Mexican Drug War: Have the Military Interventions Increased Violence?

When we start to approach this problem the first to consider is what data is available to analyze it. We are using the Military interventions listed in a non comprehensive list from may 7,2007 - nov 21, 2010) NEXOS paper: `http://www.nexos.com.mx/?P=leerarticulo&Article=1943189`.

The measure of violence that we are using is homicide rate, Y. This could be debatable, maybe a combined measure of different kinds of crimes would be better.

Assumptions:

- SUTVA
- Unconfoundedness: a very strong assumption is that we have all the covariates that could affect the homicide rates. In other words that we have all covariates, X, such that given X, W is independent of Y.
- That homicide rates, Y, are an accurate measure of violence.

The main idea is to combine synthetic and propensity score matching to address this question. What are the advantages of doing that? (and what's new?) The current proposal is to use propensity score matching to create pools of acceptable controls for each treated unit (is there some multiple comparison thing going on here? Probably not, we just cared about the observed imbalances, we are not saying anything else.). Furthermore, to get the synthetic match for treated unit T_i we can choose the weights for the units in it's control pool such that the $Y_1^T, \ldots, Y_{I_i}^{T_i}$ is best matched

There are 2213 municipalities in the initial control pool, and the numbers per treated unit are:

unit	number of municipalities	Date of first intervention	**unit**	number of municipalities	Date of first intervention
1	5	2008	**10**	10	2009
2	5	2008	**11**	8	2008
3	12	2010*	**12**	27	2007
4	15	2009	**13**	11	2010*
5	14	2007	**14**	9	2010*
6	24	2008	**15**	9	2009
7	5	2010*	**16**	10	2007
8	20	2009	**17**	6	2010*
9	18	2008	**18**	35	2008

Note that for the data that we're working with, all the regions with $*$ are eliminated from the data set since we have no post intervention information. The data collected spans 1990-2010. That leaves us with 205 treated municipalities.

1 Estimand

We want to measure the effect of the military interventions in terms of the increase in homicide rates. Following the Rubin Causal Model, let $Y_j(1)$ and $Y_j(0)$ denote the homicide rate of region j one year after it received a military intervention[1], and what it would have been at that same point in time if it hadn't received the military intervention. The estimand of interest is the average causal effect of the military intervention for the regions that were intervened. That is

$$\tau = \overline{Y}(1) - \overline{Y}(0) = \frac{\sum_j Y_j(1) - Y_j(0)}{J}.$$

[1]We can also estimate the effect two and three years post intervention. However, the uncertainty will increase because there are only three regions with 2007 interventions and an additional six with 2008 interventions.

A common approach is to assume $Y_j(1)$ is observed for all treated units. Let N_j denote the number of municipalities that correspond to region j, then

$$Y_j(1) = \sum_{i=1}^{N_j} w_{ij} Y_{ij}(1),$$

where $\text{Pop}_j = \sum_i^{N_j} \text{Pop}_{ij}$, and

$$w_{ij} = \frac{\text{Pop}_{ij}}{\text{Pop}_j}.$$

However, $Y_j(0)$ is missing for all j. Following the reasoning above,

$$Y_j(0) = \sum_{i=1}^{N_j} w_{ij} Y_{ij}(0),$$

and all $Y_{ij}(0)$ are unobserved.

2 Matching Procedure and "Naive" Analysis

We attempt to use the information of all other municipalities to estimate each $Y_{ij}(0)$ to obtain an estimate $Y_j(0)$. How will we do that? The idea is to use propensity score matching to identify good matches for each treated municipality.

