

# Spatial Beta-Convergence Forecasting Models:

Evidence from Municipal Homicide Rates in Colombia

Felipe Santos-Marquez  
Master's student (M2)  
Graduate School of International Development  
Nagoya University, JAPAN

Prepared for the 57th Annual Conference of JSLA

[ slides available at: <https://felipe-santos.rbind.io/>  
& <https://jsla-2020.netlify.app/> ]

## Motivation:

- Policy development and management should consider not only current challenges but also the potential challenges in the near future.
- Persistent income differences, differences in health indicators and in "general" regional inequality in Colombia.
- Scarce academic literature on the **forecasting of crime** at the municipal level in Colombia.

## Research Objective:

- To build a forecasting model that can predict homicide rates 4 years into the future
- Analyze the spatial distribution of the forecasted homicide rates in 2022

## Methods:

- Exponential smoothing
- ARIMA models
- STAR models
- Classical convergence framework (Barro and Sala-i-Martin 1992)

## Main Results:

1. **Beta and Spatial-beta forecasting models** are more accurate than other conventional forecasting models.
2. **768 out of 1120 Colombian municipalities** are expected to reach the 2022 national SDG target. (**23 extra municipalities when compared to 2018**)
3. **Increasing spatial polarization of the data.** higher crime areas are more likely to be surrounded by high crime areas and conversely for low-low cluster formation.
4. **Crime spillovers are forecasted** for municipalities in the pacific region and positive spillovers are forecasted in the north and north-eastern parts of the country.

# **Outline of this presentation**

**1. Data description** administrative levels in Colombia and homicide rates.

**2. Forecasting models**

- Exponential smoothing, ARIMA and STAR models

**3. Beta convergence models**

- Beta convergence and forecasting

**4. Main results and prediction:**

- Best models and forecasted data

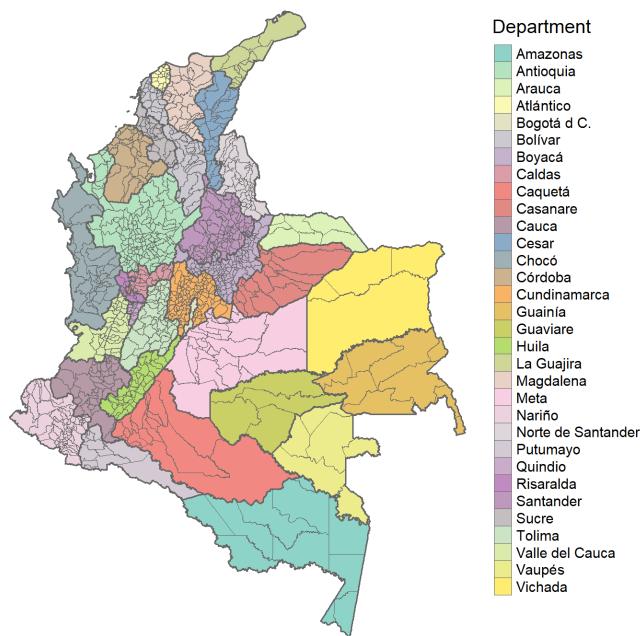
**5. Policy discussion**

- The Colombian National Development Plan 2018-22

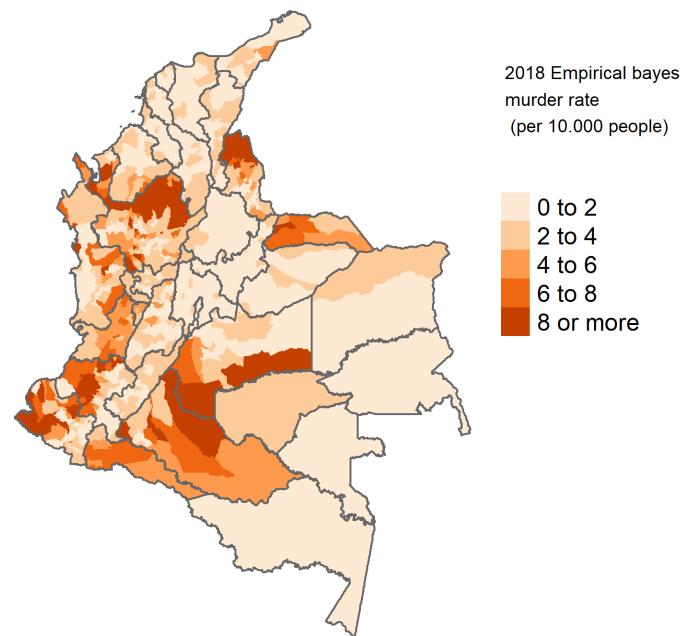
**6. Concluding Remarks**

# Colombian administrative levels and the spatial distribution of crime

States and Municipalities



Crime



(In Japan in 2018 about 0.02 per 10,000 people)

