

Tarea 2

Adrián Javier Martínez Rodríguez

4/11/2021

Tarea 2

Pregunta 1

Table 1: Tabla 1

Dependent Variable: Model:	(1)	(2)	weeksm1 (3)	(4)	(5)
<i>Variables</i>					
(Intercept)	19.27*** (0.4190)	16.68*** (2.336)	24.94*** (5.992)	12.41*** (3.012)	13.95*** (1.327)
kidcount	-5.538*** (0.0539)	-4.433*** (0.9822)	-7.961*** (2.552)	-2.612** (1.274)	-3.267*** (0.5394)
agem1	1.402*** (0.0135)	1.339*** (0.0569)	1.538*** (0.1444)	1.237*** (0.0730)	1.274*** (0.0331)
agefstm	-1.428*** (0.0163)	-1.344*** (0.0764)	-1.612*** (0.1947)	-1.206*** (0.0982)	-1.256*** (0.0439)
black	11.20*** (0.1937)	11.02*** (0.2520)	11.59*** (0.4607)	10.72*** (0.2858)	10.83*** (0.2139)
hispan	0.4735*** (0.1795)	0.1682 (0.3250)	1.143 (0.7277)	-0.3351 (0.3954)	-0.1539 (0.2328)
othrace	3.147*** (0.2069)	3.007*** (0.2413)	3.455*** (0.3851)	2.775*** (0.2634)	2.858*** (0.2183)
<i>Fit statistics</i>					
Observations	254,654	254,654	254,654	254,654	254,654
R ²	0.07411	0.07275	0.06759	0.06460	0.06838
Adjusted R ²	0.07409	0.07273	0.06757	0.06458	0.06836
F-test (1st stage)		868.30	128.40	517.92	2,670.8
Ivs		samesex	twoboys	twogirls	multi2nd

Heteroskedasticity-robust standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Pregunta 2

Podemos observar en la tabla 1 que los coeficientes de kidcount son significativos al 10, 5 y 1% tanto para la estimación por MCO y por 2sls.

Para el coeficiente estimado en la columna 1, que una madre tenga un hijo más, ceteris paribus, está relacionado con una caída de 5.53 en el número de semanas que la madre trabajó en 1979.

Para el coeficiente de la segunda columna, que una madre tenga un hijo más, *ceteris paribus*, está asociado a una disminución de -4.43 en el número de semanas que la madre trabajó en el año de 1979.

La media de semanas trabajadas en 1979 por las madres es de 19.018. Por lo tanto, los efectos mencionados en los dos párrafos anteriores implicarían que la madre trabaja menos en un 29% y 23% de la media, lo cual es una disminución significativa.

Pregunta 3

Para que un instrumento sea de utilidad para una estimación, necesitamos que éste esté altamente correlacionado con nuestra variable explicativa. Para verificar la relevancia de nuestro instrumento, podemos estimar la siguiente ecuación

$$\text{kidcount}_i = \beta_0 + \gamma \times \text{samesex} + \beta \mathbf{X}_i$$

donde *samesex* es nuestro instrumento y \mathbf{X} contiene los controles (variables exógenas). Para verificar que nuestro instrumento sea relevante necesitamos que nuestra estimación nos arroje que el coeficiente γ es significativo. De hecho, la regla dice que necesitamos un estadístico F mayor a 10, lo cual es aproximadamente igual, en el caso de un solo instrumento, a un estadístico t de 3.2.

El estadístico t encontrado haciendo la regresión mencionada es igual a 29.467, el cual es considerablemente mayor que el valor mínimo necesario de 3.2.

La relevancia del instrumento es intuitiva. Una familia cuyos dos primeros hijos son del mismo sexo pueden inclinarse a tener otro hijo para poder tener un hijo con el sexo que hace falta.

Pregunta 4

Además del supuesto de relevancia, el instrumento debe cumplir con el supuesto de exogeneidad/exclusión, es decir, que el instrumento no esté correlacionado con la variable dependiente, *weeksm1*, de manera directa (solo a través de *kidcount*).

Table 2: Tabla de Balance

Variables	Samesex = 0	Samesex = 1	Valor-p
agefstm	20.8253580	20.8377801	0.2832999
agem1	30.4040061	30.3827644	0.1134896
black	0.0522838	0.0510544	0.1611723
othrace	0.0566441	0.0560488	0.5148298

El supuesto de exogeneidad no se puede demostrar pero sí se puede dar una intuición. En primer lugar, el sexo del hijo es aleatorio, por lo tanto, es muy difícil pensar con qué variable omitida se puede correlacionar.

