pesar de las mejoras en la estimación LSDV, T= 30 no es suficiente como para ignorar el sesgo de LSDV.

**Ejercicio 6.**

*Repetir el Ejercicio 1 con = 1, T= 7, N= 100, = 0,8. Comparar los resultados con los obtenidos por AB, Tabla 1, página 284. ¿Cuáles son las conclusiones? (Nota: Ahora, se tienen que reportar también los resultados para la estimación de ).*

En la tabla 11, para cada método de estimación (LSDV, AB-GMM1, AB-GMM2, BB-GMM1, AH, Kiviet), se presentan la media, el desvío estándar, el RMSE y el tamaño del test (con un nivel de significatividad del 1%) de la estimación de y del modelo (8), considerando los parámetros = 0,8, = 1, N= 100 y T= 7 y considerando 1000 replicaciones.

**Tabla 11.** *Simulaciones mediante estimaciones por LSDV, AB-GMM1, AB-GMM2, BB-GMM1, AH y Kiviet (con = 0,8, = 1, N= 100 y T= 7 y con 1000 replicaciones).*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Resultado** | **LSDV** | **AB-GMM1** | **AB-GMM2** | **BB-GMM1** | **AH** | **Kiviet** |
| Media | 0,728 | 0,768 | 0,766 | 0,853 | 0,995 | 0,810 |
| Desvío estándar | 0,019 | 0,073 | 0,079 | 0,042 | 7,162 | 0,054 |
| RMSE | 0,075 | 0,079 | 0,085 | 0,068 | 7,161 | 0,055 |
| Tamaño del test | 92,7 | 0,7 | 0,8 | 22,0 | 0,0 | 0,0 |
| Media | 1,020 | 0,995 | 0,982 | 0,880 | 1,066 | 1,011 |
| Desvío estándar | 0,040 | 0,059 | 0,062 | 0,093 | 2,327 | 0,055 |
| RMSE | 0,045 | 0,059 | 0,065 | 0,152 | 2,327 | 0,056 |
| Tamaño del test | 2,2 | 1,5 | 1,2 | 17,0 | 0,0 | 2,2 |

Fuente: Elaboración propia.

Con esta configuración de parámetros, es posible comparar con los resultados obtenidos por BB, Tabla 2(a), página 131. Se puede observar que, tanto para como para , las medias y los desvíos estándar son semejantes, debiéndose las pequeñas diferencias, probablemente, a la cantidad de replicaciones (AB usan 100 y, acá, se usan 1000) y a la semilla utilizada.

**Ejercicio 7.**

*Repetir el Ejercicio 1 con = 0, T= 4, N= 100, = 0,8. Comparar los resultados con los*

*obtenidos por BB, Tabla 2(a), página 131.*

En la tabla 12, para cada método de estimación (LSDV, AB-GMM1, AB-GMM2, BB-GMM1, AH, Kiviet), se presentan la media, el desvío estándar, el RMSE y el tamaño del test (con un nivel de significatividad del 1%) de la estimación de del modelo (8), considerando los parámetros = 0,8, = 0, N= 100 y T= 4 y considerando 1000 replicaciones.

**Tabla 12.** *Simulaciones mediante estimaciones por LSDV, AB-GMM1, AB-GMM2, BB-GMM1, AH y Kiviet (con = 0,8, = 0, N= 100 y T= 4 y con 1000 replicaciones).*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Resultado** | **LSDV** | **AB-GMM1** | **AB-GMM2** | **BB-GMM1** | **AH** | **Kiviet** |
| Media | 0,287 | 0,205 | 0,185 | 0,895 | 0,974 | 0,403 |
| Desvío estándar | 0,057 | 1,025 | 1,044 | 0,092 | 27,976 | 0,200 |
| RMSE | 0,516 | 1,185 | 1,211 | 0,132 | 27,963 | 0,444 |
| Tamaño del test | 100 | 8,5 | 12,1 | 30,0 | 0,0 | 77,1 |

Fuente: Elaboración propia.

