Bikeshare and Traffic Fatalities: Evidence from the Boston Metropolitan Area

Joshua Rosen

Outline

- Background / Research Question
- Data
- Previous Approach
- Updated Approach
- Next Steps

Background and Research Question

- Lightweight micromobility systems, such as dock-to-dock bikeshare and dockless scooters are ubiquitous in a growing number of cities, but the impacts have not been well studied
- Between 2011 and 2014, the Boston metropolitan area launched and began an initial expansion of its first dock-to-dock bikeshare system
- Does the introduction of bikeshare have a (negative) impact on city-level traffic injuries?

Brief literature review

- Cloud, Heb, Kasinger (2022) operationalize the staggered rollout of e-scooters in 93 European cities to a roughly 10% increase in reported traffic accidents after e-scooters were introduced. Effects were largest in cities with limited cycling infrastructure, and no effects were identified in cities with existing infrastructure.
- Fisherman, Schepers (2016) find that the introduction of a bikeshare system is associated with a reduction in cycling injury risk.

Data

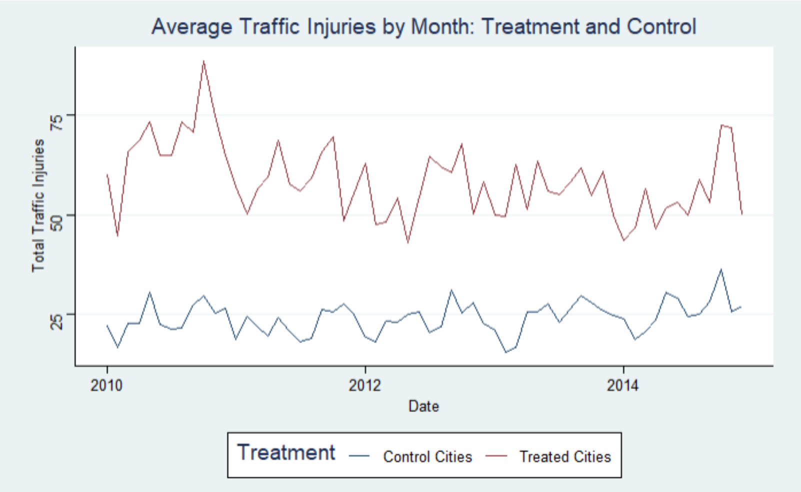
Aggregated and merged month/city level data for the following variables:

– Traffic injuries

– Aggregate bikeshare trips

Data includes 7 cities: 4 cities that did not receive bikeshares during the treatment period, and 3 cities that received bikeshare (staggered treatment dates)

Trends



- Decreasing trend absent of intervention (not quite clear in this plot)

Previous Empirical Framework

- Goal: identify the causal effect of bikeshare on traffic injuries using bayesian instrumental variables with random effects

Refresher on Instrumental variables:

$$x_i = \gamma + z_i\delta + \eta_i$$

$$y_i = \alpha + \hat{x}_i\beta + \epsilon_i$$

Where the two equations represent the first and second stages respectively. Here, α and γ are intercepts, and δ and β are regression coefficients. We also note that in the second stage, x_i is replaced with $\hat{x}_i$ to indicate the fitted value of our endogenous treatment variable generated by estimating the first stage with instrument z_i. Critically, the errors η_i and ϵ_i are correlated, and we are thus unable to only estimate the second stage regression.

First stage specification

$$Trips_i = \phi + \nu_{0,cm[i]} + \gamma_0 T_i + (\gamma_1 + \nu_{1,m[i]}) T_i + (\delta_0 + \nu_{2,c[i]}) T_i \times After_i + (\delta_1 + \nu_{3,m[i]}) T_i \times After_i$$

$$\nu_{0,cm[i]} | \sigma_{\nu_0} \quad i.i.d \quad N(0, \sigma^2_{\nu_0})$$

$$\nu_{1,m[i]} | \sigma_{\nu_2} \quad i.i.d \quad N(0, \sigma^2_{\nu_2})$$

$$\nu_{2,c[i]} | \sigma_{\nu_3} \quad i.i.d \quad N(0, \sigma^2_{\nu_3})$$

$$\nu_{3,c[i]} | \sigma_{\nu_1} \quad i.i.d \quad N(0, \sigma^2_{\nu_4})$$

- Where $\nu_{c[i]}$ and $\nu_{m[i]}$ allow the relationship between T_i and $T_i \times After_i$ to vary by city-level and month-level, respectively

