Intervention Analysis of NIH Datasets

Contents

TIRF All Hours

The time series is comprised on the monthly total number of deaths over all 24 hours for each day. Driver deaths only. The effect of the following interventions is modelled:

Cause of IV	**Date**	***T***
Graduated Licensing	June 1994	30
Administrative Suspensions	December 1996	60
Remedial Requirement & Increase in License Suspension	October 1998	82

A step intervention model defined by,

$$z_t = \mu + \delta_1 \xi_t^{(1)} + \delta_2 \xi_t^{(2)} + N_t \tag{1}$$

where N_t is the error term. Based on the pre-intervention data we assume initially that N_t is normal and independent, so ordinary multiple linear regression can be used. The intervention series are defined by,

$$\xi_t^{(1)} = \begin{cases} 0 & t < 30 \\ 1 & t \geq 30 \end{cases}$$

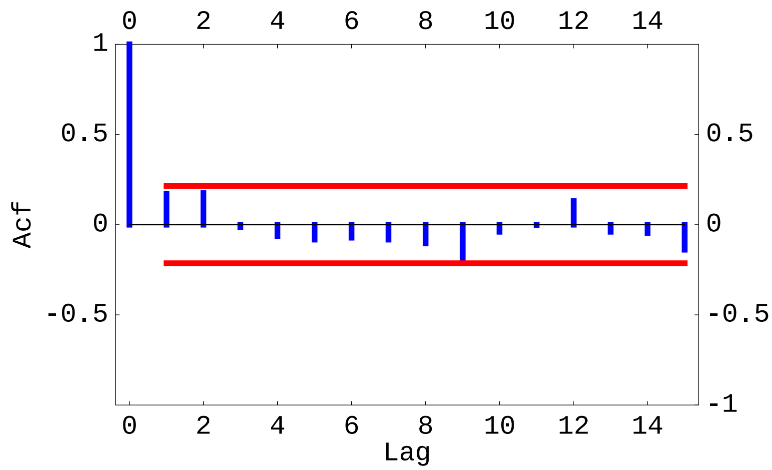
and

$$\xi_t^{(2)} = \begin{cases} 0 & t < 60 \\ 1 & t \geq 60 \end{cases}$$

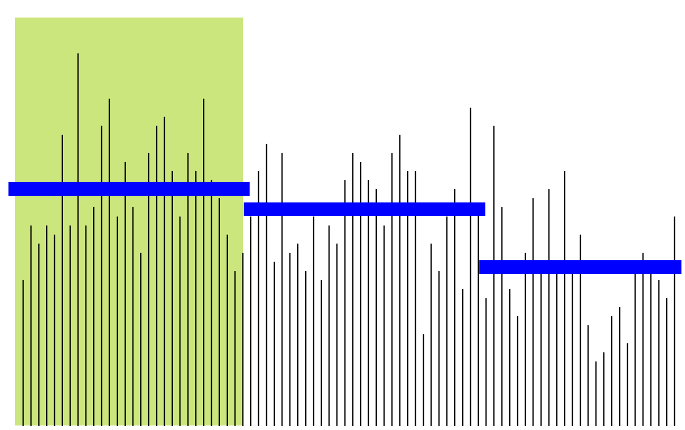
The regression result are tabulated below. Note that the PValue column is for a two-sided test. So on a one-sided test, the first intervention, Graduated Licensing is significant at about 8.5%. However Administrative Suspensions has had a much bigger effect as can be seen from the parameter estimate of -6.39 vs only -2.24.

	Estimate	**SE**	**TStat**	**PValue**
1	26.1034	1.15367	22.6265	0.
ξ1	-2.23678	1.61788	-1.38254	0.170606
ξ2	-6.38667	1.6824	-3.79616	0.000282553

As a check on the adequacy of the assumption for N_t the autocorrelation of the residuals is examined. The autocorrelation coefficients at lags one and two are nearly significant at 5%.



The graph below visually summarizes the data and the effect of the interventions. The three thick vertical lines show respectively from left to right μ, $\mu + \delta_1$, and $\mu + \delta_1 + \delta_2$.



Autocorrelated Error Model

From the autocorrelation plot it appears that there is some positive autocorrelation effect at lags one and two and possibily at the seasonal lag 12 as well. So some models were fit to eqn. (1) which allow for autocorrelation in the error term, N_t. Two models were fit,

$$N_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \tag{2}$$

and

$$N_t = (1 - \theta_1 \mathfrak{B} - \theta_2 \mathfrak{B}^2)(1 - \Theta_1 \mathfrak{B}^{12}) \tag{3}$$

where $a_t \sim \text{NID}(0, \sigma_a^2)$ and $\mathfrak{B}$ denotes the backshift operator on t, $\mathfrak{B}\, a_t = a_{t-1}$.