Con esta configuración de parámetros, es posible comparar con los resultados obtenidos por BB, Tabla 2(a), página 131. Se puede observar que la media y el desvío estándar son semejantes, debiéndose las pequeñas diferencias, probablemente, a la cantidad de replicaciones y a la semilla utilizada.

**Ejercicio 8.**

*Comparando los resultados de los ejercicios 3, 6 y 7, ¿se puede sacar alguna conclusión acerca del comportamiento de los estimadores para las dimensiones del panel en 3 versus la dimensión en 6 y 7?*

Comparando los resultados de los ejercicios 3, 6 y 7, acerca del comportamiento de los diferentes estimadores, se puede mencionar lo siguiente:

* LSDV: son preferibles las dimensiones del ejercicio 3 (N= 30 y T= 20), ya que se comporta mejor el tamaño del test.
* AB-GMM1, AB-GMM2 y BB-GMM1: son preferibles las dimensiones del ejercicio 6 (N= 100 y T= 7), ya que se compartan mejor la media, el desvío estándar y el tamaño del test.
* AH y Kiviet: son preferibles las dimensiones del ejercicio 3 (N= 30 y T= 20), ya que se compartan mejor la media, el desvío estándar y el RMSE.

**TERCERA PARTE: Elasticidad Precio de Demanda de los Cigarrillos.**

*En este ejercicio, se pide estimar una serie de modelos de panel para tratar de calcular la elasticidad precio de la demanda de los cigarrillos. Los datos, “cigs.dta”, tienen las siguientes columnas:*

* *sales: ventas de cigarrillos en paquetes per cápita.*
* *pimin: precio mínimo en estados cercanos por paquete de cigarrillos.*
* *ndi: ingreso disponible per cápita.*
* *cpi: índice de precios al consumidor (1983=100).*
* *pop16: población mayor a los 16 años.*
* *pop: población total.*
* *price: precio por paquete de cigarrillos.*
* *year: año.*
* *state: estado.*

**Ejercicio 1.**

*La elasticidad precio de demanda, , es el cambio porcentual en las ventas de cigarrillos dividido por el cambio porcentual en el precio de los cigarrillos. Sea las ventas de cigarrillos en el período t y el precio de los cigarrillos en el período t, entonces, la elasticidad precio de demanda es:*

*=*

*Para estimar la elasticidad en un modelo lineal, aplicar logaritmos a las ventas y al precio y estimar por POLS:*

*log ()= + log () + + + + (9)*

*Reportar la estimación de la elasticidad precio de demanda. ¿Qué puede estar potencialmente mal en esta estimación?*

En la tabla 13, se presenta la estimación por POLS del modelo (9). Se puede observar que la estimación de la elasticidad precio de la demanda es -0,839, lo que indica que, ante una variación de 1% en el precio de los cigarrillos, las ventas de cigarrillos disminuyen, en promedio, 0,839%, *céteris páribus*.

Por otra parte, lo que puede estar potencialmente mal en esta estimación es que no se considera la posible presencia de heterogeneidad individual no observable (en este caso, efectos específicos de cada estado) ni tampoco la posibilidad de que sea un modelo dinámico (en este caso, que log () se considere como variable explicativa), con lo cual, de ser relevantes estas variables, el estimador POLS será sesgado e inconsistente.

**Tabla 13.** *Estimación por POLS del modelo (9).*

Linear regression Number of obs = 1,380

F(4, 1375) = 76.11

Prob > F = 0.0000

R-squared = 0.2196

Root MSE = .1987

------------------------------------------------------------------------------

| Robust

ln\_sales | Coefficient std. err. t P>|t| [95% conf. interval]

