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** - stable unit treatment value assumption
  - **No hidden values of treatments** A broad definition of what "military intervention"- any mili means in this context helps us think of a two level treatment: receiving a military intervention (defined as ... see paper(that have resulted in deaths?)) or not receiving it.
  - **No interference between units** The main idea is to group neighboring regions that have received military interventions in such a way that distances make the "no interference" assumption more reasonable. For treated regions that are side to side were also assessed in terms of neighboring geographic situation such as lack of highways connecting them The last homicide rate that we have corresponds to 2010. That eliminated some of the interventions mentioned in the Nexos paper. Following this reasoning 2213 municipalities were included in the initial control pool. Our 13 treated regions are shown in Table **??**
- **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. Unfortunately we didn't get experts to guide most of our decisions. However, we did get to interact with a couple of them and made our covariate choices based on the information received and our understanding of the relevant information. Our covariates include: location, political party before Calderón, income 2006, education and medical information at 2005, percentage of indigenous language speakers, 2006 homicide rate at the municipality level, and GDP, Homicide Rate and Population at the state level.
- **Missing Data:** there were few treated units that had one covariate (Doctors per medical unit) missing, we exactly matched on that and Political Party in power before Calderón.
- 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

0.1 How to make SUTVA more feasible?

The main idea is to group neighboring regions that have received military interventions (that have resulted in deaths) in such a way that distances make no interference more reasonable or geographic situation such as lack of highways connecting them `http://mx.kalipedia.com/kalipediamedia/geografia/media/200805/11/geomexico/20080511klpgeogmx_4_Ges_SCO.png` or big mountains between them, or closeness to where the crops or smuggling routes are `http://utopiaguatemala.files.wordpress.com/2011/11/mexico_routes_in_466.gif` (can we find such data?). We proposed the following units (the bold names are the ones that actually received an intervention):

1. **Tijuana**, Playas de Rosarito, Tecate, Ensenada (Baja California)

2. **Nogales**, Sric, Tubutama, Magdalena, Imuris, Santa Cruz (Sonora)

3. **Madera***, Temsachic, Casas Grandes, Zaragoza, Gmez Faras, Matach (Chihuahua) Ycora, Arivechi, Sahuaripa, Ncori Chico, Huachinera, Bacerac, Bacadhuachi (Sonora)

4. **Jurez, Chihuahua**, Ascensin, Ahumada, Guadalupe, Praxedis G. Guerrero, Coyame del Socol, Aldama, Aquiles Serdn, Rosales, Satev, Dr. Belisario Dominguez, Santa Isabel, Riva Palacio, Namiquipa, Buenaventura (Chihuahua)

5. **Pnuco ***, Pueblo Viejo, Tampico Alto, El Higo, Tempoal, Ozuluama de Mascareas (Veracruz), **Tampico**, Ciudad Madero, Altamira, Gonzlez, El Mante (Tamaulipas), Ebano,Tamun, San Vicente Tancuayalab (San Luis Potos)

6. **Nuevo Laredo*, Guerrero*, Mier*, Miguel Alemn*, Daz Ordaz*, Reynosa, Ro Bravo, Matamoros***, Camargo, Valle Hermoso, Mndez, San Fernando (Tamaulipas), Dr. Cos, Anhuac, Gral. Bravo, China, Vallecillo, Pars, Agualeguas (Nuevo Len), **Melchor Ocampo***, Los Herreras, Los Aldamas, Gral Trevio, Cerralvo (Nuevo Len)

7. **Bustamante***, Villaldama, Mina, Lampazos de Naranjo (Nuevo Len), Candela (Coahuila)

8. **Santa Catarina*, Guadalupe, Jurez*, General Zuazua**, Higueras, Marn, Pesquera, Apodaca, General Escobedo, Carmen,Salinas Victoria, Cinega de Flores, Cadereyta Jimnez, Santiago, Monterrey, San Pedro Garza Garca, Garca, San Nicols de los Garza, (Nuevo Len), Ramos Arizpe, Arteaga (Coahuila)

9. **Villa de Cos, Fresnillo, Tepetongo***, Pnuco, Calera, General Enrique Estrada, Jerez , Susticacn, Valparaso, Sombrerete, Sain Alto, Ro Grande, Caitas de Felipe Pescador, General Francisco R. Murgua, Mazapil, Guadalupe,Villanueva, Monte Escobedo (Zacatecas), Villa de Ramos, Santo Domingo (San Luis Potos), Huejcar, Santa Mara de los ngles (Jalisco)

10. Tel de Gonzlez Ortega, Mezquital, Trinidad Garca de la Cadena, Moyahua de Estrada, Juchipila, Apozol, Santa Mara de la Paz, Benito Jurez (Zacatecas), Tequila, San Martn de Bolaos (Jalisco)

11. Rincn de Romos, Tepezal, Coso, San Jos de Gracia, Pabelln de Arteaga (Aguascalientes), Cuauhtmoc, Genaro Codina, Luis Moya (Zacatecas)