## Data:

- Total number of **homicides** and in Colombia per year from 2003 until 2018 (**data taken from Municipal data-set Universidad de los Andes, Bogota, Colombia**).
- Data is aggregated at the municipal and departmental levels.
- Population census and estimates for states and municipalities (data from the same dataset).
- Raw rates computed

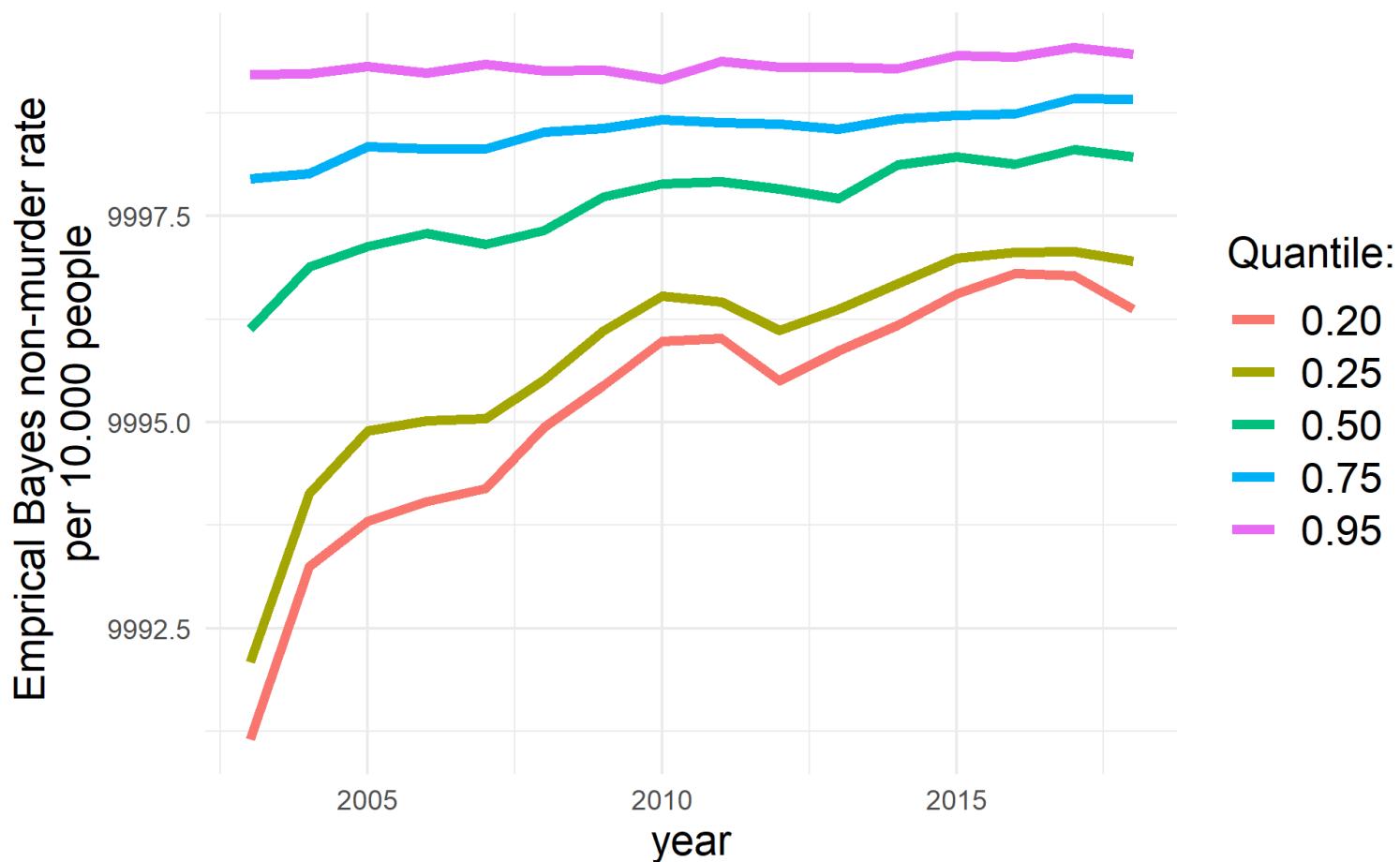
$$\text{raw rates} = \text{crimes}/\text{population}$$

- Non crime rates computed

$$NCR = 10000 - \text{raw rate} * 10000$$

- **Survival rates** are chosen because positively defined variables are a **standard** in the convergence literature.
- The variables are transformed from raw rates to Empirical Bayes rates in order to control for variance instability.