For the first model,

	Estimate	**SE**	**P-Value**
μ	25.832	0.872	0
δ_1	-1.734	2.025	0.392
δ_2	-6.655	2.102	0.002
θ_1	-0.148	0.107	0.165
θ_2	-0.202	0.107	0.058

And for the second model,

	Estimate	**SE**	**P-Value**
μ	25.937	0.961	0
δ_1	-2.127	2.079	0.306
δ_2	-6.408	2.148	0.003
θ_1	-0.141	0.107	0.189
θ_2	-0.195	0.107	0.069
Θ_1	-0.124	0.108	0.252

Diagnostic checks indicate that both models give acceptable fits. The first model is slightly more parsimonious and may be preferred on that grounds. But the second model has a coefficient, δ_1, which is more in line with the previous regression results.

It may be concluded that after allowing for the effect of autocorrelation, the signficance level of the first intervention due Graduated Licensing is increased and in the non-seasonal model, its effect is also lessed. The second intervention due to Administrative Suspensions remains statistically significant.

Late-Night TIRF Time Series with All Drinking Classes

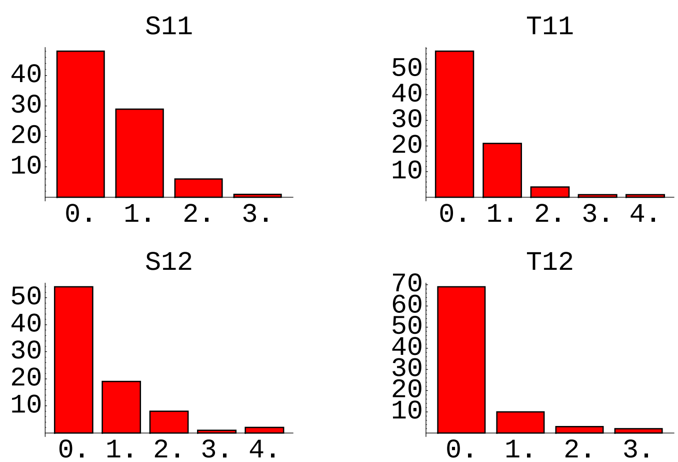
Introduction

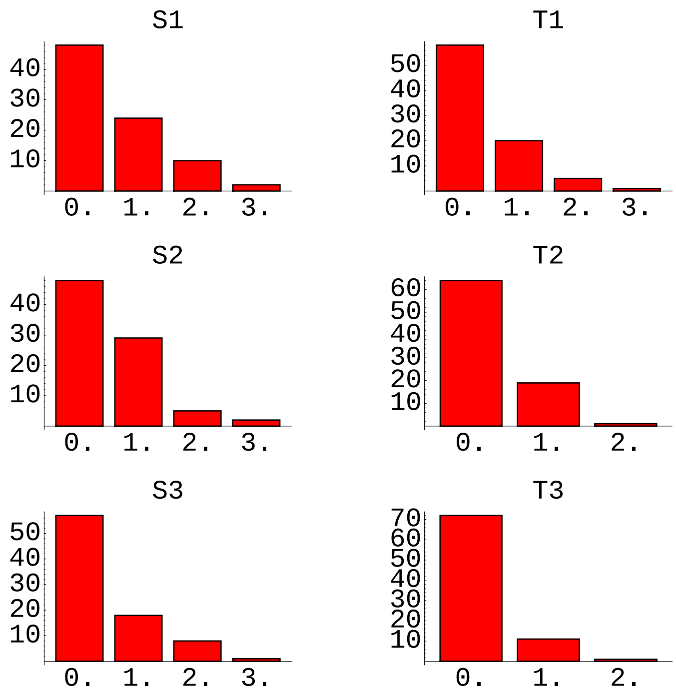
TIRF, Traffic Injury Research Foundation, provided this data on fatal car accidents in Ontario from January 1,1992 to December 31, 1998. This data are for automobile driver deaths only. The data for drinking classes "yes", "no" and "unknown" are combined in this analysis.

Ten time series were created from the TIRF dataset corresponding the two weekgroup variables SunWed and ThuSat and the five hour one hour periods beginning at 11PM, 12AM, 1AM, 2AM and 3AM. For brevity we will refer to these time series using the codes S11, S12, S1, S2, S3, T11, T12, T1, T2 and T3. The time series were aggregated to a monthly level starting January 1992 and running to December 1998. There are $n = 84$ consecutive observations in total for each time series. In our analysis of these latenight time series we are primarily interested in testing to see if a change occurred starting effective with May 1996, the 53rd observation.

Bar Chart Summaries

The TIRF late-night time series are comprised of small numbers mostly zeros. The Bar Charts look very similar to data from a Poisson distribution. However the data for T11, S12 and T12 are over-dispersed as is confirmed by the Poisson dispersion test.





Autocorrelation Analysis

The sample autocorrelation at lag k is defined by,

$$r_k = \frac{\sum_{t=k+1}^{n} (z_t - \bar{z})(z_{t-k} - \bar{z})}{\sum_{t=k+1}^{n} (z_t - \bar{z})^2}, \; k = 0, \; 1, \; 2, \; \ldots$$

provides a fairly robust test for possible serial dependence even for data which is highly discrete such as the TIRF late-night time series. If there is possible dependence we would expect it to be strongest at lag one or possibly at the seasonal lag of 12.

