Analyzing effect of compulsory seat belts on road mortality

Henry Svedberg Fabian Eriksson Måns Wallström

Abstract

In this study the effect of introducing compulsory frontseat seatbelts on drivers killed in traffic was analyzed using a Poisson- and a negative binomial regression model, including trend and autoregressive terms and a covariate measuring distance travelled. By evaluation using pareto-smoothed leave one out cross validation, the negative binomial model was found to fit the data best, and its estimated effect of compulsory seatbelts on drivers killed was approximately -8 percent, with a 90 percent credible interval of (-0.14, -0.02).

1. Introduction

The goal of this report is to analyze how the introduction of compulsory seat belts affected road accident mortality in the UK. In January 1983, the country introduced a new law requiring the wearing of seatbelts. By analyzing data from before the law and from after a law, its effect can be estimated.

Due to the data being a count date time series, the analysis will be done by fitting a Poisson autoregression on the number of mortalities each month conditioned on whether the law is in effect as well as other covariates. Also, a negative binomial model will be used to compare with the Poisson model.

2. Data

The data set was originally from the UK Department of Transport in 1984. It is monthly data on the total of drivers killed, as well as some additional variables. From this data set, three variables are analyzed as presented in table 1.

Table 1. Variables used in Model

Variable	**Type**	**Description**
DriversKilled	Discrete	n drivers killed
Distance	Continuous	Dist. driven (Mm)
Law	Binary	Seatbelt law in effect

The first variable, DriversKilled, represents the number of drivers killed in car accidents each month during the time period. The variable is discrete since it consists of count data which will affect the choice of models. Distance represents the total distance driven in the UK in 1/1000 km during the time period. This is included since it is reasonable to assume that the number of drivers killed will increase if there is more driving during a month. Including the variable will thus control for the possibility that the distance driven differed during the months the law was not in effect and the months the law was in effect. The final variable, Law, is a binary variable indicating whether the law is in effect during the month.

Table 2. Descriptive Statistics

	mean	min	max	sd
DriversKilled	122.80	60.00	198.00	25.38
Distance	14.99	7.68	21.63	2.94
Law	0.12	0.00	1.00	0.33

Table 2 indicates that on average 122.8 drivers died in car accidents each month during the time period with a minimum of 60 and a maximum of 198. The law was only in effect for 12% of the months and an average of 14,990 km were driven each month.

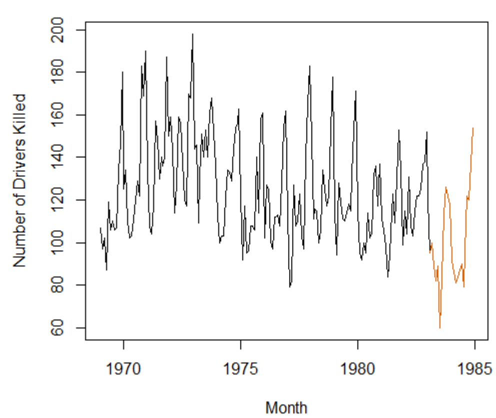

Figure 1. Drivers killed in accidents by month

Figure 1 indicates that number of drivers killed per month varies over the period in a seemingly seasonal pattern with lower numbers of deaths during the summer months. Moreover, there seems to be a weak decreasing linear trend which would suggest that the number of drivers killed in car accidents is decreasing over time. The orange section indicates the period in which the law was in effect.

3. Models

Two common choices for regression models on count data are the Poisson distribution and negative binomial distribution. Both models can be understood through a generalized linear models (GLM) framework. With GLM, the aim is to model the conditional expected value $E[y|X] = g^{-1}(\eta) = g^{-1}(\beta' X)$ as linearly related to the covariates through some link function. As the parameters of the Poisson and negative binomial distributions have positive support, a common choice of link function is the log function.

The GLM model can be extended to a time series setting by including autoregressive- and trend components in the model. The autoregressive component is included by regressing y on a set of lagged observations of y. ACF and PACF plots for the detrended data are presented in the appendix. The PACF spikes at 1 and 2 lags, as well as the seasonal pattern in the autocorrelations, leads to y_{t-1}, y_{t-2} and y_{t-12} being included in the models. As a log link is used, the lagged components are included on a log scale.

The plot of DriversKilled presented in figure 1 suggests that there could be a weak deterministic trend in the time series. Thus, a linear trend component is also included in the models. If the linear trend is ignored, it could bias the results by deflating or inflating the coefficient for the Law variable. The complete model is then $E[y_t|y_t^*, t, X] = \exp(\eta_t)$ where η_t is

$$\eta_t = \alpha + \gamma' y_t^* + \delta \log(t) + \beta' X)$$

With α as the model intercept, the γ coefficients for the log lagged observations, δ as the trend coefficient, and the β coefficients for the covariates. y_t^* represents the logged lag observations.

