Gunshots and Turf Wars

Inferring Gang Territories from Administrative Data[*]

Brendan Cooley[†] Noam Reich[‡]

December 12, 2019

Abstract

Street gangs are conjectured to engage in violent territorial competition. This competition can be difficult to study empirically as the number of gangs and the division of territory between them are usually unobserved to the analyst. However, traces of gang conflict manifest themselves in police and administrative data on violent crime. In this paper, we show that the frequency and location of shootings are sufficient statistics for the territorial partition under mild assumptions about the data generating processes for gang-related and non-gang related shootings. We then show how to estimate this territorial partition from a panel of geolocated shooting data. We apply our method to analyze the structure of gang territorial competition in Chicago using victim-based crime reports from the Chicago Police Department. The method reveals both the number of gangs in operation in Chicago and their territorial boundaries.

[*]Thanks go here.

[†]Ph.D. Candidate, Department of Politics, Princeton University

[‡]Ph.D. Candidate, Department of Politics, Princeton University

Introduction

Literature

Data

Model

Primitives and Assumptions

There are N districts in the city ($i, j \in \mathcal{N} = \{1, ..., N\}$). r_i residents live in each district. The city is also inhabited by K gangs ($k, \ell \in \mathcal{K} = \{1, ..., K\}$). Each gang is endowed with a m_k soldiers. A partition function $\pi : \mathcal{N} \to \{0, \mathcal{K}\}$ assigns territories to the gangs that control them, where $\pi(i) = 0$ indicates the absence of any gang activity. $\mathcal{N}_k$ is the set of territories controlled by gang k and $n_k = |\mathcal{K}_k|$ the number of territories controlled by gang k. The set of unoccupied territories is $\mathcal{K}_0$. We are interested in estimating the number of groups, K, and the territorial partition, π.

We observe data on geo-located shootings for T periods, indexed $\{1, ..., T\}$. We hold the above quantities constant over time. There are three types of shootings that occur in the city – inter-gang, intra-gang, and non-gang. Let y_i^t denote non-gang related shootings in district i during period t and x_i^t denote gang-related shootings in the same district-period. Non-gang shootings are committed by residents with probability η_i and are independent across districts. Then, the expected number of shootings in district i is $\eta_i r_i$ with variance $\psi_i = \eta_i(1 - \eta_i)r_i$.[1]

Gang-related shootings are determined by the geographic distribution of gang activity and the state of relations between and within gangs. We assume the probability a given soldier from gang k is operating in territory i is constant and given by n_k^{-1}. Members of the same gang sometimes commit violence against one another. The probability a member of gang k shoots a member of his own gang during period t is given by ξ_k^t. Assumption 1 states that the expected likelihood of such violence is non-zero.

Assumption 1: $\mathrm{E}[\xi_k^t] > 0$ for all $k \neq 0$ and $\xi_0^t = 0$ for all t.

We also assume that conflict within gangs is unrelated to within-gang conflict between other gangs.

Assumption 2: $\mathrm{E}[\xi_k^t \xi_\ell^t] - \mathrm{E}[\xi_k^t]\mathrm{E}[\xi_\ell^t] = 0$ for all $k \neq \ell$.

We impose no other restrictions on the distribution of intra-gang shocks. The possibility of intra-gang violence allows us to distinguish between territories owned by the same

[1] In other words, non-gang shootings are distributed i.i.d. binomial.

gang and territories whose owners exclusively war with one another.[2]

Gangs also war with one another with varying intensity. The probability a member of gang k shoots a member of gang ℓ during period t is $\epsilon^t_{k\ell}$. We make two assumptions on the distribution of these inter-gang shocks. First, we assume they are quasi-symmetric. This requires that any increase in the likelihood that members of gang k shoot members of gang ℓ is accompanied by a proportionate increase in reciprocal violence. Notably, we allow this retaliation propensity to vary at the level of the gang but not the gang-dyad.

Assumption 3: $c_k\epsilon^t_{k\ell} = c_\ell\epsilon^t_{\ell,k}$ with the normalization $c_1 = 1$. If $k = 0$ or $\ell = 0$ then $\epsilon^t_{k\ell} = 0$ for all t.

Second, we assume inter-gang shocks are independent across gang dyads.[3]

Assumption 4: $\mathrm{E}\left[\epsilon^t_{k,\ell}\epsilon^t_{m,n}\right] - \mathrm{E}\left[\epsilon^t_{k,\ell}\right]\mathrm{E}\left[\epsilon^t_{m,n}\right] = 0$ for $m, n \notin \{k, \ell\}$.