The standard deviation of the lag one autocorrelation coefficient is $1 / \sqrt{84}$ and the benchmark significance limits are $1.96 / \sqrt{84} \doteq 0.213$.

The table below gives r_1 and we see that there is no evidence of an autocorrelation in these time series at lag one.

S11	0.0430784
S12	0.127372
S1	0.181458
S2	−0.0325027
S3	0.0587406
T11	−0.0990359
T12	−0.0242566
T1	−0.0032918
T2	−0.00352113
T3	0.0210707

The table below shows r_{12}, only SunWed-2AM window shows significant seasonal correlation at the 5% level.

S11	0.00960531
S12	−0.0303431
S1	−0.0431046
S2	**0.255263**
S3	−0.0127046
T11	0.0508326
T12	−0.0449036
T1	0.0368116
T2	−0.0985915
T3	0.0127013

In view of these results, we may assume that time series are approximately statistically independent.

Poisson Modelling

Poisson Dispersion Test

We test if the pre-intervention data (ie. the first 52 observations) are approximately Poission distributed. Let z_t, $t = 1$, ..., 52 denote the values in the series. Then the Poisson dispersion test is based on the statistic,

$$d = \frac{\sum_{i=1}^{n} (z_t - \bar{z})^2}{\bar{z}}$$

where $\bar{z}$ is the sample mean and $n = 52$. Under the null hypothesis that the data are independent Poisson random variables, d, is distributed approximately as χ^2 on $n-1$ df. The table below suggests that the data for the SunWed group are Poisson but there is a strong indications over over-dispersion in **S12, T11** and **T12**.

	d	p-value
S11	38.	0.911302
S12	**76.6667**	**0.0115418**
S1	45.8	0.679696
S2	48.1034	0.589411
S3	54.8947	0.329282
T11	**71.2222**	**0.0321581**
T12	**78.8182**	**0.00747674**

T1	56.12	0.288949
T2	42.5	0.795739
T3	47.	0.633224

■ Poisson Model

We will use the notation $z_t \sim \text{IPo}(\lambda_t)$, $t = 1, \ldots, n$ to mean that the random variables z_t, $t = 1, \ldots, n$ are independently distributed Poisson random variables with means λ_t. Then our intervention analysis model may be written, $z_t \sim \text{IPo}(\lambda_t)$, $t = 1, \ldots, n$, where $n = 84$ and

$$\lambda_t = \begin{cases} \lambda, & t = 1, \ldots, 52 \\ \lambda + \delta, & t = 53, \ldots, 84 \end{cases}$$

The null hypothesis of no effect is then $H_0 : \delta = 0$. The exact log likelihood function for our model may be written as

$$\mathcal{L}(\lambda, \delta) = \sum_{t=1}^{n} z_t \log(\lambda_t) - \sum_{t=1}^{n} \lambda_t$$

This function may be maximized numerically to obtain the maximum likelihood estimates for λ and δ, which may be denoted by $\hat{\lambda}$ and $\hat{\delta}$. Then the null hypothesis $H_0 : \delta = 0$ may be tested using a likelihood ratio test and a confidence interval for the parameter δ may be given. Under the null hypothesis $H_0 : \delta = 0$ the loglikelihood function simplies to

$$\mathcal{L}(\lambda, 0) = \sum_{t=1}^{n} z_t \log(\lambda) - n\,\lambda$$

which is maximized with $\hat{\lambda}_0 = \bar{z}$ where $\bar{z}$ denotes the sample mean. The likelihood ratio statistic may be written,

$$R = 2\,(\max_{\lambda,\delta}\ \mathcal{L}(\lambda, \delta) - \max_{\lambda}\ \mathcal{L}(\lambda, 0)).$$