3.1. The Poisson Model

The Poisson distribution's mass function is defined as

$$\text{Po}(y|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Where k is 0 or any natural number. For the Poisson distribution, its mean is its parameter λ. Thus, the autoregressive Poisson model can be expressed as $\lambda_t = \exp(\eta_t)$. The likelihood for the conditional mean is then

$$L(\lambda|y, y^*, t, X) = \prod_{t=13}^{N} \frac{\lambda_t^{y_t} e^{-\lambda_t}}{y_t!}$$

In the Bayesian setting, priors are defined for each parameter resulting in the posterior distribution

$$p(\lambda|y, y^*, t, X)) \propto L(y, y^*, t, X) p(\alpha) p(\gamma) p(\delta) p(\beta)$$

3.2. The Negative Binomial Model

The negative binomial distribution's mass function with a suitable regression parameterization is defined as

$$\text{NB}(y|\mu, \phi) = \binom{y + \phi - 1}{y} \left(\frac{\mu}{\mu + \phi}\right)^y \left(\frac{\phi}{\mu + \phi}\right)^\phi$$

with μ as its mean parameter and ϕ as a dispersion parameter. This negative binomial parameterization has mean $E[y] = \mu$ and variance $V[y] = \mu + \frac{\mu^2}{\phi}$. As ϕ increases, the variance will converge towards μ reducing the negative binomial model to a Poisson model.

Using a log link again, drivers killed can be modeled as $\mu_t = \exp(\eta_t)$ with likelihood $L = L(\mu, \phi|y, y^*, t, X)$

$$L = \prod_{t=13}^{N} \binom{y_t + \phi - 1}{y_t} \left(\frac{\mu_t}{\mu_t + \phi}\right)^{y_t} \left(\frac{\phi}{\mu_t + \phi}\right)^\phi$$

resulting in the posterior distribution

$$p(\mu, \phi|y, y^*, t, X)) \propto L(y, y^*, t, X) p(\alpha) p(\gamma) p(\delta) p(\beta) p(\phi)$$

3.3. Priors

The intercept can be interpreted as the conditional base rate mean number of drivers killed. Since the intercept parameter can be any value on the real line, a suitable prior is the normal distribution. To get a crude prior mean for the intercept, the log of the number of drivers killed in the UK in 1966 divided by 12 and multiplied by 0.25 is used. 7,985 people were killed in traffic in 1966 resulting in an approximate mean for the intercept of 5.11. Since this is a rough estimate of the prior mean, and as the intercept is effected by the inclusion of covariates, a diffuse prior is used by setting the standard deviation to 10. A normal standard deviation of 10 will ensure that intercepts considerably lower and higher than 5.11 are considered.

For the autoregressive parameters, a $N(0, 0.5)$ prior is used. This reflects the belief that the autoregressive parameters are likely to be in the range between -1 and 1. The trend could be non-existent, positive or negative. Thus a prior distribution for δ should include values above and below 0. Since the strength and direction of the trend is uncertain, a $N(0, 10)$ prior is set for δ.

The law coefficient, β_1, will estimate the relative change in drivers killed when the new law is in effect. Since the decrease cannot be over 100 percent, and we would not expect an increase over 100 percent, we set a $N(0, 0.5)$ prior, conservatively assuming no prior effect of the law. The same reasoning is used to set a $N(0, 0.5)$ for the Distance parameter β_2.

For the negative binomial distribution, the dispersion parameter ϕ is restricted to be positive. Since there is no prior information about the dispersion in the number of drivers killed, no prior is specified for the parameter.

In summary, the priors for both models are defined as:

$$\alpha \sim N(5.11, 10) \qquad \gamma_i \sim N(0, 0.5), \forall\, i$$
$$\delta \sim N(0, 10) \qquad \beta_i \sim N(0, 0.5), \forall\, i$$

4. Results

4.1. Model fit

Both the Poisson- and negative binomial autoregressive models are fit using stan. Both models are run for 4 000 iterations using 4 chains with 2000 iterations being discarded as warmup. The expected values of the posterior parameters for the Poisson model are presented in table 3 and for the negative binomial model in table 4.