# Global trend of crime



# Exponential smoothing methods

In general terms, these methods use a **weighted average of past observations** in order to draw a forecast of future observations

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1 - \alpha)y_{T-1} + \alpha(1 - \alpha)^2y_{T-2} + \dots$$

# ARIMA models

An autoregressive model of order p

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

A moving average model considers past forecast errors in a linear model:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

An autoregressive model, a moving average model and the differentiating of the data can be combined into a non-seasonal ARIMA model:

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

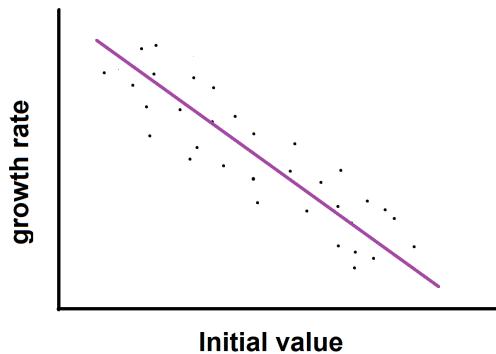
## Space time Autoregressive models (STAR)

$$y_{it} = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \phi_{kl} \sum_{j=1}^N w_{ij}^{(l)} y_{jt-k} + a_{it}$$

- $y_{it}$  is the variable of interest at location  $i$  during time  $t$
- $a_{it}$  is the random normal error associated with that location and time.
- $y_{jt-k}$  are the time lagged variables at location  $j$
- $w_{ij}^{(l)}$  is the distance weight matrix of order (l).

These models have been **successfully used to forecast crime data in the United States**. See Shoesmith GL (2013) Space-time autoregressive models and forecasting national, regional and state crime rates.

# Beta convergence and forecasting (catch-up process)



When studying regional data sets, Barro and Sala-i Martin (1995) proposes the following equation:

$$\log \frac{y_{iT}}{y_{i0}} = \alpha - [1 - e^{-\beta T}] \cdot \log(y_{i0}) + w_{i,0T}$$

Solving the equation for  $y_{iT}$ :

$$\log(y_{iT}) = \alpha + e^{-\beta T} \cdot \log(y_{i0}) + w_{i,0T}$$

Therefore, a plausible 4-year ahead forecast is:

$$\log(\hat{y}_{i(t+4)|t}) = \hat{\alpha}_{t+4|t} + \hat{\beta}_{t+4|t} \cdot \log(y_{it}) + w_{i,t}$$

# Spatial Beta convergence model

- Spatial beta convergence models have been used to model the dynamics of crime (**Santos-Marquez and Mendez, 2020**)
- In order to simplify the computations instead of a SAR (spatial autoregressive model) or a SEM (spatial error model), **an SLX model (Spatially Lagged X)** is proposed to **forecast homicide**.
- the spatial equivalent of the equation in the previous slide is:

$$\log \frac{y_{iT}}{y_{i0}} = \alpha - [1 - e^{-\beta T}] \cdot \log(y_{i0}) + \theta W \cdot \log(y_{i0}) + \epsilon_{i,0T}$$

Therefore, a plausible 4-year ahead forecast is:

$$\log(\hat{y}_{i(t+4)|t}) = \hat{\alpha}_{t+4|t} + \hat{\beta}_{t+4|t} \cdot \log(y_{it}) + \hat{\theta}_{t+4|t} W \cdot \log(y_{it}) + \epsilon_{i,t}$$

Where three coefficients need to be computed:  $\alpha$ ,  $\beta$  and  $\theta$ .

