A Spatio-Temporal Mixture Model for Urban Crimes

Un Modello Mistura per Dati Spazio-Temporali Relativi alla Criminalità

Ferretti Angela, Ippoliti Luigi and Valentini Pasquale

Abstract This paper considers the determinants of severe crimes at the census-tract level in Pittsburgh, Pennsylvania. We develop a mixture panel data model to describe the number of severe crimes that allows for temporal as well as spatial correlation, together with significant heterogeneity across census tracts. We use traditional Bayesian mixtures admitting uncertainty about the number of groups. We focus on pooling regression coefficients across clusters, implying that census-tracts belonging to the same cluster are similar. The clustering is done in a data-based fashion.
Abstract *In questo articolo ci proponiamo di studiare le determinanti dei reati gravi verificatisi nei distretti della città di Pittsburgh (Pennsylvania). A tal fine, si propone una mistura di modelli di regressione per dati panel che consente di cogliere la correlazione temporale e spaziale, nonchè l'eterogeneità tra i distretti. Assumendo come incognito il numero delle componenti della mistura, il modello consente di pervenire ad una classificazione in cui i gruppi si distinguono per diversi profili di covariate.*

Key words: Crime places, Distribution dynamics, Spatio-temporal models, Dirichlet process

---

Ferretti Angela
DeC, University "G.d'Annunzio" of Chieti-Pescara, e-mail: angela.ferretti29@gmail.com

Ippoliti Luigi
DeC, University "G.d'Annunzio" of Chieti-Pescara e-mail: luigi.ippoliti@unich.it

Valentini Pasquale
DeC, University "G.d'Annunzio" of Chieti-Pescara e-mail: pasquale.valentini@unich.it

1 Introduction

Any individual behaviour is a product of interaction between the person and the setting [1]. In recent years, the spatio-temporal urban distribution of crimes is receiving growing attention not only from researchers (criminologists, sociologists, economists, geographers, etc) but also from law enforcement agencies. In particular, in [11] it is highlighted a need to "integrate geographic and temporal representation and analyses" and in [10] it is stated that "the most under-researched area of spatial criminology is that of spatio-temporal crime patterns". In this paper, we aim at addressing such needs by proposing a mixture panel data model for high-dimensional urban crime count data. The model allows to include temporal and spatial effects, socio-economic census tract characteristics and random effect components to take care of the heterogeneity existing across census tracts. Additionally, our model extends a "traditional" panel data model in many ways. For example, model coefficients are pooled across clusters, implying that census tracts belonging to the same cluster are similar. However the number of clusters is not known in the data but it is determined by the Dirichlet process.

The paper is organized as follows. In section 2 we present our model and the econometric methodology. In section 3 we describe our empirical dataset and discuss the empirical results.

2 Model Specification

Let y_{it} be the number of Part I offenses in census tract i $(i = 1,...,N)$ at time t $(t = 1,...,T)$. Let us assume that, conditionally on the mean, the y_{it}'s are mutually independent with Poisson distribution, $y_{it} \sim Po\,(\exp(\eta_{it}))$. The logarithm of the conditional mean is given as follows:

$$\eta_{it} = \nu + \phi\eta_{it-1} + \rho\sum_{l=1}^{N} W_{il}\eta_{lt} + x_{it}'\beta + \delta_{it} \quad (1)$$

where x_{it} is a vector including strictly exogenous variables, ν is the intercept, β is a vector of regression parameters, η_{it-1} is the temporally lagged value of η_{it}, W_{il} is a generic element of a matrix W reflecting contiguity relations between the N census tracts, ρ is a scalar parameter reflecting the strength of spatial dependence, and δ_{it} can be related to a set of common determinants as $\delta_{it} = \xi_t'\gamma_i + \varepsilon_{it}$ where $\varepsilon_{it} \sim N(0,\sigma^2)$. The linear combination $\xi_t'\gamma_i$ transfers the contemporaneous correlation from the errors to the conditional expectation part of the model. Since in our empirical analysis we choose to include the first latent factor (principal component) as common regressor ξ_t, γ_i can be interpreted as a parameter vector of factor loadings specific for each census tract.

We consider a generalization of model (1) as used in [9], where the number of Part I offenses is modeled as a mixture of C^* unobserved clusters, whose coefficients

are pooled only across census tracts having similar characteristics. Particularly, the cluster-specific coefficients imply that census tracts belonging to the same cluster are defined by common effects, while census tracts belonging in different clusters have structural differences in their severe crimes' determinants.
Let $r_i = j$ index that census tract i belongs to cluster j, for $j = 1,...,C^*$ clusters in total, and with a respective probabilities $\pi_{i1} = P[r_i = 1], \pi_{i2} = P[r_i = 2], ..., \pi_{iC^*} = P[r_i = C^*]$, where $0 \leq \pi_{i1}, \pi_{i2}, ..., \pi_{iC^*} \leq 1$ and $\sum_{j=1}^{C^*} \pi_{ij} = 1$. The model we use can be written as

$$\eta_{it} = \nu_j + \phi_j \eta_{it-1} + \rho \sum_{l=1}^{N} W_{il} \eta_{lt} + x'_{it} \beta_j + \delta_{it} \tag{2}$$

$$if \quad r_i = j \qquad\qquad j = 1,...,C^*;$$

where ϕ_j and $\beta_j, j = 1,2,...,C^*$ are cluster specific coefficients.