Una forma de analizar si el instrumento está correlacionado con alguna característica no observable es por medio de una tabla de balance. Al observar la tabla, podemos ver que la media entre los grupos que tienen *samesex* == 1 y *samesex* == 0 son iguales estadísticamente hablando. Esto es, las madres que tienen sus primeros dos hijos con el mismo sexo y las que no, no tienen, en promedio, características distintas. Esto nos puede dar indicios de que tal vez no hay no observables correlacionados con *z*, aunque, repito, esto no se puede demostrar.

Pregunta 5

Para hacer la estimación del coeficiente de *kidcount* por medio de la forma reducida y de la primera etapa.

La forma reducida es

$$\text{weeksm1}_i = \gamma_0 + \gamma_1 \text{samesex}_i + \eta \mathbf{X}_i + \tau_i$$

donde \mathbf{X} son los controles *agem1*, *agefstm*, *black*, *hispan* y *othrace*.

El coeficiente que nos importa de la regresión es γ_1 . El valor de γ_1 es -0.3766973.

La primera etapa es

$$\text{kidcount}_i = \beta_0 + \gamma \times \text{samesex} + \beta \mathbf{X}_i + \theta_i$$

El coeficiente de interés de esta regresión es γ . El valor de γ es 0.0849798.

Para obtener el mismo coeficiente de *kidcount* que se obtuvo en la pregunta 1 cuando se utiliza como instrumento *samesex*, se tiene que hacer lo siguiente

$$\alpha = \gamma_1 / \gamma$$

El valor de α es igual a -4.4327882 (encontramos el mismo coeficiente estimado en la pregunta número 1 columna 2).

Pregunta 6

(a)

En las columnas 3, 4 y 5 se pueden observar los resultados de las estimaciones de 2SLS utilizando como instrumentos las variables *twoboys*, *twogirls* y *multi2nd*. Los coeficientes estimados implican que, ceteris paribus, el número de semanas que una madre trabajó en 1979 disminuye en 7.96, 2.61 y 3.26 cuando los instrumentos son *twoboys*, *twogirls* y *multi2nd* respectivamente. El signo del efecto es consistente con lo encontrado en el modelo de la columna 2. Sin embargo, las magnitudes difieren mucho dependiendo del instrumento utilizado. Esto se debe a los diferentes mecanismos por los cuales el instrumento afecta a *kidcount*. Por ejemplo, tal vez las familias tienen una preferencia por hijos varones. De hecho, si checamos los resultados de las primeras etapas, que los dos primeros hijos sean mujeres tiene un efecto en magnitud más grande (0.077) sobre *kidcount* que cuando se utiliza que los dos primeros hijos sean hombres (0.037). Por lo tanto, como los diferentes instrumentos afectan a la variable de interés por medio de distintos mecanismos es de esperar que los coeficientes obtenidos para *kidcount* sean distintos dependiendo del instrumento, pues el origen de la variación es distinta.

(b)

Cuando el IV es *samesex*, un individuo es complier si al tener hijos del mismo sexo decide tener un hijo más. Cuando el IV es *twoboys*, un individuo es complier si decide tener otro hijo más cuando sus dos primeros hijos fueron varones. Cuando el IV es *twogirls*, un individuo es complier cuando decide tener otro hijo dado que sus dos primeros hijos fueron mujeres. Por último, cuando el IV es *multi2nd*, un individuo es complier si tiene más hijos dado que el segundo parto fue múltiple.

(c)

La variable *morekids* nos indica si la madre tuvo más de dos hijos. Por lo tanto, podemos decir que una mujer toma el tratamiento si *morekids* es igual a 1. Por ejemplo, si *samesex* = 1, entonces la mujer va a tener más hijos y *morekids* sería igual a 1 (en el caso de que sea una complier).

