Bikeshare and Traffic Fatalities - A Case Study in Bayesian Time Series

December 5, 2022

Abstract

In this paper I explore how Bayesian time series models can be utilized to observe the causal effect of bikeshare introduction on traffic injuries in the Boston metropolitan area. While this paper fails to derive a causal interpretation, it nonetheless provides a case study in adapting Bayesian time series methods for the prediction of city-level traffic injuries across multiple groups. In all, I iterate on a simplistic ARMA(1,1) specification to estimate an ARMA model where the autoregressive term is defined hierarchically.

Keywords: Bayesian hierarchicaly modeling, autoregressive moving average, time series, causal inference, bikeshare

1 Introduction and Theoretical Approach

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. In particular, understanding the causal effect of bikeshare on traffic fatalities represents a critical issue for policymakers across the world. As a result, this case study attempts to derive the causal relationship between bikeshare and traffic fatalities by examining the staggered introduction of the Bluebikes bikeshare program in the Boston metropolitan area. To answer this question, this paper explores a series of Bayesian time series methods and provides a full Bayesian workflow. In all, while this exercise proves a futile investigation into a critical empirical question, the following paper nevertheless provides an important study on Bayesian time series.

To answer this question, I operationalize the launch and initial expansion of the Boston metropolitan area's Bluebikes bikeshare system between 2011 and 2014. In the four year span, Bluebikes slowly expanded into four cities: Boston, Brookline, Cambridge, and Somerville. Based on this time range, data on traffic fatalities at the city-month level was obtained via the state of Massachusetts' Crash Data Portal and utilized for this analysis.

This paper notably diverges from traditional quasi-experimental approaches to causal inference in favor of time series modeling. This decision is in large part due to significant threats to identification such as possible spill-over effects created in situations where the stable unit treatment value assumption may be violated locally. As a result, comparing effects across treated and control cities would likely produce biased estimates. Instead, this paper attempts to glean a rudimentary but nevertheless informative view of the causal impact of bikeshare's introduction on traffic injuries via Bayesian time series. In particular, various ARMA models are utilized to predict city-level traffic injury trends based on data leading up to the introduction of bikeshare in the city c. Therefore, if no other conditions vary at intervention point t_i, we can attribute the forecast error in traffic injuries to be result of Bluebike program's introduction.

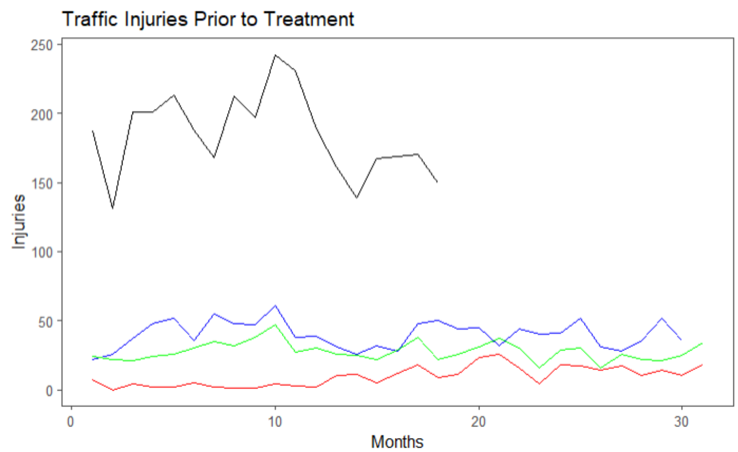


Figure 1: City-level Injury Trends

2 Initial Model

As an initial approach, I model traffic injuries at the city-month level using an ARMA(1,1) model. In doing so, I make the assumption that the current value at each month is a linear function of the previous value and current and past errors. In particular, I estimate the following model:

$$\nu_t = \mu + \phi * y_{t-1} + \theta * \epsilon_{t-1}$$
$$\epsilon_t = y_t - \nu_t$$
$$\mu \sim \mathcal{N}(0, 10)$$
$$\phi \sim \mathcal{N}(0, 100)$$
$$\theta \sim \mathcal{N}(0, 100)$$
$$\sigma \sim \text{Student-t}(3, 0, 29.7)$$
$$\epsilon_t \sim \mathcal{N}(0, \sigma)$$