The expected number of gang-related shootings in district i during period t can then be calculated as

$$x^t_i = \underbrace{\frac{m_{\pi(i)}}{n_{\pi(i)}}\mathrm{E}[\xi^t_{\pi(i)}]}_{\text{intra-gang}} + \underbrace{\sum_{k\neq\pi(i)}\frac{m_k}{n_{\pi(i)}}\mathrm{E}[\epsilon^t_{k,\pi(i)}]}_{\text{inter-gang}}$$

The total number of shootings in district i during period t is

$$v^t_i = x^t_i + y^t_i$$

Covariance Structure

In the proceeding section we will show that the covariance in shootings across districts is informative about the number of groups and the territorial partition. Let $a_{ij} = \mathrm{Cov}[v^t_i, v^t_j]$ Proposition 1 describes the covariance structure of our model. A derivation of this quantity can be found in Appendix A.

Proposition 1: The covariance in shootings between districts i and j is

$$a_{ij} = \begin{cases} \sum_{k\neq\pi(i)}\left(\left(\frac{m_k}{n_{\pi(i)}}\right)^2\mathrm{Var}[\epsilon^t_{\pi(i),k}]\right) + \left(\frac{m_{\pi(i)}}{n_{\pi(i)}}\right)^2\mathrm{Var}[\xi^t_{\pi(i)}] + \psi_i & \text{if } i = j \\ \sum_{k\neq\pi(i)}\left(\left(\frac{m_k}{n_{\pi(i)}}\right)^2\mathrm{Var}[\epsilon^t_{\pi(i),k}]\right) + \left(\frac{m_{\pi(i)}}{n_{\pi(i)}}\right)^2\mathrm{Var}[\xi^t_{\pi(i)}] & \text{if } \pi(i) = \pi(j) \\ \frac{m_{\pi(i)}}{n_{\pi(j)}}\frac{m_{\pi(j)}}{n_{\pi(i)}}\frac{c_{\pi(j)}}{c_{\pi(i)}}\mathrm{Var}[\epsilon^t_{\pi(i),\pi(j)}] & \text{if } \pi(i) \neq \pi(j) \\ 0 & \text{otherwise} \end{cases}$$

[2]Alternatively, we could assume that gangs fight at least two other groups with positive probability. We view this assumption as less restrictive.

[3]Of course, the intensity of conflict between any two gangs is almost certainly affected by the broader conflict environment. This assumption is made for purposes of model tractability. In future work, we plan to model the genesis of conflict shocks and perhaps relax this assumption.

Corollary 1 states that violence will covary constantly for all pairs of districts controlled by the same gang.

Corollary 1 (Block Structure):

1. If $\pi(i) = \pi(j) = k$ and $i \neq j$ then $a_{ij} = b_{kk}$ constant for all i, j.
2. If $\pi(i) = k$ and $\pi(j) = \ell$ with $\ell \neq k$ then $a_{ij} = b_{k\ell}$ constant for all i, j.

Let $A_{N \times N} = (a_{ij})_{\{i,j \in \mathcal{N}\}}$ be the covariance matrix.[4] Let $A(k, \ell)_{n_k \times n_\ell} = (a_{ij})_{\{i,j | \pi(i)=k, \pi(j)=\ell\}}$ be the submatrix where the row districts are controlled by k and the column districts are controlled by ℓ. If the partition function π is known then the rows and columns of this matrix can be permuted to reveal the block structure described in Corollary 1. To reveal the block structure, we rearrange district identifiers in accordance with their territorial assignment. Let f be a bijection that maps $\mathcal{N}$ to itself. Specifically,

$$f : \begin{cases} \mathcal{K}_k \to \left\{ \sum_{\ell=1}^{k-1} (n_\ell) + 1, \ldots, \sum_{\ell=1}^{k} (n_\ell) \right\} & \text{if } k \geq 1 \\ \mathcal{K}_0 \to \left\{ \sum_{\ell=1}^{K} (n_\ell) + 1, \ldots, N \right\} & \text{if } k = 0 \end{cases}$$