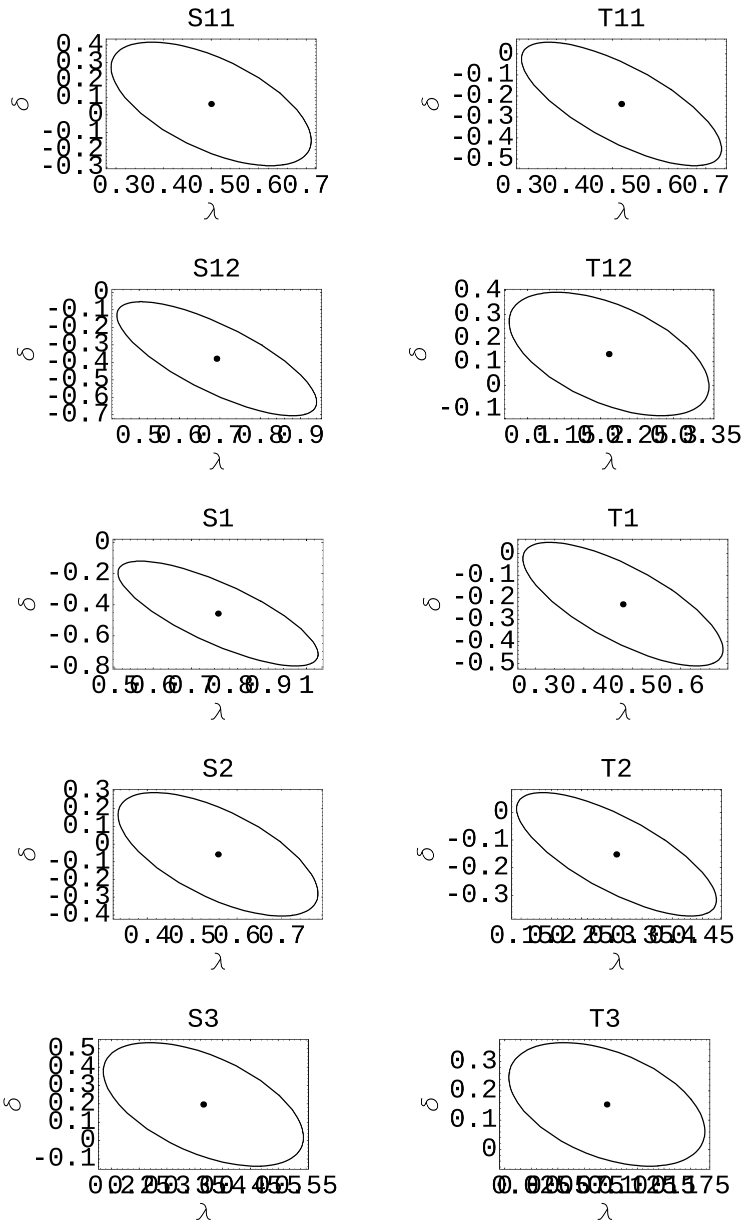
Under $H_0 : \delta = 0$, R is χ^2- distributed on 1 df. From the table below we see that there is evidence that $\delta \neq 0$ for the S12, S1, T1 and T2 windows at the 10% level. In all these cases $\delta < 0$.

■ Fitted Parameters, Standard Errors, R and p-value

	λ	seλ	δ	seδ	R	p-value
S11	0.5	0.098	0.062	0.165	0.146	0.702
S12	**0.692**	**0.115**	**-0.38**	**0.152**	**5.661**	**0.017**
S1	**0.769**	**0.122**	**-0.457**	**0.157**	**7.627**	**0.006**
S2	0.558	0.104	-0.058	0.162	0.124	0.725
S3	**0.365**	**0.084**	**0.197**	**0.157**	**1.701**	**0.192**
T11	**0.519**	**0.1**	**-0.238**	**0.137**	**2.78**	**0.095**
T12	0.212	0.064	0.132	0.122	1.284	0.257
T1	**0.481**	**0.096**	**-0.231**	**0.131**	**2.865**	**0.091**
T2	**0.308**	**0.077**	**-0.151**	**0.104**	**1.944**	**0.163**
T3	**0.096**	**0.043**	**0.154**	**0.098**	**2.914**	**0.088**

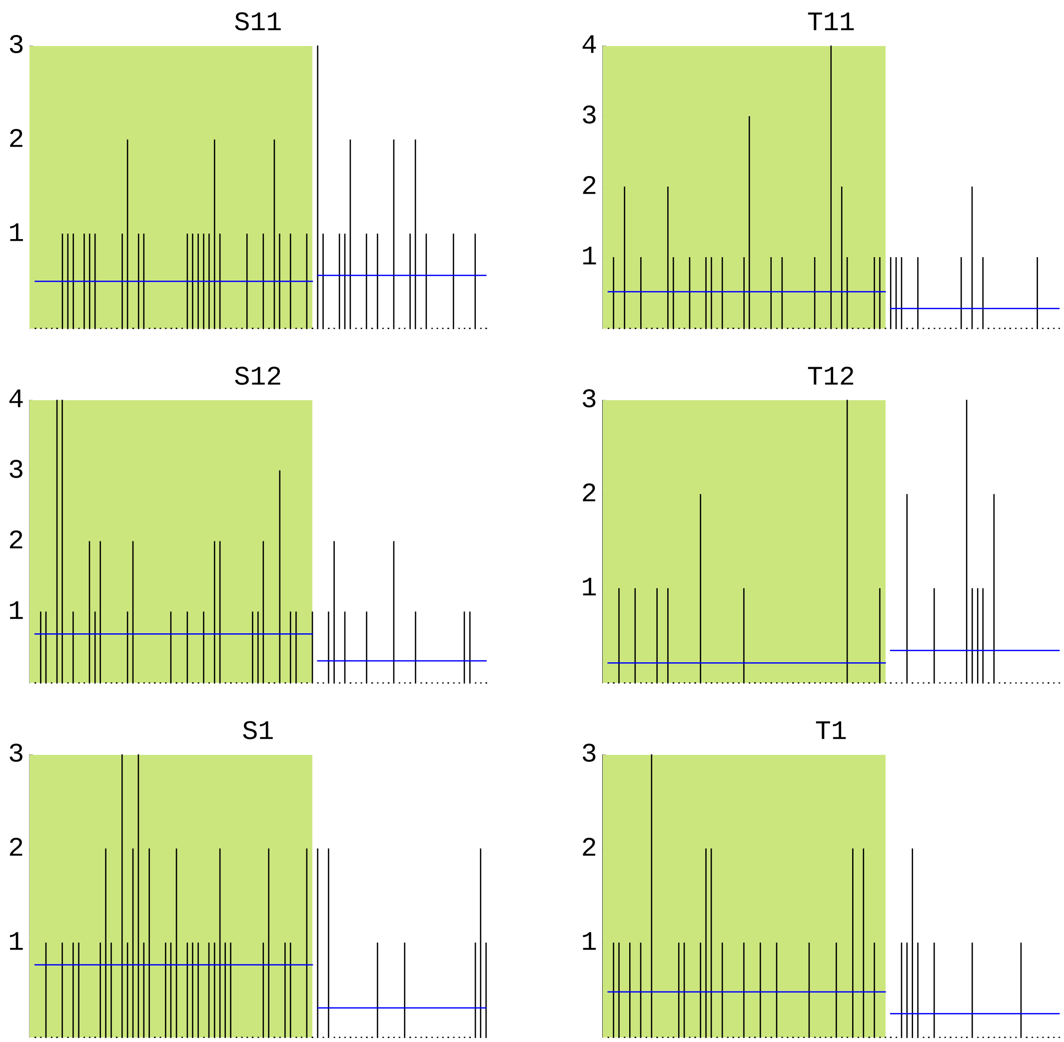
The p-value reported is for a two-sided test. A one-sided test would seem to be more appropriate so in this case the p-value in the above table should be halved.

