A Frequentist and Bayesian Approach to Modeling the Impact of a Lower BAC

Landon Mecham

University of Utah

2020-04-15

Introduction

The BAC is one of the simplest and most universal standards for governments to enforce drunk driving. It is crucial for law enforcement to keep impaired drivers off the road, as well as help drivers to be aware of their alcohol consumption[1]. The BAC is the legal measure of alcohol intoxication, according to the percentage of alcohol in an individual's blood. Prior to the 1980s, very little legislation was in place to keep impaired drivers off the roads but this became a National issue as collisions and associated deaths grew and laws were put in place to combat the rise[2]. A BAC of 0.10 was the first legal threshold for most states. Utah became the first state to reduce the legal BAC to 0.08 in 1983. After success in Utah, as well has an increasing number of alcohol related deaths in much of the U.S., most remaining states lowered the BAC to 0.08. As of 2001, 49 states enacted a BAC of 0.08[3] (with the exception of Massachusetts).

Alcohol related collisions continued to decrease nationwide from the early 2000s, but recently have recently begun to plateau[5]. For this reason, public health professionals are asking how we can continue the trend towards zero alcohol related vehicle fatalities. In January 2018, the National Academies of Sciences, Engineering and Medicine (NASEM) formed a committee to identify potential strategies to reduce the number of alcohol related fatalities. They found several potential strategies, one of which was lowering the BAC to 0.05[6].

Once again, Utah has pioneered a new BAC limit, following NASEM's recommendation, by reducing the state-wide BAC from 0.08 to 0.05 on December 30, 2018. State officials believe that that even a BAC of 0.05 is too high to be driving and that lowering the limit will prompt impaired individuals even further to stay off the roads. While controversial, the success or failure of this new policy could have big implications on other U.S. state's actions to combat drunk driving. I'll be analyzing state collision records to determine if this new limit has in fact reduced DUI related accidents in Utah.

Intervention analysis on policy implementation can be difficult because we often can't randomize who does and does not get the treatment – in our case the 0.05 BAC threshold. There are often several unobservable factors that could motivate our metric of interest to move in any direction, giving a false signal. Proper analysis to ensure we've controlled for (or ruled out) such factors is vital for causal inference. Previous research on the impact of BAC has primarily used frequentist statistics, specifically Box-Jenkins methods. I will use both a frequentist Box-Jenkins and a Bayesian state-space approach and compare the results.

Preliminary Discussion

While Utah is the first U.S. state to enact a 0.05 BAC, most states have lowered their BAC from 0.10 to 0.08 within the last 30 years. There have been vast amounts of research to determine the efficacy of the 0.08 reduction. Fell and Scherer (2017) performed meta-analysis on these studies and found 14 suitable studies (12 of which were conducted in the United States). They combined and standardized results and found that lowering the BAC from .10 to .08 resulted in a 9.1% decrease in the rates of fatal alcohol-related craches[9]. It should be noted that the study with the greatest impact was based on data from Canada and the authors note that policy/cultural differences could mean that Canada sees a larger impact than the U.S. for the same BAC reduction.

Kaplan and Prato (2007) studied the impact of lowering the BAC to 0.08 in 22 U.S. jurisdictions over a period of 15 years starting in 1990. They looked at alcohol-related single-vehicle crashes within these jurisdictions and found a statistically significant decrease[7]. Additionally, they found that female and elderly drivers were more compliant to the new law than men and younger drivers. They used a poisson regression model and accounted for state-specific effects.

Similarly, Apsler, Harding, and Klien (1999) studied the impact on fatal crash rate of lowering the BAC to 0.08 in 11 states from 1982 to 1994. They developed state-specific ARIMA models on impaired driver related traffic fatalities and found mixed results among the 11 states[8]. It should be noted, they included Utah's move to a 0.08 BAC in 1983 in their analysis and found no significant decrease in driver-impaired fatalities. They did note that Utah's alcohol related crash rate was substantially lower than the national average and that lowering it even further would have been very difficult. Their study showed that the 0.08 law in California was one of the most successful with a significant decrease of 33 high-BAC related crashes per month when they implemented the law in 1990.

Voas, Tippets, and Taylor (2002) studied the impact of the .08 law in Illinois using an ARIMA model and found a 14% decrease in fatal crashes. Using similar methodologies surrounding states increased by 3% over the same time period[11].

Fell and Scherer (2017) also performed meta-analysis on studies lowering the BAC to .05 or lower (all studies and data outside the United States). They found 11 studies meeting their criteria and after combining and standardizing the results they estimated that a reduction to .05 would result in 11% fewer fatal alcohol related crashes[9]. The most similar study to the intervention in Utah was that of Henstridge, Homel, and Mackay (1997). They studied the impact of a BAC reduction change in New South Wales and found a significant decrease in fatal crashes using an ARIMA approach[12].

Methodology

Intervention analysis is critical in understanding the impact of any policy; however, identifying the best approach can be a difficult task. The gold standard of causal inference is to run a randomized control trial. In a randomized trial we can separate the data into two groups – control and treatment – where the control receives no intervention. As long as the control is chosen randomly and there is no “spillover effect”, the causal effect of the treatment can be easily inferred. With large-scale interventions it’s often infeasible to run such a trial as it could be too costly, unethical, or even impossible to create a proper control group. The control in a randomized trial works as a counterfactual representing what we would have expected from the treatment group had there been no intervention. When we lack a clear counterfactual, we rely on econometrics. I consider methods from both frequentist and Bayesian statistics and compares the results.

Frequentist Approach

The majority of previous research on BAC impact has used a frequentist approach. I will follow the approach outlined in Enders (2014) where a Box-Jenkins models are used to 1) estimate the true underlying data-generating process of the series and 2) estimate the impact to the series post-intervention.

More specifically, Enders (2014) outlines the following steps:

> **Step 1:** Use the longest data span (i.e. either the pre- or the post-intervention observations) to find a plausible set of ARIMA models. Ensure that the $\{y_t\}$ sequence is stationary. If the sequence is non-stationary you should perform unit root tests on the longest span.
>
> **Step 2:** Estimate the various models over the entire sample period, including the effect of the intervention.
>
> **Step 3:** Perform Diagnostic checks of the estimated equations. This is particularly important since we've merged the observations from pre- and postintervention periods. The model chosen should have the following characteristics:
>
> 1. The coefficients should be of "high quality". (i.e. statistically significant, convergent y_t implied)
> 2. The residuals should be white noise
> 3. The model should outperform plausible alternatives

After performing these steps, the coefficient on the binary "treatment" variable should be evaluated in terms of significance and magnitude and will provide an estimate of the impact of the intervention.

