

Problem Set 4:

Variables instrumentales

ECONOMÍA APLICADA



Universidad de
San Andrés

Alumnos:

JUAN DIEGO BARNES

FRANSISCO LEGASPE

RODRIGO MARTIN

DIEGO FASAN

Profesor: MARTÍN ROSSI

Tutores: GASTÓN GARCÍA ZAVALETA

TOMÁS PACHECO

1. Chinese presence

Primero definimos a *Chinese presence* como una variable *dummy* que toma el valor uno si la población china en el municipio es mayor que cero. Además excluimos a los municipios que pertenecen al distrito de Ciudad de México para el análisis.

2. Estadística descriptivas

A continuación presentamos la tabla de estadísticas descriptivas de las variables presentes en el modelo. Se encuentran variables de presencia de carteles, población china y marginalización como las de mayor relevancia para el estudio, y otros controles como factores climáticos, geográficos, y de migración, que serán útiles para descartar otras posibles causas del desarrollo en cada localidad. Esta tabla replica la "Table 1" del paper, sin embargo, puede verse algunas diferencias en los valores obtenidos, lo cual está relacionado a diferencias con la base de datos utilizada en el paper.

Tabla 1: Summary statistics

	mean	sd	min	max
Cartel presence 2010	0.28	0.45	0.00	1.00
Cartel presence (2005)	0.09	0.29	0.00	1.00
Chinese presence	0.15	0.35	0.00	1.00
Chinese Population	4.43	36.84	0.00	1183.00
Marginalization	0.04	0.99	-2.23	5.03
Per capita tax revenue (USD)	135.63	281.98	0.01	6480.58
German presence	0.10	0.30	0.00	1.00
Poppy suitability	0.09	0.29	0.00	1.00
Distance to US (km)	745.36	266.26	6.69	1349.71
Distance to Mexico City (km)	477.78	374.48	0.00	2282.00
Distance to closest port (km)	948.22	267.18	0.00	1336.00
Head of state	0.01	0.11	0.00	1.00
Population in 2015	44,317.47	130,762.30	87.00	1,827,868.00
Surface (sq km)	829.69	2154.31	2.21	53,256.17
Average temperature (°C)	20.05	3.94	10.49	29.08
Average precipitation (mm)	90.41	51.77	6.45	338.36
Local population growth	0.23	1.68	-1.00	54.00
Population in 1930	2,662.07	7,639.55	0.00	179,336.00
Observations	2331			

3. Cartel presence

La siguiente tabla intenta comprobar la relevancia del instrumento para las distintas especificaciones de presencia de carteles. Muestra la relación parcial del instrumento con la variable a instrumentar, utilizando efectos fijos a nivel estado. Y incluyendo en las columnas (4) y (6) un set de controles adicionales. Los coeficientes de presencia china son significativos en los primeros 3 casos al 1 %, y al 5 % bajo la última especificación.

Tabla 2: Robustness Check: Alternative Definitions for Cartel Presence

	Cartel Presence 2010		Cartel Presence 2005	
	(3)	(4)	(5)	(6)
Chinese Presence	0.135*** (0.0242)	0.060*** (0.0204)	0.212*** (0.0348)	0.089** (0.0358)
State Dummies	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Clusters	30	29	30	29
Observations	2315	2041	2314	2040

Notes: Standard errors clustered at the state level are shown in parentheses. The unit of observation is a municipality. The unit of cluster is a state. All models are estimated using Ordinary Least Squares. The set of controls includes German presence, Poppy suitability, Average temperature, Average precipitation, Surface, Population in 1930, Distance to U.S., Distance to Mexico City, Distance to closest port, and Head of state. *Significant at the 10 % level. **Significant at the 5 % level. ***Significant at the 1 % level.

4. Cartel presence and marginalization

Tabla 3: IV estimates: cartel presence and marginalization.

