

---

Economía Aplicada: Problem Set N<sup>o</sup>4

---

Milton Bronstein      Felipe García Vassallo      Santiago López      Franco Riottini

## Ejercicio 2

En la Tabla 1 mostramos las estadísticas descriptivas (media, desvío estándar, mínimo y máximo) de las variables en la base de datos. Vemos que hubo un aumento empírico significativo de la media de presencia de carteles en distritos. En sólo 5 años, el 9 % de los distritos había presencia de carteles, mientras que en 2010, la media aumenta al 30 %. En cuanto a la presencia de inmigrantes, vemos que en 1930 se registraron inmigrantes chinos en el 15 % de los municipios observados, aunque la media de inmigrantes chinos viviendo en un municipio mexicano en 1930 es bastante pequeña, con menos de 5 personas en promedio. También se observa que, en promedio, se registró presencia de inmigrantes alemanes en el 10 % de las jurisdicciones en la base de datos.

Tabla 1 - Estadística Descriptiva

	Mean	SD	Min	Max
Cartel presence 2010	0.30	0.46	0.00	1.00
Cartel presence 2005	0.09	0.29	0.00	1.00
Chinese presence	0.15	0.36	0.00	1.00
Chinese population	4.73	38.92	0.00	1183.00
Marginalization	-0.06	0.95	-2.22	5.03
Per capita tax revenue	137.28	272.41	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 U.S. (km)	735.48	255.01	7.09	1337.27
Distance to Mexico City (km)	463.19	388.30	16.00	2282.00
Distance to closest port	944.69	272.52	0.00	1336.00
Head of state	0.02	0.12	0.00	1.00
Population in 2015	49413.56	136577.71	132.00	1677678.00
Surface	804.25	2056.66	2.21	53256.17
Average temperature	19.86	4.08	10.49	29.08
Average precipitation	87.69	49.03	6.45	338.36
Population in 1930	2777.60	7765.62	0.00	179336.00

## Ejercicio 3

En este ejercicio replicamos las columnas 3-6 de la Tabla 5 del paper. En las columnas 1 y 3 de nuestra Tabla 2 regresamos la presencia de carteles en municipios en 2005 y en 2010 solamente usando como variable explicativa la presencia de inmigrantes chinos en 1930, además de dummies por Estado. Encontramos resultados casi idénticos a los del paper tanto en magnitud como en significatividad estadística. En las columnas 2 y 4

corremos las mismas regresiones, pero utilizando el conjunto de controles provistos en la base. En este caso el coeficiente asociado a nuestra variable de interés tiene exactamente la misma magnitud que en el paper, pero la significatividad estadística de nuestros coeficientes es al 5 %, mientras que en el paper es al 1 %. Esto se puede deber a que no contamos con las mismas variables.

Tabla 2 - Regresiones por OLS

	(1)	(2)	(3)	(4)
	Cartel presence 2010	Cartel presence 2010	Cartel presence 2005	Cartel presence 2005
Chinese presence	0.216*** (6.27)	0.101** (2.78)	0.130*** (5.37)	0.0589** (2.96)
State dummies	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Clusters	31	30	31	30
Observations	2439	2161	2440	2162

*t* statistics in parentheses

\*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$

## Ejercicio 4

En la Tabla 3 replicamos la tabla 7 del paper incluyendo dummies por Estado. En esta tabla obtenemos resultados bastante distintos de los que se muestran en el paper. Si bien en las 5 especificaciones la presencia de carteles tiene un efecto negativo sobre el índice de marginalización, los coeficientes tienen distintas magnitudes (sin una dirección clara sobre las diferencias), y además en nuestro caso la tercera especificación (en la que se excluyen los distritos más de 100 kilómetros de la frontera con Estados Unidos) no es estadísticamente significativa, mientras que en el paper sí lo es. Por otra parte, con la excepción de la primera especificación, los niveles de significatividad estadística de nuestras regresiones son más bajos.