Para poder calcular la proporción de compliers necesitamos asumir que no existen defiers en nuestra muestra. Para calcular la proporción podemos hacer la siguiente regresión

$$P(\text{morekids} = 1 | \text{samesex}) = \beta_0 + \beta_1 \text{samesex} + u_i$$

Entonces, si *samesex* == 0,

$$P(\text{morekids} = 1 | \text{samesex} = 0) = \beta_0 = P(\text{always}) + P(\text{defier})$$

Si *samesex* == 1,

$$P(\text{morekids} = 1 | \text{samesex} = 1) = \beta_0 + \beta_1 = P(\text{always}) + P(\text{complier})$$

Por último, la probabilidad/proporción de never es igual a

$$P(Never) = 1 - P(always) - P(defier) - P(complier)$$

Notemos que tenemos un sistema con 3 ecuaciones, pero 4 incógnitas (no tiene solución). Si asumimos que no hay defiers en nuestra muestra, entonces $P(defier) = 0$ y nuestro sistema se puede identificar.

Además, al asumir que no existen defiers, podemos identificar la proporción de compliers con la regresión anterior, pues el cambio entre $P(morekids = 1|samesex = 1)$ y $P(morekids = 1|samesex = 0)$ se debe SOLO a los compliers, pues los always taker siempre escogen “tratarse”.

Table 3: Probabilidades

	samesex	twoboys	twogirls	multi2nd
P(Complier)	0.0675253	0.0322259	0.0581361	0.0581361
P(Always)	0.3464248	0.3719836	0.3666498	0.3666498
P(Never)	0.5860499	0.5957905	0.5752141	0.5752141

(d)

En el podcast de “Do Baby Girls Cause Divorce?” mencionan que las familias tienen una preferencia por hijos varón. Entonces, podemos pensar que una familia cuyos dos primeros hijos son mujeres, va a querer tener un hijo más para conseguir un varón.

Para probar la hipótesis del podcast podemos basarnos en la tabla de probabilidades donde se utiliza como instrumento a twoboys y twogirls. Ahí podemos observar que cuando los dos primeros hijos son mujeres, hay una mayor proporción de complier que cuando los dos primeros hijos son hombres. Esto significa que una mayor cantidad de familias deciden tener otro hijo cuando sus primeros hijos son mujeres que cuando sus primeros hijos son hombres, lo cual puede ser una señal de preferencias por hijos varones.

Pregunta 7

(a)

Al comparar el resultado de la tabla 2 columna 1 con los resultados de la tabla 1 columna 3 y 4, podemos observar que el resultado utilizando dos instrumentos (-4.018) se encuentra entre los resultados estimados utilizando un solo instrumento (-2.61 y -7.96). Esto es intuitivo, pues al incorporar las variaciones de dos instrumentos, podemos pensar que las predicciones de la primera etapa para kidcount van a ser una combinación de las predicciones de kidcount con tan solo un instrumento.

(b)

Los resultados de la tabla 2 columna 2 y 3 son muy similares, -3.554 y -3.502 respectivamente. La única diferencia es que en la estimación de la columna 3 no se incorporaron controles. Al no incorporar los controles, éstos se adhieren a la parte de no observables. Por lo tanto, que las magnitudes de los efectos no sean muy distintos (prácticamente son iguales) se debe a la exogeneidad de kidcount causada por los instrumentos. Como los padres no pueden elegir el sexo o si el embarazo es de mellizos, es muy poco probable que haya una correlación con los controles que genere un sesgo en el coeficiente y, por ende, resultados diferentes entre las especificaciones de la columna 2 y 3.

Pregunta 8

Para la estimación de GMM se utilizó Stata.

(a)

Table 4: Tabla 2

Dependent Variable: Model:	(1)	(2)	weeksm1 (3)	(4)	(5)
<i>Variables</i>					
kidcount	-4.018*** (0.9506)	-3.554*** (0.4734)	-3.502*** (0.4972)	-3.44*** (0.469)	-3.536*** (0.473)
<i>Fit statistics</i>					
Observations	254,654	254,654	254,654	254,654	254,654
F-test (1st stage)	464.10	1,776.7	1,563.1		
Método	2SLS	2SLS	2SLS	GMM	GMM
Controles	Sí	Sí	No	Sí	Sí
IVs	twoboys twogirls	samesex multi2nd	samesex multi2nd	twoboys twogirls multi2nd	samesex multi2nd
<i>Heterocedastic standard-errors in parentheses</i>					
<i>Signif. Codes: ***: 0.01, **: 0.05, *: 0.1</i>					

Los resultados de ambas estimaciones por GMM son muy parecidas tanto en magnitud como en errores estándar. Sin embargo, los errores estándar obtenidos en la columna 4 son menores, por lo cual, por términos de eficiencia, podríamos decir que la estimación de la columna 4 es el mejor.

