

Tarea II

Marco Ramos

Contents

I.	2
II.	2
III.	2
IV.	4
V.	4
VI.	5
A)	6
B)	6
C)	6
D)	7
VII.	8
A)	8
B)	8
VIII.	9
A)	9
B)	9
IX.	10
X.	10
XI.	11

I.

Table 1: Un instrumento

	(1)	(2)	(3)	(4)	(5)
Variable dependiente:	<i>weeksm1</i>	<i>weeksm1</i>	<i>weeksm1</i>	<i>weeksm1</i>	<i>weeksm1</i>
<i>kidcount</i>	-3.466*** (0.0528)	-4.433*** (0.9822)	- -	- -	- -
<i>N</i>	254,654	254,654	-	-	-
<i>F</i> 1a etapa	-	868.296***	-	-	-
Método	MCO	IV	-	-	-
Instrumento	-	<i>samesex</i>	-	-	-

Todas las regresiones controlan por *agem1*, *agef*, *stm*, *black*, *hispan*, *othrace*

Errores robustos entre parentesis

*p<0.1; **p<0.05; ***p<0.01

II.

Para la columna 1, el MCO, un aumento de uno en la cantidad de hijos a partir de 2 está relacionado con una disminución de 3.466 en las horas de trabajo semanales de la madre (en 1979), todo lo demas constante. Para la columna 2, un aumento de uno en la cantidad de hijos está relacionado con una disminución de 4.43 en las horas de trabajo semanales de la madre (en 1979) para las mujeres que tuvieron a los dos primeros hijos del mismo sexo, todo lo demas constante.

Ambos coeficientes son estadísticamente significantes al 99% y ambos efectos son de tamaño considerable pues componen el 6% y el 8% del máximo posible de horas que las mujeres de la muestra trabajaron (52 horas).

III.

El supuesto de relevancia se refiere a que nuestro instrumento debe estar asociado con nuestra variable endógena. Este supuesto es muy importante porque permite que nuestro instrumento identifique a nuestra variable endógena y en consecuencia la variable instrumental tenga un buen desempeño. En este caso, nuestro instrumento es un instrumento fuerte pues la regresión del primer paso tiene una F muy grande (que podemos ver en la tabla anterior) y significativa al 99%.

Asimismo, a continuación mostramos el primer paso de la estimación de VI. Notamos que *samesex* tiene un coeficiente positivo y estadísticamente significativo al 99%. Asimismo, cuenta con un estadístico t de 28.3 ($t = \frac{.085}{.003}$) lo cuál indica una fuerte asociación.

Table 2:

	<i>Dependent variable:</i>
	kidcount
samesex	0.085*** (0.003)
agem1	0.056*** (0.0005)
agefstm	-0.076*** (0.001)
black	0.164*** (0.007)
hispan	0.276*** (0.006)
othrace	0.127*** (0.007)
Constant	2.299*** (0.014)
Observations	254,654
R ²	0.106
Adjusted R ²	0.106
Residual Std. Error	0.728 (df = 254647)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

IV.

El otro supuesto es el de exclusion y se refiere a que nuestro instrumento solo tiene un efecto en nuestra variable objetivo a través de la variable endógena X. Esto con el fin de construir una aproximación exógena de nuestra variable endógena X y corregir por la correlación entre u y x. Para corroborar este supuesto podemos comparar la regresión estructural contemplando al mismo tiempo nuestra variable endógena y el instrumento frente a la misma regresión estructural sin la variable endógena.

Table 3:

	<i>Dependent variable:</i>	
	weeksm1	
	(1)	(2)
kidcount	-5.542*** (0.057)	
samesex	0.094 (0.084)	-0.377*** (0.085)
agem1	1.402*** (0.014)	1.089*** (0.014)
agefstm	-1.429*** (0.016)	-1.007*** (0.016)
black	11.199*** (0.189)	10.289*** (0.193)
hispan	0.475*** (0.177)	-1.057*** (0.179)
othrace	3.148*** (0.200)	2.442*** (0.204)
Constant	19.229*** (0.435)	6.486*** (0.422)
Observations	254,654	254,654
R ²	0.074	0.040
Adjusted R ²	0.074	0.040
Residual Std. Error	21.042 (df = 254646)	21.424 (df = 254647)
F Statistic	2,911.996*** (df = 7; 254646)	1,773.779*** (df = 6; 254647)
<i>Note:</i>		*p<0.1; **p<0.05; ***p<0.01

En esta regresión podemos observar que en presencia de kidcount el coeficiente de samesex no es distinto a cero de manera estadísticamente significativa.