12. Sinaloa, Badiraguato, Choix, El Fuerte, Ahome, Guasave, Salvador Alvarado, Mocorito, Culiacn (Sinaloa), Guadalupe y Calvo, Morelos (Chihuahua), Tamazula (Durango) Santiago Papasquiaro*, Pueblo Nuevo, Tamazula, Otez, San Dimas, Nuevo Ideal, Canatln, Coneto de Comonfort, El Oro, Tepehuanes, Topia, Durango, Canelas,Mezquital (Durango), Huajicori (Nayarit), Rosario, Concordia (Sinaloa)

13. Rosamorada*, Tepic*, Tecuala, Santiago Ixcuinta, Tuxpan, Ruz, Del Nayar, Acaponeta, Tecuala, San Blas, Xalisco, Santa Mara del Oro, (Nayarit)

14. La Piedad*, Numarn, Zinparo, Churintzio, Ecuandureo, Yurcuaro (Michoacn), Pnjamo (Guanajuato), Ayotln, Degollado (Jalisco) Celaya, Cortazar, Salvatierra, Tarimoro, Apaseo el Alto, Apaseo el Grande, Comonfort, Santa Cruz De Juventino Rosas, Villagrn (Guanajuato)

15. Apatzingn, Buenavista, Aguililla, Tumbiscato, La Huacana, Mgica, Parcuaro, Tanctaro, Tepalcatepec (Michoacn) Coahuayana*, Aquila, Chinicuilla (Michoacn), Ixtlahacn, Tecomn, Colima (Colima)

16. Taxco*, Teloloapan, Arcelia, San Miguel Totolapan, Acapulco, Ajuchitln del Progreso, San Marcos, Juan R. Escudero, Tecoanapa ,Chilpancingo de los Bravo, Coyuca de Bentez, Atoyac de lvarez, Tcpan de Galeana, Coyuca de Cataln, Pungarabato, Tlapehuala, Tlalchapa, General Canuto A. Neri, Pedro Ascencio Alquisiras, Ixcateopan de Cuauhtmoc, Tetipac,Pilcaya, Benito Jurez, Buenavista de Cullar, Iguala de la Independencia, Cocula, Cuetzala del Progreso, Apaxtla, General Heliodoro Castillo (Guerrero), Tlatlaya, Amatepec, Sultepec, Zacualpan (Mxico), Amacuzac,Tetecala, Coatln del Ro(Morelos)

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

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

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.

[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.

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 $var(\hat{\tau})$. We used two approaches:

- **Neyman Variance**: As usual $Var(\overline{Y}(1)) = Var(Y(1))/I$. Now,

$$\begin{aligned} Var(\overline{Y}(0)) &= E(Var(\overline{Y}(0)|Y_i(0))) + Var(E(\overline{Y}(0)|Y_i(0))) \\ &E() \end{aligned}$$

- **Binomial Variance**:It can be argued that $Y_j(x) \sim Bin(Pop_j, p_j(x))$. In this context the estimand is the difference of $\bar{Y}(x) = \frac{p_1(x)+p_2(x)+\cdots+p_J(x)}{J} \times 100000^2$, i.e. the average homicide rate under intervention x. Hence, $Var(Y_j(x)) = 10^{10} \times p_j(x)(1-p_j(x))/Pop_j$, and assuming independence of the regions we get that $Var(\bar{Y}(x)) = \frac{10^{10}\sum_{j=1}^{J} p_j(x)(1-p_j(x))/Pop_j}{J^2}$. Furthermore, $V\hat{a}r(\bar{Y}(x)) = \frac{10^{10}\sum_{j=1}^{J} \hat{p_j}(x)(1-\hat{p_j}(x))Pop_j}{J^2}$. (this is not the potential outcome way of estimating this!! We would usually use the Neyman estimate of the variance)

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.

Images/InitBalance1.pdf

Figure 1: Histograms to compare distributions in original pools.

Images/InitBalance2.pdf

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

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.

4 Results

Following this reasoning 2213 municipalities were included in the initial control pool. Our 13 treated regions are:

unit	**Region**	number of municipalities	Date of first intervention	Within Region Effect	SD Bin	SD Neyman
1	**Tijuana**	5	2008	20.89	1.49	12.87
2	**Nogales**	5	2008	36.97	5.37	33.51
4	**Juárez**	15	2009	195.11	3.84	88.33
5	**Pánuco**	14	2007	-0.90	0.92	0.39
6	**Reynosa**	24	2008	0.87	0.87	1.23
8	**Guadalupe**	20	2009	-5.06	0.59	0.89
9	**Villa de Cos, Fresnillo**	18	2008	-1.35	1.24	0.44
10	**Teúl de González Ortega,**	10	2009	13.83	5.52	9.21
11	**Rincón de Romos**	8	2008	-4.80	2.55	1.07
12	**Sinaloa, Badiraguato, Pueblo Nuevo**	27	2007	4.33	1.19	1.15
15	**Celaya**	9	2009	2.84	1.42	1.82
16	**Apatzingán**	10	2007	17.50	3.23	1.81
18	**Acapulco**	35	2008	12.45	1.47	1.86
	Average	205	-	19.86	0.78	0.89

[2]Note that the homicide rate is defined as the number of homicides in 100000 habitants.

Images/ZoomHom.pdf

Figure 3: Zooming in to Homicide Rate

Images/MEloveplot.pdf