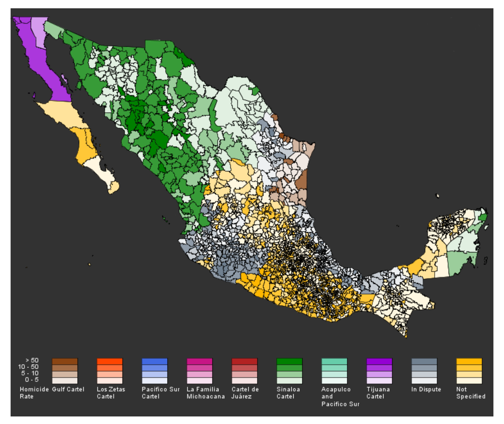
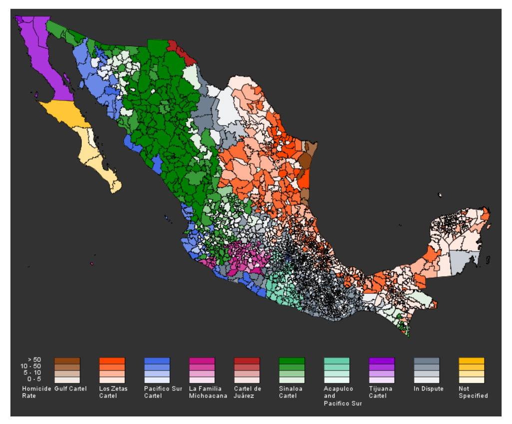
Visualization and Causal Inference of the Mexican Drug War

Valeria Espinosa and Joseph Kelly
vespinos@fas.harvard.edu, kelly2@fas.harvard.edu
Statistics Department, Harvard University

Visualize the problem



(a) 2007



(b) 2010

We attempt 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 ($W = 1$),

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

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]).
  - **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.

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

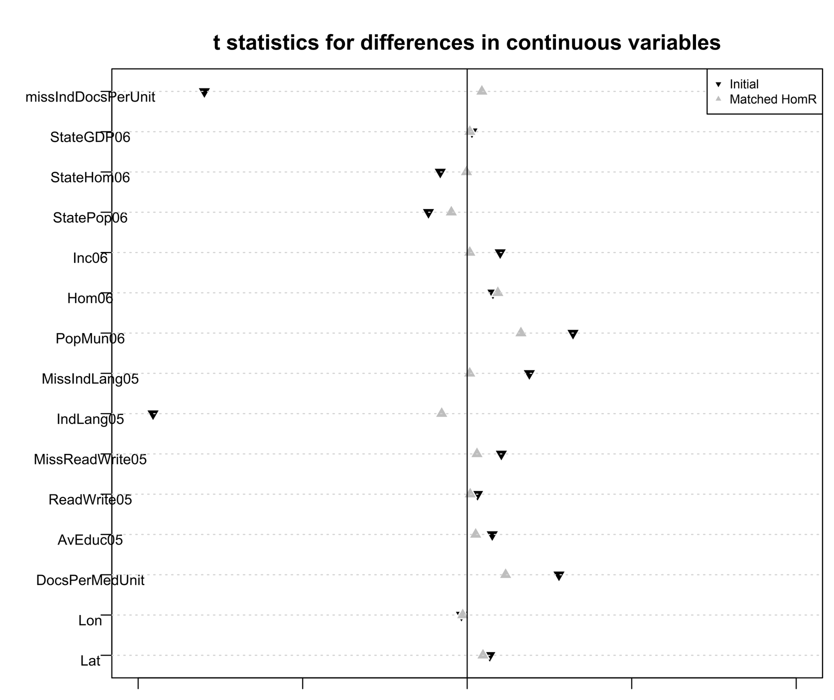
$$\hat{\tau} = \frac{\sum_j Y_j(1)}{J} - \frac{\sum_{j=1}^{J} \hat{Y}_j(0)}{J} = \bar{Y}(1) + \bar{Y}(0).$$

We know that $var(\hat{\tau})$ is largest under additivity of potenital outcomes. In that case $var(\hat{\tau}) = var(\bar{Y}(1)) + var(\bar{Y}(0))$. We use that over estimate to get confidence intervals.Now,