(b)

Al comparar los resultados por GMM y por 2SLS encontramos que los coeficientes son casi similares al igual que los errores estándar. Esto se debe a que 2SLS es una estimación particular del método GMM. De hecho, podemos obtener los mismos coeficientes si en lugar de usar una matriz de pesos ajustada (para el GMM) utilizamos una no ajustada. Además, las pequeñas diferencias que podamos encontrar pueden deberse a que GMM resuelve de manera numérica las ecuaciones de momentos, mientras que 2SLS asume que las ecuaciones de los momentos se cumplen y con base en eso obtiene los estimadores.

Pregunta 9

Con base en el resultado de la tabla 1, escogí repetir la estimación para la variable workedm utilizando como instrumento la variable multi2nd, que fue el instrumento con el estadístico F más grande de la tabla 1 y tabla 2.

Los resultados de la estimación los podemos encontrar en la tabla 3 columna 1. La variable workedm es una dummy que tiene un valor igual a 1 si la madre tuvo un trabajo con paga en el año de 1979. Un incremento de una unidad en kidcount, ceteris paribus, está relacionado con una disminución de 6.8 puntos porcentuales en la probabilidad de que la madre trabaje a cambio de un salario.

Pregunta 10

Para esta pregunta se corrió una regresión de dos etapas utilizando como variable instrumental multi2nd y solo para las madres que reportan trabajar. Los resultados se encuentran en la columna 2 de la tabla 3. Los resultados nos indican que, ceteris paribus, un incremento en el número de hijos en una unidad reduce en 788 dólares el salario laboral semanal de la madre. El salario semanal promedio de las madres que trabajan es igual a 11,791, lo que implicaría que el efecto de tener un hijo más es aproximadamente un 7% del promedio.

Pregunta 11

En la columna 3 podemos observar las estimaciones donde se incluye una interacción entre kidcount y hispan. Los instrumentos utilizados fueron multi2nd y multi2nd*hispan. Tal y como se ha encontrado a lo largo de la tarea, tener un hijo disminuye, ceteris paribus, el número de semanas que las madres trabajan. Ahora, el coeficiente de la interacción entre kidcount y hispan nos dice la diferencia del efecto de kidcount sobre madres hispanas y de otras razas. El cambio en el número de semanas para una madre hispana cuando se incrementa en uno el número de sus hijos es igual a -4.187, mientras que el cambio en el número de semanas para una madre no hispana ante un incremento de 1 en el número de hijos es de -3.009. La diferencia entre ambos efectos es igual al coeficiente de la interacción (-1.178). Por lo tanto, tener un hijo más tiene un mayor efecto sobre el número de semanas trabajadas para las madres hispanas. Sin embargo, este coeficiente no es significativo, por lo cual, la diferencia no es estadísticamente distinta a 0.

Table 5: Tabla 3

Dependent Variables: Model:	workedm (1)	incomem (2)	weeksm1 (3)
<i>Variables</i>			
(Intercept)	0.5649*** (0.0317)	-2,364.3** (959.7)	15.22*** (1.413)
kidcount	-0.0678*** (0.0129)	-787.6* (402.1)	-3.009*** (0.5552)
agem1	0.0221*** (0.0008)	469.2*** (22.59)	1.260*** (0.0367)
agefstm	-0.0264*** (0.0010)	52.56* (30.96)	-1.295*** (0.0475)
black	0.1930*** (0.0045)	5,806.7*** (143.9)	
hispan	-0.0103* (0.0054)	368.9** (169.3)	3.610 (7.344)
othrace	0.0472*** (0.0049)	3,193.5*** (180.8)	
kidcount:hispan			-1.178 (2.631)
<i>Fit statistics</i>			
Observations	254,654	134,513	254,654
F-test (1st stage)	2,670.8	1,604.7	
F-test (1st stage), kidcount			1,339.1
F-test (1st stage), kidcount:hispan			425.56
Método	2SLS	2SLS	2SLS
IV	multi2nd	multi2nd	multi2nd multi2nd*hispan

Heteroskedasticity-robust standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*