To follow the guidelines for observational studies we will first clarify what the analysis protocol will be, that will determine the way the balance checks will be performed to choose an estimated propensity score that leads to an acceptable balance. Let M_{ij} be the number of municipalities matched to the ith municipality in region j. Let

$$\text{PopM}_{ij} = \sum_{k=1}^{M_{ij}} \text{PopM}_{ijk}$$

denote the total population of all M_{ij} municipalities matched to the ith treated municipality in region j. Then

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

where $v_{ijk} = \frac{\text{PopM}_{ijk}}{\text{PopM}_{ij}}$. Therefore,

$$\hat{Y}_j(0) = \sum_{i=1}^{N_j} w_{ij} \hat{Y}_{ij}(0) = \sum_{i=1}^{N_j} w_{ij} \sum_{k=1}^{M_{ij}} v_{ijk} Y_{ijk}(0),$$

and

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

Now, the fact that we are using weighted averages is relevant for the estimate of $var(\hat{\tau})$.

- Option 1

  $$var(\hat{\tau}) = S^2(1)/J + S^2(0)/J$$

  where $S(x)^2$ is the variance of $Y_j(x)$, and $\tilde{w}ijk = w_{ij} v_{ijk}/J$. Hence,

  $$v\hat{a}r(\hat{\tau}) = s^2(1) \sum_{i=1}^{N_j} w_{ij}^2 + s^2(0) \sum_{j=1}^{J} \sum_{i=1}^{N_j} \sum_{k=1}^{M_{ij}} \tilde{w}_{ijk}^2,$$

where

$$s^2(1) = \sum_{j=1}^{J}\sum_{i=1}^{N_j} w_{ij}(Y_{ij}(1) - \overline{Y}(1))^2$$

is the sample variance of the potential outcomes under treatment, and

$$s^2(0) = \sum_{j=1}^{J}\sum_{i=1}^{N_j}\sum_{k=1}^{M_{ij}} \tilde{w}_{ijk}(Y_{ijk}(0) - \hat{\overline{Y}}(0))^2,$$

is the sample variance of the potential outcomes under control, where the means are the weighted means.

$$\hat{\overline{Y}}(0) = \sum_{j=1}^{J}\sum_{i=1}^{N_j}\sum_{k=1}^{M_{ij}} \tilde{w}_{ijk}Y_{ijk}(0),$$

and

$$\hat{\overline{Y}}(1) = \sum_{j=1}^{J}\sum_{i=1}^{N_j} w_{ij}Y_{ij}(1).$$

- Option 2 (the one that makes most sense at the moment is)

$$var(\hat{\tau}) = \sum_{j=1}^{J}\sum_{i=1}^{N_j}(w_{ij}/J)^2 S^2(1) + \sum_{j=1}^{J}\sum_{i=1}^{N_j}\sum_{k=1}^{M_{ij}} \tilde{w}^2_{ijk}S^2(0)$$

  where $S(1)^2$ is the variance of the homicide rate for all units under treatment, $Y_ijk(1)$, and $S(0)^2$ is the variance of the homicide rate for all units under control, $Y_ijk(0)$. **Maybe here it should be $Y_j(x)$ instead of $Y_ijk(x)$.**

- Option 3. In the Imbens and Rubin book it is suggested to calculate the conditional variance for each unit using matching within treated groups and within control groups. Do the simple things work?

2.1 Balance Checks

These weights (right now it is option 2) were used to assess balance.

We used matchIt to exactly match on missingness pattern (the only variable with values is "Doctors per Medical Unit") and Party at the municipality level. The initial balance, shown in figures 1 and 1, seems reasonable for the main covariates (not including higher order terms or interactions - these haven't been checks. Also, which interactions and higher order terms would be of interest?). There are some things to note though, the scale on the y axis is very different in the treated and control groups. However, there seems to be good overlap to find matches. One concern is the number of Homicides in 2006, the domain of the treated municipalities seems larger than that of the controls.

We further explore the overlap for homicides

In some tries of matching we observe that the major imbalance is at the 'Homicide 2006' covariate (should we have homicide rate or do the weights take care of that?)

3 Comments

Perhaps a more comprehensive way of approaching this problem is to analyze the homicide rate time series using Synthetic Matching.