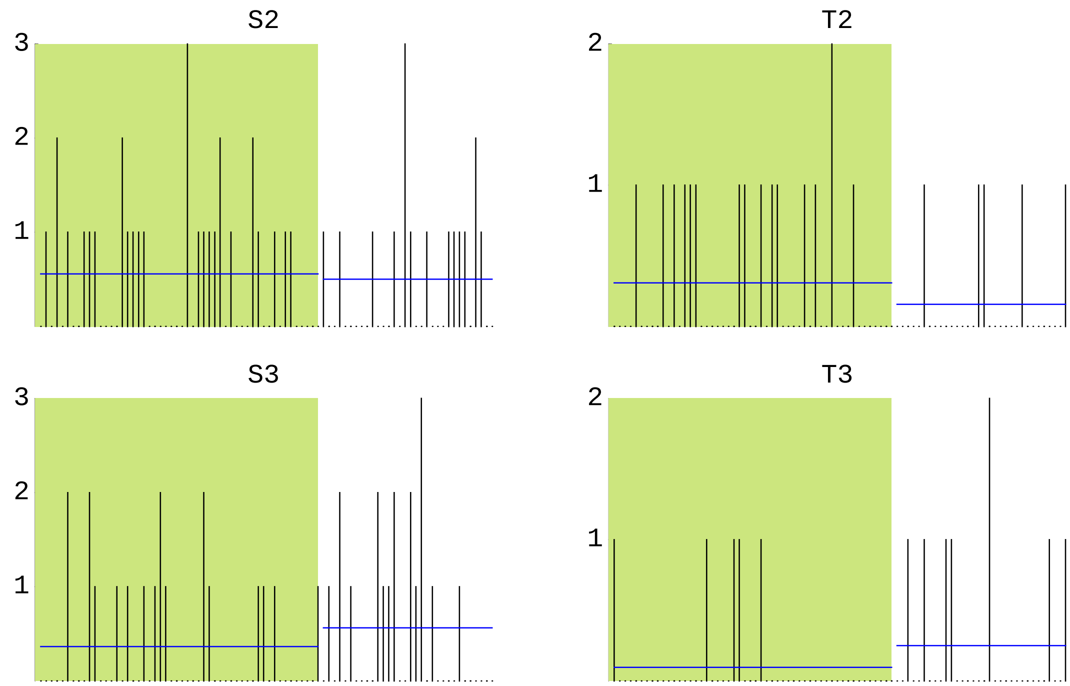
■ 90% Confidence Ellipses for λ and δ



■ Fitted Values and Visualization

The expected values of z_t in our model is given $\in\{z_t\} = \lambda_t = \lambda + \delta\,\xi_t$. The expected value is shown as a blue line in the graphs below.