Bayesian Approach

Bayesian statistics have some clear advantages when dealing with intervention analysis. Unlike frequentist, Bayesian methods allow the parameters to vary while holding the data fixed. This facilitates a great framework for modelling uncertainty by returning entire distributions instead of single point estimates. Additionally, we're able to incorporate prior knowledge into the model and combine that with the data at hand to update our beliefs. The Bayesian approach I'll use models structural time series equations, also commonly referred to as "state-space" or "dynamic linear models". These models don't require Bayesian methods, but it is a natural fit for the structure of modelling and provides many advantages.

Structural Time Series Models. Structural time series models decompose the series into different components. Brodersen Et al. (2015) define structural time series models by the following two 2 equations:

$$(1.1) \quad y_t = Z_t^T \alpha_t + \varepsilon_t,$$

$$(1.2) \quad \alpha_{t+1} = T\alpha_t + R_t \eta_t$$

where $\varepsilon_t \sim N(0,\sigma_t^2)$ *and* $\eta_t \sim N(0,Q_t)$ and are independent and identically distributed random variables[20]. (1.1) is the observation equation and (2.2) is the unobserved state vector where (1.1) is linked to (2.2) via α_t. The remaining variables are defined by the following:

Z_t -- *design matrix (k endogenous × k states × n)*
T_t -- *transition matrix* (k states × k states×n)(k states × k states×n)
R_t -- *control matrix* (k states × k posdef × n) (k states × k posdef × n)

These components easily capture several types of trends and seasonality and do not require the series to be stationary. Additionally, each component can either be modeled as a function of time or stochastically (e.g. random walk with drift) making the models very flexible and dynamic. For modeling the impact of lowering the BAC on collisions, $\{y_t\}$ is weekly DUI collisions. I consider 2 state-space approaches: (1) A primary approach as well as (2) a robustness check:

1. Develop several plausible state-space models (or representations of α) and determine which "best" fits the data by evaluating the R-Squared, Harvey's GoF statistic, RMSE, and graphical posterior predictive checks. I'll include a treatment indicator in the "best" model as an exogenous regressor in the design matrix (Z_t) indicating post-intervention (i.e. after 2018-12-30). I evaluate the treatment effect using the inclusion probability and highest posterior density interval on the treatment coefficient.

2. Using the "best" state-space representation from approach 1, I'll follow the framework outlined by Brodersen Et Al. (2015) which develops a set of counterfactuals for post-intervention and compares them to the observed data. This approach is similar to a Bayesian, state-space, difference in differences

design and requires a control that wasn't treat. For this I'll use non-DUI related collisions as done in other BAC related studies.

For the primary model I consider the following state components:

1. Local Linear Trend & Seasonality
2. Seasonality only
3. Local Linear Trend only
4. Semi Local Linear Trend only
5. Semi Local Linear Trend & Seasonality

Local Linear Trend. The local linear trend model is ideal for short-term forecasts due to its ability to adapt quickly to local variation; however, this can produce unreliable long-term forecasts with large intervals. The equations for the local linear trend model are:

(1.3) $\mu_{t+1} = \mu_t + \delta_t + \eta_{\mu,t}$

(1.4) $\delta_{t+1} = \delta_t + \eta_{\delta,t}$

where $\eta_{\mu,t} \sim N\ (0,\sigma^2_{\mu})$ and $\eta_{\delta}+\sim N\ (0,\sigma^2_{\delta}\ )$. We can see that u_t represents the value of the series at time t and δ_t is the slope. The slope then, is time dependent and hence very flexible. From these equations, we can see that both the slope and the mean of the trend follow random walks.

Semi Local Linear Trend. This is preferred for long-term forecasting. It differs from the *local linear trend* in that it assumes a random walk for the mean, but a stationary

AR(1) process for the slope of the series. This results in a more stable model but also less flexible. It is defined by the following equations:

$$(1.5) \quad \mu_{t+1} = \mu_t + \delta_t + \varepsilon_t$$

$$(1.6) \quad \delta_{t+1} = D + \varphi(\delta_t - D) + \eta_t$$

where $\varepsilon t \sim N(0, \sigma_\mu)$ and $\eta_t \sim N(0, \sigma_\delta)$. The long-term slope is represented by D meaning that δ_t will converge to D as t increases. Φ is the memory of the slope with values close to 0 only allowing for short deviations from the long-term mean[20].

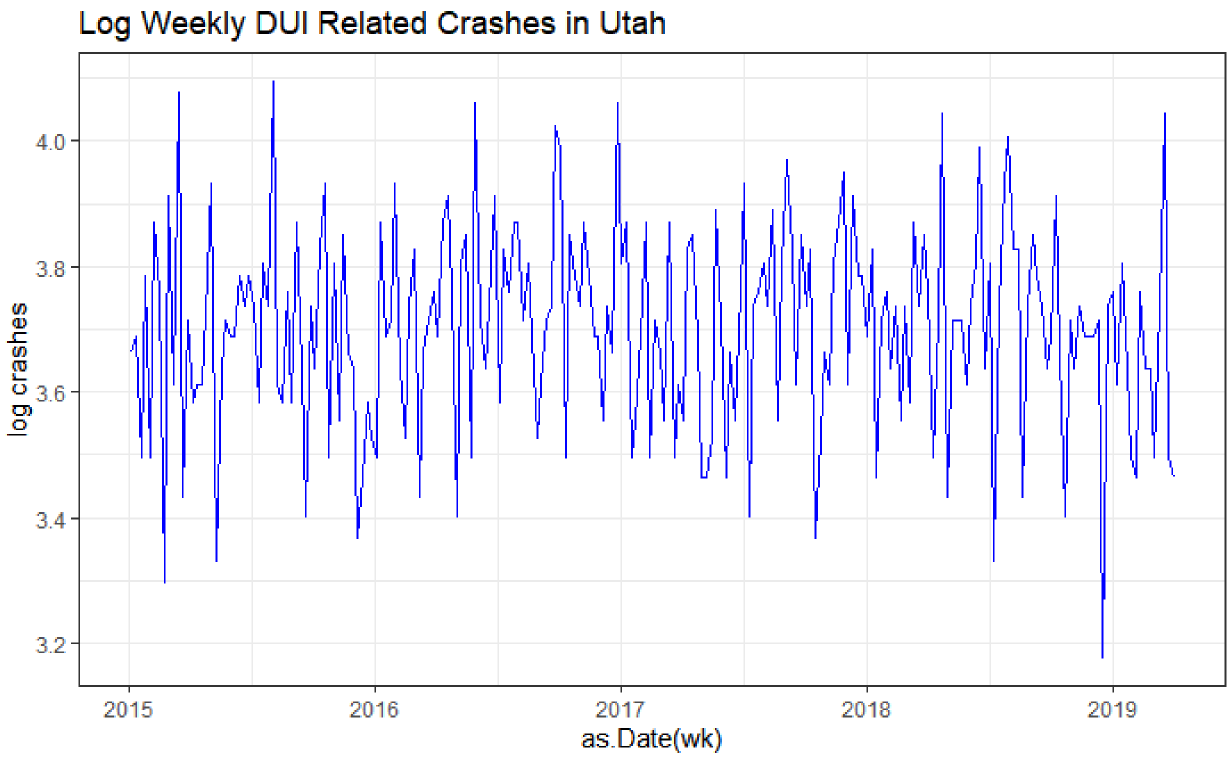
Seasonality. The seasonality component can be adjusted to fit several different time frequencies and is represented by the following equation:

$$(1.7) \quad \gamma_{t+1} = \sum_{s=0} \gamma_{t-s} + \eta_{\gamma,t}$$

where s is the total number of seasons and γ_t is the joint contribution to the series. For example, my data is weekly so we'd set S = 52 for each week of the year. This is analogous to including an indicator for each season (i.e. 52) but instead of leaving 1 out – as we would usually do in a regression model – we include a constraint that they all sum to 0 so that the model is not overparametrized. This framework allows for flexibility in the seasonal impact as the pattern can slowly evolve[20].

Empirical Results

I use weekly DUI related automobile collisions in Utah from 2015-01-01 to 2019-03-31 provided by the Utah Department of Transportation (UDOT). This allows for a full quarter of post-intervention data since the BAC was lowered on 2018-12-30. It should be noted that at the time this data was collected (Jan. 2020) UDOT had not yet finalized the data for 2019 and could not confirm whether or not Q1-2019 data would change as data is updated. In both models, I take the log of weekly crashes.



Frequentist Method

Step 1. The first step is to find a plausible set of ARIMA models using the longest span of data which is pre-intervention. I start by finding the order of integration of the series. By visual inspection and the nature of the data it seems there is some seasonality. Since seasonality is suspected, it's important to check for seasonal unit roots. Just as a trend can either be stochastic or deterministic, seasonally can also take these forms. Deterministic seasonality is introduced by systematic cycles such as weather or effects of holidays. This can easily be removed by seasonal adjustments such as the inclusion of seasonal dummy variables. Stochastic seasonality is not constant and instead is a function of the series at the previous season. This can cause the series to become non-stationary by introducing seasonal unit roots. Seasonal unit roots need to be identified and accounted for by seasonal differencing.

The test statistics derived from standard unit root tests such as the Dickey Fuller are invalid if seasonal unit roots are present. Instead, I will use the HEGY test – so called after the authors Hylleberg, Engle, Granger, and Yoo (1990)[16]. The HEGY tests for unit roots at each seasonal frequency both individually and collectively. It is analogous to running the following regression (assuming quarterly data for simplicity)[17]:

$$(2.1)\ \Delta^4 Y_t = \alpha + \beta_t + \sum_{j=2} b_j Q_{jt} + \sum_{i=1} \pi_i W_{it-1} + \sum_{l=1} \eta\ \Delta^4 Y_{t-l} + \varepsilon_t$$

Where Q_{jt} is a seasonal indicator variable and W_{it} are given by the following:

$$(2.2)\ W_{1t} = (1 + B)(1 + B^2)Y_t$$

$$(2.3)\ W_{2t} = -(1 – B)(1 + B^2)Y_t$$

$$(2.4)\ W_{3t} = -(1 – B)(1 + B)Y_t$$

$$(2.5)\ W_{4t} = -B(1 – B)(1 + B)Y_t = W_{3t-1}$$

Now, we run the t tests $\pi_1 = 0$, $\pi_2 = 0$, and F test $\pi_3 = \pi_4 = 0$. If any test is not rejected this indicates the presence of a seasonal unit root and we need to take a seasonal difference and run the test again. Note, if each test is rejected then we can conclude the series is stationary and we can move forward without differencing. This example is specific to quarterly seasonality, but Hernandez et al (2001) have applied this same methodology to weekly seasonality[18]. Table 1 displays the HEGY test statistics on the weekly collision series.

Table 1: HEGY Test Statistics on Level

	test statistic	p-value			test statistic	p-value	
t_1	-2.68	0.061		F_31:32	1.85	0.110	
t_2	-1.16	0.236		F_33:34	2.24	0.073	
F_3:4	1.82	0.113		F_35:36	2.09	0.085	
F_5:6	3.30	0.025	*	F_37:38	1.92	0.102	
F_7:8	1.46	0.169		F_39:40	1.57	0.151	
F_9:10	2.64	0.048	*	F_41:42	4.89	0.005	**
F_11:12	4.18	0.009	**	F_43:44	5.53	0.003	**
F_13:14	1.93	0.102		F_45:46	3.84	0.014	**
F_15:16	0.34	0.582		F_47:48	1.91	0.103	
F_17:18	2.81	0.041	*	F_49:50	1.94	0.101	
F_19:20	2.81	0.041	*	F_51:52	0.41	0.531	
F_21:22	0.83	0.331		F_2:52	3.67	0.253	
F_23:24	1.13	0.243		F_1:52	3.78	0.000	***
F_25:26	2.49	0.056					
F_27:28	0.42	0.525					
F_29:30	3.59	0.018	**				

Under the null hypothesis of each test a seasonal unit root exists. Of the 29 test statistics we reject the null in 10. This indicates the presence of multiple seasonal unit roots and we need to take a seasonal difference of the data and run the HEGY test again. I take a seasonal difference by taking $X_t = X_t – X_{t-s}$ where s is the number of seasonal frequencies (in our case 52). Table 2 displays the HEGY test statistics after taking one seasonal difference.