	Dependent variable: Marginalization				
	(1)	(2)	(3)	(4)	(5)
Cartel Presence	-2.090*** (0.343)	-2.742*** (0.906)	-3.273** (0.724)	-2.768*** (1.332)	-2.733*** (0.918)
F-test	36.98 (0.000)	6.12 (0.020)	4.12 (0.05)	6.62 (0.026)	6.04 (0.021)
State dummies	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Clusters	30	29	29	28	29
Observations	2,439	2040	1,986	2022	2038

Notes: Standard errors clustered at the state level are shown in parentheses. The unit of observation is a municipality. The unit of cluster is a state. All models are estimated using Two Stages Least Squares. Cartel presence is instrumented using Chinese presence F-test is the F-test of excluded instruments (p-values are shown in parentheses). In Columns (2) to (4) the set of controls includes German presence, Poppy suitability, Average temperature, Average precipitation, Surface, Population in 1930, Distance to U.S., Distance to Mexico City, Distance to closest port, and Head of state. Column (5) further controls for Local population growth. Column (3) excludes municipalities located more than 100 km from U.S. border. Column (4) excludes municipalities located in the state of Sinaloa. *Significant at the 10 % level. **Significant at the 5 % level. ***Significant at the 1 % level.

En esta tabla vemos los resultados de regresar la presencia de carteles contra el índice de marginalización. Como la presencia de carteles es endógena, se instrumenta la variable con “presencia china en 1930” y se estima utilizando 2SLS. Podemos observar que en todos los casos existe un efecto negativo de la presencia de carteles sobre la marginalización, siendo este efecto significativo al 1 % en todos los casos, excepto bajo la especificación (3) donde es significativo al 5 %, diferencia probablemente justificada en el menor poder estadístico al contar con menos observaciones.

Además podemos observar que el test de exclusión F aplicado sobre la *first stage*, es significativo al 5 % en todos los casos, de modo que el instrumento se puede considerar relevante, dada la especificación adoptada de *cartel presence*.

La tabla 4 muestra la regresión de presencia de carteles contra variables asociadas a la marginalización. De la misma forma que en la tabla Tabla 2, se estima usando 2SLS. La mayoría de las variables se correlación negativamente con la presencia de carteles, es decir que estos reducen el analfabetismo (en promedio un 14,6 % significativo al 1 %), no tener el primario hecho (en promedio un 34,3 % significativo al 1 %), la falta de servicios en el hogar (a excepción del no acceso al agua que correlaciona positivamente aunque no es significativo) y correlación negativamente también con los bajos salarios (en promedio un 27,2 % y significativo al 5 %).

Tabla 4: Cartel presence and marginalization components.

	Illiteracy	Without primary	Without toilet	Without electricity	Without water	Overcrowding	Earthen floor	Low salary
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cartel presence	-15.21*** (3.555)	-39.53** (15.35)	-6.240 (4.836)	-5.171* (3.136)	2.708 (8.360)	-24.31** (10.21)	-7.718 (8.039)	-30.91** (12.80)
State dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	29	29	29	29	29	29	29	29
Observations	2,314	2,040	2,040	2,040	2,040	2,040	2,040	2,040

Notes: Standard errors clustered at the state level are shown in parentheses. The unit of observation is a municipality. The unit of cluster is a state. All models are estimated using Two Stages Least Squares. Cartel presence is instrumented using Chinese presence. The set of controls includes German presence, Poppy suitability, Average temperature, Average precipitation, Surface, Population, Distance to U.S., Distance to Mexico City, Distance to closest port, and Head of state. *Significant at the 10 % level. **Significant at the 5 % level. ***Significant at the 1 % level.

5. Exogeneidad

La exogeneidad del instrumento es nuestro supuesto de identificación, por lo que no lo podemos testear. Si supieran que el instrumento es exógeno, o suponiendo que esto es cierto, podríamos testear la exogeneidad de la presencia del cartel mediante el test de Hausman.

El test compara los beta del método de variables instrumentales, lo cuales son consistentes (Dado que supusimos la exogeneidad del instrumento), y los coeficientes estimados por MCO los cuales son consistentes y de varianza mínima si la presencia de carteles es exógena. Bajo esta lógica el test de Hausman se define como:

$$H = (\beta^{OLS} - \beta^{IV})[V^{IV}(\beta) - V^{OLS}(\beta)](\beta^{OLS} - \beta^{IV}) \sim \chi^2$$

En nuestro caso el estadístico toma un valor de $H = 6,12$, lo que corresponde a un p-valor igual a 1, por lo que no podemos rechazar la hipótesis nula de que el estimador de MCO es consistente. Hay que considerar que estamos testeando algo con la hipótesis nula a favor, de modo que no rechazar no implica que esta sea cierta.