# Time series forecast accuracy and cross-validation

## How can the best forecasting model be chosen?

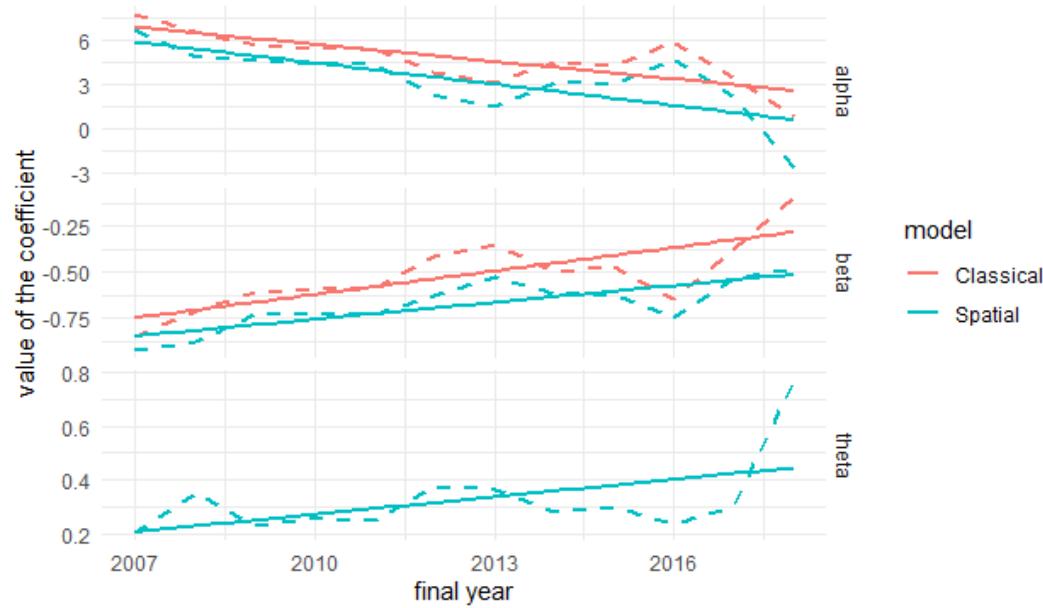
Mean absolute error:  $MAE = \text{mean}(|e_t|)$

Root mean squared error:  $RMSE = \sqrt{\text{mean}(e_t^2)}$

## Cross-validation of forecasts



# The prediction of the coefficients of the beta models



Linear regressions are used to compute the coefficients in:

$$\log(\hat{y}_{i(t+4)|t}) = \hat{\alpha}_{t+4|t} + \hat{\beta}_{t+4|t} \cdot \log(y_{it}) + \hat{\theta}_{t+4|t} W \cdot \log(y_{it}) + \epsilon_{i,t}$$

$$\log(\hat{y}_{i(t+4)|t}) = \hat{\alpha}_{t+4|t} + \hat{\beta}_{t+4|t} \cdot \log(y_{it}) + w_{i,t}$$