Table 2: HEGY Test Statistics on Seasonally Differenced Data

	test statistic	p-value			test statistic	p-value	
t_1	-3.50	1.000		F_31:32	3.68	0.013	**
t_2	-1.18	0.991		F_33:34	3.72	0.013	**
F_3:4	3.55	0.013	**	F_35:36	3.22	0.014	**
F_5:6	4.45	0.012	**	F_37:38	4.56	0.011	**
F_7:8	0.62	0.018	**	F_39:40	3.67	0.013	**
F_9:10	2.53	0.015	**	F_41:42	6.59	0.000	***
F_11:12	6.10	0.008	**	F_43:44	3.99	0.013	**
F_13:14	5.57	0.009	**	F_45:46	4.19	0.012	**
F_15:16	3.42	0.014	**	F_47:48	3.19	0.014	**
F_17:18	4.97	0.011	**	F_49:50	2.97	0.014	**
F_19:20	0.63	0.018	**	F_51:52	1.70	0.017	**
F_21:22	5.24	0.010	**	F_2:52	6.99	0.304	
F_23:24	3.66	0.013	**	F_1:52	7.13	0.000	***
F_25:26	5.11	0.010	**				
F_27:28	1.39	0.017	**				
F_29:30	0.99	0.017	**				

We now reject at nearly all seasonal frequencies and can conclude that the series is integrated of seasonal order 1. Remember, after determining stationarity through the HEGY test it’s no longer necessary to check for standard unit roots. We can now move forward with the modelling approach.

I will consider both ARIMA and SARIMA models. SARIMA models not only look at effects of the most recent lagged data (e.g. t – 1, t – 2, …) but also at the effect of values at the same seasonal frequency (e.g. t – s). I’ll look at the ACF and PACF displayed in tables figures 2 & 3 to determine a set of plausible models.

Figure 2: Seasonally Differenced ACF

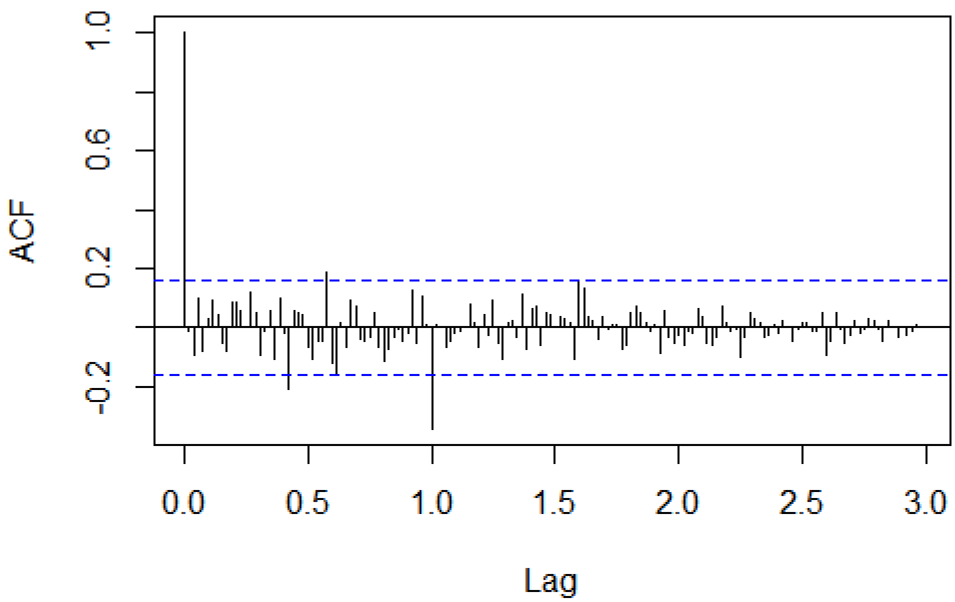
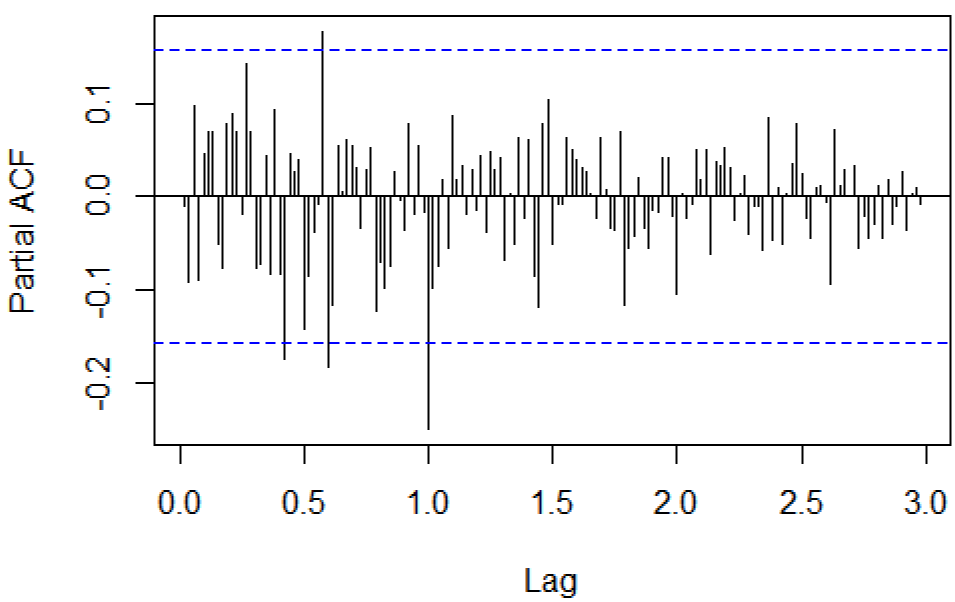


Figure 3: Seasonally Differenced ACF



Figures 2 & 3 show a significant correlation at lag 52 in both the ACF and PACF which strongly points to a SAR(1). I will include both a SAR(1) and a SMA(1) in the list of plausible models as well as a pure noise model that doesn't include any MA or AR terms.