Alternativamente realizamos el test mediante el comando *estat endogenous*, donde la hipótesis nula es la exogeneidad de la variable. En este caso si se encuentra evidencia significativa para rechazar la hipótesis nula de que la presencia de carteles es exógena, con un p-valor de 0.003. De modo que podemos ver que el test de Hausman podría llevarnos a cometer un error si no somos rigurosos con la interpretación estadística del test.

6. Test de sobreidentificación (Sargan o J-test).

Partiendo de la especificación básica, descrita en (1), agregamos la variable “german presence” como instrumento. Luego realizamos el test para sobre identificación de Sargan, el cual arroja un p-valor de 0.36 por lo que no podemos rechazar la hipótesis nula de los instrumentos en conjunto son exógenos. Debemos considerar que estamos realizando un test de hipótesis, con la hipótesis nula a favor por lo que el no rechazarla no nos permite asegurar que esta es cierta bajo ningún nivel de confianza.

7. Crítica

Podemos discutir la exogeneidad del instrumento, a partir de pensar que existen factores en el término de error que afectan la presencia de Chinos en 1930 por estado. La elección de los inmigrantes chinos sobre donde ubicarse puede estar correlacionándose con el instrumento. Los autores intentan captar este efecto utilizando la población de inmigrantes alemanes, como proxy de características que afectan a la decisión de ubicación de los inmigrantes. Pero podemos pensar que las preferencias de los chinos no necesariamente es igual a la de los alemanes, por lo que esta variable solo estaría controlando parte de la endogeneidad pero no eliminándola.

El argumento principal es que la población china sufre otras características como inmigrante diferentes a las de la población alemana. Como detalla el paper, hay evidencia que muestra que durante esos años existió una corriente discriminatoria contra la población china, más conocida como “sinofobia”. Lo que argumentamos es un posible reordenamiento de la población china hacia estados/ciudades más amigables con el origen de esta inmigración. Dado que la variable Presencia de Chinos solo toman el año 1930, creemos que pudo haber desplazamientos internos años previos, justificadas en el grado de discriminación de los municipios. Más precisamente desde 1882 que es el año que el trabajo destaca como el inicio de una fuerte ola migratoria a México, de manera exógena a partir de el Acta de Exclusión China celebrada en ese año en Estados Unidos.

En resumen, si consideramos que la discriminación puede afectar a la marginalización (fragmenta la sociedad, impide la cooperación y, podría generar violencia y desigualdad), entonces es posible pensar en que la discriminación hacia los chinos (sinofobia), incluida en el termino de error, está correlacionada con el instrumento. Por este motivo, la presencia china en un estado es posiblemente endógena.

Apéndice. Código PS4 (STATA)

2023-08-17

```

/*****
Semana 5 y 6: Fuentes de sesgo e imprecisión

Universidad de San Andrés
Economía Aplicada
*****/

Barnes, Fasan, Legaspe y Martin
/*****
Este archivo sigue la siguiente estructura:

*****/

* 0) Set up environment
*=====*/
clear all
global main "C:\Users\Usuario\Desktop\MAESTRIA\Economia Aplicada\TPs\PS4"
global input "$main/input"
global output "$main/output"

cd "$main"

* Open data

use "$input/poppy.dta", clear
*=====
* 1) Generen la variable "Chinese presence" tal como lo especifica el paper.
gen chinese_presence = .
replace chinese_presence = 1 if chinos1930hoy > 0 & chinos1930hoy !=.
replace chinese_presence = 0 if chinos1930hoy == 0

*=====
* 2) Estadística descriptiva de las variables relevantes (como en la Tabla 1)
drop if estado == "México"

* Generamos labels para hacerla mas facil de leer
label var cartel2010 "Cartel presence (2010)"
label var cartel2005 "Cartel presence (2005)"
label var chinese_presence "Chinese presence"
label var chinos1930hoy "Chinese Population"
label var IM_2015 "Marginalization"
label var Impuestos_ "Per capita tax revenue"
label var dalemanes "German presence"