Table 3. Posterior statistics for the Poisson model

	variable	mean	q5	q95	rhat
1	α	1.13	0.64	1.61	1.00
2	γ_{lag1}	0.450	0.37	0.53	1.00
3	γ_{lag2}	-0.086	-0.16	-0.01	1.00
4	γ_{lag12}	0.424	0.36	0.49	1.00
5	δ	-0.055	-0.078	-0.032	1.00
6	β_{law}	-0.078	-0.120	-0.037	1.00
7	β_{dist}	0.092	0.003	0.016	1.00

The expected value of the posterior distribution of β_{law} is -0.078 rounded according to its Monte Carlo standard error. This suggests that the mean number of DriversKilled is affected by a multiplicative factor of $\exp(-0.078)$ which is roughly a decrease by 8% when the law is in effect controlling for autoregressiveness, a constant trend, and the distance driven. A 90% credible interval for the parameter $(-0.120, -0.037)$, which can be interpreted as with 90 percent probability, that the mean of the parameter is contained within that interval.

Unlike the Poisson model, the negative binomial has a dispersion parameter, ϕ, which has a mean of 98 in its posterior distribution. As the negative binomial approaches a Poisson when ϕ increases, a ϕ of 98 suggests that there is not a lot of overdispersion in the conditional distribution of DriversKilled. This is a reasonable result considering the similarity in model fit between the Poisson- and negative binomial models.

The negative binomial model fit is similar to the Poisson model fit with differences in the thousandth place. The expected value of the posterior distribution of β_{law} is approximately -0.08 with a 90% credible interval of approximately $(-0.14, -0.02)$. This corresponds to roughly an 8% decrease in the mean number of DriversKilled when the law is in effect controlling for autoregressiveness, a constant trend, and the distance driven.

Table 4. Posterior statistics for the negative binomial model

	variable	mean	q5	q95	rhat
1	α	1.17	0.4	1.9	1.00
2	γ_{lag1}	0.452	0.33	0.57	1.00
3	γ_{lag2}	-0.09	-0.21	0.03	1.00
4	γ_{lag12}	0.413	0.31	0.51	1.00
5	δ	-0.055	-0.091	-0.018	1.00
6	β_{law}	-0.08	-0.14	-0.02	1.00
7	β_{dist}	0.0089	-0.002	0.019	1.00
8	ϕ	98	70	133	1.00

4.2. Model Convergence

To evaluate model convergence, trace plots, and the $\hat{R}$ measure are considered for both models. Figure 4 and 5 in the appendix indicate that the chains for all the parameters are well-mixed and no clear trends indicate that the parameters have converged. $\hat{R}$ estimates the ratio of within-chain variance and between-chain variance (Gelman et al., 2014, p. 285) with a ratio close to 1 indicating that the chains are converged. Using a cut-off rule of 1.01 (Vehtari et al., 2021, p. 4), it is clear that each parameter in table 3 and 4 are suitably converged.

4.3. Model Comparison

To compare the models, Pareto-smoothed importance sampling leave-one-out cross-validated (PSIS-LOO) expected log pointwise predictive density (elpd) is computed for both models. Elpd indicates how well the model will perform in terms of prediction out of sample. In table 5, elpd and deviance are presented for both models where deviance is defined as $-2(\text{elpd})$. A higher elpd and lower deviance indicate better model fit. If the resulting $\hat{k}$ parameters from the Pareto distribution are too high (over 0.7), it may indicate that the predictive accuracy of the model is overestimated (Vehtari et al., 2017, p.5). In this case, all $\hat{k}$ values are under 0.7. Figures are included in the appendix.

Table 5. Comparison using PSIS LOO

	elpd	Deviance
Poisson	-804	1608
Negative Binomial	-764.3	1528.7

Since the elpd for the negative binomial model is higher than for the Poisson model, the negative binomial autoregressive model is a better fit for the data out-of-sample. This suggests that including the dispersion model results in better predictions of the number of drivers killed even though the overdispersion is low.

4.4. Effect of the law

Given that the negative binomial autoregressive model is a better fit for the variable DriversKilled, the effect of the law is considered using the negative binomial model. The posterior distribution from 8 000 draws of β_{law} is presented in figure 2.

The posterior distribution is symmetric and centered around -0.08. To compute the probability that the law had an effect, the probability that $P(\beta_{\text{law}}) < 0$ is computed using the posterior distribution. 7 870 of the 8 000 posterior draws are below 0 suggesting that there is a 0.984 probability that the effect of the law is lower than 0 when controlling for autoregression, the linear trend, and distance driven under the assumption of a negative binomial process.