V.

Podemos generar el coeficiente de variable instrumental de *kidcount* con valor de -4.43 dividiendo el coeficiente del instrumento de la forma reducida entre el coeficiente del instrumento en la regresión de primera etapa.

$$Y = -0.377 * X + \sum_{i=j}^k \beta_i W_i$$

$$X = 0.085 * Z + \sum_{i=j}^k \beta_i W_i$$

$$\beta_{IV} = -0.377/0.085 = -4.43$$

Asimismo, también podemos estimar el coeficiente de la regresión estructural usando nuestra x de la regresión de primera etapa (etiquetada como *kdcount_inst*).

Table 4:

	<i>Dependent variable:</i>		
	(Etapla 1) kidcount	(Reducida) weeksm1	(Estructural IV) weeksm1
samesex	0.085*** (0.003)	-0.377*** (0.085)	
kdcount_inst			-4.433*** (0.999)
agem1	0.056*** (0.0005)	1.089*** (0.014)	1.339*** (0.058)
agefstm	-0.076*** (0.001)	-1.007*** (0.016)	-1.344*** (0.078)
black	0.164*** (0.007)	10.289*** (0.193)	11.017*** (0.253)
hispan	0.276*** (0.006)	-1.057*** (0.179)	0.168 (0.329)
othrace	0.127*** (0.007)	2.442*** (0.204)	3.007*** (0.240)
Constant	2.299*** (0.014)	6.486*** (0.422)	16.679*** (2.379)
Observations	254,654	254,654	254,654
R ²	0.106	0.040	0.040
Adjusted R ²	0.106	0.040	0.040
Residual Std. Error (df = 254647)	0.728	21.424	21.424
F Statistic (df = 6; 254647)	5,009.233***	1,773.779***	1,773.779***

Note:

*p<0.1; **p<0.05; ***p<0.01

VI.

A)

Table 5: Un instrumento

	(1)	(2)	(3)	(4)	(5)
Variable dependiente:	<i>weeksm1</i>	<i>weeksm1</i>	<i>weeksm1</i>	<i>weeksm1</i>	<i>weeksm1</i>
<i>kidcount</i>	-3.466*** (0.0528)	-4.433*** (0.9822)	-7.961*** (2.552)	-2.612** (1.2743)	-3.267*** (0.5394)
<i>N</i>	254,654	254,654	254,654	254,654	254,654
<i>F</i> 1a etapa	-	868.296***	128.405***	517.92***	2670.76***
Método	MCO	IV	IV	IV	IV
Instrumento	-	<i>samesex</i>	<i>twoboys</i>	<i>twogirls</i>	<i>multi2nd</i>

Todas las regresiones controlan por *agem1*, *agef*, *stm*, *black*, *hispan*, *othrace*

Errores robustos entre parentesis

*p<0.1; **p<0.05; ***p<0.01

De la tabla podemos observar que parece que todas las estimaciones se elaboraron con base en un buen instrumento (por el alto valor de cada *F*); qué todos los coeficientes son estadísticamente significantes y negativos; sin embargo, algo muy interesante es su tamaño pues el de la estimación con el instrumento de *twoboys* produce un coeficiente de casi 3 veces el tamaño de la estimación con *twogirls* como instrumento. Esto es evidencia de un efecto diferenciado por género en el que las mamás dejan de trabajar más horas una vez tenidos dos hijos que si hubieran tenido dos hijas. Probablemente por la hipótesis del podcast cuyo argumento se desarrolla en otros incisos. Asimismo notamos que cuando usamos el instrumento *multi2nd*, el coeficiente que obtenemos es muy similar al que obtuvimos con MCO, lo que puede indicar que no hay un sesgo considerable que controlar para las parejas que tuvieron gemelos en su segundo y tercer hijo. Personalmente a priori esperaba que los coeficientes serían similares y que la desviación de cada uno sería mayor en IV. Lo que no esperaba era la diferencia entre la estimación para IV *twoboys* con *twogirls*. Finalmente, los resultados no son comparables pues dependiendo la variable instrumental que se está usando es el grupo específico del cual el β indica el LATE.