Where ν_t and ϵ_t are the predicted values and error term at time t. The parameters of the model are mean μ, autoregressive coefficient ϕ, moving average coefficient θ, and noise scale σ. The priors on these parameters are assumed to be normally distributed with means of 0 and standard deviations of 100 (flat priors). The likelihood of the error is also assumed to be normally distributed with a mean of 0 and a standard deviation equal to

σ.[1]

Put explicitly, the specified model estimates the current value of the time series ν_t as a linear function of the previous value y_{t-1} and the current and past errors ϵ_{t-1}. Flat priors are utilized on all parameters to provide a neutral starting point for the estimation of the model parameters. Similarly, this choice in priors is also useful since the data itself carries more information than prior knowledge. Nevertheless, the risks in assigning flat priors are suboptimal estimates and overly conservatives model parameters. Most notably, the simplistic initial ARMA(1,1) model does not incorporate a hierarchical approach and thus fails to leverage any potential information provided by other city-level data.

The order of both the autocorrelation and moving average terms of the initial ARMA model was determined by analyzing the autocorrelation function (ACF) and partial autocorrelation function (PACF) for each of the four cities, as displayed in figures 2-5. In particular, plotting the autocorrelation function r_k indicates that monthly variation in traffic injuries is essentially random and uncorrelated with previous months. Further, the plots reveal no trends in seasonality. In other words, the patterns are stationary (and thus do not require any differencing that would be produced by a more complicated ARIMA model), and timing is predictable.

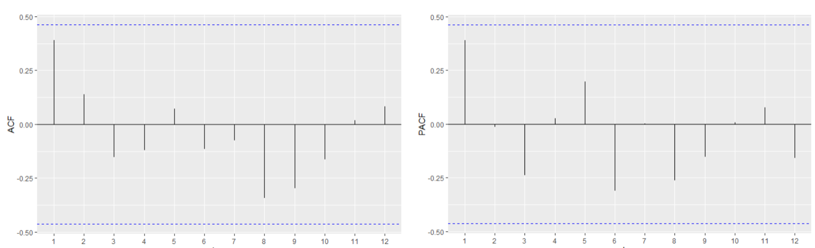
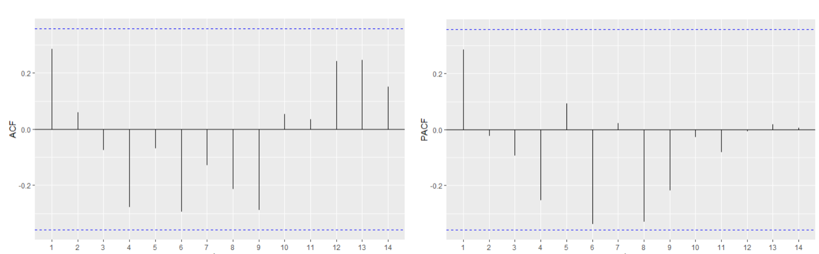


Figure 2: Boston ACF and PACF



[1] Priors on sigma vary for each modeled city.

Figure 3: Cambridge ACF and PACF

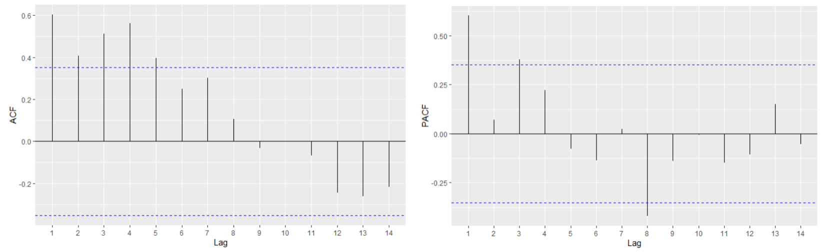


Figure 4: Brookline ACF and PACF

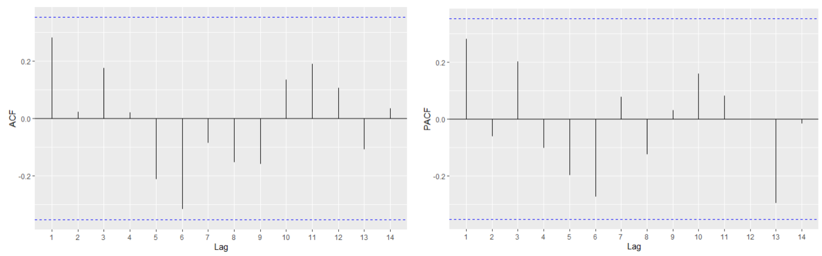


Figure 5: Somerville ACF and PACF

2.1 Model Validation

In this section, I provide a brief overview of the methods utilized to validate the ARMA(1,1) model. For brevity, all model checks are solely demonstrated on the city of Boston.