Steps 2 & 3: Step 2 is to run the three models in consideration on the full series ranging from 2015-01-01 to 2019-03-31 and in step 3 I run the models through a series of quality checks to choose a "best" model. The quality checks are 1) statistically significant coefficients, 2) convergence of $\{y_t\}$, 3) white noise residuals, and 4) the model should outperform plausible alternatives in terms of the AIC and BIC. Table 3 summarizes the quality metrics for each of the models.

Table 3: Model Quality Checks

	SAR(1)	**SMA(1)**	**Pure Noise**
Estimate (P-Value)	sar1: -0.49 (0.000) impact: -0.03 (0.449)	sma1: -0.99 (0.223) impact: -0.04 (0.421)	impact: -0.04 (0.46)
AIC	-60.87	-0.71	-0.31
BIC	-51.47	-0.61	-0.25
Q-Statistic			
lag 5 p-value	0.41	0.21	0.08
lag 10 p-value	0.43	0.48	0.17
lag 15 p-value	0.13	0.2	0.11
lag 20 p-value	0.15	0.34	0.05
lag 25 p-value	0.11	0.31	0.03

Table 3 shows that the sar1 coefficient is statistically significant and not equal to 1. The ma1 coefficient is insignificant and essentially 1 which implies non-convergence of $\{y_t\}$. The SMA(1) model has the lowest AIC and BIC while the pure noise model has the highest. Both the SAR(1) & SMA(1) fail to reject the null hypothesis white noise residuals at all lag levels indicating that they are capturing the signal produced; However, the pure noise model does not. The SAR(1) model is the only model that passes all quality checks while maintaining a relatively low AIC and BIC. I'll move forward with the SAR(1) as the "best" model.

Since the intervention treatment variable (i.e. "impact") is insignificant with a two-sided p-value of 0.45 then we can conclude that the treatment had no significant effect on weekly collisions.

Bayesian Method

Approach 1: To find the primary state-space model to use for inference I'll evaluate using the following:

1. RMSE on a holdout
2. RMSE on one-step ahead predictions
3. Harvey's Goodness of Fit Metric
4. R-Squared
5. Graphical Posterior Predictive Checks

The RMSE is evaluated on a holdout and is created by training each model on a shortened dataset ranging from 2015-01-01 to 2018-10-21, then sampling from the posterior predictive distribution to obtain mean sample forecasts through the remainder of the year. I'll calculate the RMSE on those predictions versus what was actually observed. Evaluation on out-of-sample performance helps in selecting a robust model with reliable forecasts. This evaluation will help determine if our model is simply overfit to our particular dataset. Additionally, this gives us a better measure of long-term forecast ability of the models. To ensure that the model performs well on the remainder of the dataset I'll sample from the posterior predictive distribution to obtain one-step-ahead predictions. This is essentially the prediction at each point in time t in our dataset if we'd used all the data up to t-1 to build the model. This gives us an idea of the short-term forecasting ability.

The Harvey's Goodness of Fit Metric is analogous to the R-Squared for regression models. It differs I t benchmarks to a random walk with drift instead of the mean. Harvey (1989) argues that this is a better comparison for state-space models[19].

The evaluation metrics previously mentioned fail to take full advantage of the Bayesian methodology. Graphical posterior predictive checks compare the observed values a series to random draws from the posterior predictive distribution. Instead of comparing a single predictive estimate, we can compare a distribution of predictions for each data point. This is based on the simple idea that if the model fits well then, we should be able to use samples from the posterior distribution to mirror the observed data. I will be using plots to ensure that this holds. If there are obvious anomalies, I will exclude the given model from the set.

Table 4 shows the multiple evaluation metrics we'll be comparing

Table 4

	Rsquared	HarveyGOF	RMSE (1-step ahead predictions)	RMSE (holdout)
Model1	11.7%	42.2%	0.181	0.172
Model2	9.1%	48.3%	0.171	4.497
Model3	-0.6%	48.5%	0.171	0.173
Model4	5.6%	49.9%	0.169	0.168
Model5	16.5%	43.5%	0.179	0.162

Model 5 has the highest R-squared while Model 4 has the highest Harvey GOF statistic. Models 4 & 5 have the lowest have the lowest RMSE on the holdout while Model 4 maintains a slightly lower RMSE on the full dataset. The only difference between Models 4 & 5 is that Model 5 includes seasonality in the state component of the model. We use graphical PPCs to compare Models 4 & 5.

First, we'll perform visual residual diagnostics on the one-step-ahead predictions on the full models using draws from the posterior predictive distribution. Figures 5.1 and 5.2 show normal Q-Q plots and we see that both models produce distributions that follow a normal distribution. Figures 5.3 and 5.4 show the distribution of autocorrelations at each lag using box plots. Both models also pass this test showing no single autocorrelation with a median above 0.1 in absolute value. We can conclude that the residuals are normally distributed and independent of each other for both models.