Tabla 3 - Regresiones por IV - Marginalización

	(1)	(2)	(3)	(4)	(5)
	Marginalization	Marginalization	Marginalization	Marginalization	Marginalization
Cartel presence 2010	-2.002*** (-5.93)	-2.388** (-3.06)	-0.850 (-1.17)	-2.420** (-3.14)	-2.382** (-3.02)
F-test	39.38 (0.000)	7.74 (0.000)	0.65 (0.000)	8.27 (0.000)	7.64 (0.000)
State dummies	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes
Clusters	31	30	5	29	30
Observations	2439	2161	54	2143	2159

*t* statistics in parentheses

\*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$

Luego, en la Tabla 4 replicamos la Tabla 8 del paper, en la que se regresan los distintos componentes del índice de marginalización sobre nuestra variable de interés (la presencia de carteles en 2010) y los controles y dummies por Estado. En este caso nos encontramos nuevamente que nuestros coeficientes estimados para el analfabetismo y la población sin educación primaria son más altos que los estimados en el paper. Por otro lado, en nuestra especificación la presencia de carteles no es significativa para explicar municipios sin electricidad, mientras que en el paper es significativa al 10 %, y también con signo negativo. Como en el paper, encontramos que una mayor presencia de carteles está correlacionada con un menor hacinamiento, aunque en nuestro caso si bien nuestro coeficiente es de mayor magnitud, la significatividad es más baja. Como en el paper, la presencia de carteles no tiene poder para explicar la cantidad de personas que viven en casas con pisos de tierra. Por último, la presencia de carteles está asociada a un menor porcentaje de la población viviendo en localidades pequeñas, así como a una menor cantidad de población que gana menos de 2 salarios mínimos, tal como sucede en el paper. Nuevamente, nuestros coeficientes tienen una mayor magnitud, y son menos estadísticamente significativos que en el paper.

Tabla 4 - Regresiones por IV - Componentes de la marginalización

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Illiteracy	Without primary	Without toilet	Without electricity	Without water	Overcrowding	Earthen floor	Small localities	Low salary
Cartel presence 2010	-21.77*	-34.34**	-6.100	-4.595	3.271	-19.99*	-7.041	-100.7*	-27.20*
	(-2.44)	(-2.70)	(-1.45)	(-1.73)	(0.45)	(-2.15)	(-1.01)	(-2.58)	(-2.51)
State dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	30	30	30	30	30	30	30	30	30
Observations	2161	2161	2161	2161	2161	2161	2161	2161	2161

*t* statistics in parentheses

\*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$

## Ejercicio 5

La exogeneidad, tanto en variables instrumentales como en mínimos cuadrados ordinarios, no es testeable, sino que es el supuesto crucial que debemos hacer en inferencia causal. Sin embargo, realizamos el test de Hausman y obtenemos que no podemos rechazar la hipótesis nula (obtenemos un p-valor de 1) de que no hay diferencias sistemáticas entre las estimaciones por IV y OLS. Este test consiste en primero hacer las estimaciones del modelo para variables instrumentales y mínimos cuadrados ordinarios. Luego, se comparan coeficiente a coeficiente para observar las diferencias existentes entre ellos. Tanto bajo hipótesis nula como bajo hipótesis alternativa las estimaciones realizadas por IV y OLS son consistentes. Así como la estimación realizada por mínimos cuadrados ordinarios es eficiente. Dado que no ha podido rechazarse, según el resultado del test de Hausman, sería conveniente realizar estimaciones por OLS porque son más eficientes sin dejar de ser consistentes.

## Ejercicio 6

En el test de sobreidentificación utilizamos como instrumentos para la presencia de cartel en 2010 la presencia de inmigrantes alemanes y chinos en 1930. Nos quedamos con los residuos, los regresamos sobre las variables explicativas y los instrumentos. El resultado del test es 0,445 por lo que habría poca evidencia rechazar la hipótesis nula de que todos los instrumentos son válidos. En este test, bajo hipótesis nula, ambos instrumentos (las presencias de inmigrantes chinos y alemanes) son variables exógenas.

