

VI: The Slave Trade and the Origins of Mistrust in Africa

1. Breve descripción de la base de datos

Esta base de datos incluye las respuestas de un survey a la confianza de las personas Africanas, sus características principales y de sus grupos étnicos. Las variables principales son:

- **trust_neighbors**: que indica el nivel de confianza (entre 0 y 3) que la persona entrevistada tiene en sus vecinos
- **ln_export_area**: indica las exportaciones historicas del grupo etnico de la persona entrevistada (en log)
- **distsea**: es la distancia histórica del grupo etnico de la persona entrevistada al litoral (en 1000 kms). Se nota que hay grande variación

```
. sum trust_neighbors ln_export_area distsea
```

Variable	Obs	Mean	Std. dev.	Min	Max
trust_neig~s	20,698	1.738477	1.010104	0	3
ln_export_~a	21,702	.5346949	.9436631	0	3.65603
distsea	21,702	439.892	311.4725	1.249979	1252.683

2. Regresión de impacto sin VI

Cada esclavo negociado/km² reduce en 10.5% (Coefficient: -0.1050591) la confianza hacia los demas. Para un p-value de 0.000 este es un resultado significativo, para un nivel de significancia de 5%. Además, algunas variables controladas presentan un resultado no significativo (p-value > 0.05), por lo que se considera que no tienen impacto en el resultado obtenido. Para este caso la metodología OLS puede no ser una buena metodología para evaluar la causalidad porque puede existir un sesgo de endogeneidad. Si la variable independiente es una variable endógena, o sea, la variable independiente afecta la variable dependiente y la dependiente afecta la independiente, se dice que existe un sesgo de endogeneidad. De este modo, como la variable independiente está correlacionada con el error, el efecto de la variable independiente en la variable dependiente está sesgado. Como la exportación de esclavos viola uno de los derechos básicos de la humanidad, esta afecta la confianza que las personas tienen por los demas ya que no les proporcionan unas condiciones de vida dignas. Sin embargo, la confianza o desconfianza entre los demas tambien podría estar influyendo en el comercio de esclavos, ya que con la desconfianza, la gente está más alerta y se protege de los posibles abusos/tráfico que podría sufrir, y además al alejar a la gente, podría perjudicar la creación de un negocio de comercio de esclavos. Así, que la confianza también puede estar influyendo en el comercio de esclavos y creando una endogeneidad que hace que la regresión OLS esté sesgada y no determine la causalidad.

	trust_neighbors	Coefficient	Std. err.	t	P> t	[95% conf. interval]
	ln_export_area	-.1050591	.0112409	-9.35	0.000	-.1270923 - .0830259

3. Análisis del instrumento propuesto

El método de la variable instrumental se basa en alguna fuente de variación externa (variable exógena) para determinar el estado del tratamiento y resolver el sesgo de endogeneidad. Así que hay dos supuestos que se necesitan de verifícas:

- **Supuesto de relevancia:** Una variable instrumental (Z) debe influir en un solo estimador, es decir, la variable instrumentada (X), pero estando fuera del control del participante/ empresa y no está relacionada con sus otras características. Es decir, la variable instrumental debe estimar la variable instrumentada. Este es lo supuesto de más importancia
- **Supuesto de exclusión:** Una variable instrumental sólo puede estar relacionada con la variable dependiente a través de la variable instrumentada, no estando relacionada de forma directa con Y (o sea con el error de la regresión)

En la situación propuesta, intuitivamente se verifica el supuesto de relevancia. Por un lado, cuanto menor sea la distancia al litoral (suponiendo que el litoral es donde se ubica el mercado) los costos asociados al comercio serán menores también, lo que aumentaría la oferta y comercio del grupo étnico. Por otro lado, la mayor exposición también puede aumentar la procura por ese grupo étnico. Ambos mecanismos hacen aumentar la cantidad de esclavos comercializados.

Cuanto al supuesto de exclusión, nos es tan claro que la distancia y la confianza son totalmente exógenas. Por ejemplo, la distancia al litoral (y aislamiento consecuente) puede relacionarse con el sentido de comunidad y confianza social del grupo étnico por las generaciones siguientes. Grupos más lejos del litoral (de donde se desarrolla, normalmente, más el comercio y la sociedad), podrían crear mayores lazos de confianza. Sin embargo, la probabilidad de esto realmente influir en la confianza con varias generaciones de distancia es baja y la metodología con esta VI puede lograr resultados más robustos que no usar VI.