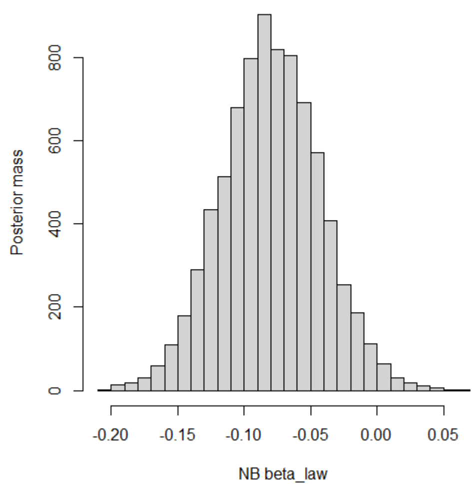

Figure 2. Posterior distribution for the β_{law} parameter in the NB model

5. Conclusion

A comparison of the Poisson autoregressive model and the negative binomial autoregressive model suggests that the negative binomial model performs better in terms of elpd. This could be due to the data exhibiting some dispersion with a computed expected ϕ of 98. However, since the dispersion is quite low, the actual estimates of the Poisson- and negative binomial model are quite similar and the Poisson model would not be a too unreasonable model for the variable DriversKilled.

Based on the negative binomial model, the report indicates that there is a 98.5% probability that the front seat seatbelt regulation was associated with a decrease in the absolute number of drivers killed in car accidents in the UK. The expected posterior effect is an approximate 8% decrease in deaths each month. However, the study only looked at a time period from 1969 to 1985 so it is uncertain whether the number of drivers killed continued to be lower further into the future.

While the model is a good fit in terms of $\hat{R}$, the model depends on stationarity assumptions that haven't been formally analyzed. The analyzed data shows patterns of seasonal variation which may cause non-stationarity. A more proper treatment of the seasonality, such as modeling the 12th difference of the time series, may have improved the reliability of the results.

Furthermore, only one covariate, distance driven, is controlled for in the analysis. Other covariates such as the average monthly petrol price and the number of injured people were included in the data set but not analyzed. Including these covariates could result in a better model by controlling for more covariates. Furthermore, it is possible that other covariates not in the data set such as the number of accidents would improve the analysis if available.

Furthermore, while reducing the total number of deaths is a good metric, it is not obvious that the reduction in number of deaths is directly due to the increased use of seatbelts. The introduction of new legislation could have affected the behavior of drivers by reminding them of road safety thus reducing deaths due to more safety-conscious behavior. A beta regression model could be used to estimate the law's effect on the proportion of accidents with a deadly outcome.

References

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2014), *Bayesian data analysis*, Texts in statistical science series, 3. ed. edn, Chapman & Hall/CRC, Boca Raton, Fla.

Vehtari, A., Gelman, A. and Gabry, J. (2017), 'Practical bayesian model evaluation using leave-one-out cross-validation and waic', *Statistics and computing* **27**(5), 1413–1432.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. and Bürkner, P.-C. (2021), 'Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of mcmc', *arXiv.org* .

Appendix

ACF and PACF

Figure 3 displays a time series for the variable DriversKilled alongside its ACF and PACF plots. The plots suggest the time series is autocorrelated with its first and second lag along with a twelfth lag to capture the seasonal component.

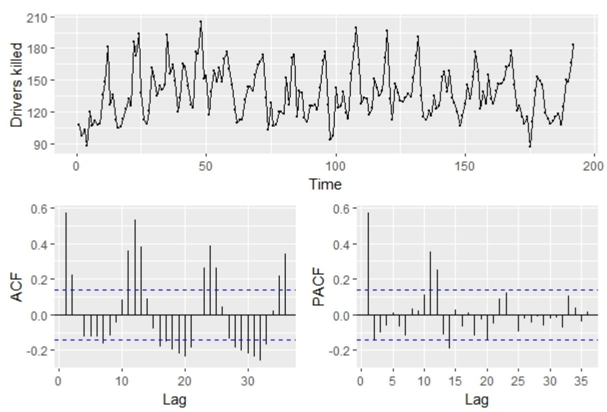

Figure 3. Drivers killed time series and correlogram, linear trend subtracted

MCSE

Table 6 and 7 summarize the Monte Carlo standard errors for estimated quantiles in the Poisson and negative binomial model. In the report, estimates are only displayed up to the last zero decimal as any decimals beyond that are influenced by the use of MCMC.