Negative Binomial Modelling

■ Negative Binomial Distribution

Actual count data are often over-dispersed, that is, they fail the Poission dispersion test. In this case, the negative binomial distribution provides a more flexible alternative than the Poisson distribution for modelling discrete random variables. Suppose there is an unobserved random variable E having a gamma distribution with mean 1 and variance $1/\theta$ and that conditional on E, the random variable Y has a Poisson distribution with mean μE. Then Y has a negative binomial distribution and its density function may be written,

$$f(y; \theta, \mu) = \frac{\Gamma(\theta + y)}{\Gamma(\theta)\, y!} \, \frac{\mu^y \theta^\theta}{(\mu + \theta)^{\theta + y}}$$

The mean and variance of Y are given by $\mathcal{E}\{Y\} = \mu$ and $\text{Var}\{Y\} = \mu + \mu^2/\theta$. Notationally we may denote this distribution by NB(μ, θ).

■ Generalized Linear Models

The generalized linear model provides an alternative and more general statistical model for these data. GLM's are frequently used for regression modelling of non-Gaussian data such as data arising from the binomial, lognormal or negative binomial distributions. Given independently distributed z_t, $t = 1, \ldots, n$ and possibly p covariates of interest $x_{t,j}$, $t = 1, \ldots, n$, $j = 1, \ldots, p$ the GLM may be defined. There are three components to a GLM:

(i) the statistical density or probability function, $f(z_t;\ \mu_t;\ \theta)$, where θ denotes distributional parameters, $\mu_t = \mathcal{E}\{z_t\}$ and it is assumed that μ_t depends on the distribution parameter or parameters as well as the covariates.

(ii) the linear predictor which depends on the covariates linearly,

$$\eta_t = \sum_{j=1}^{p} \alpha_t \, x_{t,j}$$

(iii) the link function, $\eta_t = \ell(\mu_t)$.

The standard GLM algorithm is based on Iteratively Reweighted Least Squares (IRLS) and this algorithm provides a good approximation to the more exact maximum likelihood method. Using *Mathematica* it is possible to obtain the exact maximum likelihood estimates which are preferable to the IRLS estimates.

■ Model Formulation

For $j = 1$, we take $x_{t,1} = 1$, $t = 1, \ldots, n$ which corresponds to the overall mean. The intervention is represented by,

$$x_{t,2} = \xi_t = \begin{cases} 0 & t \leq 48 \\ 1 & t > 49 \end{cases}$$

It is assumed that $z_t \sim \text{NB}(\lambda_t,\ \theta)$ where

$$\log(\lambda_t) = \lambda + \delta \, \xi_t$$

that is, the link function is taken to be logarithmic.