# Cross-validation results (Beta and spatial Beta)

**Table 1** Forecast accuracy cross validation

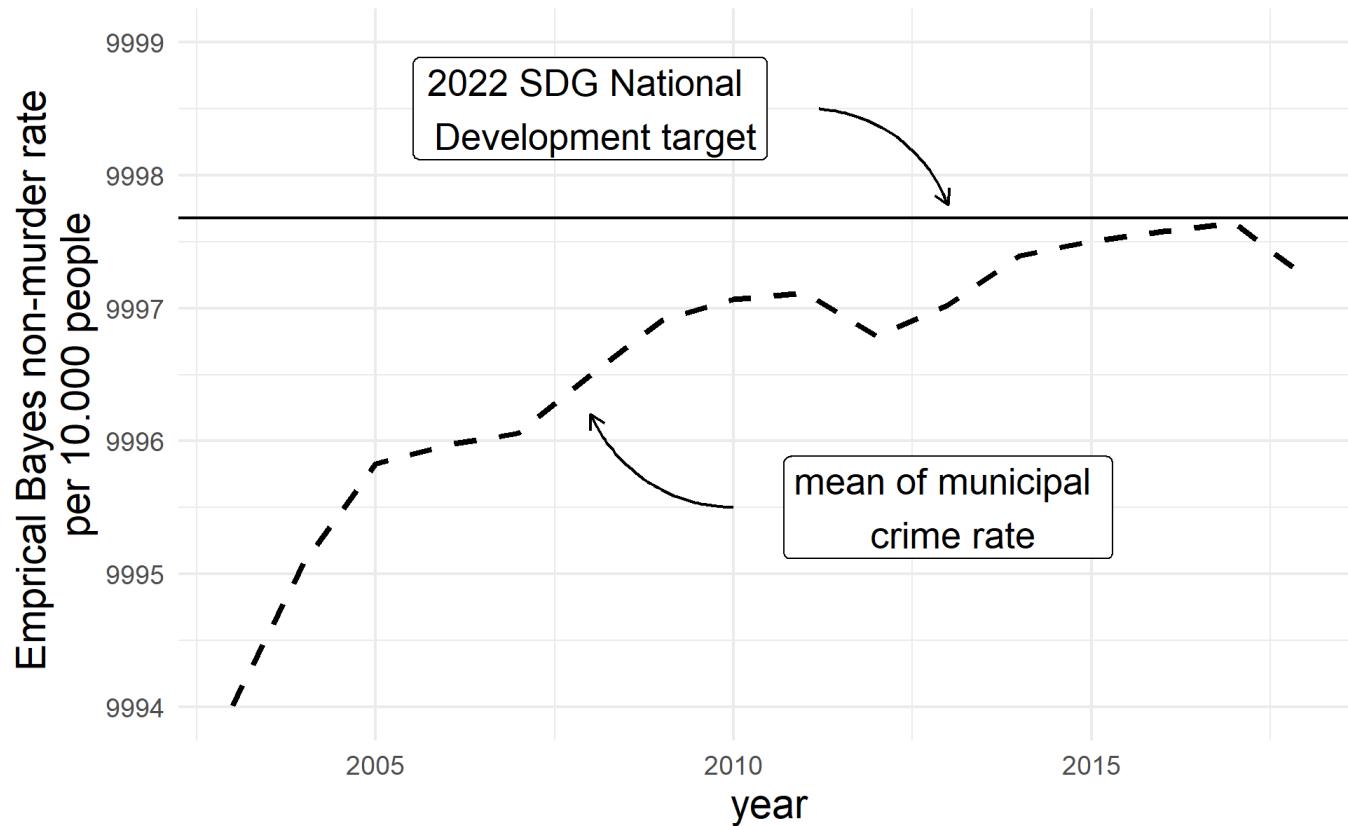
method	2011	2012	2013	2014	2015	2016	2017	2018
ETS (MAE)	2.83	2.31	1.49	1.46	1.43	1.38	1.37	1.42
ETS (RMSE)	5.74	4.66	2.40	2.36	2.33	2.44	2.14	2.66
ARIMA (MAE)	2.34	1.94	1.54	1.57	1.60	1.51	1.53	1.50
ARIMA (RMSE)	3.68	3.43	2.45	2.49	2.66	2.56	3.09	2.79
BETA (MAE)	1.83	2.36	2.21	1.84	1.51	1.19	1.02	1.31
BETA (RMSE)	2.29	3.42	2.66	2.32	2.14	2.18	1.72	2.80
Spatial B (MAE)	1.63	2.53	1.94	1.56	1.31	1.15	1.03	1.23
Spatial B (RMSE)	2.14	3.50	2.39	2.05	1.97	1.97	1.67	2.65
STAR (MAE)	2.49	2.46	2.19	2.23	2.37	2.32	2.37	2.59
STAR (RMSE)	2.80	3.31	2.55	2.53	2.76	2.71	2.79	3.54

**Table 2** Forecasts yearly cross validation

method	MAE	RMSE
ETS	1.71	3.34
ARIMA	1.69	2.93
Beta	1.66	2.49
Beta Spatial	1.55	2.35
STAR	2.38	2.89

# Forecasting homicide rates for 2022

- In the Colombian National Development Plan 2018-2022 DNP (2018) there are national targets for a variety of SDG related variables for the year 2022.
- In terms of homicides the baseline for 2018 is 2.58 (per 10.000 people) and the 2022 target is 2.32 per 10.000 people (nmr= 9997.68)