-------------+----------------------------------------------------------------

ln\_price | -.8385905 .0522926 -16.04 0.000 -.9411723 -.7360086

ndi | .0000319 4.77e-06 6.69 0.000 .0000225 .0000412

cpi | .0075624 .0007855 9.63 0.000 .0060216 .0091032

pop16 | -2.53e-06 1.43e-06 -1.78 0.076 -5.33e-06 2.65e-07

\_cons | 7.410038 .1602936 46.23 0.000 7.095591 7.724484

------------------------------------------------------------------------------

**Ejercicio 2.**

*Una característica importante puede ser la heterogeneidad entre estados. Sin embargo, no se sabe si se quiere modelar esta heterogeneidad como un efecto fijo o aleatorio. Realizar un contraste de hipótesis para determinar qué modelización usar. Interpretar el resultado del contraste a niveles de significación usuales.*

En la tabla 14, se presenta el test de Hausman (con la opción *sigmamore*). Al realizar este contraste de hipótesis para determinar qué modelización usar, y considerando un nivel de significatividad del 1%, se encuentra que 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.

**Tabla 14.** *Test de Hausman.*

---- Coefficients ----

| (b) (B) (b-B) sqrt(diag(V\_b-V\_B))

| est\_fe est\_re Difference Std. err.

-------------+----------------------------------------------------------------

ln\_price | -.3984733 -.4111023 .012629 .0033814

ndi | -.0000481 -.0000454 -2.67e-06 4.50e-07

cpi | .0105834 .0104635 .00012 .0000558

pop16 | 7.77e-06 6.33e-06 1.44e-06 2.37e-06

------------------------------------------------------------------------------

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

chi2(3) = (b-B)'[(V\_b-V\_B)^(-1)](b-B)

= 36.66

Prob > chi2 = 0.0000

En la tabla 15, se presenta la estimación por FE del modelo (9). Por un lado, se puede observar que la estimación de la elasticidad precio de la demanda es -0,398, lo que indica que, ante una variación de 1% en el precio de los cigarrillos, las ventas de cigarrillos disminuyen, en promedio, 0,398%, *céteris páribus*. Por otro lado, se observa que este valor es considerablemente menor (en valor absoluto) que en la estimación por POLS, indicando que la omisión de los efectos fijos por estado genera un sesgo negativo en este coeficiente.

**Tabla 15.** *Estimación por FE del modelo (9).*

Fixed-effects (within) regression Number of obs = 1,380

Group variable: state Number of groups = 46

R-squared: Obs per group:

Within = 0.4163 min = 30

Between = 0.0194 avg = 30.0

Overall = 0.0491 max = 30

F(4,45) = 22.51

corr(u\_i, Xb) = -0.2658 Prob > F = 0.0000

(Std. err. adjusted for 46 clusters in state)

------------------------------------------------------------------------------

| Robust

ln\_sales | Coefficient std. err. t P>|t| [95% conf. interval]

-------------+----------------------------------------------------------------

ln\_price | -.3984733 .0598727 -6.66 0.000 -.5190632 -.2778834

ndi | -.0000481 .0000101 -4.74 0.000 -.0000685 -.0000277

cpi | .0105834 .0013546 7.81 0.000 .0078551 .0133118

pop16 | 7.77e-06 .0000178 0.44 0.664 -.000028 .0000436

\_cons | 5.967983 .1848623 32.28 0.000 5.595651 6.340315

-------------+----------------------------------------------------------------

sigma\_u | .20556957

sigma\_e | .09935482

rho | .81064 (fraction of variance due to u\_i)

------------------------------------------------------------------------------

**Ejercicio 3.**

*Se empieza a creer que no sólo la heterogeneidad es importante, sino también la dinámica, así que se decide estimar el siguiente modelo:*

*log ()= + log () + log () + + + + .*

*Aplicar la transformación de diferencias finitas al modelo del ejercicio anterior. Sea = log () - log () la variable dependiente de una estimación de mínimos cuadrados en dos etapas usando el procedimiento de Anderson-Hsiao. Reportar los resultados y cualquier diferencia con las estimaciones anteriores.*

En la tabla 16, se presenta la estimación por IV (Anderson-Hsiao) del modelo (9). Se puede observar que ninguna variable es estadísticamente significativa a los niveles usuales de significatividad.