Figure 6 reveals the results of a prior predictive check. In this process, I simulated solely from the priors and likelihood of the ARMA(1,1) model and compare the results to the actual data. Here I display the residual between the simulated and actual data. The results below, however, indicate significant issues with the underlying parameterization of the initial model. In particular, it is likely that both (a) the prior distributions of the model parameters (flat priors) are chosen poorly and do not accurately reflect the data generating process, and (b) the model itself is misspecified causing the simulated data to be a poor fit to the actual data (thereby leading to a large residual). These significant issues thus inform the construction of the model specified in the second section of this paper.

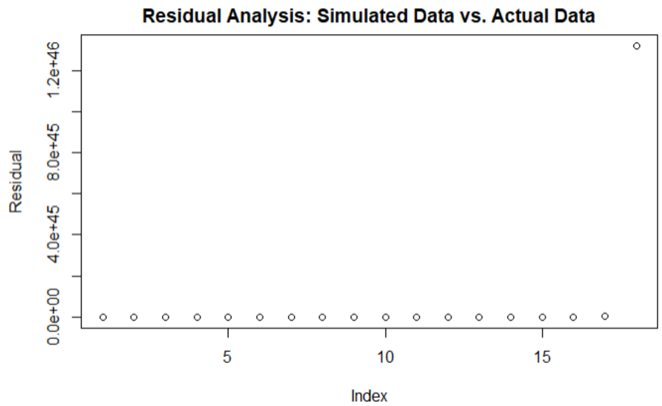


Figure 6: Prior Predictive Check

2.2 Results

Despite poor simulated results, the models nevertheless achieve the necessary $\hat{R}$ and S_{eff} to justify moving forward – albeit skeptically.

Using the previously specified ARMA(1,1) model, I predict an additional 10 months of data following the introduction of bikeshare. As stated, the original intention of this approach was to conclude a causal effect by observing the difference between the observed values and the forecast. However, due to poor model accuracy, it is unlikely that the forecast error depicts an unbiased, accurate assessment of the unobserved counterfactual trend. Instead, I interpret the forecast error for each city as a method for assessing model accuracy.

I first define forecast error explicitly in the form $\epsilon_t = y_t - \hat{y}_t$ where y_t is the observed value at time t. Since this measure will be taken over a period of 10 simulated months, I further define forecast error to be the *mean* forecast error specified as $\bar{\epsilon} = \frac{1}{T}\sum_{t=1}^{T} \epsilon_t$.

Further, due to differences in the order of magnitude between cities, the forecast error must be either (a) assessed using a scale invariant method, or (b) not compared across cities. One common scale invariant measure of forecast error is mean absolute percentage error (MAPE). While this metric solves the issue of variation in unit measurement, the measure nonetheless involves a number of unavoidable complications. Primarily, measures based on percentage errors have the disadvantage of returning an undefined value if $y_t = 0$ for any t in the period (Hyndman & Athanasopoulos, 2018). Fortunately for a very real policy problem, and unfortunately for the modeling approach, the total monthly traffic injuries in Brookline hit zero and thus prevent the use of mean absolute percentage error. Instead, I estimate mean absolute error ($mean(|e_t|)$) to calculate the forecast error across the four cities, with the results displayed in figure 11.

