Visualization and Causal Inference of the Mexican Drug War

Valeria Espinosa and Joseph Kelly

vespinos@fas.harvard.edu, kelly2@fas.harvard.edu

Department of Statistics, Harvard University

Problem Description

The presidency of Felipe Calderón (2006-2012) has been characterized for the war against organized crime, raising many questions regarding security and violence. We attempt to visualize and analyze homicide rates at the municipality level, and link this to information obtained about the association of drug cartels to municipalities with goal to answer whether **homicide rates increase significantly after a military intervention**.

Estimand

Let $Y_i(1)$ denote the homicide rate change in region i from 2006 to one year after receiving a military intervention, and $Y_i(0)$ what it would have been if it hadn't received it (Rubin Causal Model). Our estimand is the average causal effect of the military intervention, W, for the regions that were intervened ($W = 1$),

$$\tau = \overline{Y}(1) - \overline{Y}(0) = \frac{\sum_{i=1}^{I} Y_i(1) - Y_i(0)}{I}.$$

Let N_i denote the number of municipalities that correspond to region i, then

$$Y_i(1) = \sum_{j=1}^{N_i} w_{ij} Y_{ij}(1) \text{ and } Y_i(0) = \sum_{j=1}^{N_i} w_{ij} Y_{ij}(0),$$

$$\text{where } w_{ij} = \frac{\text{Pop}_{ij}}{\text{Pop}_i} \text{ and } \text{Pop}_i = \sum_{j}^{N_i} \text{Pop}_{ij}.$$

However, $Y_i(0)$ and $Y_{ij}(0)$ are missing $\forall i, j$.

Key Assumptions

- **SUTVA:No hidden values of treatments** Broad definition of treatment: at least one municipality in the region received an intervention between 2007-2010, or not ([2]).
- **SUTVA:No interference between units** Grouped close regions that received an intervention, and their neighboring municipalities to make the "no interference" assumption more reasonable.
- **Unconfoundedness** We assume we have all covariates, $\mathbf{X}$, such that given $\mathbf{X}$, treatment assignment is independent of $\mathbf{Y}$.
- **Missing Data** One covariate had missing values. We exactly matched on missingness pattern and Political Party in municipality before Calderón.
- **Appropriateness of response variable** We assume Y is an adequate measure of violence.

Estimation

The control pool consists of 2213 municipalities. There are 13 treated regions considered (205 municipalities). Propensity score matching was used to identify 5 control municipalities that look like each treated ones, and estimate $Y_{ij}(0)$ and $Y_i(0)$. Let M_{ij} be the number of municipalities matched to the jth municipality in region i, and $\text{PopM}_{ij} = \sum_{k=1}^{M_{ij}} \text{PopM}_{ijk}$ is the sum of their populations. Then,

$$\hat{Y}_{ij}(0) = \sum_{k=1}^{M_{ij}} v_{ijk} Y_{ijk}(0), \text{ where } v_{ijk} = \frac{\text{PopM}_{ijk}}{\text{PopM}_{ij}}.$$

Therefore,

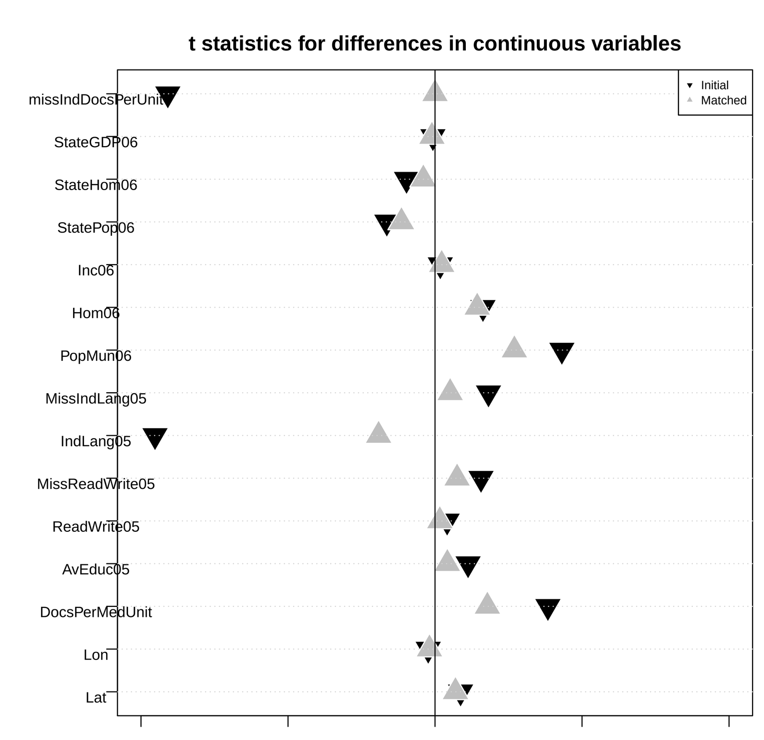
$$\hat{Y}_i(0) = \sum_{j=1}^{N_i} w_{ij} \hat{Y}_{ij}(0) = \sum_{j=1}^{N_i} w_{ij} \sum_{k=1}^{M_{ij}} v_{ijk} Y_{ijk}(0) = \sum_{j=1}^{N_i} \sum_{k=1}^{M_{ij}} \tilde{w}_{ijk} Y_{ijk}(0) \text{ with } \tilde{w}_{ijk} = w_{ij} v_{ijk},$$