**Tabla 16.** *Estimación por IV (Anderson-Hsiao) del modelo (9).*

Instrumental variables 2SLS regression Number of obs = 1,288

Wald chi2(5) = .

Prob > chi2 = .

R-squared = .

Root MSE = .21724

------------------------------------------------------------------------------

| Robust

D.ln\_sales | Coefficient std. err. z P>|z| [95% conf. interval]

-------------+----------------------------------------------------------------

ln\_sales |

LD. | 4.956136 3.604764 1.37 0.169 -2.109071 12.02134

|

ln\_price |

D1. | -.176714 .1431834 -1.23 0.217 -.4573483 .1039204

|

ndi |

D1. | .0000923 .000067 1.38 0.168 -.000039 .0002236

|

cpi |

D1. | -.0009513 .0026053 -0.37 0.715 -.0060575 .0041549

|

pop16 |

D1. | 7.10e-06 .0000204 0.35 0.728 -.0000329 .0000471

------------------------------------------------------------------------------

Instrumented: LD.ln\_sales

Instruments: D.ln\_price D.ndi D.cpi D.pop16 L2.ln\_sales

**Ejercicio 4.**

*Reestimar el modelo usando el estimador de Arellano-Bond. ¿Se encuentra alguna diferencia con las estimaciones anteriores? Interpretar los resultados.*

En las tablas 17 y 18, se presentan las estimaciones por GMM One-Step y GMM Two-Step, respectivamente, del modelo (9). Por un lado, se puede observar que, cuando se estima por GMM One-Step (GMM Two-Step), la estimación de la elasticidad precio de la demanda es -0,094 (-0,113), lo que indica que, ante una variación de 1% en el precio de los cigarrillos, las ventas de cigarrillos disminuyen, en promedio, 0,094% (0,113%), *céteris páribus*. Por otro lado, se observa que este valor es sustancialmente menor (en valor absoluto) que en las estimaciones anteriores (POLS y FE), indicando que la omisión de log () como variable explicativa genera un sesgo negativo en este coeficiente. Por último, esta nueva variable explicativa, log (), es estadísticamente significativa al 1%.

**Tabla 17.** *Estimación por GMM One-Step (Arellano-Bond) del modelo (9).*

Dynamic panel-data estimation, one-step difference GMM

------------------------------------------------------------------------------

Group variable: state Number of obs = 1288

Time variable : year Number of groups = 46

Number of instruments = 32 Obs per group: min = 28

Wald chi2(0) = . avg = 28.00

Prob > chi2 = . max = 28

------------------------------------------------------------------------------

| Robust

ln\_sales | Coefficient std. err. z P>|z| [95% conf. interval]

-------------+----------------------------------------------------------------

ln\_sales |

L1. | .9975072 .0460666 21.65 0.000 .9072184 1.087796

|

ln\_price | -.0935868 .0323942 -2.89 0.004 -.1570782 -.0300953

ndi | .000016 3.63e-06 4.42 0.000 8.91e-06 .0000231

cpi | -.0016501 .0003375 -4.89 0.000 -.0023117 -.0009886

pop16 | .0000646 .000038 1.70 0.089 -9.75e-06 .000139

------------------------------------------------------------------------------

Instruments for first differences equation

Standard

D.(ln\_price ndi cpi pop16)

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

L(1/29).L.ln\_sales collapsed

------------------------------------------------------------------------------

Arellano-Bond test for AR(1) in first differences: z = -5.04 Pr > z = 0.000

Arellano-Bond test for AR(2) in first differences: z = 2.09 Pr > z = 0.037

------------------------------------------------------------------------------

Sargan test of overid. restrictions: chi2(27) = 255.58 Prob > chi2 = 0.000

(Not robust, but not weakened by many instruments.)

Hansen test of overid. restrictions: chi2(27) = 40.47 Prob > chi2 = 0.046

(Robust, but weakened by many instruments.)