## Ejercicio 7

Como resaltan los autores, "una proporción de los chinos que ingresaban a México probablemente se sintieron atraídos por las oportunidades económicas locales, pero otros simplemente fueron desviados hacia el sur del Río Grande por la Ley de Exclusión Estadounidense." La principal crítica que podemos hacerle al argumento de exogeneidad es que los Chinos no se ubicaron de manera aleatoria en los distintos municipios mexicanos. Sea por mejores condiciones económicas o por mayores posibilidades para poder traficar inmigrantes ilegales a través de la frontera, los Chinos que se radicaron en México post-ley de exclusión, pueden haberse autoseleccionado en ciertos municipios. La forma que tienen los autores de controlar por condiciones económicas (el tamaño de la población en 1930) no nos parece suficiente para depurar a la presencia de inmigrantes chinos de su endogeneidad.

Por otro lado, utilizar la presencia de inmigrantes alemanes como control puede tampoco ser del todo útil para el argumento de exogeneidad. Podríamos pensar que la expresión de los lugares de asentamiento de estas dos poblaciones no diluye que los inmigrantes chinos eligieran las áreas más prósperas, sino que podría ser simplemente una cuestión de preferencias reveladas por características inobservables para los investigadores.

Todo esto implica que características socioeconómicas difíciles de observar a principios del Siglo XX que pueden haber guiado de forma no aleatoria la presencia de inmigrantes chinos esté sesgando los resultados positivos que muestra la presencia de carteles sobre outcomes socioeconómicos actuales.



```

/*****
*
*                               Semana 5: Variables Instrumentales
*
*                               Universidad de San Andrés
*                               Economía Aplicada
*                               2022
*
>
*****/
*                               Bronstein           García Vassallo           López           Riottini
/*****/
Este archivo sigue la siguiente estructura:

0) Configurar el entorno

1) Generar variable "Chinese presence"

2) Estadística descriptiva

3) Replicación de regresiones Tabla 5

4) Replicación de Tabla 7 y 8

5) Testear exogenenidad del instrumento

6) Test de sobreidentificación

*****/

* 0) Configurar el entorno
*=====

global main "C:\Users\felip\Documents\UdeSA\Maestría\Aplicada\Problem-Sets\PS 4"
global input "$main/input"
global output "$main/output"

cd "$main"

use "$input/poppy.dta", clear

* 1) Variable "Chinese presence"
*=====

gen chinese_pres= chinos1930hoy>0
replace chinese_pres=. if chinos1930hoy==.

* 2) Estadística descriptiva
*=====

* Primero pongo bien las labels

label variable cartel2005 "Cartel presence 2005"
label variable cartel2010 "Cartel presence 2010"
label variable chinese_pres "Chinese presence"
label variable suitability "Poppy suitability"
label variable distancia_km "Distance to U.S. (km)"
label variable distkmDF "Distance to Mexico City (km)"
label variable mindistcosta "Distance to closest port"
label variable Impuestos_pc_mun "Per capita tax revenue"
label variable chinos1930hoy "Chinese population"
label variable pob1930cabec "Population in 1930"
label variable capeestado "Head of state"

* Ahora hago el summ de las variables de interes

estpost summarize cartel2010 cartel2005 chinese_pres chinos1930hoy IM_2015 Impuestos_p
> c_mun dalemanes suitability distancia_km distkmDF mindistcosta capeestado POB_TOT_201
> 5 superficie_km TempMed Anual PrecipAnual med pob1930cabec, listwise
esttab using "$output/Table 1.tex", cells("mean(fmt(2)) sd(fmt(2)) min max") ///
collabels("Mean" "SD" "Min" "Max") nomtitle nonumber replace label