and

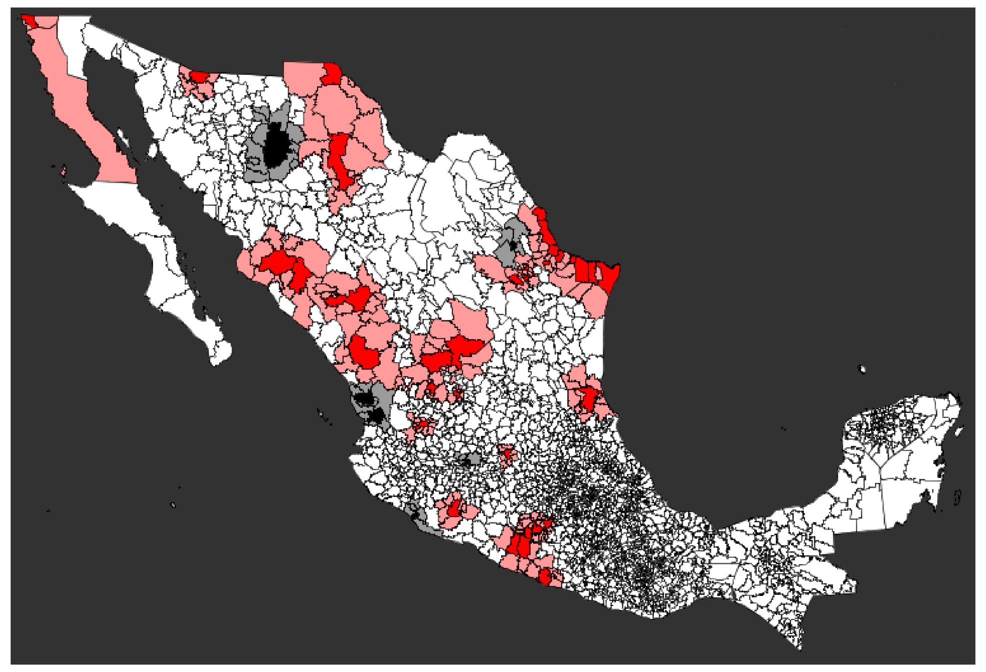
$$\hat{\tau} = \frac{\sum_j Y_j(1)}{J} - \frac{\sum_{j=1}^{J} \hat{Y}_j(0)}{J} = \overline{Y}(1) + \overline{Y}(0).$$

We know that $var(\hat{\tau})$ is largest under additivity of potenital outcomes. In that case $var(\hat{\tau}) = var(\overline{Y}(1)) + var(\overline{Y}(0))$. We use that over estimate to get confidence intervals.We estimate the variance assuming $Y_i(x) \sim Pois(p_i(x) * Pop_i)$ for $x = 0, 1$. Hence $var(\overline{Y})$