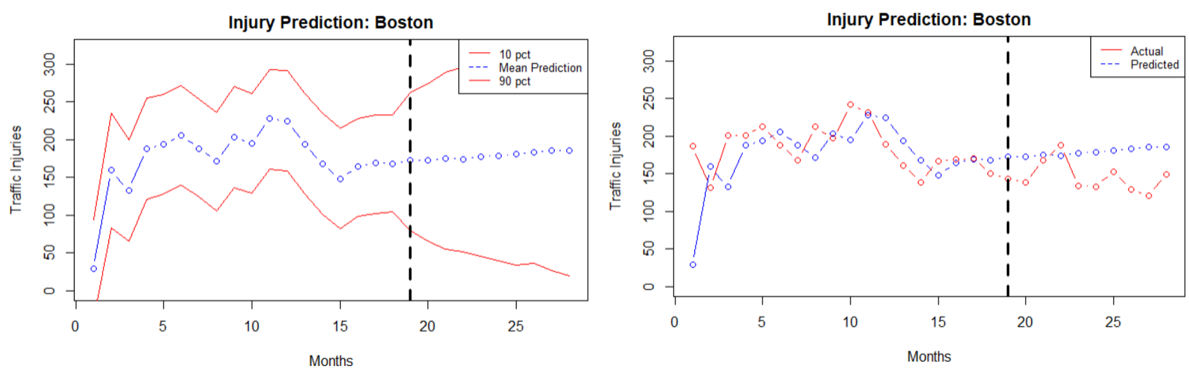


Figure 7: Boston ARMA(1,1)

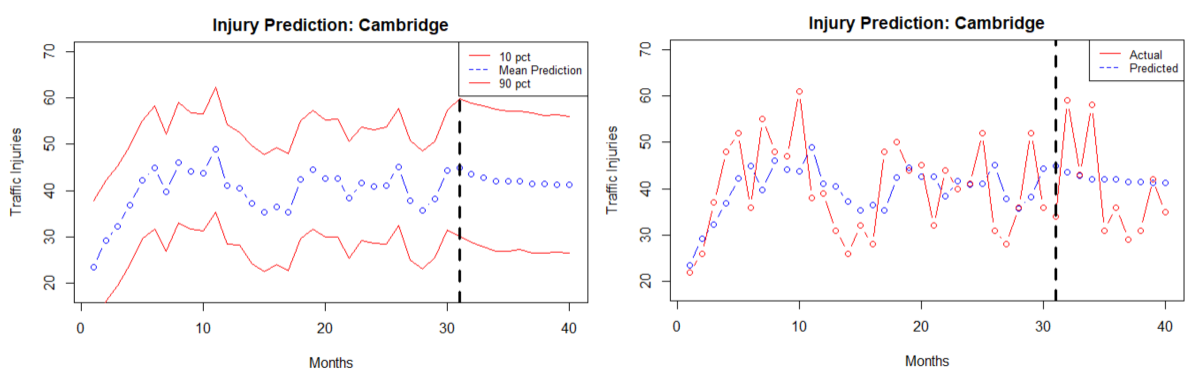


Figure 8: Cambridge ARMA(1,1)

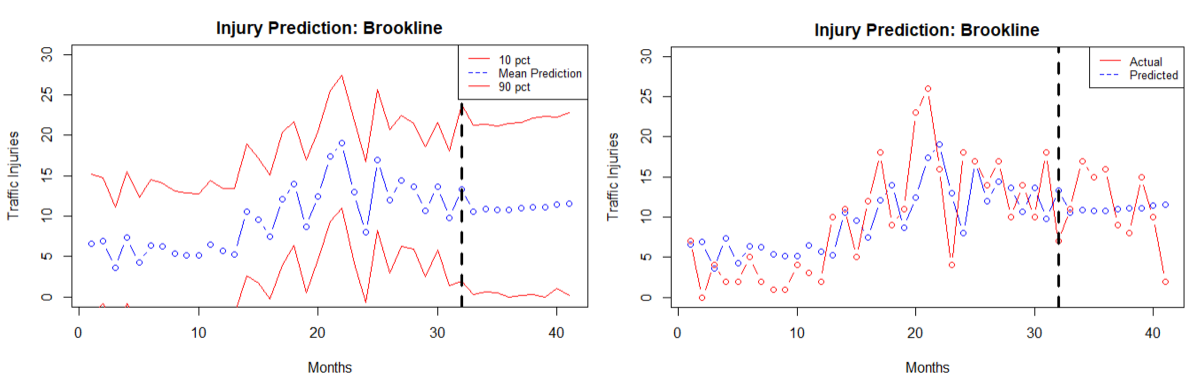


Figure 9: Brookline ARMA(1,1)