Se puede verificar lo primer supuesto (de relevancia) analizando la relación entre el instrumento (distancia) y la instrumentada (exportaciones):

. reg ln_export_area distsea \$baseline_controls \$colonial_controls									
Source	SS	df	MS	Number of obs	=	20,020			
				F(75, 19944)	=	526.72			
Model	12366.5035	75	164.886713	Prob > F	=	0.0000			
Residual	6243.33859	19,944	.313043451	R-squared	=	0.6645			
				Adj R-squared	=	0.6633			
Total	18609.842	20,019	.929608974	Root MSE	=	.5595			
	ln_export_area	Coefficient	Std. err.	t	P> t	[95% conf. interval]			
	distsea	-.0009658	.0000181	-53.23	0.000	-.0010014 - .0009303			

Para un nivel de significancia de 5%, como el p-value es menor que el nivel de significancia, se rechaza la hipótesis nula, por lo tanto el instrumento es significativo. Además, se comprueba que la covarianza entre estas dos variables es distinta de cero (-0.4291), por lo tanto se comprobó el supuesto de relevancia, es decir, el instrumento es relevante.

4. Regresión con variable instrumental

Cada esclavo negociado/km² aumenta en 0.4% (Coefficient: 0.0041742) la confianza hacia los demás. Sin embargo este resultado no es significativo porque el p-value de 0.900 es mayor que el nivel de significancia de 5%.

Diferencias en Diferencias: El Caso de Bangladesh

Basándose en la ayudantía describa que indican las variables *Trat* y *Post*

La base de datos está se encuentra a nivel individual, pero el análisis se hace a nivel del hogar, así que es necesario identificar los hogares tratados (indicado por la variable **Trat**). Entonces, necesitamos identificar a los hogares (nh) que tiene al menos un integrante (masculino o femenino) de la familia participando del programa:

```
gen P = 0
```

```
replace P = 1 if dfmfd == 1 | dmmfd == 1
```

Además, solo nos interesa las observaciones tratadas para el segundo período (año 1998), después del tratamiento:

```
gen T98 = 0
```

```
replace T98 = 1 if P == 1 & year == 1
```

La variable **Trat** es 1 si hay al menos una observación tratada en el año de 1998:

```
egen Trat = max(T98), by(nh)
```

Por otro lado, la variable **Post** es una variable dummy para el segundo periodo de tiempo, que indica si la observación pertenece a este período:

```
gen Post = 1 if year == 1
```

```
replace Post = 0 if year == 0
```

El estimador de Diferencias en Diferencias

Para replicar la regresión es necesario generar la variable *TratPost* que es el término de interacción (*Trat* * *Post*), equivalente a una variable ficticia igual a 1 para las observaciones en el grupo de tratamiento, en el segundo período. Luego se replica la regresión, sin controles, a través del comando:

```
reg lnexpfd Trat Post TratPost egg agehead educhead
```

Así se obtiene los siguientes resultados:

. reg lnexpfd Trat Post TratPost

Source	SS	df	MS	Number of obs	=	1,652
Model	7.22029583	3	2.40676528	F(3, 1648)	=	21.65
Residual	183.183447	1,648	.111155004	Prob > F	=	0.0000
				R-squared	=	0.0379
Total	190.403743	1,651	.115326313	Adj R-squared	=	0.0362
				Root MSE	=	.3334

lnexpfd	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
Trat	-.0645155	.024174	-2.67	0.008	-.1119305	-.0171005
Post	.1033665	.0273591	3.78	0.000	.0497043	.1570287
TratPost	.0299295	.0341872	0.88	0.381	-.0371254	.0969844
_cons	8.053182	.0193458	416.28	0.000	8.015237	8.091127

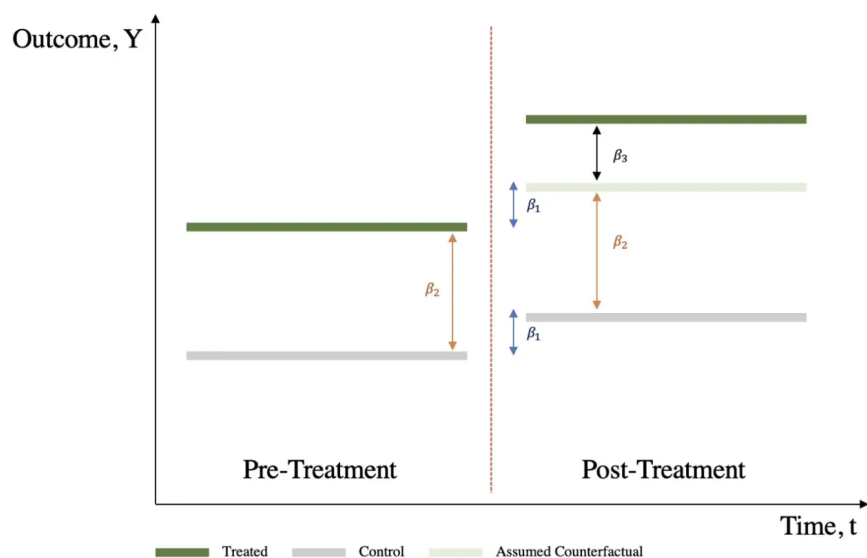
Los coeficientes se pueden interpretar como siendo:

Post (β_1): Cambio medio en y del primer al segundo periodo de tiempo que es común a ambos grupos. El cambio es significativo, y fué de 10%.