```

```

label var capestado "Head of state"
label var pob1930cabec "Population 1930"
label var cartel2010 "Cartel presence 2010"
label var mindistcosta "Distance to closest port"
label var capestado "Head of state"
label var growthperc "Local population growth"

* Armamos una unica tabla de descriptivas
estpost summarize cartel2010 cartel2005 chinese_p chinos1930 IM_2015 Impuestos_
dalemanes tempopium distancia_km distkmDF mindistcosta capestado POB_TOT_2015
superficie_km TempMed PrecipAnual growthperc pob1930cabec

esttab using "$output/Table 1.tex", cells("mean sd min max") ///
nomtitle nonumber replace label

*=====
*3) Replicamos las regresiones hechas por OLS de las columnas 3 a 6 de la Tabla5.

* Columna 3
regress cartel2005 chinese_presence i.id_estado, cluster(id_estado)
outreg2 using "$output/Table 2.rtf", replace label

* Columna 4
regress cartel2005 chinese_presence dalemanes tempopium TempMed PrecipAnual
superficie_km pob1930cabec distancia_km distkmDF mindistcosta capestado
i.id_estado, cluster(id_estado)

outreg2 using "$output/Table 2.rtf", append label

* Columna 5
regress cartel2010 chinese_presence i.id_estado, cluster(id_estado)
outreg2 using "$output/Table 2.rtf", append label

* Columna 6
regress cartel2010 chinese_presence dalemanes tempopium TempMed PrecipAnual
superficie_km pob1930cabec distancia_km distkmDF mindistcosta capestado
i.id_estado, cluster(id_estado)

outreg2 using "$output/Table 2.rtf", append label

*=====
*4)Reproducimos las Tabla 7 y 8 utilizando la presencia del cartel en 2010"
*-----*
* TABLA 7
*-----*

*col 1
ivregress 2sls IM_2015 i.id_estado (cartel2010=chinese_presence), cluster(id_estado)
outreg2 using "$output/Table 3.rtf", replace label
estat firststage

```

```

*col 2
ivregress 2sls IM_2015 dalemanes tempopium TempMed PrecipAnnual superficie_km
pob1930cabec distancia_km distkmDF mindistcosta capestado
i.id_estado (cartel2010=chinese_presence), cluster(id_estado)

outreg2 using "$output/Table 3.rtf", append label
estat firststage

*col 3
preserve
drop if distancia_km < 100
ivregress 2sls IM_2015 dalemanes tempopium TempMed PrecipAnnual superficie_km
pob1930cabec distancia_km distkmDF mindistcosta capestado
i.id_estado (cartel2010=chinese_presence), cluster(id_estado)

outreg2 using "$output/Table 3.rtf", append label
estat firststage

restore

*col 4
preserve
drop if estado == "Sinaloa"
ivregress 2sls IM_2015 dalemanes tempopium TempMed PrecipAnnual superficie_km
pob1930cabec distancia_km distkmDF mindistcosta capestado
i.id_estado (cartel2010=chinese_presence), cluster(id_estado)

outreg2 using "$output/Table 3.rtf", append label
estat firststage

restore

*col 5
ivregress 2sls IM_2015 dalemanes tempopium TempMed PrecipAnnual superficie_km
pob1930cabec distancia_km distkmDF mindistcosta capestado growthperc
i.id_estado (cartel2010=chinese_presence), cluster(id_estado)

outreg2 using "$output/Table 3.rtf", append label
estat firststage

*-----*
* TABAL 8
*-----*

*col 1
ivregress 2sls ANALF_2015 i.id_estado (cartel2010=chinese_presence),
cluster(id_estado)

outreg2 using "$output/Table 4.rtf", replace label