Then, let $P_{N \times N} = (p_{ij})_{\{i,j \in \mathcal{N}\}}$ be a permutation matrix with $p_{ij} = 1$ if $f(i) = j$ and $p_{ij} = 0$ otherwise. Let $\bar{A} = PAP$ denote the permuted covariance matrix. Then,

$$\bar{A} = \begin{bmatrix} \overbrace{\begin{bmatrix} A(1,1)_{n_1 \times n_1} \end{bmatrix}}^{\pi(i)=1} & \overbrace{\begin{bmatrix} A(1,2)_{n_1 \times n_2} \end{bmatrix}}^{\pi(i)=2} & \cdots & \overbrace{\begin{bmatrix} A(1,K)_{n_1 \times n_K} \end{bmatrix}}^{\pi(i)=K} & \mathbf{0} \\ \begin{bmatrix} A(2,1)_{n_2 \times n_1} \end{bmatrix} & \begin{bmatrix} A(2,2)_{n_2 \times n_2} \end{bmatrix} & \cdots & \vdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \mathbf{0} \\ \begin{bmatrix} A(K,1)_{n_K \times n_1} \end{bmatrix} & \cdots & \cdots & \begin{bmatrix} A(K,K)_{n_K \times n_K} \end{bmatrix} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \underbrace{\begin{bmatrix} A(0,0)_{n_0 \times n_0} \end{bmatrix}}_{\pi(i)=0} \end{bmatrix}$$

Figure 1 shows a schematic representation of this permutation. In the right column blocks and bottom row blocks are districts that are not controlled by any gang. These exhibit no covariance with other districts because the only shootings that occur there are from residents, and these are i.i.d. across districts. Along the block-diagonal are districts owned by the same gang. Shootings within a gang's territory covary for two reasons. First, shocks to within-gang relations (ξ_k^t) are shared by all districts controlled by a given gang. Second, members of gang k operating in these districts share equally the risk of attacks

[4]Note also that this matrix is symmetric and positive definite.

that comes from *all* gang wars in which k is a belligerent ($\epsilon^t_{k,\ell}$). On the off block-diagonal are covariances produced through *specific* gang wars. For example, k, ℓ block of the matrix is positive whenever $\mathrm{E}[\epsilon^t_{k,\ell}] > 0$, or there is a positive probability of conflict between gangs k and ℓ. These reason that shootings in the districts controlled by gangs k and ℓ covary is because inter-gang shocks generate retaliatory violence (Assumption 3).

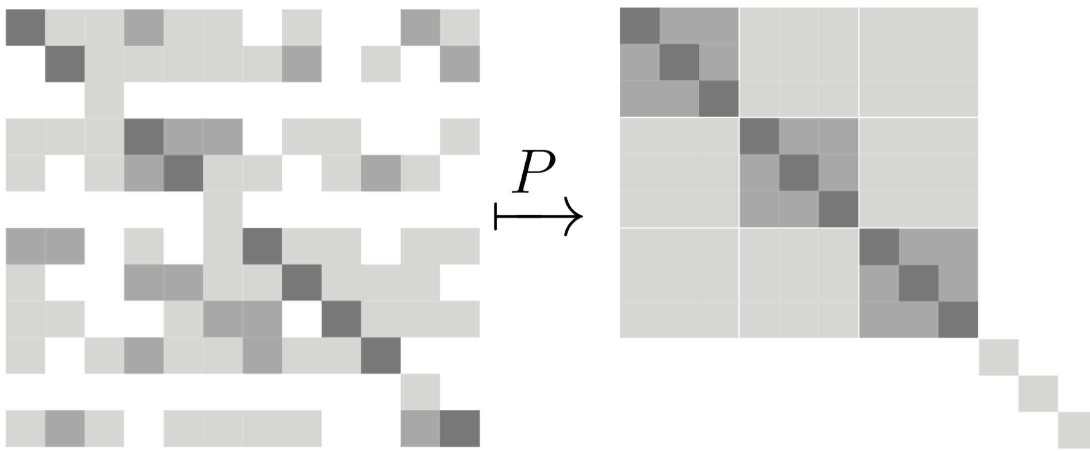

Figure 1: The input covariance matrix A is shown in the left panel. Applying the transformation PAP produces the block diagonal structure shown in the right panel.

This covariance matrix can be compactly represented as a function of our estimands, K and π. Let $\Psi = \text{diag}(\psi_1, \ldots \psi_N)$ and $Q = A - \Psi$. Let $B_{K+1 \times K+1} = (b_{k\ell})_{\{k,\ell \in \mathcal{K}\}}$ store the constant block covariance values defined in Corollary 1 and note that $b_{k0} = 0$ for all k. Finally, let $\Theta_{N \times K+1} = (\theta_{ik})_{\{i \in \mathcal{N}, k \in \mathcal{N}\}}$ be a membership matrix with $\theta_{ik} = 1$ if $\pi(i) = k$ and 0 otherwise. Then,

$$Q = \Theta B \Theta^T$$

.

