Multivariate Temporal Modeling of Crime with Dynamic Linear Models

Jarad Niemi and Nate Garton

Iowa State University

August 6, 2020

Funded, in part, by the Center for Statistics and Applications in Forensic Evidence (CSAFE):
https://forensicstats.org/

Outline

- Background
  - Hypothesis: burglaries $\rightarrow$ criminal trespass
- Chicago Crime Data
- Multivariate DLM
  - Linear trend
  - Fourier form seasonal model
- Bayesian Estimation in Stan
- Results
  - Standard crime trend analysis
  - Posterior (partial) correlations
- Discussion

Background

https://www.chicagomag.com/Chicago-Magazine/May-2014/Chicago-crime-rates/:

- Anecdotal evidence of misreporting
- Pressure to down-grade crimes
- Selective reporting

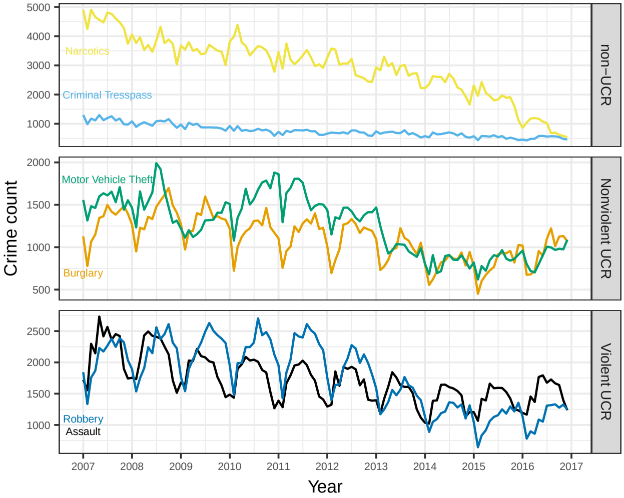
https://www.economist.com/democracy-in-america/2014/05/22/deceptive-numbers: specific hypothesis:

> *burglaries (tracked in UCR) are being misclassified as criminal trespasses (non-UCR)*

FBI's Uniform Crime Reporting (UCR) program:

> *[crime] data received from more than 18,000 city, university and college, county, state, tribal, and federal law enforcement agencies voluntarily participating in the program.*

Chicago crime data



Multivariate dynamic linear model

Let $Y_t = (Y_{1,t}, \ldots, Y_{C,t})$ where $Y_{c,t}$ is the log count for crime type c in month t.

Multivariate dynamic linear model:

$$\begin{array}{lrll} \text{observation:} & Y_t & = F_t\theta_t + \epsilon_t, & \epsilon_t \stackrel{ind}{\sim} N_6\left(0, \Sigma_\epsilon\right) \\ \text{evolution:} & \theta_t & = G_t\theta_{t-1} + \delta_t, & \delta_t \stackrel{ind}{\sim} N_*\left(0, \Sigma_\delta\right). \end{array}$$

where

$$\begin{array}{llll} F_t & = F & = & I_{6\times 6} \otimes (1,0,1,0,...,1,0)_{(2+2q)\times 1} \\ G_t & = G & = & I_{6\times 6} \otimes \text{blockdiag}(G_0, G_1, ..., G_q)_{(2+2q)\times(2+2q)}. \end{array}$$

The evolution is decomposed into a

$$\text{Linear trend: } G_0 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

and

$$\text{Seasonality: } G_j = \begin{bmatrix} \cos(\omega_j) & \sin(\omega_j) \\ -\sin(\omega_j) & \cos(\omega_j) \end{bmatrix}, \qquad \omega_j = 2\pi j/s$$

where $s = 12$ (due to 12 months) and $q = 4$ is the number of harmonics.

Bayesian Estimation in Stan

Decompose covariance matrices: $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_6)\Omega\text{diag}(\sigma_1, \ldots, \sigma_6)$.

Priors (mutually independent):

- $\sigma_j \sim Ca^+(0, 1)$
- $p(\Omega) \propto 1$ implies slight peak at 0 correlation
- $\theta_0 \sim N_{6(2+2q)}(0, 10^7\text{I})$

Full posterior:

$$p(\theta, \Sigma_\epsilon, \Sigma_\delta | Y) \propto \left[\prod_{t=1}^{T} N(Y_t; F\theta_t, \Sigma_\epsilon) N(\theta_t; G\theta_{t-1}, \Sigma_\delta)\right] \left[\prod_j Ca^+(\sigma_j; 0, 1)\right] N(\theta_0; 0, 10^7 I)$$