B)

Compliers

- Para el modelo 2 (con IV *samesex*) el complier serían mujeres que tuvieron hijos adicionales si solo si los dos primeros hijos fueron del mismo sexo.
- Para el modelo 3 (con *twoboys*) el complier son mujeres que tuvieron hijos adicionales solo si los dos primeros hijos fueron varones.
- Para el modelo 4 (con *twogirls*) el complier son las mujeres que tuvieron hijos adicionales si solo si los dos primeros hijos fueron mujeres.
- Para el modelo 5 (con *multi2nd*) el complier son las mujeres que tuvieron hijos adicionales solo si tuviesen gemelos en el segundo y tercer hijo.

C)

Para hacer el cálculo siguiente haremos uso de la siguientes desigualdades que se derivan de las posibles categorizaciones de individuos.

$$\begin{aligned}
P(Defier) &\leq P(T = 1|Z = 0) + P(T = 0|Z = 1) \\
P(AlwaysT) &\leq P(T = 1|Z = 1) + P(T = 1|Z = 0) \\
P(NeverT) &\leq P(T = 0|Z = 0) + P(T = 0|Z = 1) \\
P(Complier) &\leq P(T = 1|Z = 1) + P(T = 0|Z = 0)
\end{aligned}$$

De ahí que obtengamos las siguientes 2 ecuaciones:

$$\begin{aligned}
P(T = 1|Z = 1) &= P(AlwaysT) + P(Compliers) \\
P(T = 1|Z = 0) &= P(AlwaysT) + P(Defiers)
\end{aligned}$$

Donde Z se refiere a nuestro instrumento y T en este caso se refiere a si la persona ha tenido más de 2 hijos, para lo cual aprovechamos que ya existe dicha variable (*morekids*).

Ahora, podemos despejar la probabilidad de ser cada tipo de participante solo si asumimos el supuesto de que la probabilidad de ser defier es cero. De lo contrario jamás podríamos averiguar la probabilidad de ser un *always taker* en la segunda ecuación y en consecuencia no podríamos despejar la probabilidad de ser un *complier* de la primera ecuación. El *defier* sería aquella persona que teniendo valor de 1 en el instrumento, no ha tenido más de 2 hijos o que teniendo valor de cero en el instrumento, sí ha tenido más de dos hijos. Como estos casos también podrían referirse a un *never taker* y a un *always taker*, respectivamente, de ahí la conveniencia del supuesto.

Table 6: Porcentaje

Modelo:	(1)	(2)	(3)	(4)	(5 naive)	(5)
<i>always takers</i>	-	34.6 %	37.2 %	36.7 %	37.53 %	9.4 %
<i>compliers</i>	-	06.7 %	03.2 %	05.8 %	62.47 %	14.4 %
<i>never takers</i>	-	59.6 %	59.6 %	57.5 %	00.0 %	76.1 %
<i>defiers</i>	-	00.0 %	00.0 %	00.0 %	00.0 %	00.0 %
<i>N</i>	254,654	254,654	254,654	254,654	254,654	254,654
Método	MCO	IV	IV	IV	IV	IV
Instrumento	-	<i>samesex</i>	<i>twoboys</i>	<i>twogirls</i>	<i>multi2nd</i>	<i>multi2nd</i>

Todas las regresiones controlan por *agem1*, *agef*, *stm*, *black*, *hispan*, *othrace*

Para 5 naive se uso *morekids* como T

Para 5 se uso una variable ad hoc (más de 3 hijos) como T

D)

Sí es posible que sean compatibles. Por ejemplo, el hecho de que el coeficiente de *kidcount* sea mayor con la VI *twoboys* puede indicarnos que las mamá compliers en este caso al contar con la seguridad de un matrimonio (y el ingreso del esposo asegurado para el hogar) puede dejar de trabajar más tiempo. Por el contrario, el coeficiente (de menor tamaño) con la IV *twogirls* puede indicarnos que las mujeres compliers están en una posición donde no pueden dejar de trabajar tanto, probablemente debido a que están solteras, divorciadas o el marido abandonó a la familia.

Table 7: Varios instrumentos

	(1)	(2)	(3)	(4)	(5)
Variable dependiente:	<i>weeksm1</i>	<i>weeksm1</i>	<i>weeksm1</i>	<i>weeksm1</i>	<i>weeksm1</i>
<i>kidcount</i>	-4.018*** (0.9506)	-3.554*** (0.4734)	-3.502*** (0.497)	- -	- -
<i>N</i>	254,654	254,654	254,654	-	-
<i>F</i> 1a etapa	464.098***	1776.748***	1563.088 ***	-	-
Método	IV	IV	IV	GME	GME
Instrumento	<i>twoboys</i> <i>twogirls</i>	<i>samesex</i> <i>multi2nd</i>	<i>samesex</i> <i>multi2nd</i>		
Controles	Sí	Sí	No	Sí	Sí

Todas las regresiones controlan por *agem1*, *agef*, *stm*, *black*, *hispan*, *othrace*

Errores robustos entre parentesis

*p<0.1; **p<0.05; ***p<0.01

VII.