Readers may recognize this structure as similar in form to a stochastic blockmodel (Holland, Laskey, and Leinhardt 1983). In such models, nodes are partitioned into groups and interact with members of other groups with some latent probability determined by their group membership. These latent probabilities can be expressed in a *connectivity matrix* akin to our B. If counts of these interactions are observed, the partition function and connectivity matrix can be estimated using spectral clustering (Jin 2015; Lei and Rinaldo 2015).

Here, we do not observe directly these interactions, and our B matrix does not have this simple interpretation. However, under the assumptions of our model, the spatial covariance in shootings mirrors the structure of the stochastic blockmodel, as in Trebbi and Weese (2019). We can therefore employ existing methods to estimate our model using these data.

Estimation

We will first show how to estimate the territorial partition, described by the matrix Θ, holding the number of groups, K, fixed. We will then proceed to estimate K using cross validation, following Chen and Lei (2018). Let $J = K + 1$ for convenience.

Territorial Partition

We observe the sample analogue to A,

$$\tilde{A} = \mathrm{E}[A] + \Phi$$

where $\Phi = (\phi_{ij})_{\{i,j \in \mathcal{N}\}}$ is a noise matrix with $\mathrm{E}[\phi_{ij}] = 0$ for all i, j. Note that

$$\begin{aligned} Q - \mathrm{diag}(Q) &= \mathrm{E}[A] - \mathrm{diag}(\mathrm{E}[A]) \\ &= \tilde{A} - \Phi - \mathrm{diag}(\mathrm{E}[A]) \\ \Phi - \mathrm{diag}(\Phi) &= \left(\tilde{A} - \mathrm{diag}(\tilde{A})\right) - (Q - \mathrm{diag}(Q)) \end{aligned}$$

Let $\mathbb{R}_{+}^{J\times J}$ be the set of all $J \times J$ symmetric matrices with non-negative entries, $\mathbb{D}^{J\times J}$ be the set of all $J \times J$ diagonal matrices and let $\mathbb{M}^{N\times J}$ be the set of all membership matrices.[5] A moment estimator for Θ and B satisfies

$$(\hat{\Theta}, \hat{B}) = \underset{B \in \mathbb{R}^{J\times J}, \Theta \in \mathbb{M}^{N\times J}}{\arg\min} \|\Phi - \mathrm{diag}(\Phi)\|_F \tag{1}$$

where $\|M\|_F = \left(\sum_i \sum_j M_{ij}^2\right)^{\frac{1}{2}}$ is the Frobenius norm.

We estimate these quantities using spectral clustering.[6] These methods exploit the eigenstructure of Q. If there are K gangs in the city, Q will have J positive eigenvalues. Let $\Delta = \mathrm{diag}(\sqrt{n_1}, \ldots, \sqrt{n_J})$ so that $\Delta B \Delta$ normalizes the connectivity matrix by the number of territories controlled by each group. Q can then be written as

$$\begin{aligned} Q &= \Theta B \Theta^T \\ &= \Theta \Delta^{-1} \Delta B \Delta \Delta^{-1} \Theta^T \\ &= \Theta \Delta^{-1} Z \Lambda Z^T \Delta^{-1} \Theta^T \\ &= \Theta X \Lambda X^T \Theta^T \end{aligned}$$

following Lei and Rinaldo (2015) (Lemma 2.1), where $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_J)$ stores the nonzero eigenvalues of the normalized connectivity matrix with $|\lambda_1| \geq \cdots \geq |\lambda_J| > 0$

[5] These have binary entries with rows summing to 1.

[6] Luxburg (2007) provides an overview of this family of methods.

and $Z_{N\times J}$ stores the associated eigenvectors. Therefore, $Z\Lambda Z^T = \Delta B\Delta$ is the eigendecomposition of the normalized connectivity matrix. Because $\Theta\Delta^{-1}$ is an orthonormal matrix, the rows of ΘX remain orthogonal and $Q = U\Lambda U^T$ is an eigendecomposition of Q with $U = \Theta X$.