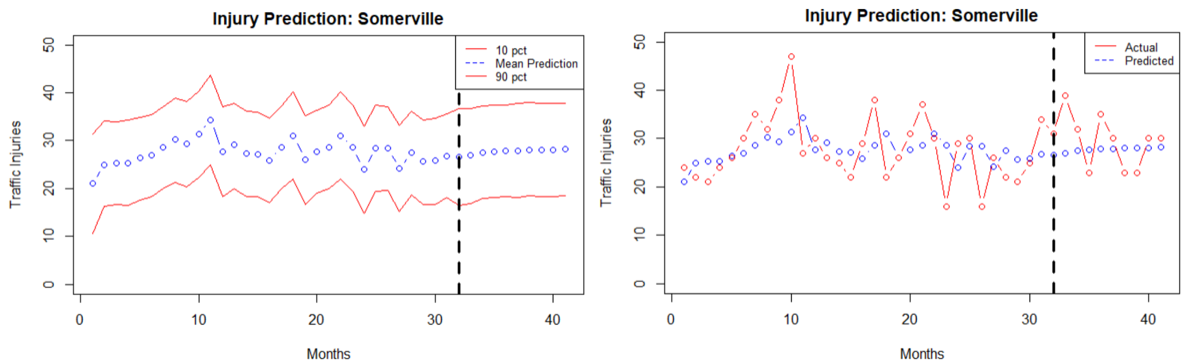


Figure 10: Somerville ARMA(1,1)

In all, the results produced by the simplistic, initial ARMA(1,1) model indicate a necessity to adopt a more complex approach. In the following section, I therefore leverage data from multiple cities to provide a more accurate estimation.

Period	Boston	Cambridge	Brookline	Summerville
Before	31.11	7.35	4.10	4.94
After	35.56	8.94	4.10	4.88

Figure 11: MAE by City

3 Estimating The Autoregressive Term Hierarchically

In the initially specified model, by estimating city-level trends both (a) independently from other cities, and (b) over a relatively short period, the approach fails to properly reduce the significant uncertainty associated with the sample autocorrelations. In this next section, I therefore attempt to iterate on the initial model and incorporate a hierarchical structure to (1) improve estimates of the autoregressive term, (2) increase flexibility, (3) better capture the data generating process, and (4) better quantify uncertainty.

In particular, I now estimate the autoregressive term of the ARMA model hierarchically. By modeling the autoregressive term hierarchically, I allow the model to incorporate additional information to minimize uncertainty. Specifically, this process enables estimates of the autoregressive term to (1) be informed by observed data, and (2) vary across group levels to produce more robust, accurate predictions. I therefore specify the following model:

$$
\begin{aligned}
y_t &= a_t + b_t y_{t-1} + \epsilon_t \\
y_t &\sim \mathcal{N}(a_t + b_t y_{t-1}, \sigma_e) \\
a_t &\sim \mathcal{N}(a_{t-1}, \sigma_a) \\
b_t &\sim \mathcal{N}(b_{t-1}, \sigma_a) \\
a_1 &\sim \mathcal{N}(\mu_a, \sigma_a) \\
b_1 &\sim \mathcal{N}(0, \sigma_a) \\
\mu_a &\sim \mathcal{N}(0, \sigma_a)
\end{aligned}
$$

where y_t measures monthly variation in city-level traffic injuries, a_t is the autoregressive term with mean a_{t-1} and standard deviation σ_a, σ_e is the standard deviation of the error term, and μ_a is the mean of the autoregressive term across all groups. As intended, the autoregressive term is given a normal prior and allowed to vary across different groups such that estimates are informed by observed data. Similarly, the mean of the autoregressive term across all groups μ_a is set as the highest level of the hierarchy. In contrast, the moving average term is not estimated hierarchically with a higher-level parameter controlling its distribution.

Figure 12 displays the results of this model for the city of Boston. The plot compares the hierarchically estimated approach with the initially specified ARMA(1,1) model and the true injury count prior to intervention. By hierarchically estimating the autoregressive term, we are both able to produce a significantly more accurate point estimate, as well as massively reduce the bounds on the interval estimates. Further, this approach preserves the reliable S_{eff} and $\hat{R}$ scores necessary for a dependable Bayesian model.