```

```
* Dropeo las obs que sean de distrito federal
drop if estado=="Distrito Federal"
```

```
* 3) Replicación de regresiones
```

```
*=====*
```

```
reg cartel2010 chinese_pres i.id_estado, cluster(id_estado)
est store ols1
```

```
reg cartel2010 chinese_pres dalemanes suitability TempMed_Anual PrecipAnual_med superf
> icie_km pobl930cabec distancia_km distkmDF mindistcosta capestado i.id_estado, clust
> er(id_estado)
est store ols2
```

```
reg cartel2005 chinese_pres i.id_estado, cluster(id_estado)
est store ols3
```

```
reg cartel2005 chinese_pres dalemanes suitability TempMed_Anual PrecipAnual_med superf
> icie_km pobl930cabec distancia_km distkmDF mindistcosta capestado i.id_estado, clust
> er(id_estado)
est store ols4
```

```
* Exporto tabla a Latex
```

```
esttab ols1 ols2 ols3 ols4 using "$output/EJ_3.tex", replace label keep(chinese_pres)
```

```
* 4) Replicación de Tabla 7
```

```
*=====*
```

```
ivregress 2sls IM_2015 (cartel2010=chinese_pres) i.id_estado, cluster(id_estado)
est store iv1
qui testparm*
estadd scalar p_value = r(p)
estat firststage
mat fstat = r(singleresults)
estadd scalar fs = fstat[1,4]
```

```
ivregress 2sls IM_2015 (cartel2010=chinese_pres) dalemanes suitability TempMed_Anual P
> recipAnual_med superficie_km pobl930cabec distancia_km distkmDF mindistcosta capesta
> do i.id_estado, cluster(id_estado)
est store iv2
qui testparm*
estadd scalar p_value = r(p)
estat firststage
mat fstat = r(singleresults)
estadd scalar fs = fstat[1,4]
```

```
ivregress 2sls IM_2015 (cartel2010=chinese_pres) dalemanes suitability TempMed_Anual P
> recipAnual_med superficie_km pobl930cabec distancia_km distkmDF mindistcosta capesta
> do i.id_estado if distancia_km<100, cluster(id_estado)
est store iv3
qui testparm*
estadd scalar p_value = r(p)
estat firststage
mat fstat = r(singleresults)
estadd scalar fs = fstat[1,4]
```

```
ivregress 2sls IM_2015 (cartel2010=chinese_pres) dalemanes suitability TempMed_Anual P
> recipAnual_med superficie_km pobl930cabec distancia_km distkmDF mindistcosta capesta
> do i.id_estado if estado!="Sinaloa", cluster(id_estado)
est store iv4
qui testparm*
estadd scalar p_value = r(p)
estat firststage
mat fstat = r(singleresults)
estadd scalar fs = fstat[1,4]
```

```

ivregress 2sls IM_2015 (cartel2010=chinese_pres) dalemanes suitability TempMed_Anual P
> recipAnual_med superficie_km pob1930cabec distancia_km distkmDF mindistcosta capesta
> do growthperc i.id_estado, cluster(id_estado)
est store iv5
qui testparm*
estadd scalar p_value = r(p)
estat firststage
mat fstat = r(singleresults)
estadd scalar fs = fstat[1,4]

*Exporto tabla a Latex
esttab iv1 iv2 iv3 iv4 iv5 using "$output/EJ 4.a.tex", replace label keep(cartel2010)
> scalar(F p_value) stats(fs p_value N, fmt(2 3 0))

* Replicación Tabla 8

ivregress 2sls ANALF_2015 (cartel2010=chinese_pres) dalemanes suitability TempMed_Anua
> l PrecipAnual_med superficie_km pob1930cabec distancia_km distkmDF mindistcosta cape
> stado i.id_estado, cluster(id_estado)
est store iv6

ivregress 2sls SPRIM_2015 (cartel2010=chinese_pres) dalemanes suitability TempMed_Anua
> l PrecipAnual_med superficie_km pob1930cabec distancia_km distkmDF mindistcosta cape