The noise matrix Φ will distort the eigenvalues of $\tilde{A}$ away from zero. As $T \to \infty$, however, this noise matrix becomes small and the eigenvalues that take nonzero values due to noise will shrink toward zero. We therefore eigendecompose $\tilde{A} - \text{diag}(\tilde{A})$ into

$$\tilde{A} - \text{diag}(\tilde{A}) = \tilde{U}\tilde{\Lambda}\tilde{U}^T$$

with $\tilde{\Lambda} = \text{diag}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_J)$ and $|\tilde{\lambda}_1| \geq \cdots \geq |\tilde{\lambda}_J| > |\tilde{\lambda}_i|$ for $i \notin \{1, \ldots, J\}$. Then, the problem in 1 can be reformulated as

$$\begin{aligned}\left(\hat{\Lambda}, \hat{X}, \hat{\Theta}\right) &= \underset{\Lambda\in\mathbb{D}^{J\times J}, X\in\mathbb{R}^{J\times J}, \Theta\in\mathbb{M}^{N\times J}}{\arg\min} \|\tilde{U}\tilde{\Lambda}\tilde{U}^T - \left(\Theta X\Lambda X^T\Theta^T - \text{diag}(Q)\right)\|_F \\ &\approx \underset{\Lambda\in\mathbb{D}^{J\times J}, X\in\mathbb{R}^{J\times J}, \Theta\in\mathbb{M}^{N\times J}}{\arg\min} \|\tilde{U}\tilde{\Lambda}\tilde{U}^T - \Theta X\Lambda X^T\Theta^T\|_F\end{aligned}$$

Setting $\hat{\Lambda} = \tilde{\Lambda}$, the problem reduces to

$$\left(\hat{X}, \hat{\Theta}\right) = \underset{X\in\mathbb{R}^{J\times J}, \Theta\in\mathbb{M}^{N\times J}}{\arg\min} \|\Theta X - \tilde{U}\|_F \tag{2}$$

which can be solved via K-means clustering on the leading eigenvectors of $\tilde{A} - \text{diag}(\tilde{A})$ where Θ are the cluster memberships and X are the cluster centroids. An estimate for B can then be recovered as

$$\hat{B} = \hat{X}\hat{\Lambda}\hat{X}^T \tag{3}$$

Shootings in districts without gangs will exhibit no covariance in expectation with shootings in districts in which gangs operate, $\text{E}[b_{0k}] = 0$ for all $k \neq 0$. Once we have estimated B, we can therefore isolate the cluster corresponding to no gang activity by finding the row of B with the smallest values, formally

$$\min_{k\in\{1,\ldots,J\}} \|(B - \text{diag}(B))^{(k)}\|_2 \tag{4}$$

where $M^{(k)}$ is the kth row of M and $\|M^{(k)}\|_2$ is the Euclidean vector norm.

As discussed in the previous section, our model differs slightly from the stochastic block model. Where we observe between district covariance matrix, these models instead work with a binomial matrix of interaction counts between nodes (districts). Efforts to prove the consistency of spectral estimators therefore derive asymptotics as the number of nodes grows large.[7] Intuitively, the off-diagonal entries of our empirical covariance matrix

[7] Lei and Rinaldo (2015), for example, show that the spectral estimator is approximately consistent for Θ. As the number of groups grows large, the estimator misclassifies a vanishing proportion of nodes with probability approaching one.

converge to the off diagonal entries of Q as T grows large. In the limit, then $\tilde{U} \to \Theta X$ and K-means should not have trouble isolating distinct clusters in $\tilde{U}$. We rely on this heuristic for esimation, as in Trebbi and Weese (2019).

Number of Gangs

We rely on the cross-validation approach described in Chen and Lei (2018) to estimate the number of gangs operating in the city. For each trial $\tilde{K}$, this method iteratively splits the covariance matrix into V rectangular subsets for testing. It then estimates Θ and B on $V-1$ subsets and calculates the predictive loss on the square subset of the covariance matrix held out for testing. The $\tilde{K}$ that minimizes predictive loss is chosen as $\hat{J} = \hat{K} + 1$. Chen and Lei (2018) provide no theoretical guarantees against overestimating J and in practice, we find that predictive loss stochastically decreases as $\tilde{K}$ grows larger. We therefore select the first $\tilde{K}$ for which predictive loss does not decrease for $\tilde{K} + 1$ as our estimate for $\hat{J}$, averaged over many trial runs of the estimator. Let $\bar{L}_{\tilde{K}}(\tilde{A})$ be the average predictive loss on $\tilde{A}$ when $J = \tilde{K}$ and let $\delta = \{\delta_1, \ldots, \delta_{\bar{K}}\}$ be a sequence of changes in the predictive loss where $\delta_k = \bar{L}_k(\tilde{A}) - \bar{L}_{k+1}(\tilde{A})$. Our estimator for J selects

$$\hat{J} = \arg\min_k \{k \mid \delta_k < 0\}_{k \in \{1,\ldots,\bar{K}\}} \tag{5}$$

We now describe how this loss function is constructed. Let $\mathcal{V} = \{1, \ldots, V\}$ be the set of V cross validation folds, $\mathcal{N}_v \subset \mathcal{N}$ disjoint sets with $\cup_{v \in \mathcal{V}} \mathcal{N}_v = \mathcal{N}$, and $\mathcal{N}_{-v} = \cup_{u \neq v \in \mathcal{V}}$. Let $M^{(u,v)}$ denote the submatrix of M consisting of the rows in u and the columns in v.