# Forecasting homicide rates for 2022

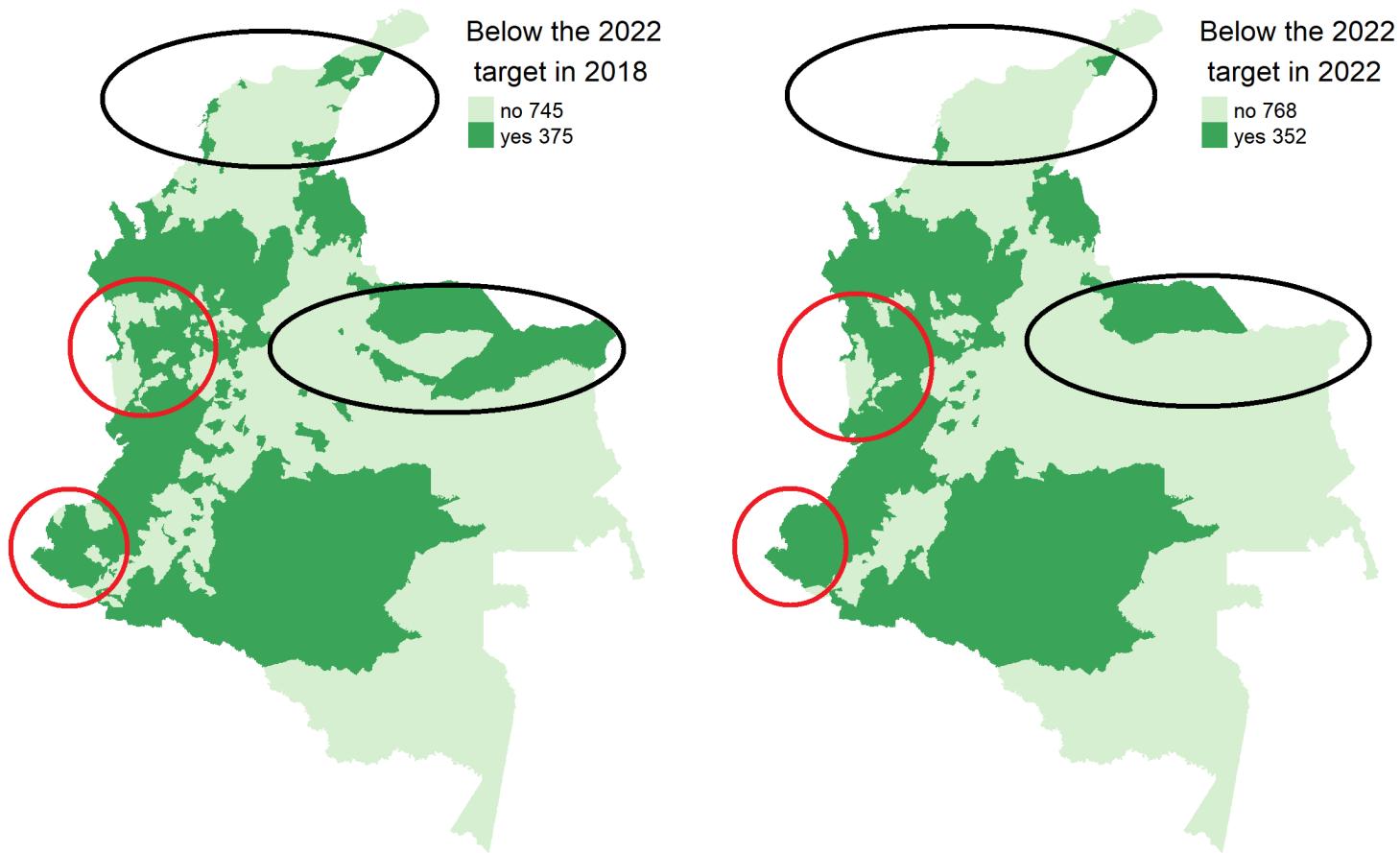
## (2022 Target = 9997.68)

	nmr 2003	nmr 2018	nmr 2022
Mean	9994.01	9997.23	9997.61
Std.Dev	7.08	3.29	3.04
Min	9885.76	9954.94	9963.42
Median	9996.13	9998.22	9998.58
Max	9999.82	9999.84	10000.41

A noticeable error in this forecasting model is the fact that a total of 68 municipalities have a predicted non crime rate slightly above 10000 (WHICH IS PRACTICALLY IMPOSSIBLE)