Figure 5.1: Model 4

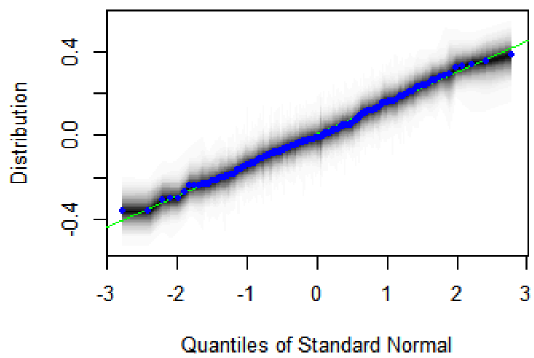


Figure 5.2: Model 5

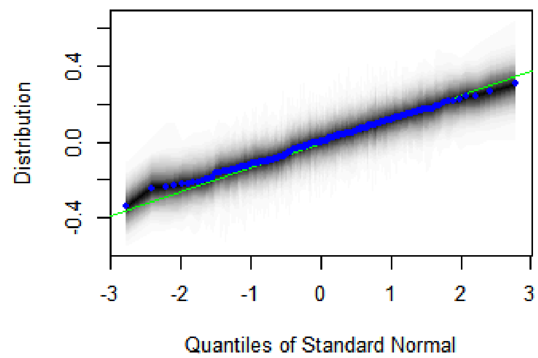


Figure 5.3: Model 4

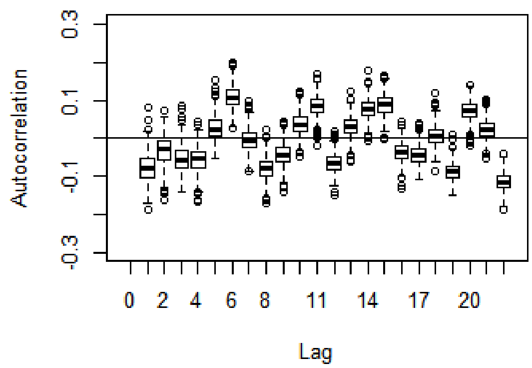
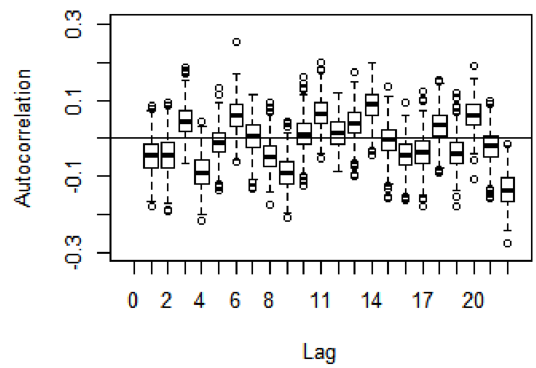


Figure 5.4: Model 5



Now, to better understand the predictive relationship we'll plot the 20-step-ahead predictive distribution against observed values using the shortened models used for the holdout metrics.

Figure 4.1: Model 4

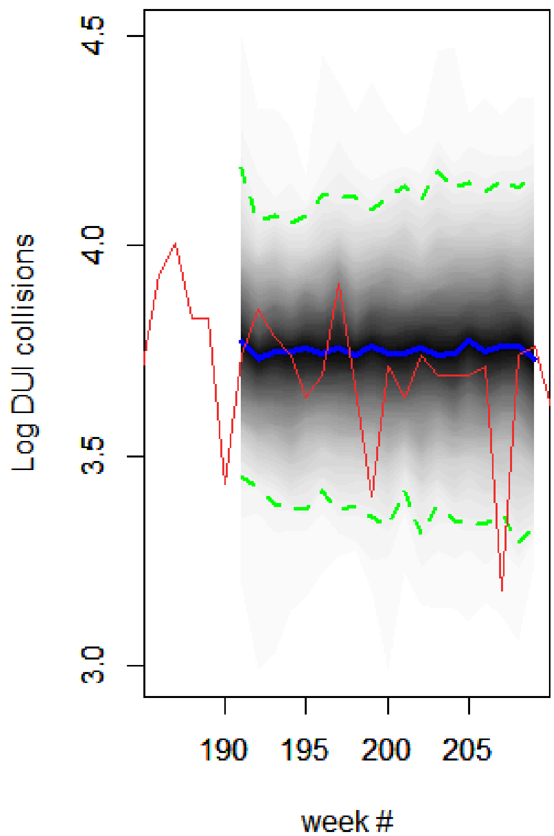
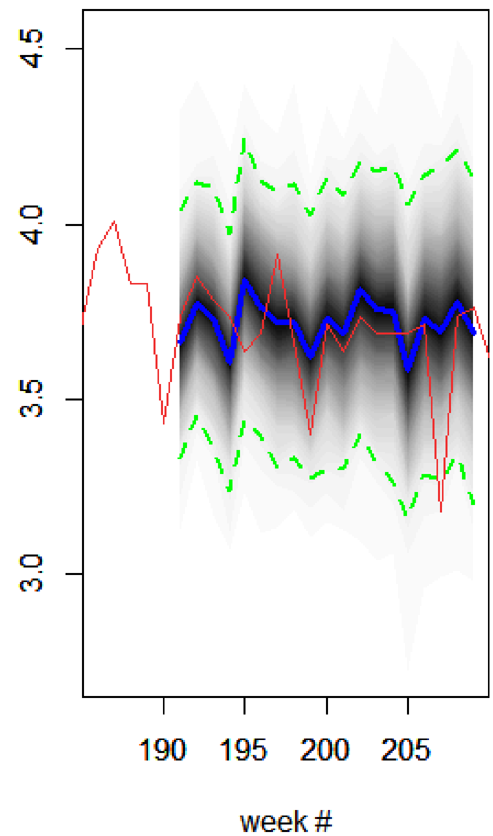
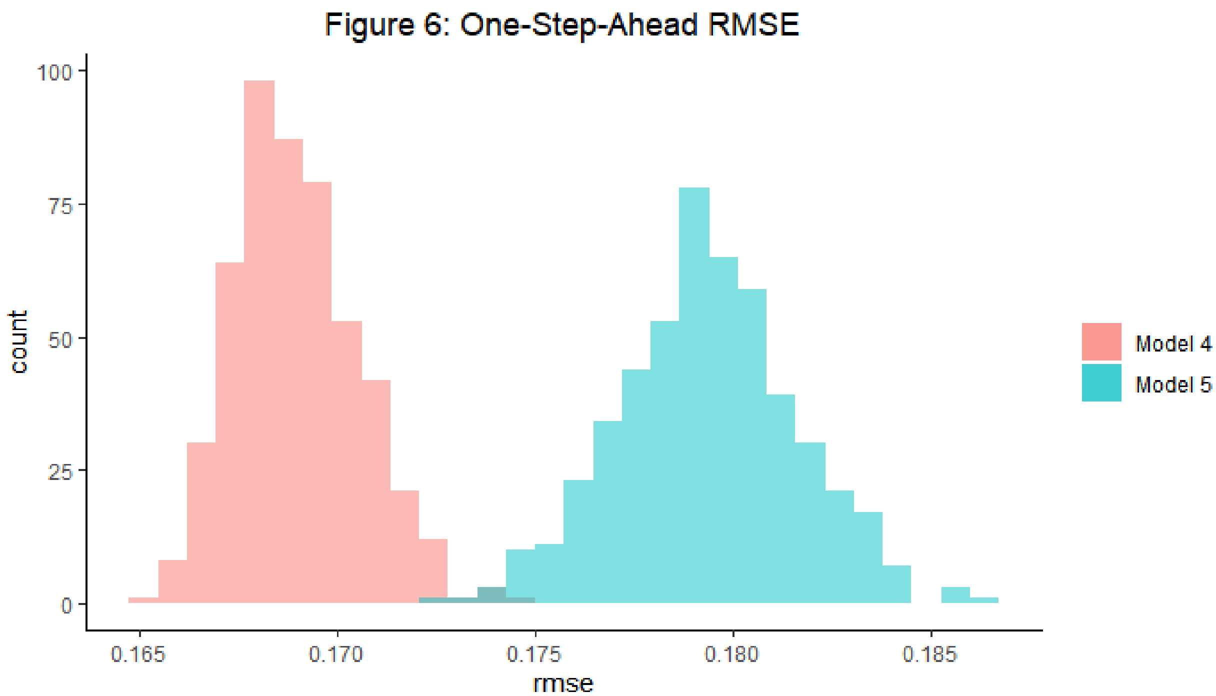