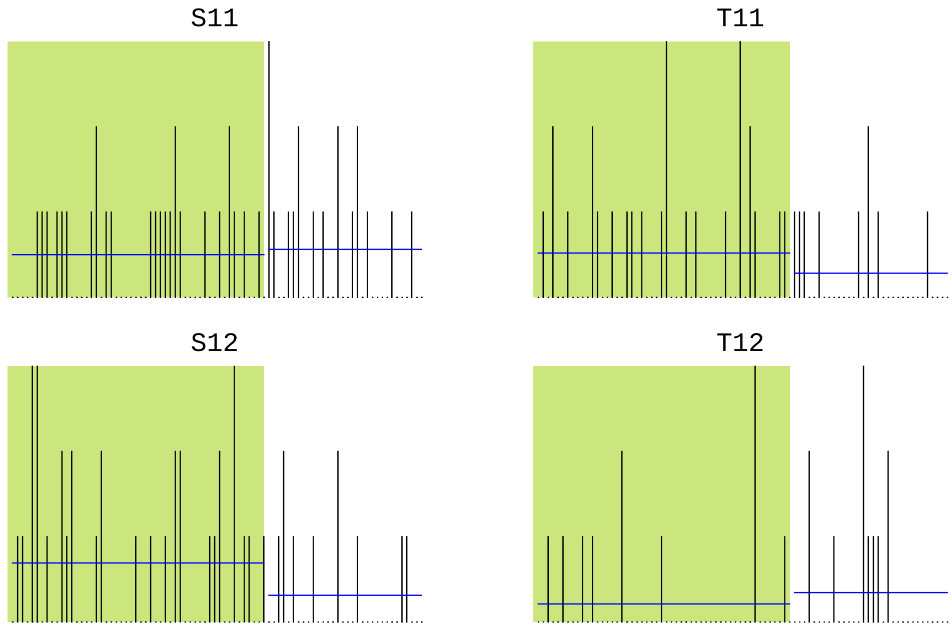
■ Maximum Likelihood Estimates

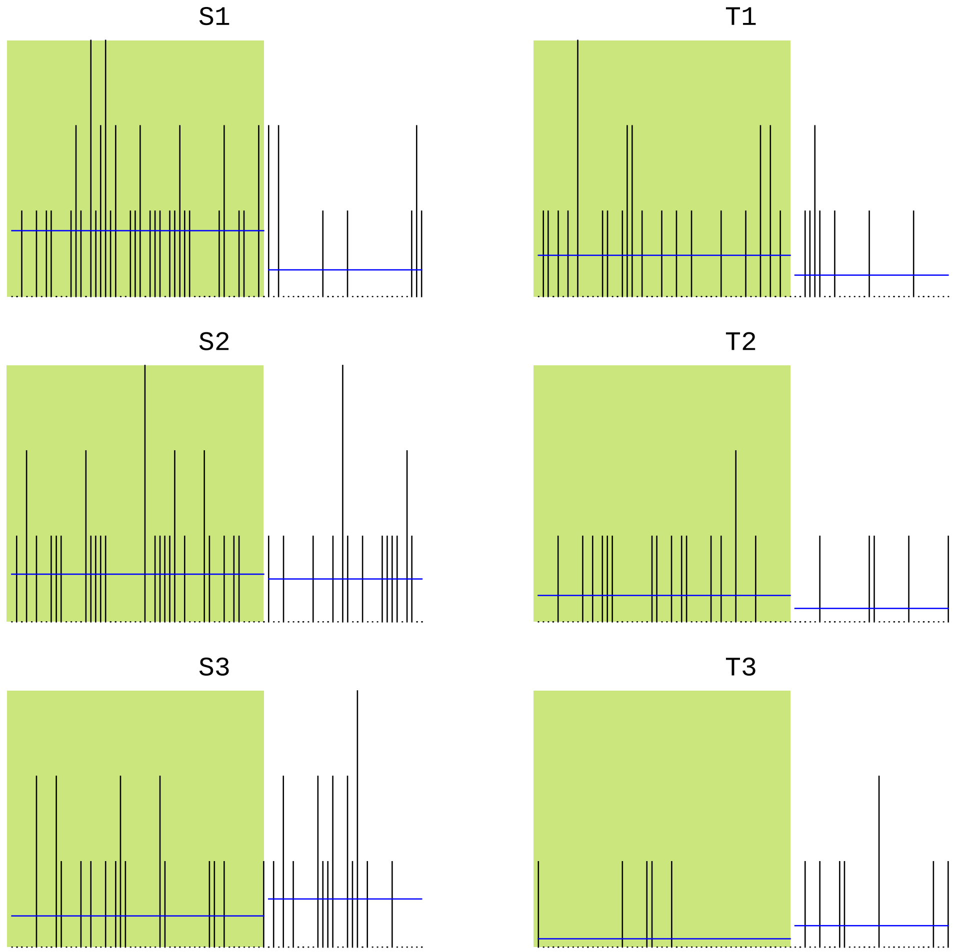
	λ	δ	θ	R	p-value
S11	-0.693	0.118	148533.	0.146	0.702
S12	**-0.368**	**-0.795**	**1.407**	**4.038**	**0.044**
S1	**-0.262**	**-0.901**	**23233.4**	**7.529**	**0.006**
S2	-0.584	-0.109	430.549	0.119	0.73
S3	-1.007	0.431	2.565	1.426	0.232
T11	-0.655	-0.613	1.953	2.269	0.132
T12	-1.553	0.485	0.399	0.757	0.384
T1	**-0.732**	**-0.654**	**6.288**	**2.661**	**0.103**
T2	**-1.179**	**-0.678**	**313336.**	**1.944**	**0.163**
T3	**-2.338**	**0.951**	**1054.07**	**2.913**	**0.088**

The p-value reported is for a two-sided test. A one-sided test would seem to be more appropriate so in this case the p-value in the above table should be halved.

■ Visualization

The expected values of z_t in the negative binomial regression model is given by $\mathbb{E}\{z_t\} = e^{\lambda+\delta\,\xi_t}$. As expected the impact of the interventions is almost the same as that given the by Poisson model.





Comparison with Normal Regression Modelling

The late night time series are comprised of very small numbers and these number clearly violate the assumption of normality as the bar charts made clear. However normal regression model would be expected to be fairly robust against such departures as shown by Hjort (1994). It is of interest to compare our previous analyses using Poisson and Negative Binomial regression with standard normal regression. In the standard normal regression we may formulate our step intervention model,

$$z_t = \mu + \delta\, \xi_t + N_t \qquad (1)$$

where N_t is the error term. Based on the pre-intervention data we assume initially that N_t is normal and independent, so ordinary multiple linear regression can be used. The intervention series are defined by,

$$\xi_t = \begin{cases} 0 & t < 53 \\ 1 & t \geq 53 \end{cases}$$

The term N_t represents the disturbance or error term and it has been tentatively identified as Gaussian white noise, that is $N_t = a_t$, where $a_t \sim \mathrm{NID}(0, \sigma^2)$.

The following table is in quite close agreement with the results from the Poisson analyses. However only 6 interventions are detected on a one-sided test at the 10% level whereas previously there were 7 and significance levels are larger suggesting this analysis is not quite as sensitive as the Poisson analysis. There is almost no difference though in many cases such as for T2 and T3.