2.1 Dirichlet Process Mixture Model

One of features of model (2) is the number of clusters C^*. In order to address to this question, in this paper we adopt a truncated Dirichlet process model to define the prior over the mixing probabilities based on some (large) upper bound C (see [4]).

Denote the respose variable $\eta_i = (\eta_{i1}, \eta_{i2}, \ldots, \eta_{iT})'$ and a set of covariates $z_{it} = \left(1, \eta_{it-1}, \sum_{l=1}^{N} W_{il} \eta_{lt}, x'_{it}\right)'$ observed at time t for the ith individual. With no loss of generality here, we rewrite model (2) as $\eta_{it} = z'_{it} \tau_j + \varepsilon_{it}$ where $\tau_j = (\nu_j, \rho, \phi_j, \beta'_j)'$ and $\delta_{it} = \varepsilon_{it}$. Then the density of the mixture is

$$f(\eta_i | z, \theta) = \sum_{j=1}^{C} \pi_j \left(\prod_{t=1}^{T} f_j \left(\eta_{it} | z_{it}, \psi_j \right) \right)$$

where $\theta = \{\alpha, \pi_{1:C}, \psi_{1:C}\}$, $\psi_j = \{\tau_j, \sigma_j^2\}$, $\sum_{j=1}^{C} \pi_j = 1$ with $0 \leq \pi_j \leq 1$, f_j are the C component densities and α is a precision parameter of the Dirichlet process. The mixture model can be realized through the configuration indicators r_i for each observation η_i with prior $P(r_i = j | \pi) = \pi_j$, so that we obtain the standard hierarchical model:

$$(\eta_i | r_i = j, z_i, \psi_j) \sim f_j(\eta_i | \psi_j). \qquad (\psi_j | G) \sim G \qquad (G | \alpha, G_0) \sim DP(\alpha, G_0). \tag{3}$$

where $G(\cdot)$ is an uncertain distribution function, $G_0(\cdot)$ is the prior mean of $G(\cdot)$ and $\alpha > 0$ the total mass, or precision of the DP. From the *Pólya Urn Scheme*,

$$\psi_j | \psi_1, \psi_2, ..., \psi_{j-1} \sim \frac{\alpha}{j-1+\alpha} G_0(\cdot) + \frac{1}{j-1+\alpha} \sum_{k=1}^{j-1} \delta_{\psi_k}(\cdot) \tag{4}$$

where $\delta_{\psi_k}(\cdot)$ is the point mass distribution at ψ_k. The truncated Dirichlet process prior is such that

$$\pi_1 = V_1, \qquad \pi_j = V_j \times \prod_{i=1}^{j-1}(1-V_i), \tag{5}$$

$j > 1$, where V_i has a Beta distribution $Be(1,\alpha)$, $i < C$ indipendently over i and $V_C = 1$. Prior specification for each component j ($j = 1,\dots,C$) is completed with the following distribution,

$$G_0(\tau_j, h_j^{-1}) = N(\tau_j|\tau_0, h_j^{-1})Ga(h_j|a,b) \tag{6}$$

where $h_j^{-1} = \sigma_j^2$; and with a Gamma prior $\alpha \sim Ga(\varsigma_1, \varsigma_2)$. Placing a prior on α ([4]) allows us to draw inferences about the number of mixture components through the role of α of the *Pólya Urn Scheme* as the prior number of observations in each component.

3 Application

In this section we apply model (2) to study the determinants of census tract severe crimes in Pittsburgh, Pennsylvania. We first describe the full dataset and next we give details on empirical results.

3.1 Data

The crime dataset that we used includes monthly (January 2008 to December 2013) counts of Part I and Part II offenses for each of the 138 2000 census tracts in Pittsburgh, Pennsylvania. Part I offenses, also known as index crimes, regroup serious felonies in the following eight categories: criminal homicide, forcible rape, robbery, aggravated assault, burglary, larceny-theft (except motor vehicle theft), motor vehicle theft and arson. Part I offenses consist of the number of offenses in these categories that are known to law enforcement. Part II offenses include 21 categories of non-serious felonies and misdemeanors for which only arrest data were collected. A more detailed description of these variables are provided in [7].