We can construct estimates for Θ from a rectangular subset of $\tilde{A}$, $\tilde{A}^{(\mathcal{N}_{-v}, \mathcal{N})}$. As shorthand, let $\tilde{A}^{(-v,v)} = \tilde{A}^{(\mathcal{N}_{-v}, \mathcal{N})}$. Then,

$$Q^{(-v,v)} = \Theta^{(-v,v)} B \Theta$$

and

$$\begin{aligned}\left(Q^{(-v,v)}\right)^T Q &= \Theta B^T \left(\Theta^{(-v,v)}\right)^T \Theta^{(-v,v)} B \Theta^T \\ &= \Theta B^T \left(\Delta^{(-v,-v)}\right)^2 B \Theta^T\end{aligned}$$

. An eigendecomposition of this matrix (whose eigenvectors are the right singular vectors of $Q^{(-v,v)}$) can be clustered as above to produce estimates for Θ, which we'll call $\hat{\Theta}(v)$. Then, we can construct $\hat{B}(v)$ by averaging over off-diagonal values of the clusters of the rectangular covariance matrix (excluding the rows in $\mathcal{N}_v$)

$$\hat{B}_{k,\ell} = \begin{cases} \frac{\sum_{i \in \hat{\mathcal{N}}_{-v,k}, j \in \hat{\mathcal{N}}_\ell} \tilde{A}_{ij}}{\hat{n}_{v,k} \hat{n}_\ell} & \text{if } k \neq \ell \\ \frac{\sum_{i,j \in \hat{\mathcal{N}}_{-v,k}, i \neq j} A_{ij} + \sum_{i \in \hat{\mathcal{N}}_{-v,k}, j \in \hat{\mathcal{N}}_{v,k}} A_{ij}}{(\hat{n}_{-v,k} - 1)\hat{n}_{-v,k} + \hat{n}_{-v,k} \hat{n}_{v,k}} & \text{if } k = \ell \end{cases}$$

as in Chen and Lei (2018) Equation 5. Now we can create predicted values for A where

$$\hat{A}(v) = \hat{\Theta}(v)\hat{B}(v)\left(\hat{\Theta}(v)\right)^T$$

The predicted loss for the held out block of the covariance matrix can then be calculated as

$$L_v(\tilde{A}, \hat{A}(v)) = \left\|\left(\tilde{A}^{(v,v)} - \text{diag}(\tilde{A}^{(v,v)})\right) - \left(\hat{A}(v)^{(v,v)} - \text{diag}(\hat{A}(v)^{(v,v)})\right)\right\|_F$$

The average loss for a trial value $\tilde{K}$ is then

$$\bar{L}_k(\tilde{A}) = \frac{1}{V}\sum_{v=1}^{V} L_v(\tilde{A}, \hat{A}(v))$$

. A sequence δ can then be constructed for values of $k \in \left\{1, ..., \bar{K}\right\}$ allowing us to implement our estimator for J (Equation 5).

To summarize, our cross validation algorithm proceeds as follows:

1. For each $k \in \left\{1, ..., \bar{K}\right\}$,
    - Randomly split districts into folds $\mathcal{N}_1, \ldots, \mathcal{N}_V$.
    - For each fold, estimate $\hat{\Theta}(v)$ and $\hat{B}(v)$.
    - For each fold, calculate the predictive loss on $\tilde{A}^{(v,v)}$, $L_v(\tilde{A}, \hat{A}(v))$
    - Average the predictive loss across folds, $\bar{L}_k(\tilde{A})$.
2. Construct the sequence of changes in predictive loss, δ.
3. Select $\hat{J}$ using Equation 5.

In practice, we repeat this algorithm many times and choose the most frequent value for $\hat{J}$ as our estimate.

An alternative set of approaches to estimating J exploit the intuition discussed in the preceding subsection regarding the eigenvalues of $\tilde{A} - \text{diag}(\tilde{A})$. As $T \to \infty$, the eigenvalues associated with noise shrink toward zero while those associated with clusters remain positive. This generates a “eigengap” between the eigenvectors associated with true clusters and those associated with noise. Ahn and Horenstein (2013) investigate this inuition and construct an estimator for the number of factors in a similar class of models. In the next section, we show that this “eigengap” presents near our estimate for $\hat{J}$, consistent with this intuition.