A)

El resultado sí era de esperar pues ahora estamos usando dos VIs, lo que quiere decir que la variable endógena está siendo estimada con base en dos instrumentos. Como esto implica que la primera etapa sería algo como $X = \beta_{z1}Z_1 + \beta_{z2}Z_2 + \sum_{i=j}^k \beta_i W_i$, en la ecuación estructural el coeficiente de X sería ahora un promedio ponderado de ambos instrumentos. En este sentido, esperaríamos que el coeficiente de este nuevo modelo existiera acotado entre los dos coeficientes del modelo 3 y 4 de la tabla 1.

$$-7.961 \leq -4.018 \leq -2.612$$

$$\beta_{IVtwoboys} \leq \beta_{IV(twoboys,twogirls)} \leq \beta_{IVtwogirls}$$

donde

$$\beta_{IV(twoboys,twogirls)} = \omega \beta_{IVtwoboys} + (1 - \omega) \beta_{IVtwogirls}$$

y

$$\omega = \frac{\beta_{z1}Cov(X, Z_1)}{\beta_{z1}Cov(X, Z_1) + \beta_{z2}Cov(X, Z_2)}$$

B)

Si era de esperarse, pues la motivación de los covariates no es cambiar el LATE sino aumentar la eficiencia del estimador, en este sentido fue exitoso pues el error estandar se ha disminuido.

Table 8: Varios instrumentos

	(1)	(2)	(3)	(4)	(5)
Variable dependiente:	<i>weeksm1</i>	<i>weeksm1</i>	<i>weeksm1</i>	<i>weeksm1</i>	<i>weeksm1</i>
<i>kidcount</i>	-4.018*** (0.9506)	-3.554*** (0.4734)	-3.502*** (0.497)	-3.452*** (0.4712)	-3.538*** (0.4747)
<i>N</i>	254,654	254,654	254,654	-	-
<i>F</i> 1a etapa	464.098***	1776.748***	1563.088 ***	-	-
Método	IV	IV	IV	GME	GME
Instrumento	<i>twoboys</i> <i>twobgirls</i>	<i>samesex</i> <i>multi2nd</i>	<i>samesex</i> <i>multi2nd</i>	<i>twoboys</i> <i>twobgirls</i>	<i>samesex</i> <i>multi2nd</i>
Controles	Sí	Sí	No	Sí	Sí

Todas las regresiones controlan por *agem1*, *agef*, *stm*, *black*, *hispan*, *othrace*

Errores robustos entre parentesis

*p<0.1; **p<0.05; ***p<0.01

VIII.

A)

Considero que es mejor estimacion el modelo 4 que el 5 porque A) tiene menor desviación estandar y mas que nada porque B) esta mejor especificado. En particular este modelo 5 puede estar mal especificado porque el instrumento *multi2nd* es incorrecto para la variable endógena *kidcount* en el sentido de que el complier de *multi2nd* son aquellas mujeres que tendrían un cuarto o más hijos solo si tuvieron gemelos en el segundo y tercero, osea son mujeres que ya tuvieron un tercer hijo si o si y por ello el coeficiente es ligeramente más grande en el modelo 5. Este detalle es evidente en el inciso de las proporciones donde la proporción de never takers del modelo tal cual está especificado sería del 0%.

B)

El valor del coeficiente y su desviación estandar es muy similar, aunque GMM logró una menor desivación estandar. Creo que es razonable que esto pasara debido a que ambos enfoques están haciendo cálculos muy similares y además que GMM es una mejor alternativa para obtener los errores adecuados en presencia de heterocedasticidad, los errores de GMM en presencia de heterocedasticidad son incluso mejores que los errores robustos.

IX.

Table 9: Más especificaciones

	(1)	(2)	(3)
Variable dependiente:	<i>workedm</i>	-	-
<i>kidcount</i>	-0.099*** (0.02263)	-	-
<i>N</i>	254,654	-	-
<i>F</i> 1a etapa	-	-	-
Método	GMM	-	-
Instrumento	<i>samesex</i>	-	-
	-	-	-
Controles	Sí	-	-

Todas las regresiones controlan por *agem1*, *agef*, *stm*, *black*, *hispan*, *othrace*

Errores robustos entre parentesis

*p<0.1; **p<0.05; ***p<0.01

De la tabla anterior podemos inferir que el efecto de tener un hijo adicional afecta la probabilidad de haber trabajado para las mamás compliers en 9.9 puntos porcentuales. Las mamás compliers son aquellas que tuvieron un hijo adicional solo si tuvieron los dos primeros hijos del mismo sexo.