Figure 4.2: Model 5



Figures 4.1 and 4.2 show that Model 4 predictions are relatively flat with no big swings in either direction. Model 5 has much more movement due to the seasonal influence. The Model 5 predictive distribution appears to move much closer with observed data than Model 4, however it occasionally predicts spikes and dips that don’t occur which is potentially costly. Model 5 does have narrower confidence intervals and appears to do much better at predicting further out than model 4.

Finally, we’ll compare the distribution of one-step-ahead RMSEs on the full models. These are calculated by 1) taking 500 random draws from the one-step ahead predictive distribution at each time frequency, and 2) calculate the RMSE overall the entire series for each draw. Figure 6 shows histograms of these RMSEs for both Model 4 and Model 5.



This provides some additional detail to the mean RMSE estimates provided in Table 4. Figure 6 shows that nearly every draw from the posterior distribution for Model 4 resulted in a lower RMSE than that of Model 5. Because of this clear separation in predictive ability, Model 4 should be used in determining impact of the treatment.

BSTS uses “spike and slab” regression for all predictors in the design matrix. This returns posterior distributions for both the probability of the variable being included in the true data-generating process (i.e. inclusion probability) and the coefficient value given inclusion. To determine whether the treatment binary has a true effect I’ll look at both the inclusion probability and 95% highest posterior density (HPD) of the coefficient. The resulting inclusion probability is 1.4% which is well below the conventional threshold of 10%. If we assume that the treatment is truly included, it's posterior mean is 0.045 with 95% HPD between -0.08 and 0.17. Since the interval includes zero we cannot conclude that the treatment had any effect on weekly collisions.

Robustness Check: Brodersen et al (2015) proposed a method to infer causal effects with state-space models that involve creating a counterfactual series that represents the expected results had no intervention occurred[20]. They then estimate the treatment effect by comparing this to the observed data.

This technique requires 3 steps. First, we estimate the state-space model by simulating draws of the parameters over the period $y_{1:n}$ where y_{n+1} is the first observation in the treatment period. If available, a control series should be included as a static regressor in the model. The control should not be affected in any way by the treatment and will represent all variables unaccounted for in the model. Second, we draw from the posterior predictive distribution to simulate $P(\acute{y}_{n+1:m}|y_{1:n})$ where m is the last observation so that $y_{n+1:m}$ represents the treated portion of the series and $\acute{y}$ is the counterfactual simulation. Third, the pointwise treatment effect is estimated by calculating $i_t = y_t - \acute{y}_t$ for all t from n+1 to m. This results in a distribution of treatment effects obtained at each time period during the treatment and we can average over the full treatment period to obtain a cumulative effect or mean weekly effect[20].

Similar to a difference in differences design, the control series should follow the treatment series closely pre-intervention and we assume that the treatment had no effect on the control. Previous research on BAC impact have used non-DUI related collisions as a control for DUI related collisions. This assumes that lowering the BAC will have no effect on non-DUI related collisions which seems reasonable.

Brodersen et al (2015) streamline this approach in the R package CausalImpact which I leverage for my analysis[20]. Using model 5 from approach 1, I run through the three steps outlined above.