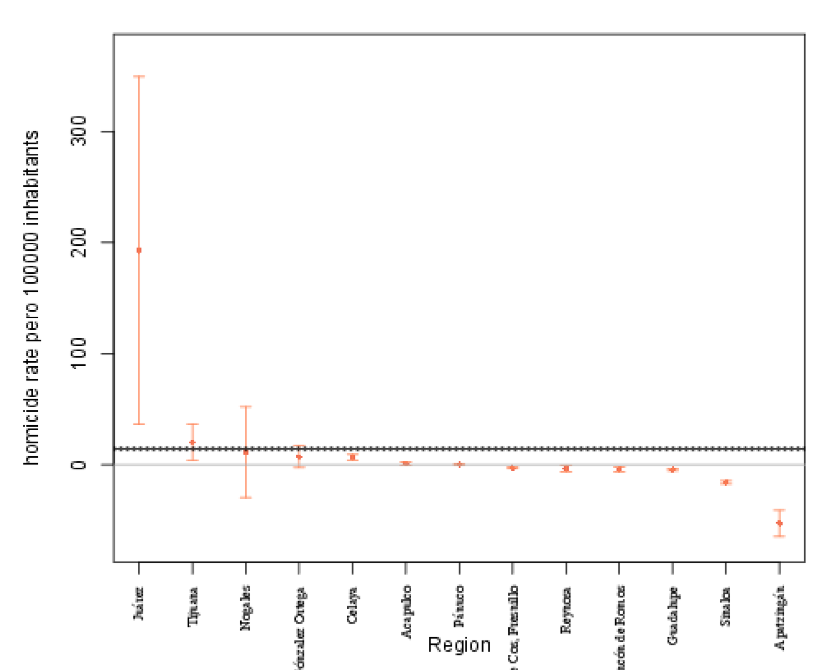
Visualization & Results



Love plot - balance checks



Interventions and SUTVA

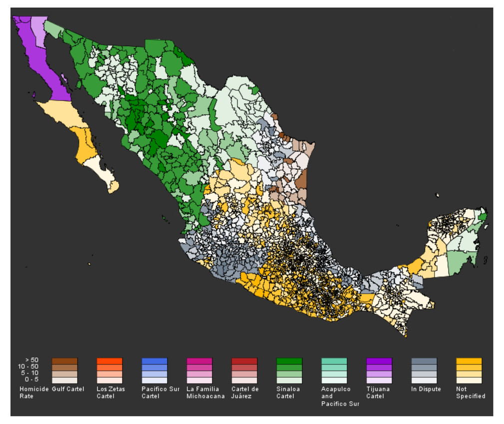


Results

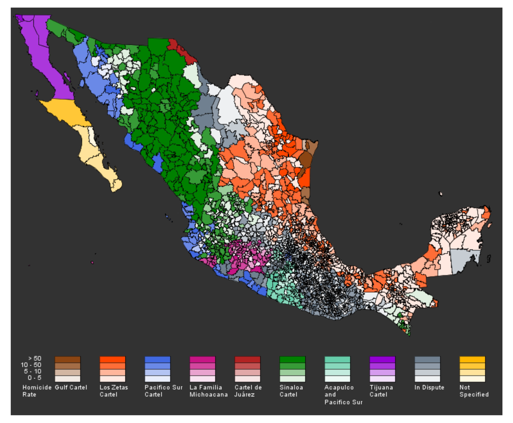
unit	Region	number of municipalities	Date of first intervention	$Y_j(1) - Y_j(0)$ (SD)
4	**Juárez**	15	2009	192.99 (79.88)
1	**Tijuana**	5	2008	20.49 (8.27)
2	**Nogales**	5	2008	11.41 (20.90)
10	**Teúl de González Ortega**	10	2009	7.32 (4.99)
15	**Celaya**	9	2009	6.74 (1.37)
18	**Acapulco**	35	2008	1.19 (0.77)
5	**Pánuco**	14	2007	0.37 (0.24)
9	**Villa de Cos, Fresnillo**	18	2008	-2.87 (0.34)
6	**Reynosa**	24	2008	-3.49 (1.48)
11	**Rincón de Romos**	8	2008	-4.10 (1.05)
8	**Guadalupe**	20	2009	-4.27 (0.58)
12	**Sinaloa, Badiraguato, Pueblo Nuevo**	27	2007	-15.84 (0.74)
16	**Apatzingán**	10	2007	-52.81 (5.97)
	$\hat{\tau}$	205	-	14.61 (23.14)

Results

Visualization and Processing



(a) 2007



(b) 2010

Processing is an open source programming language that allows the creation of dynamic graphics and tables. Due to the sheer size of the data we had collected the use of dynamic graphics ended up playing a significant role in how we would present the results of the analysis. Namely, the use of dynamic graphics allows:

- The user to navigate through some 2500 municipalities and view the homicide rate all within the same window via a map of Mexico.
- The presentation of further information to the user such as cartel presence and homicide rate over time.
- The presentation of matched municipalities and the corresponding causal effects for each region.

Conclusions

Details about processing? We're running out of room the left perhaps we can put Visualize the problem over here?

Data Sources

INEGI, CIDAC, Stratfor, Nexos

Key References & Data Source

[1] Abadie, Diamond, Hainmueller *Synthetic Control Methods for Comparative Case Studies* (2009)

[2] Escalante F, *Homicidios 2008-2009 La muerte tiene permiso* (2011)

[3] Imbens G. & Rubin D.R., (2012)

[4] Rubin D.R. ,*Matched Sampling for Causal Effects*(),