X.

En este segundo modelo usamos como IV una dicotomica que nos indica si la mujer trabajó o no y *samesex*, por lo tanto, nuestros compliers son mujeres que sí trabajaron y que si tuvieron más hijos si solo si los primeros dos fueron del mismo sexo. Notamos que el coeficiente de *kidcount* y de *hoursun* son positivos y estadísticamente significativos al 99%, lo que indica que por cada hijo y por cada hora de trabajo se va a ganar más dinero (por que se va a trabajar) sin embargo a este monto se le tiene que restar la interacción entre estas dos variables. El coeficiente de la interacción es negativo y estadísticamente significativo, lo cual es muy interesante y nos indica la existencia de un efecto negativo sobre el salario debido a las horas de trabajo con hijos, todo lo demas constante para las compliers. En resumen si hay un efecto negativo sobre el salario debido a la cantidad de hijos controlando incluso por el tiempo que trabajan las mujeres.

Table 10: Más especificaciones

	(1)	(2)	(3)
Variable dependiente:	<i>workedm</i>	<i>incomem</i>	-
<i>kidcount</i>	-0.099*** (0.02263)	1,679.921*** (581.6)	-
<i>hoursum</i>	-	584.005*** (38.6)	-
<i>kidcount:hoursum</i>	-	-88.538*** (14.2)	-
<i>N</i>	254,654	254,654	-
<i>F</i> 1a etapa	-	688.71	-
Método	GMM	IV	-
Instrumento	<i>samesex</i> <i>multi2nd</i>	<i>samesex</i> <i>work</i>	- -
Controles	Sí	Sí	-

Todas las regresiones controlan por *agem1*, *agef*, *stm*, *black*, *hispan*, *othrace*
 Errores robustos entre parentesis

work es una dicotomica que indica si la mujer trabajó

*p<0.1; **p<0.05; ***p<0.01

XI.

Finalmente, en este tercer modelo usamos como IV una dicotómica que nos indica si la mujer trabajó o no y *samesex*, por lo tanto, nuestros compliers son mujeres que sí trabajaron y que si tuvieron más hijos si solo si los primeros dos fueron del mismo sexo. Notamos que el coeficiente de *kidcount* y de *hispan* son negativos y estadísticamente significativos al 99%, lo que indica que por cada hijo y por el hecho de ser hispana, la persona va a trabajar más sin embargo a este monto se le tiene que sumar la interacción entre estas dos variables, la cual tiene el potencial de hacer la cantidad neta incluso positiva. El coeficiente de la interacción es positivo y estadísticamente significativo, lo cual nos indica la existencia de un efecto positivo sobre el la cantidad de horas trabajadas debido a la cantidad de hijos para las mujeres hispanas compliers, todo lo demas constante. En resumen sí hay un efecto posivo en la cantidad de horas trabajadas derivado de los hijos para las mujeres hispanas, aunque al final es incierto si la cantidad de horas disminuye, aumenta o se queda igual dependiendo las demas variables.

Table 11: Más especificaciones

	(1)	(2)	(3)
Variable dependiente:	<i>workedm</i>	<i>incomem</i>	<i>weeksm1</i>
<i>kidcount</i>	-0.099*** (0.02263)	1,679.921*** (581.6)	-138.346*** (1.5802)
<i>hoursum</i>	-	584.005*** (38.6)	-
<i>kidcount:hoursum</i>	-	-88.538*** (14.2)	-
<i>hispan</i>	-	-	-303.426*** (3.6575)
<i>kidcount:hispan</i>	-	-	122.310*** (1.4807)
<i>N</i>	254,654	254,654	254,654
<i>F</i> 1a etapa	-	688.71	4012.2
Método	GMM	IV	IV
Instrumento	<i>samesex</i> <i>multi2nd</i>	<i>samesex</i> <i>work</i>	<i>samesex</i> <i>work</i>
Controles	Sí	Sí	Sí

Todas las regresiones controlan por *agem1*, *agef*, *stm*, *black*, *hispan*, *othrace*

Errores robustos entre parentesis

work es una dicotómica que indica si la mujer trabajó

*p<0.1; **p<0.05; ***p<0.01