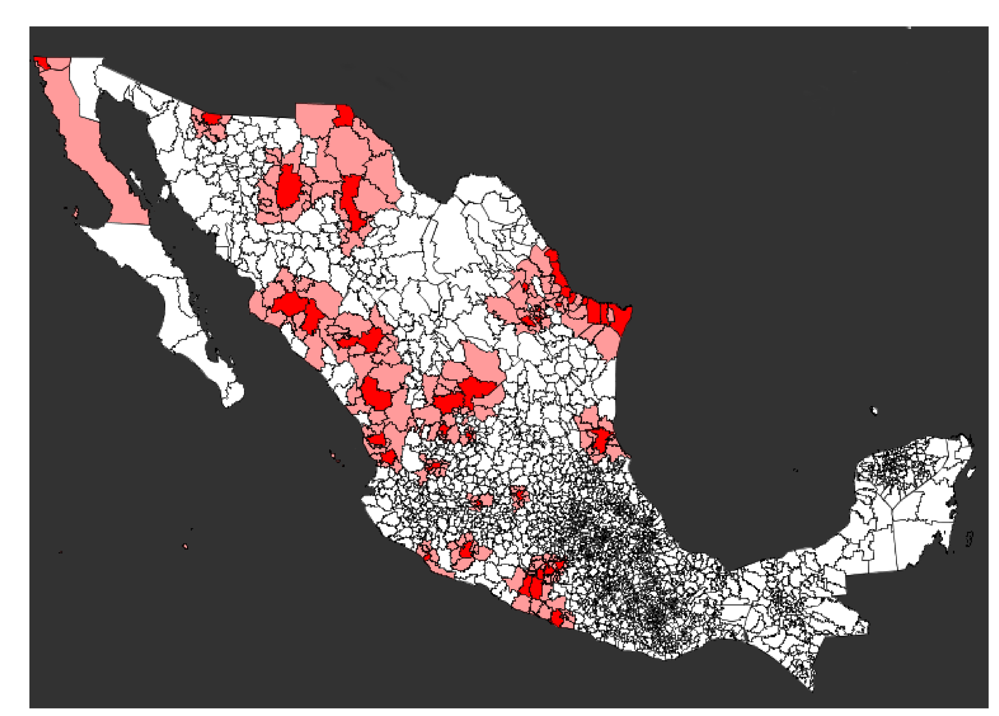
$$\begin{aligned} var(\hat{\bar{Y}}(0)) &= E(var(\hat{\bar{Y}}(0)|Y_i(0)\forall i)) + var(E(\hat{\bar{Y}}(0)|Y_i(0)\forall i)) = E(\Sigma\, var(\hat{Y}_i(0))/I) + var(\sum_i Y_i(0))/I \\ &= E(\frac{\sum_{j,k} w_{ijk}(Y_{ijk}(0) - \bar{Y}_i(0))^2}{1 - \sum_{j,k} w_{ijk}^2}) + var(Y(0))/I = \frac{\sum_{i,j,k} w_{ijk}(Y_{ijk}(0) - \bar{Y}_i(0))^2}{I(1 - \sum_{j,w} w_{ijk}^2)}) + S^2(0)/I. \end{aligned}$$

Now, $var(\hat{\bar{Y}}(1)) = S^2(1)/I$ because the all $Y_j(1)$ are observed.

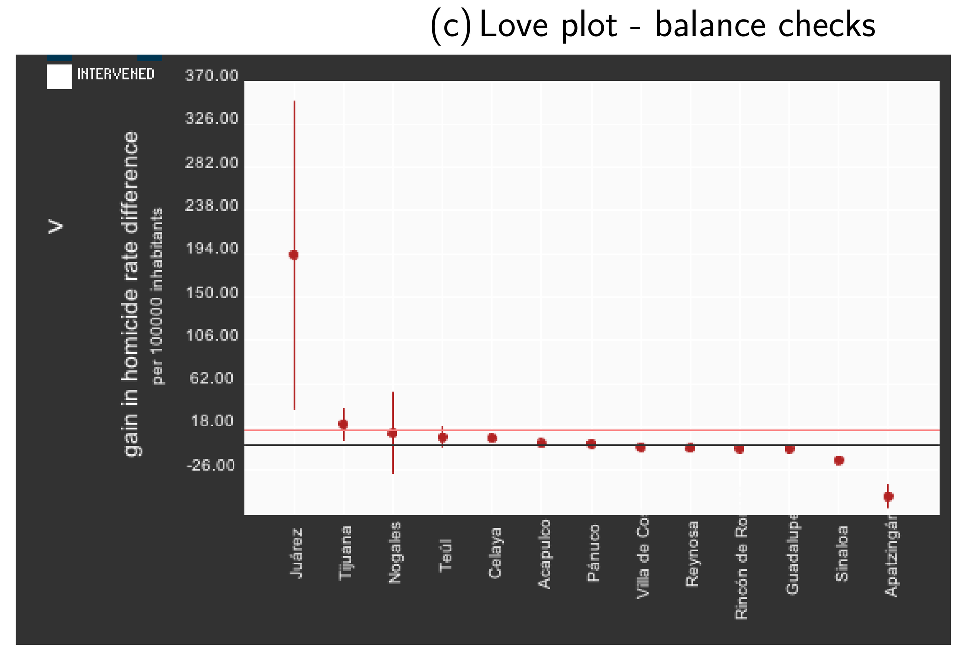
Visualization



(c) Love plot - balance checks



(d) Interventions and SUTVA



(e) 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)