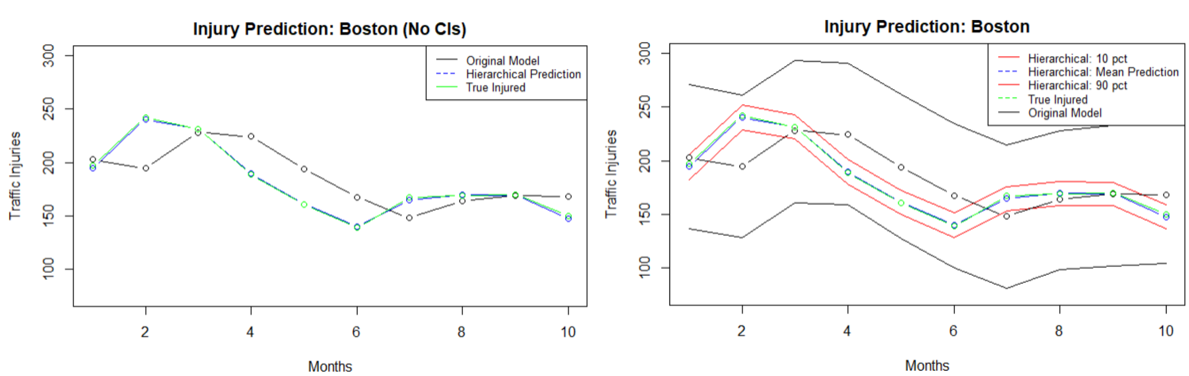


Figure 12: Boston Injury Predictions

Nevertheless, the hierarchical approach fails to mirror this accuracy when predicting forward on to unseen data. As with the ARMA(1,1) approach, it is expected for the model to produce some discrepancy between the predicted and true injury counts due to the a priori assumptions on the possible effect of bikeshare introduction. However, as displayed in Figure 13, it is unlikely that a counterfactual Boston without bikeshare introduced would see the trends predicted by the hierarchically estimated model. The difference is particularly striking in reference to the mean absolute error for both the pre-intervention period (MAE = 1.23) and the post-intervention period (MAE 65.31). For this reason, I am not comfortable making a misleading causal claim without stronger evidence in favor of model accuracy.

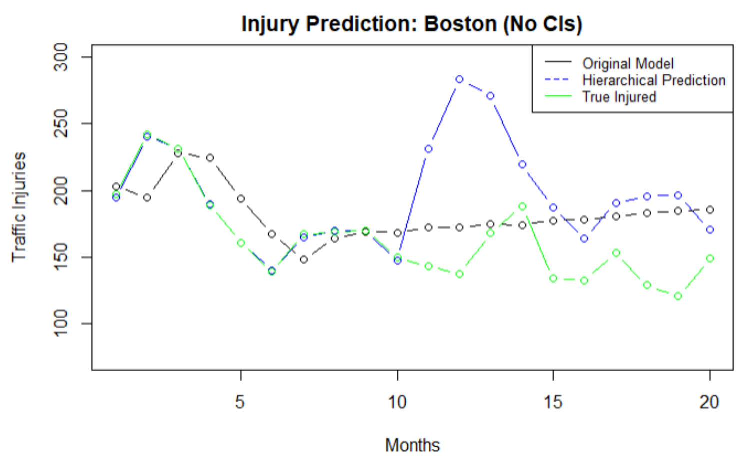


Figure 13: Boston Injury Predictions

4 Conclusion

In conclusion, this paper provides a case study on utilizing time series modeling for the prediction of traffic injuries over multiple cities in the Boston metropolitan area. While the initial goal of this paper was to derive a causal estimate via the forecast error after the introduction of the bikeshare program, neither the initial ARMA(1,1) model nor the hierarchically estimated ARMA model demonstrated high enough accuracy to confidently glean a reliable causal estimate. As a result, describing a causal effect in this situation would be misleading and potentially dishonest.

5 References

Fishman, E., and Schepers, P. Global bike share: What the data tells us about road safety. *Jorunal on Safety Research*, 56, 41-45

Hyndman, R.J., and Athanasopoulos, G. (2018) *Forecasting: principles and practice* (2nd ed.), OTexts: Melbourne, Australia

Stan Development Team. (n.d.). *Stan Reference Manual.* Retrieved December 5th, 2022

6 Appendix

Link to public Github repository containing all project materials