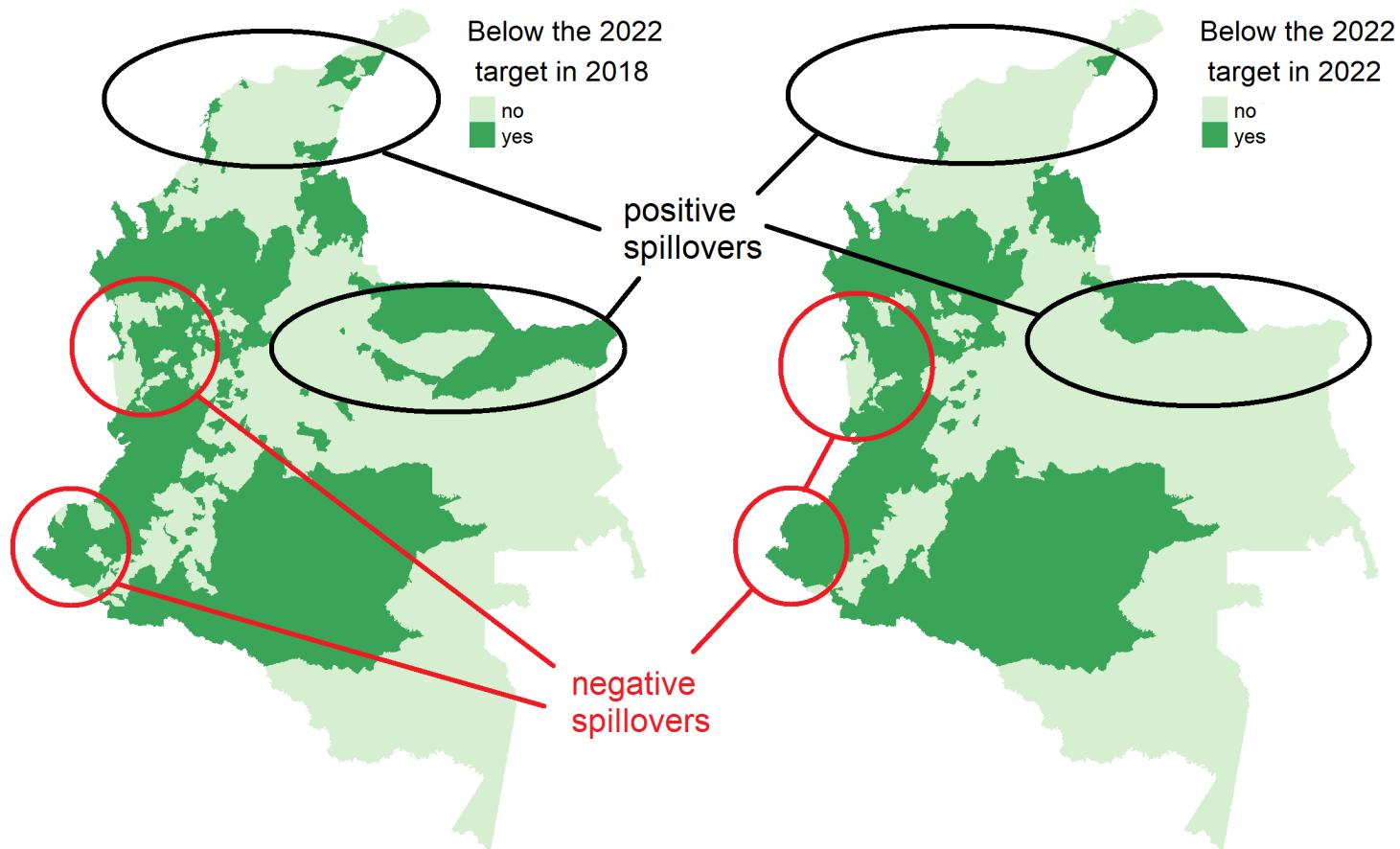
## The spatial distribution of the forecasted crime