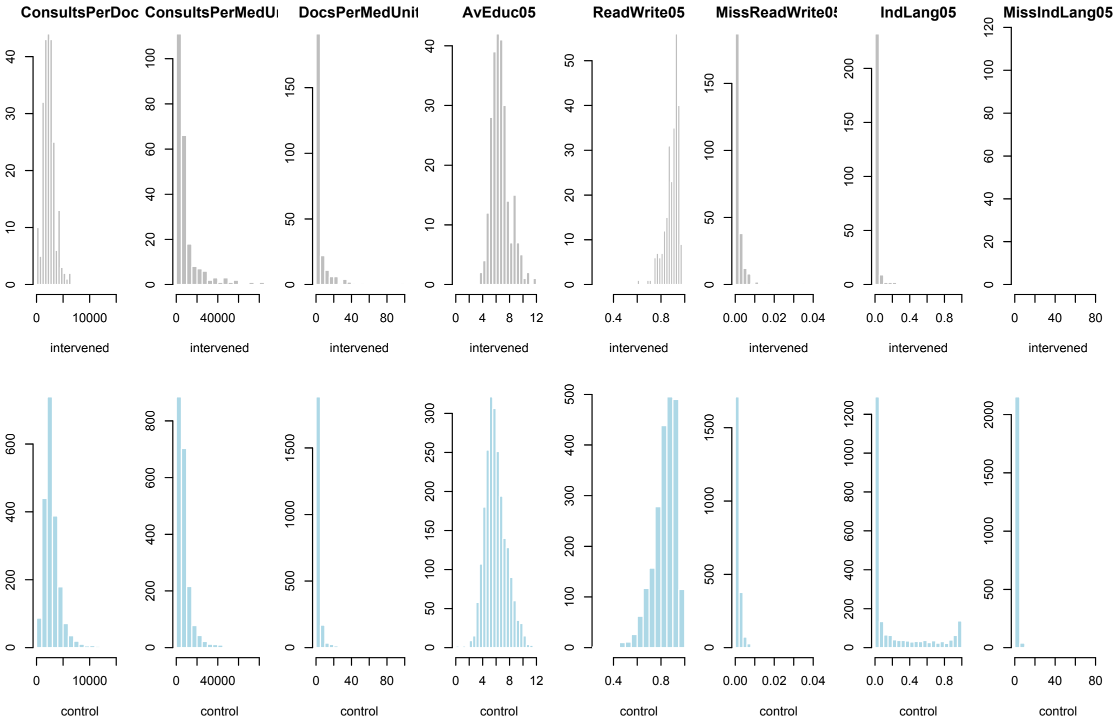

Figure 1: Histograms to compare distributions in original pools.