Trat (β_2): Diferencia media de y entre los dos grupos en el primer período uy que siguiera siendo igual si no hubiera tratamiento. Esta diferencia también es significativa y en promedio los grupo tratado tiene 6% menos de gasto per capita en comida.

TratPost (β_3): Cambio diferencial medio en y del primer al segundo periodo de tiempo del grupo de tratamiento en relación con el grupo de control, es decir el Average Treatment Effect que buscamos. Este resultado no es significativo (p-value = 0.381 > 0.05) y no hay evidencia que el tratamiento tiene impacto sobre el consumo de comida.

A continuación se presenta una interpretación gráfica de los coeficientes:



Controlando por las variables *egg*, *agehead*, *educhead* se obtienen los siguientes resultados:

. reg lnexpfd Trat Post TratPost egg agehead educhead

Source	SS	df	MS	Number of obs	=	1,652
Model	26.3279044	6	4.38798407	F(6, 1645)	=	43.99
Residual	164.075838	1,645	.099742151	Prob > F	=	0.0000
				R-squared	=	0.1383
				Adj R-squared	=	0.1351
Total	190.403743	1,651	.115326313	Root MSE	=	.31582

lnexpfd	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
Trat	-.0267591	.0231161	-1.16	0.247	-.0720992	.0185809
Post	.1685354	.0282323	5.97	0.000	.1131602	.2239105
TratPost	.0193635	.0324785	0.60	0.551	-.0443401	.083067
egg	.1455105	.0226351	6.43	0.000	.1011138	.1899072
agehead	.0022004	.0006277	3.51	0.000	.0009691	.0034316
educhead	.0289865	.0023331	12.42	0.000	.0244102	.0335627
_cons	7.528917	.0642463	117.19	0.000	7.402904	7.65493

Se concluye que el resultado sigue siendo no significativo (p-value = 0.551) por lo que el aumento en cerca de 2% en el consumo de alimentos per capita de los hogares (Coef. TratPost = 0.0193635), no puede ser interpretado con significancia.

Explicaciones sobre aleatorización en el tiempo y supuesto de Diferencias en Diferencias

Los 3 elementos necesarios para realizar una aleatorización en el tiempo

1. Panel de datos: las observaciones deben ser temporales y por individuo (empresa);
2. Grupo tratamiento y grupo control;
3. Un antes (del tratamiento) y un después (del tratamiento).

El supuesto clave que estamos estableciendo para implementar el método de Diferencias en Diferencias

El método diferencias-en-diferencias consiste en medir cambios en el tiempo de la variable de interés. Este método compara los cambios en los resultados a lo largo del tiempo entre las unidades que están inscritas en un programa (el grupo de tratamiento) y las unidades que no lo están (el grupo de control). Esto permite corregir cualquier diferencia entre el grupo de tratamiento y el grupo de control que son constantes a lo largo del tiempo.

La metodología utilizada consiste en sustraer una diferencia (primera diferencia) por otra. La diferencia en los resultados antes y después del grupo de tratamiento (**primera diferencia**) controla los factores que son constantes a lo largo del tiempo en ese grupo, ya que compara el mismo grupo consigo mismo. Para captar los factores que varían con el tiempo en el grupo de tratamiento se mide el cambio antes y después de los resultados de un grupo que no se inscribió en el programa, pero que estuvo expuesto al mismo conjunto de condiciones ambientales (grupo de control), o sea, se calcula la **segunda diferencia**.

Sin embargo, para que el sesgo de las comparaciones antes y después sea eliminado al restar la primera de la segunda diferencia, o sea, para que el método Dif-Dif sea válido, el grupo de control debe representar con exactitud el cambio en los resultados que habría experimentado el grupo de tratamiento.

Así que el supuesto clave de la metodología es que el resultado en el grupo de tratamiento se mueve en paralelo con el resultado en el grupo de control antes del tratamiento (**tendencias paralelas**). Esto permite suponer que los resultados mostrarían tendencias iguales en ausencia del tratamiento. A esto se llama el **supuesto de identificación**: ambos grupos deben de seguir tendencias paralelas, es decir, no tienen cambios en todas las variables a lo largo del tiempo. Los grupos de control y el grupo de tratamiento pueden ser distintos, pero estas diferencias se tienen que mantener constantes en el tiempo, en ausencia del tratamiento.

Esto se puede comprobar gráficamente, y analizando el gráfico a continuación, se puede ver que antes de la aplicación del tratamiento, antes del año 1998, los dos grupos presentan líneas rectas paralelas, es decir tenían una tendencia de evolución paralela antes del tratamiento, por lo tanto se verifica el supuesto de identificación.