municipalities in dark green are lagging behind



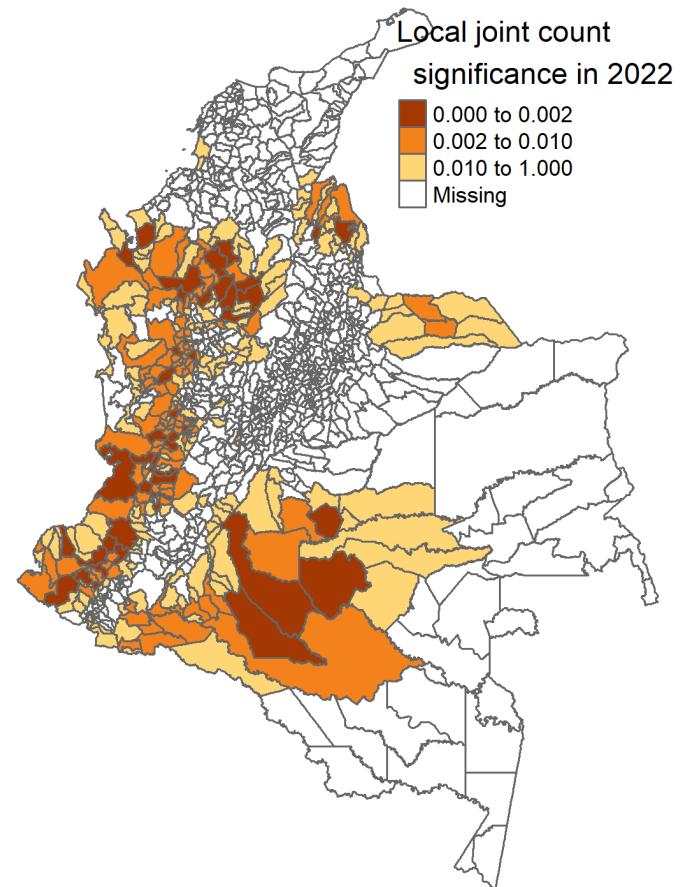
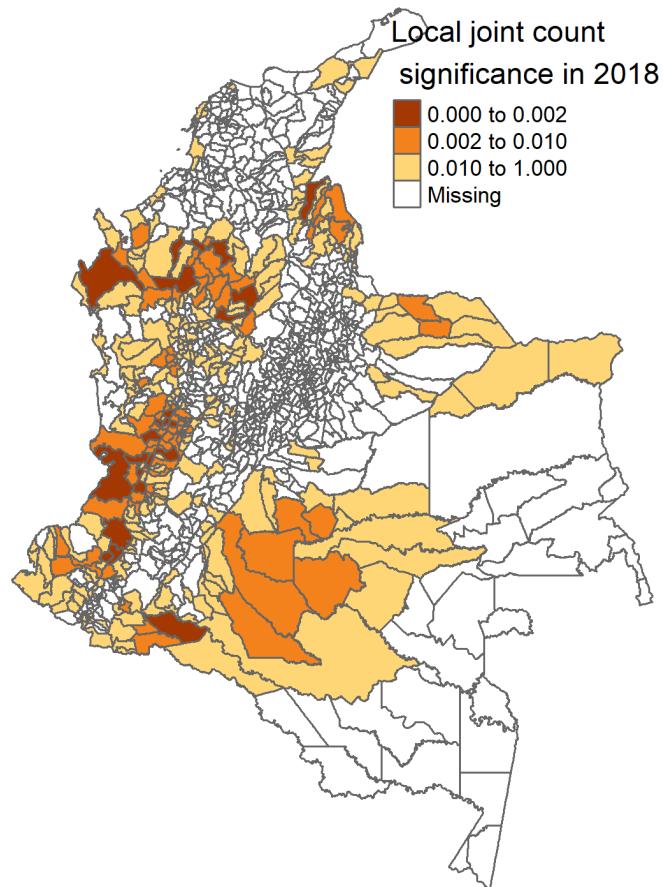
## The spatial distribution of the forecasted crime

municipalities in dark green are lagging behind



# Measuring the polarization of regional crime

Local join count statistic (Anselin (1995), Anselin and Li (2019))



# Policy discussion

- Vertical and horizontal policy coordination, spillovers and borders.
- Spatial spillovers from neighbors can have **both positive and negative effects**.
- It could be more appropriate for the formulation of development plans to have targets at the **state level or regional level**.
- For example, there is a clear negative spill-over in the municipalities in the **pacific region (in purple)**.

## Colombian natural regions



# Concluding Remarks

Uplifting results "on average":

- The dispersion of non-crime (crime) rates at the has decreased level **has decreased**.
- It was found that the forecasted mean in 2022 is still below the 2022 target.
- Nevertheless, major improvements are foreseen as **23 extra municipalities are expected to reach the SDG homicide target by 2022**. This will bring the total of municipalities that have accomplished the target to 768 out of 1120 in 2022.

## The Role of Space

- There are **positive and negative spillovers**.
- **Policy should be localized in order to reverse the creation of crime clusters.**

## Further Research

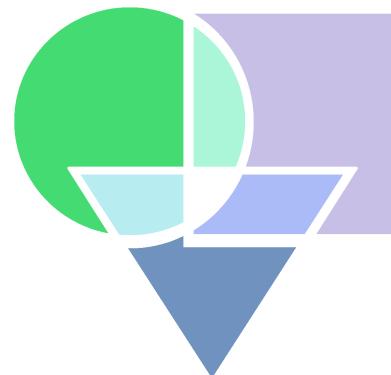
- Other determinants may help to improve the forecasting accuracy of the beta models.
- A **damped trend method** may help to correct for municipalities that are forecasted to have 0.0 crime rates (over 10.000 non crime rates)

# Thank you very much for your attention

You can find more about my research outputs on my website <https://felipe-santos.rbind.io>

If you are interested in our research please check the QuaRCS lab website

<https://quarcs-lab.org/>



QuaRCS lab

Quantitative Regional and Computational Science Lab