Table 6. Results for Poisson model MSCE

	MSCE mean	MSCE q5	MSCE q95
α	0.004	0.008	0.006
γ_{lag1}	0.0007	0.001	0.001
γ_{lag2}	0.0007	0.001	0.002
γ_{lag12}	0.0005	0.001	0.002
δ	0.0002	0.0004	0.0003
β_{law}	0.0003	0.0004	0.0009
β_{dist}	0.00006	0.0001	0.0001

Table 7. Results for Negative binomial MSCE

	MSCE mean	MSCE q5	MSCE q95
α	0.007	0.01	0.02
γ_{lag1}	0.001	0.003	0.002
γ_{lag2}	0.001	0.003	0.002
γ_{lag12}	0.008	0.002	0.002
δ	0.0003	0.0006	0.0006
β_{law}	0.0004	0.001	0.001
β_{dist}	0.00009	0.0002	0.0002
ϕ	0.26	0.31	1.15

Trace plots

Trace plots for both models are also examined to visually check for convergence of the chains.

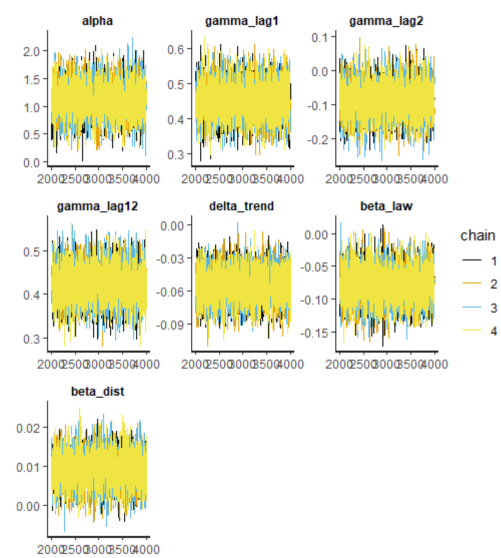

Figure 4. Trace plot, Poisson model

Figure 4 suggests that the Poisson model converged in 1000 iterations for each parameter. The chains are well mixed with no apparent trends.

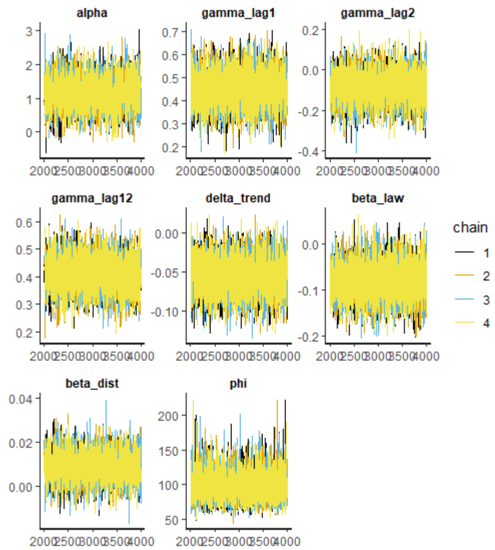

Figure 5. Trace plot, Negative binomial model

Figure 5 also suggests that the negative binomial model converged in 1000 iterations for each parameter as the chains are well mixed without any apparent trends.

In the report, convergence is further analyzed in more detail by considering computed $\hat{R}$ values.

PSIS Diagnostic Plots

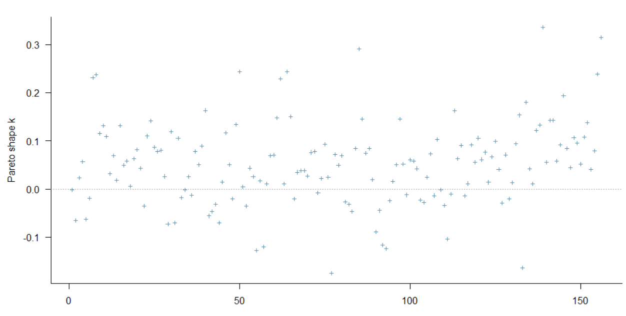

Figure 6. Tail shape estimates, Poisson model

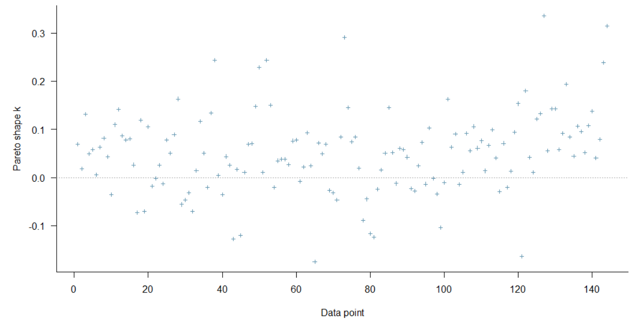

Figure 7. Tail shape estimates, Negative Binomial model