Second stage specification

$$y_i = \alpha + \eta_{0,mc[i]} + (\beta_1 + \eta_{1,c[i]})Trips_i + (\beta_2 + \eta_{2,m[1]})Trips_i \tag{6}$$
$$\eta_{0,cm[i]}|\sigma_{\eta_0} \quad i.i.d \quad N(0, \sigma^2_{\eta_0})$$
$$\eta_{1,c[i]}|\sigma_{\eta_1} \quad i.i.d \quad N(0, \sigma^2_{\eta_1})$$
$$\eta_{2,m[i]}|\sigma_{\eta_2} \quad i.i.d \quad N(0, \sigma^2_{\eta_2})$$

- Goal: capture city-specific $\eta_{1,c[i]}$ and month-specific $n_{2,m[i]}$ deviations in the relationship between the instrumented total bike trips and the outcome variable total traffic injuries

Current Empirical Framework

- Updated Goal: Utilize seasonal ARIMA for Bayesian time series
- Compare the predicted traffic injuries post bikeshare introduction to the observed traffic injuries. If no other conditions change at intervention point t_i, then we can attribute the forecast error to be the causal effect of bikeshare on traffic injuries.
- Forecast error: the difference between an observed value and its forecast (measured using MAE, RMSE, etc.)

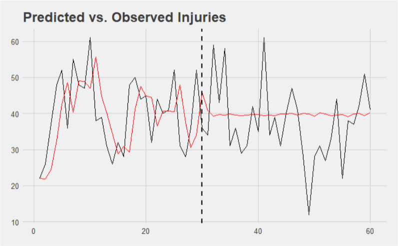
$$e_{j,T+h} = y_{T+h} - \hat{y}_{T+h|T}$$

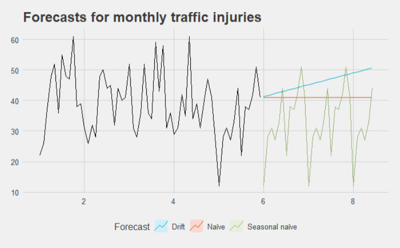
Reference Materials

- Largely relying on the 'bayesforecast' library in R
- 'bayesforecast' allows users to fit Bayesian time series models using (r)stan for full Bayesian inference

Methods matter!

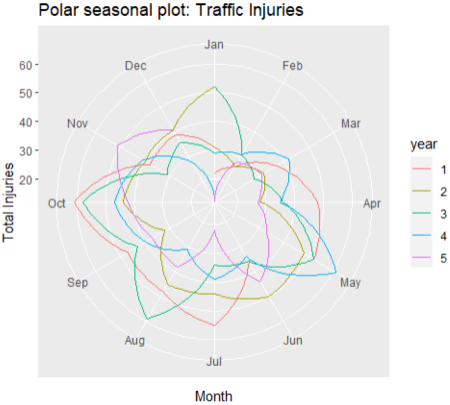
- ARIMA vs. Non-ARIMA





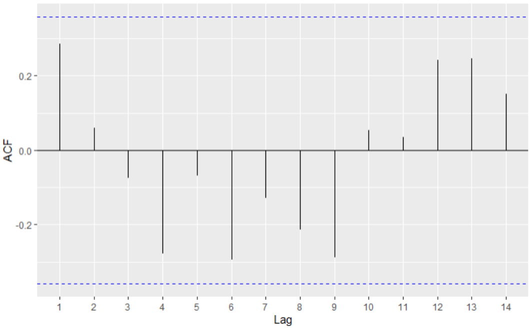
Data (ctd)

- A look at Cambridge data
- 60 months (5 years); 30 months before and after intervention



- Interestingly, here there is no clear seasonal trend (before or after treatment)

Data (ctd)



Correlogram: Autocorrelation Function

- Plotting the autocorrelation function r_k indicates that that monthly variation in traffic injuries is essentially random and is uncorrelated with previous months
- No trend or seasonality: this is white noise (stationary), so timing is predictable

Simple specification

Naive model / Random walk (forecast is equal to the last observed value):

$$\epsilon_{j,t} = y_{j,t} - \hat{y}_{j,t}$$

$$\text{where } \hat{y}_{j,t} = y_{j,t-1}$$

$$\text{and } \epsilon \sim N(0, \sigma^2)$$

ARIMA specification

- Seasonal ARIMA model

  General model: $ARIMA(p, d, q)(P, D, Q)s$

- We might estimate a model:

$$y_{j,t} = c + \phi_1 y_{j,t-1} + \ ... \ + \phi_{j,t-p} + \ ... \ + \theta_1 \varepsilon_{j,t-1} + \ ... \ + \theta_q \varepsilon_{j,t-q} + \varepsilon_{j,t}$$

Next Steps

- Set priors
- Set ARIMA parameters
- Dive into hierarchical time series