> stado i.id_estado, cluster(id_estado)
est store iv7

ivregress 2sls OVSDE_2015 (cartel2010=chinese_pres) dalemanes suitability TempMed_Anua
> l PrecipAnual_med superficie_km pob1930cabec distancia_km distkmDF mindistcosta cape
> stado i.id_estado, cluster(id_estado)
est store iv8

ivregress 2sls OVSEE_2015 (cartel2010=chinese_pres) dalemanes suitability TempMed_Anua
> l PrecipAnual_med superficie_km pob1930cabec distancia_km distkmDF mindistcosta cape
> stado i.id_estado, cluster(id_estado)
est store iv9

ivregress 2sls OVSAE_2015 (cartel2010=chinese_pres) dalemanes suitability TempMed_Anua
> l PrecipAnual_med superficie_km pob1930cabec distancia_km distkmDF mindistcosta cape
> stado i.id_estado, cluster(id_estado)
est store iv10

ivregress 2sls VHAC_2015 (cartel2010=chinese_pres) dalemanes suitability TempMed_Anual
> PrecipAnual_med superficie_km pob1930cabec distancia_km distkmDF mindistcosta capes
> tado i.id_estado, cluster(id_estado)
est store iv11

ivregress 2sls OVPT_2015 (cartel2010=chinese_pres) dalemanes suitability TempMed_Anual
> PrecipAnual_med superficie_km pob1930cabec distancia_km distkmDF mindistcosta capes
> tado i.id_estado, cluster(id_estado)
est store iv12

ivregress 2sls PL5000_2015 (cartel2010=chinese_pres) dalemanes suitability TempMed_Anua
> al PrecipAnual_med superficie_km pob1930cabec distancia_km distkmDF mindistcosta cap
> estado i.id_estado, cluster(id_estado)
est store iv13

ivregress 2sls PO2SM_2015 (cartel2010=chinese_pres) dalemanes suitability TempMed_Anua
> l PrecipAnual_med superficie_km pob1930cabec distancia_km distkmDF mindistcosta cape
> stado i.id_estado, cluster(id_estado)
est store iv14

*Exporto tabla a Latex
esttab iv6 iv7 iv8 iv9 iv10 iv11 iv12 iv13 iv14 using "$output/EJ_4.b.tex", replace la
> bel keep(cartel2010)

* 5) Testear exogenidad del instrumento (test de Hausman)
*=====*
```



```

*Estimaciones por IV
ivregress 2sls IM_2015 (cartel2010=chinese_pres) dalemanes suitability TempMed_Anuar P
> recipAnual_med superficie_km pob1930cabec distancia_km distkmDF mindistcosta capesta
> do i.id_estado
est store ivh

*Estimaciones por OLS
reg IM_2015 cartel2010 dalemanes suitability TempMed_Anuar PrecipAnual_med superficie_
> km pob1930cabec distancia_km distkmDF mindistcosta capeestado i.id_estado
est store olsh

*Test de Hausman (bajo hipótesis nula IV es consistente)
hausman ivh olsh

* 6) Test de sobreidentificación
*=====

ivregress 2sls IM_2015 (cartel2010=chinese_pres dalemanes) suitability TempMed_Anuar P
> recipAnual_med superficie_km pob1930cabec distancia_km distkmDF mindistcosta capesta
> do i.id_estado, cluster(id_estado)
predict resid1, residual

reg resid1 suitability TempMed_Anuar PrecipAnual_med superficie_km pob1930cabec distan
> cia_km distkmDF mindistcosta capeestado chinese_pres dalemanes i.id_estado, cluster(i
> d_estado)

test chinese_pres=dalemanes=0

return list
ereturn list

display chi2tail(1,2*r(F))

ivreg2 IM_2015 (cartel2010=chinese_pres dalemanes) suitability TempMed_Anuar PrecipAnu
> al_med superficie_km pob1930cabec distancia_km distkmDF mindistcosta capeestado, clus
> ter(id_estado)

*Exportar do-file a pdf
translate "$main/programs/PS 4.do" "$output/PS 4 do-file.pdf", translator(txt2pdf) rep
> lace

```