The dependent variable in our study, y_{it}, is the number of Part I offenses in census tract i for $i = 1,\dots,138$ in month t for $t = 1,\dots,72$. Potential covariates include the log of number of Part I offenses in census tract at time $t-1$, the log of Part II offenses lagged by 1 month as leading indicator and the spatially lagged state variable. In addition, in order to account for heterogeneity across census tracts, we collected data on the following 15 time-invariant socio-economic variables from the Census 2000 (US Census Bureau and Social Explorer Tables): log of median income (*Lmi*), civilian unemployment rate (*Cur*), poverty rate (*Pvr*), percentage of population with less than a high school degree (*Hdl*), percentage of population with a bachelor degree or higher (*Bdh*), rental housing units as percentage of occupied housing units (*Rhu*), percentage of households having been in the same house for more than 1 year (*Sh1*), percentage of female-headed households (*Fhh*), housing units vacancy rate (*Hvr*), percentage of total population that is African-American (*Paa*), log of total population (*Ltp*), log of population density per square mile (Lpd), dropout rate age 16–19 (*Dra*), percentage of total population under 18 ($U18$) and group quarter proportion (*Gqp*). Finally, missing values for our socio-economic covariates in 14 census tracts that do not have a regular resident population were replaced by dummies.

3.1.1 Results

By means of a Markov chain Monte Carlo approach, posterior inference was based on the last $50,000$ draws (after a burn-in of $5,000$) using every 5th member of the chain to avoid autocorrelation within the sampled values. From the computational viewpoint, we first sample η_{it} from its marginal distribution using the adaptive rejection sampling [2] and then we draw all the other parameters. Conditional on η_{it}, the full conditional posterior distributions take convenient functional forms and can be easily sampled from. Convergence of the chains of the model was monitored visually through trace plots as well as using the R-statistic of [3] on two chains simultaneously started from different initial points.

Results indicate the existence of two clusters of census tracts in Pittsburgh, with 42 census tracts (30%) belonging to cluster 1 and 96 census tracts (70%) belonging to cluster 2. Also, we note that the temporal correlation parameters referring to the Part II crime data, only impact in cluster 1. This suggests that crime hot spots may arise first as a concentration of soft crimes that later hardens into more serious crimes. Consistently, if a large number of Part I offenses happen at time $t-1$, a huge number of the same crimes will occur at time t. The spatial dependence in Part I offenses is relevant in both clusters, so we do have a spatial diffusion of certain types of crime in Pittsburgh. This result could provide a useful tool for efficient allocation of law enforcement resources.

Of the 15 socio-economic determinants of severe crimes, only the *Cur* seems to be relevant in both clusters. The positive influence of civilian unemployment rate on the number of Part I offenses confirms the social organization theory according to which bad socio-economic conditions, such as job unavailability, give rise to criminal motivation. For the rest of potential determinants we find that they can have an impact in one cluster but not the other. While *Ldp* and *Fhh* do not provide any impact in group 1, they become important in group 2. Here, census tracts with a small population size or lack of residential instability should enjoy lower number of Part I offenses. Furthermore, the variables *Lmi*, *Pvr* and *U18* appear with a negative sign in cluster 2. In contrast, *Hvr* and *Hdl* represent important determinants for the Part I offenses with positive and negative signs, respectively, only in cluster 1.

Overall, this study shows that criminal dynamics have different features across the two clusters, with differences which cannot be captured by traditional regression analyses.

References

1. Felson, M., Clarke, R.V.: Opportunity Makes the Thief: Practical Theory for Crime Prevention. In: Police Research Studies. London (1998)
2. Gelman, A.: Inference and Monitoring Convergence. In Gilks, W.R and Richardson, S. and Speigelhalter, D. (eds). Chapman & Hall:Boca Raton; 131-143 (1996)
3. Gilks, W.R. and Wild, P. : Adaptive Rejection sampling for Gibbs Sampling. Journal of the Royal Statistical Society, Series C, **41**, 337–348 (1992)
4. Ishwaran, H., James, L.: Approximate Dirichlet process computing in finite normal mixture: Smoothing and prior information. J. Comput. Graph. Stat. **11**, 508–532 (2002)
5. Kikuchi, G.: Neighborhood Structures and Crime: A Spatial Analysis. El Paso, Texas (2010)
6. Korobilis, D., Gilmartin, M.: On regional unemployment: An empirical examination of the determinants of geographical differentials in the UK. Scott. J. Polit. Econ. (2012) doi: 10.1111/j.1467-9485.2011.00575.x
7. Liesenfeld, R., Richard, J.F., Vogler, J.: Likelihood-based inference and prediction in spatio-temporal panel count models for urban crime. J. Appl. Econ. (2016) doi: 10.1002/jae.2534
8. Manolopoulou, I., Chan, C., West, M.: Selection sampling from large data ses for targeted inference in mixture modeling. Bayesian Anal. **3**, 429–450 (2010)
9. Paap, R., Franses, P.H., van Dijk, D.: Does Africa grow slower than Asia, Latin America and the Middle East? Evidence from a new data-based classification method. J. Dev. Econ. **77**, 553–570 (2005)
10. Ratcliffe, J.H.: Crime mapping: spatial and temporal challenges. Handbook of Quantitative Criminology, Piquero, A.R., Weisburd, D. (eds). Springer: New York, 524 (2013)
11. Roth, R.E., Ross, K.S., Finch, B.G., Luo, W., MacEachren, A.M.: Spatiotemporal crime analysis in U.S. law enforcement agencies: current practices and unmet needs. J. Gov. Inf. **30**, 226–240 (2013)