Results

Conclusion

Appendices

Appendix A: Covariance Derivation

$$\begin{aligned}
\mathrm{Cov}[v_{it}, v_{jt}] =& \mathrm{E}[v_{it}v_{jt}] - \mathrm{E}[v_{it}]\mathrm{E}[v_{jt}] \\
=& \mathrm{E}[(x_{it} + y_{it})(x_{jt} + y_{jt})] - \mathrm{E}[x_{it} + y_{it}]\mathrm{E}[x_{jt} + y_{jt}] \\
=& \left(\mathrm{E}[x_{it}x_{jt}] + \mathrm{E}[x_{it}y_{jt}] + \mathrm{E}[x_{jt}y_{it}] + \mathrm{E}[y_{it}y_{jt}]\right) - \\
& \left(\mathrm{E}[x_{it}]\mathrm{E}[x_{jt}] + \mathrm{E}[x_{it}]\mathrm{E}[y_{jt}] + \mathrm{E}[x_{jt}]\mathrm{E}[y_{it}] + \mathrm{E}[y_{it}]\mathrm{E}[y_{jt}]\right) \\
=& \left(\mathrm{E}[x_{it}x_{jt}] - \mathrm{E}[x_{it}]\mathrm{E}[x_{jt}]\right) + \left(\mathrm{E}[y_{it}y_{jt}] - \mathrm{E}[y_{it}]\mathrm{E}[y_{jt}]\right) \\
=& \mathrm{E}\left[\left(\frac{m_{\pi(i)}}{n_{\pi(i)}}\xi^t_{\pi(i)} + \sum_{k\neq\pi(i)} \frac{m_k}{n_{\pi(i)}}\epsilon^t_{k,\pi(i)}\right)\left(\frac{m_{\pi(j)}}{n_{\pi(j)}}\xi^t_{\pi(j)} + \sum_{\ell\neq\pi(j)} \frac{m_\ell}{n_{\pi(j)}}\epsilon^t_{\ell,\pi(j)}\right)\right] - \\
& \mathrm{E}\left[\frac{m_{\pi(i)}}{n_{\pi(i)}}\xi^t_{\pi(i)} + \sum_{k\neq\pi(i)} \frac{m_k}{n_{\pi(i)}}\epsilon^t_{k,\pi(i)}\right]\mathrm{E}\left[\frac{m_{\pi(j)}}{n_{\pi(j)}}\xi^t_{\pi(j)} + \sum_{\ell\neq\pi(j)} \frac{m_\ell}{n_{\pi(j)}}\epsilon^t_{\ell,\pi(j)}\right] + \\
& \left(\mathrm{E}[y_{it}y_{jt}] - \mathrm{E}[y_{it}]\mathrm{E}[y_{jt}]\right) \\
=& \frac{m_{\pi(i)}}{n_{\pi(i)}}\frac{m_{\pi(j)}}{n_{\pi(j)}}\underbrace{\left(\mathrm{E}\left[\xi^t_{\pi(i)}\xi^t_{\pi(j)}\right] - \mathrm{E}[\xi^t_{\pi(i)}]\mathrm{E}[\xi^t_{\pi(j)}]\right)}_{\text{I: intra-gang}} + \\
& \sum_{k\neq\pi(i)}\sum_{\ell\neq\pi(j)} \frac{m_k}{n_{\pi(i)}}\frac{m_\ell}{n_{\pi(j)}}\underbrace{\left(\mathrm{E}\left[\epsilon^t_{k,\pi(i)}\epsilon^t_{\ell,\pi(j)}\right] - \mathrm{E}[\epsilon^t_{k,\pi(i)}]\mathrm{E}[\epsilon^t_{\ell,\pi(j)}]\right)}_{\text{II: inter-gang}} + \\
& \underbrace{\left(\mathrm{E}[y_{it}y_{jt}] - \mathrm{E}[y_{it}]\mathrm{E}[y_{jt}]\right)}_{\text{III: resident violence}}
\end{aligned}$$

We can derive the piecewise equation given in Proposition 1 by considering several cases. We start from the bottom of the piecewise stack. First, assume $i \neq j$ and $\pi(i) = 0$ or $\pi(j) = 0$. Then $\mathrm{E}\left[\xi^t_{\pi(i)}\xi^t_{\pi(j)}\right] - \mathrm{E}[\xi^t_{\pi(i)}]\mathrm{E}[\xi^t_{\pi(j)}] = 0$ by Assumption 1 and $\mathrm{E}\left[\epsilon^t_{k,\pi(i)}\epsilon^t_{\ell,\pi(j)}\right] - \mathrm{E}[\epsilon^t_{k,\pi(i)}]\mathrm{E}[\epsilon^t_{\ell,\pi(j)}] = 0$ by Assumption 3. $\mathrm{E}[y_{it}y_{jt}] - \mathrm{E}[y_{it}]\mathrm{E}[y_{jt}]$ because resident shootings are i.i.d. across districts. Therefore $\mathrm{Cov}[v_{it}, v_{jt}] = 0$.