```



```

*col 2
ivregress 2sls SPRIM_2015 dalemanes tempopium TempMed PrecipAnual
superficie_km pob1930cabec distancia_km distkmDF mindistcosta capestado
i.id_estado (cartel2010=chinese_presence), cluster(id_estado)

outreg2 using "$output/Table 4.rtf", append label

*col 3
ivregress 2sls OVSDE_2015 dalemanes tempopium TempMed PrecipAnual superficie_km
pob1930cabec distancia_km distkmDF mindistcosta capestado
i.id_estado (cartel2010=chinese_presence), cluster(id_estado)

outreg2 using "$output/Table 4.rtf", append label

*col 4
ivregress 2sls OVSEE_2015 dalemanes tempopium TempMed PrecipAnual superficie_km
pob1930cabec distancia_km distkmDF mindistcosta capestado
i.id_estado (cartel2010=chinese_presence), cluster(id_estado)

outreg2 using "$output/Table 4.rtf", append label

*col 5
ivregress 2sls OVSAE_2015 dalemanes tempopium TempMed PrecipAnual superficie_km
pob1930cabec distancia_km distkmDF mindistcosta capestado
i.id_estado (cartel2010=chinese_presence), cluster(id_estado)

outreg2 using "$output/Table 4.rtf", append label

*col 6
ivregress 2sls VHAC_2015 dalemanes tempopium TempMed PrecipAnual superficie_km
pob1930cabec distancia_km distkmDF mindistcosta capestado
i.id_estado (cartel2010=chinese_presence), cluster(id_estado)

outreg2 using "$output/Table 4.rtf", append label

*col 7
ivregress 2sls OVPT_2015 dalemanes tempopium TempMed PrecipAnual
pob1930cabec distancia_km distkmDF mindistcosta capestado
i.id_estado (cartel2010=chinese_presence), cluster(id_estado)

outreg2 using "$output/Table 4.rtf", append label

*col 8
ivregress 2sls P02SM_2015 dalemanes tempopium TempMed PrecipAnual superficie_km
pob1930cabec distancia_km distkmDF mindistcosta capestado
i.id_estado (cartel2010=chinese_presence), cluster(id_estado)

outreg2 using "$output/Table 4.rtf", append label

*=====
*5) Testear Exogeneidad

*-----*
* Exogeneidad de la presencia del cartel (Hausman Test)

```

```

-----*
* regression IV
ivregress 2sls IM_2015 dalemanes tempopium TempMed PrecipAnnual superficie_km
pob1930cabec distancia_km distkmDF mindistcosta capestado
i.id_estado (cartel2010 = chinese_presence)

est store iv

* regression OLS
reg IM_2015 cartel2010 dalemanes tempopium TempMed PrecipAnnual superficie_km
pob1930cabec distancia_km distkmDF mindistcosta capestado i.id_estado

est store ols

* Hausman Test
hausman iv ols

* Alternativa
ivregress 2sls IM_2015 dalemanes tempopium TempMed PrecipAnnual
superficie_km pob1930cabec distancia_km distkmDF mindistcosta
capestado i.id_estado (cartel2010 = chinese_presence), cluster(id_estado)

estat endogenous

/* The null hypothesis is that there is not systematic difference between
the estimates.
If you reject the null hypothesis, the coefficients are different so you should
use IV, education is endogenous. If you don't reject the null hypothesis, you
should use OLS, because being the coefficients consistent, ols is the most
efficient. */

=====
*6) Sargan Test

* Estimate by IV and predict the residuals
ivregress 2sls IM_2015 tempopium TempMed PrecipAnnual superficie_km pob1930cabec
distancia_km distkmDF mindistcosta capestado
i.id_estado (cartel2010 = chinese_presence dalemanes), cluster(id_estado)

predict resid, residual

* Regress the residuals on all the exogenous variables (instruments and controls)
reg resid chinese_presence dalemanes tempopium TempMed PrecipAnnual superficie_km
pob1930cabec distancia_km distkmDF mindistcosta capestado
i.id_estado, cluster(id_estado)

* Obtain the R2 and use it to compute the statistic S=nR2
ereturn list
display chi2tail(1,e(N)*e(r2))

```

```
/* The null hypothesis is that all moment conditions are valid. If the test is
rejected, you cannot determine which the invalid moment conditions are.
In this case you cannot reject the null hypothesis: the instruments
are exogenous*/
```