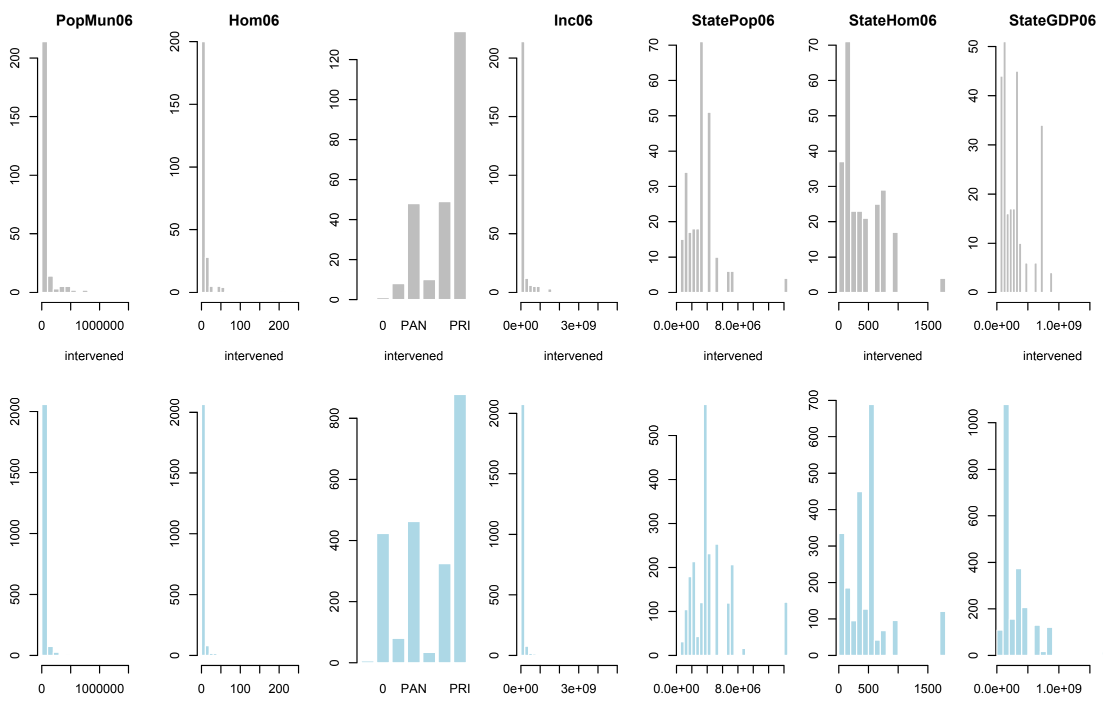

Figure 2: Histograms and Barplots to compare distributions in original pools.

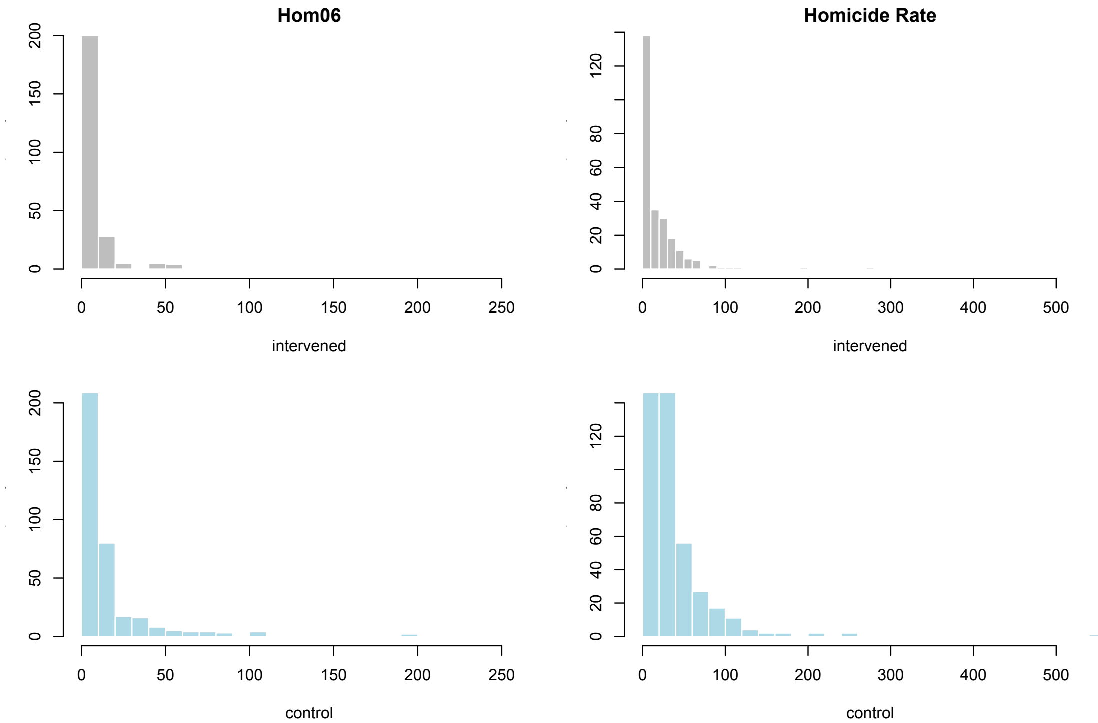

Figure 3: Zooming in to Homicide Rate