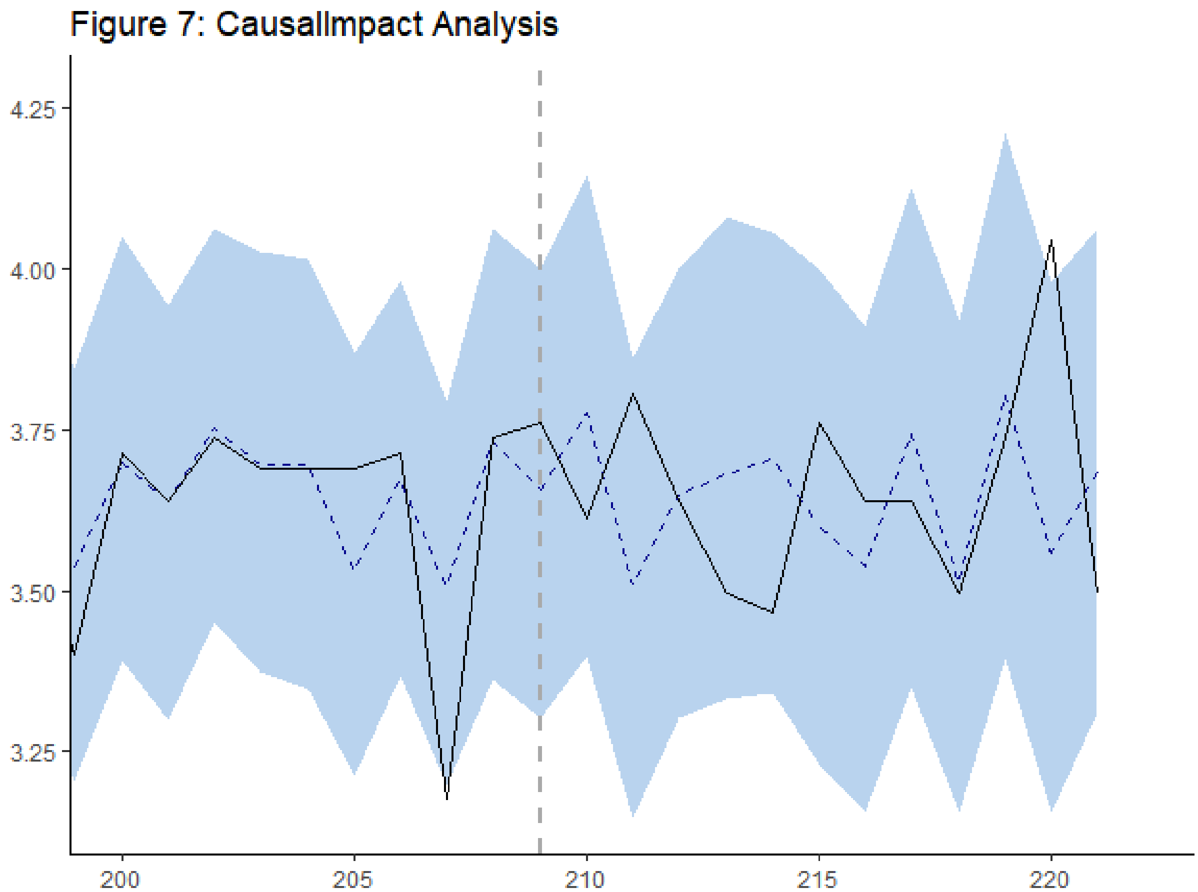


Figure 7 compares the observed data (solid line) to the counterfactual predictions (dashed line) had the BAC remained 0.08. Visually, it's clear that the observed series is well within the bounds of the counterfactual. After lowering the BAC, the post-treatment mean DUI related collisions was 3.65. We would have expected a mean of 3.65 with a 95% credible interval of [3.35, 3.96] in the post-treatment period had we not lowered the BAC. This mean and credible interval serves as our counterfactual. If we subtract the counterfactual from the actual mean of DUI related collisions post-intervention we get mean effect size of -0.001 with a 95% interval of [-0.31, 0.30]. This effect was not statistically significant and we cannot conclude that lowering the BAC had any effect on DUI related collisions.

Conclusion

All three approaches aligned in finding zero significant impact of the intervention. While data is still limited, early indicators show that the implementation of a 0.05 BAC limit in Utah has not had an impact on DUI related collisions. Since I only had Q1-2019 data available, it’s important to do further analysis as more data is released. Additionally, the policy could still be effective in keeping drivers with a BAC between 0.05 and 0.08 off the road – which opens another question as to whether those drivers are truly impaired.

Additionally, most traffic related research has normalized data by including vehicle miles travelled. This controls for potentially unaccounted for variables that increase miles driven and therefore number of collisions. While our robustness check does this in a sense by using a control series, further analysis should seek to use DUI collisions per vehicle mile travelled.

References:

(1) National Highway Traffic Safety Administration. Traffic Safety Facts 2016 data: alcohol-impaired driving. U.S. Department of Transportation, Washington, DC; 2017 Available at: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812450External Accessed 16 April 2018.

(2) https://one.nhtsa.gov/About-NHTSA/Traffic-Techs/current/ci.State-Alcohol-Cost-Fact-Sheets-Document-The-Costs-Of-Impaired-Driving-For-Each-State.print

(3) https://one.nhtsa.gov/people/injury/research/pub/alcohol-laws/08history/1_introduction.htm

(4) National Highway Traffic Safety Administration, 2005, and How to Control Your Drinking, WR Miller and RF Munoz, University of New Mexico, 1982.

(5) https://www.nap.edu/resource/24951/011718AlcoholImpairedDrivingHighlights.pdf

(6) National Academies of Sciences, Engineering, and Medicine. 2018. Getting to zero alcohol-impaired driving fatalities: A comprehensive approach to a persistent problem. Washington, DC: The National Academies Press. doi: https://doi.org/10.17226/24951.

(7) Kaplan S, Prato CG. Impact of BAC limit reduction on different population segments: a Poisson fixed effect analysis. Accident Analysis & Prevention. 2007;39:1146–1154

(8) Apsler R, Char AR, Harding WM, Klein TM. The effects of .08 BAC laws, in Series The effects of .08 BAC laws, U.S. Department of Transportation, National Highway Traffic Safety Administration; Washington, DC: 1999.

(9) Fell JC, Scherer M. Estimation of the Potential Effectiveness of Lowering the Blood Alcohol Concentration (BAC) Limit for Driving from 0.08 to 0.05 Grams per Deciliter in the United States. Alcohol Clin Exp Res. 2017;41(12):2128–2139. doi:10.1111/acer.13501

(10) Asbridge M, Cartwright J, Langille D. Driving under the influence of opioids among high school students in Atlantic Canada: prevalence, correlates, and the role of medical versus recreational consumption. Accident Analysis & Prevention. 2015;75:184–191.

(11) Voas RB, Tippetts AS, Taylor EP. The Illinois .08 law: An evaluation. Journal of Safety Research. 2002;33:73–80. [PubMed] [Google Scholar]

(12) Henstridge J, Homel R, Mackay P. The long-term effects of random breath testing in four Australian states: a time series analysis, in Series The long-term effects of random breath testing in four Australian states: a time series analysis. Federal Office of Road Safety; Canberra, Australia: 1997

(13) Enders, Walter, 1948Applied econometric time series / Walter, University of Alabama. – Fourth edition.

(14) Fell JC, Voas RB. The effectiveness of a 0.05 blood alcohol concentration (BAC) limit for driving in the United States. Addiction. 2014 Jun; 109(6):869-74.

(15) Scott, S. Fitting Bayesian structural time series with the bsts R package, unofficial google data science review. 2017

(16) Hylleberg, S. Engle, R.F., Granger, C.W.J., and Yoo, B.S., Seasonal integration and cointegration. 1990. Journal of Econometrics, 44: pages 215:238

(17) http://www.ams.sunysb.edu/~zhu/ams586/Seasonality.pdf

(18) Hernández C, Juan J, Fernández V, Francisco, M. Observaciones anómalas y contrastes de raíz unitaria en datos semanales. 2001 Estudios de economia aplicada

(19) Harvey, A. Diagnostic Checking of Unobserved-Components Time Series Models. 1992 Siem Jan Koopman Journal of Business & Economic Statistics, Vol. 10, No.4, 317:389

(20) Brodersen K, Fabian G. Koehler J, Remy N, Scott S. Inferring causal impact using Bayesian structural time-series. 2015. Annals of Applied Statistics 247:274.

(21) Scott S, https://cran.r-project.org/web/packages/bsts/bsts.pdf