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

iv(ln\_price ndi cpi pop16)

Hansen test excluding group: chi2(23) = 39.20 Prob > chi2 = 0.019

Difference (null H = exogenous): chi2(4) = 1.27 Prob > chi2 = 0.866

**Tabla 18.** *Estimación por GMM Two-Step (Arellano-Bond) del modelo (9).*

Dynamic panel-data estimation, two-step difference GMM

------------------------------------------------------------------------------

Group variable: state Number of obs = 1288

Time variable : year Number of groups = 46

Number of instruments = 32 Obs per group: min = 28

Wald chi2(0) = . avg = 28.00

Prob > chi2 = . max = 28

------------------------------------------------------------------------------

| Corrected

ln\_sales | Coefficient std. err. z P>|z| [95% conf. interval]

-------------+----------------------------------------------------------------

ln\_sales |

L1. | .946648 .0786545 12.04 0.000 .792488 1.100808

|

ln\_price | -.112763 .0423575 -2.66 0.008 -.1957822 -.0297438

ndi | .0000137 5.07e-06 2.69 0.007 3.73e-06 .0000236

cpi | -.0014292 .0004688 -3.05 0.002 -.0023481 -.0005103

pop16 | .000102 .0000635 1.61 0.108 -.0000225 .0002264

------------------------------------------------------------------------------

Instruments for first differences equation

Standard

D.(ln\_price ndi cpi pop16)

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

L(1/29).L.ln\_sales collapsed

------------------------------------------------------------------------------

Arellano-Bond test for AR(1) in first differences: z = -4.74 Pr > z = 0.000

Arellano-Bond test for AR(2) in first differences: z = 2.21 Pr > z = 0.027

------------------------------------------------------------------------------

Sargan test of overid. restrictions: chi2(27) = 255.58 Prob > chi2 = 0.000

(Not robust, but not weakened by many instruments.)

Hansen test of overid. restrictions: chi2(27) = 40.47 Prob > chi2 = 0.046

(Robust, but weakened by many instruments.)

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

iv(ln\_price ndi cpi pop16)

Hansen test excluding group: chi2(23) = 39.20 Prob > chi2 = 0.019

Difference (null H = exogenous): chi2(4) = 1.27 Prob > chi2 = 0.866

**Tabla 19.** *Tabla comparativa de estimaciones del modelo (9) (POLS, FE, IV, GMM1, GMM2).*

--------------------------------------------------------------------------------------------

(1) (2) (3) (4) (5)

POLS FE IV GMM (One-S~) GMM (Two-S~)

--------------------------------------------------------------------------------------------

ln\_price -0.839\*\*\* -0.398\*\*\* -0.0936\*\*\* -0.113\*\*\*

(0.0523) (0.0599) (0.0324) (0.0424)

ndi 0.0000319\*\*\* -0.0000481\*\*\* 0.0000160\*\*\* 0.0000137\*\*\*

(0.00000477) (0.0000101) (0.00000363) (0.00000507)

cpi 0.00756\*\*\* 0.0106\*\*\* -0.00165\*\*\* -0.00143\*\*\*

(0.000785) (0.00135) (0.000338) (0.000469)

pop16 -0.00000253\* 0.00000777 0.0000646\* 0.000102

(0.00000143) (0.0000178) (0.0000380) (0.0000635)

LD.ln\_sales 4.956

(3.605)

L.ln\_sales 0.998\*\*\* 0.947\*\*\*

(0.0461) (0.0787)

D.ln\_price -0.177

(0.143)

D.ndi 0.0000923

(0.0000670)

D.cpi -0.000951

(0.00261)

D.pop16 0.00000710

(0.0000204)

\_cons 7.410\*\*\* 5.968\*\*\*

(0.160) (0.185)

--------------------------------------------------------------------------------------------

N 1380 1380 1288 1288 1288

r2 0.220 0.416

--------------------------------------------------------------------------------------------

Standard errors in parentheses

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