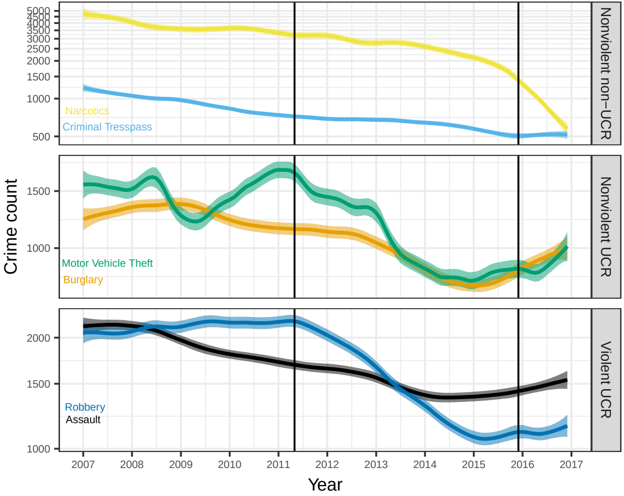
Stan (`rstan`): No-U-turn HMC sampler targets:

$$p(\Sigma_\epsilon, \Sigma_\delta | Y) = \int p(\theta, \Sigma_\epsilon, \Sigma_\delta | Y) d\theta$$

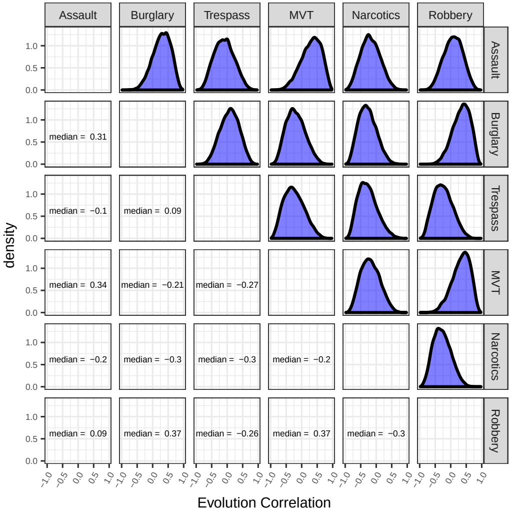
Use forward-filtering, backward-sampling (FFBS) algorithm (`dlm`) to sample

$$p(\theta | \Sigma_\epsilon, \Sigma_\delta, Y).$$

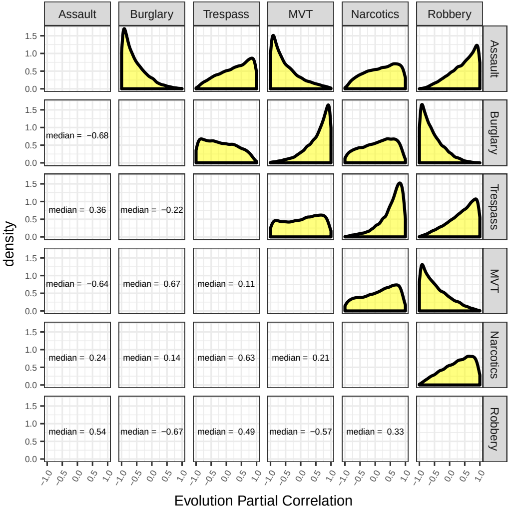
Standard crime trend analysis



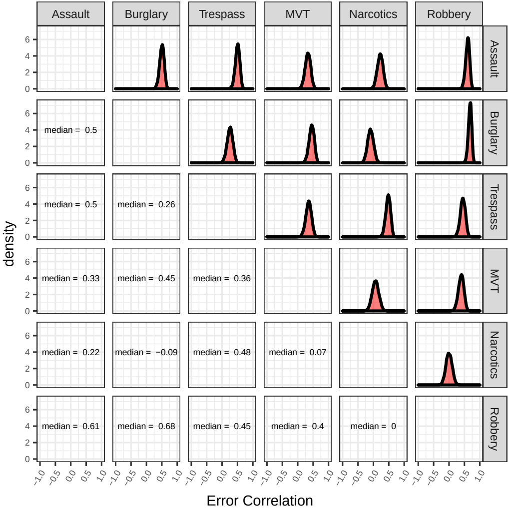
Evolution correlation



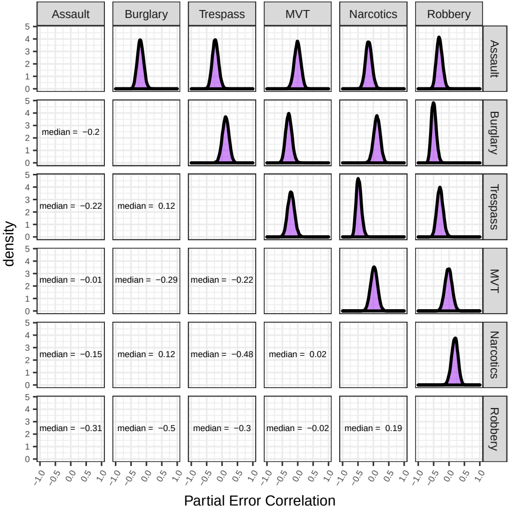
Evolution partial correlation



Error correlation



Error partial correlation



Summary

Built a multivariate DLM to model Chicago crime data to investigate a hypothesis that burglaries were being mis-classified as criminal trespasses, but did not find much evidence for this hypothesis.

These slides are available at

- https://github.com/jarad/JSM2020
- http://www.jarad.me/research/presentations.html
- https://github.com/330nategarton13/CrimeDLM.RPackage

Thank you!

Other links:

- https://www.youtube.com/jaradniemi
- https://twitter.com/jaradniemi