Now consider $i \neq j$ and $\pi(i) \neq \pi(j)$ and $\pi(i), \pi(j) \neq 0$. $\pi(i) \neq \pi(j) \implies \mathrm{E}\left[\xi^t_{\pi(i)}\xi^t_{\pi(j)}\right] - \mathrm{E}[\xi^t_{\pi(i)}]\mathrm{E}[\xi^t_{\pi(j)}] = 0$ by Assumption 2. By Assumption 3, $\epsilon^t_{\pi(i),\pi(j)} = \frac{c_{\pi(j)}}{c_{\pi(i)}}\epsilon^t_{\pi(j),\pi(i)}$. By Assumption 4, $\mathrm{E}\left[\epsilon^t_{k,\pi(i)}\epsilon^t_{\ell,\pi(j)}\right] - \mathrm{E}[\epsilon^t_{k,\pi(i)}]\mathrm{E}[\epsilon^t_{\ell,\pi(j)}] = 0$ whenever $k \neq \pi(j)$ and $\ell \neq \pi(i)$. Therefore, $\mathrm{Cov}[v_{it}, v_{jt}] = \frac{m_{\pi(i)}}{n_{\pi(j)}}\frac{m_{\pi(j)}}{n_{\pi(i)}}\frac{c_{\pi(j)}}{c_{\pi(i)}}\mathrm{Var}[\epsilon^t_{\pi(i),\pi(j)}]$ where $\mathrm{Var}[\epsilon^t_{\pi(i),\pi(j)}] = \mathrm{E}\left[\left(\epsilon^t_{\pi(i),\pi(j)}\right)^2\right] - \mathrm{E}\left[\epsilon^t_{\pi(i),\pi(j)}\right]^2$.

Next, let $i \neq j$ and $\pi(i) = \pi(j)$. Here, $\mathrm{E}\left[\xi^t_{\pi(i)}\xi^t_{\pi(j)}\right] - \mathrm{E}[\xi^t_{\pi(i)}]\mathrm{E}[\xi^t_{\pi(j)}] = \mathrm{Var}[\xi^t_{\pi(i)}]$. By

Assumption 4, $\mathrm{E}\left[\epsilon^t_{k,\pi(i)}\epsilon^t_{\ell,\pi(j)}\right] - \mathrm{E}[\epsilon^t_{k,\pi(i)}]\mathrm{E}[\epsilon^t_{\ell,\pi(j)}] = 0$ whenever $k \neq \ell$. Therefore, the intergang sum condenses to

$$\left(\frac{m_k}{n_{\pi(i)}}\right)^2 \mathrm{Var}[\epsilon^t_{\pi(i),k}]$$

.

Finally, if $i = j$ then $\pi(i) = \pi(j)$. The within district variance is ψ_i. Otherwise, these districts inherit the covariance structure derived in the preceding paragraph. This yields the first component of the piecewise function.

References

Ahn, Seung C., and Alex R. Horenstein. 2013. "Eigenvalue Ratio Test for the Number of Factors." *Econometrica* 81 (3): 1203–27.

Chen, Kehui, and Jing Lei. 2018. "Network Cross-Validation for Determining the Number of Communities in Network Data." *Journal of the American Statistical Association* 113 (521): 241–51.

Holland, Paul W., Kathryn Blackmond Laskey, and Samuel Leinhardt. 1983. "Stochastic blockmodels: First steps." *Social Networks* 5 (2): 109–37.

Jin, Jiashun. 2015. "Fast Community Detection by SCORE." *The Annals of Statistics* 43 (1): 57–89.

Lei, Jing, and Alessandro Rinaldo. 2015. "Consistency of Spectral Clustering in Stochastic Block Models." *The Annals of Statistics* 43 (1): 215–37.

Luxburg, Ulrike von. 2007. "A tutorial on spectral clustering." *Statistics and Computing* 17 (4): 395–416.

Trebbi, Francesco, and Eric Weese. 2019. "Insurgency and Small Wars: Estimation of Unobserved Coalition Structures." *Econometrica* 87 (2): 463–96.