		Estimate	SE	TStat	PValue
S11	1	0.5	0.0954831	5.23653	1.23913×10^{-6}
	ξ	0.0625	0.1547	0.404007	0.687259

		Estimate	SE	TStat	PValue
S12	1	0.692308	0.122467	5.65303	2.22552×10^{-7}
	ξ	**-0.379808**	**0.198419**	**-1.91417**	**0.0590865**

		Estimate	SE	TStat	PValue
S1	1	0.769231	0.106216	7.24214	2.16742×10^{-10}
	ξ	**-0.456731**	**0.17209**	**-2.65403**	**0.00955079**

		Estimate	SE	TStat	PValue
S2	1	0.557692	0.100219	5.56474	3.21682×10^{-7}
	ξ	-0.0576923	0.162373	-0.355307	0.723272

		Estimate	SE	TStat	PValue
S3	1	0.365385	0.0991671	3.68453	0.000409787
	ξ	0.197115	0.160669	1.22684	0.223393

		Estimate	SE	TStat	PValue
T11	1	0.519231	0.103242	5.02926	2.85058×10^{-6}
	ξ	**-0.237981**	**0.167271**	**-1.42273**	**0.158609**

		Estimate	SE	TStat	PValue
T12	1	0.211538	0.0891536	2.37274	0.0199973
	ξ	0.132212	0.144445	0.915305	0.362715

		Estimate	SE	TStat	PValue
T1	1	0.480769	0.0905745	5.308	9.26536×10^{-7}
	ξ	**-0.230769**	**0.146748**	**-1.57256**	**0.119672**

		Estimate	SE	TStat	PValue
T2	1	0.307692	0.0636884	4.83122	6.22699×10^{-6}
	ξ	**-0.151442**	**0.103187**	**-1.46765**	**0.146025**

		Estimate	SE	TStat	PValue
T3	1	0.0961538	0.0541851	1.77454	0.0796849
	ξ	**0.153846**	**0.08779**	**1.75243**	**0.0834372**

References

Hjort, N.L. (1994), The exact amount of t-ness that the normal model can tolerate, *Journal of the American Statistical Association* 89, 665-675.

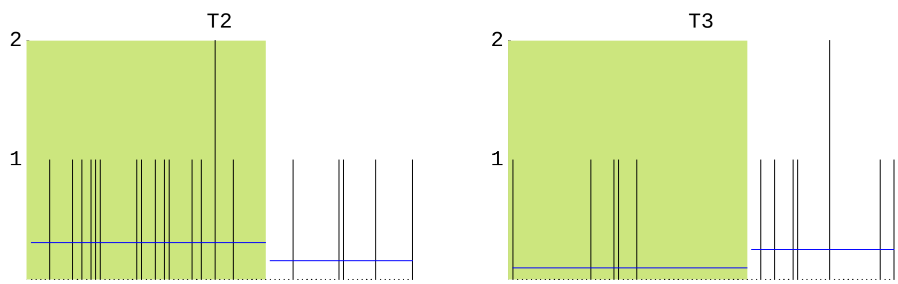
■ TIRF Conclusion

The clearest evidence of a possible effect due to May 1996 legislation in the TIRF data we modelled seems to be with the TIRF late-night all fatalities data.

The Traffic Injury Research Foundation, provided data on all car accidents in Ontario from January 1,1992 to December 31, 1998 in which the driver was killed. Our objective is to see if there has been any change since May 1996 when new legislation extending bar closure time to 2AM came into effect. Drunk driving is known to be more common during the early morning hours on Fridays, Saturdays, and Sundays than other weekdays so the data was aggregated into two week groups, denoted by S and T. Two hourly windows from 2AM to 3AM and 3AM to 4AM were compared. The time series in this windows do not exhibit significant autocorrelation or over-dispersion relative to the Poisson distribution so a Poisson distribution with mean function,

$$\lambda_t = \begin{cases} \lambda, & t = 1, \ldots, 52 \\ \lambda + \delta, & t = 53, \ldots, 84 \end{cases}$$

was fit to the data by exact maximum likelihood. According to the *temporal shift hypothesis*, the extended bar hours may be expected to have caused a shift in drunk driving from the 1AM to 2AM window to the 2AM to 3AM window. In the plot below the observed fatalities per month for the T weekgroup are shown by vertical lines equal to the number of fatalities which is either 0, 1 or 2 for this data. The shaded area indicates the pre-May 1996 period and the blue horizontal line is the estimated mean. The drop in T2 is significant at the 10% level and increase in T3 is significant at 5% level.



MTO Fatal, DRCOND

The time series comprises all fatal accidents. The data were aggregated to monthly time series from January 1992 to December 1998. So there are $n = 84$ observations. May 1996 corresponds to observation #53 in this series. There are 16 time series corresponding to the variables for hour window (11PM, 12AM, 1AM, 2AM), weekgroup (SunWed, ThuSat) and driver condition (sober, drunk). The codes used and their defintions are given in the table below.

#	Code	Definition
1	PM11SS	11PM SunWed Sober
2	AM12SS	12AM SunWed Sober
3	AM1SS	1AM SunWed Sober
4	AM2SS	2AM SunWed Sober
5	PM11TS	11PM ThuSat Sober
6	AM12TS	12AM ThuSat Sober
7	AM1TS	1AM ThuSat Sober
8	AM2TS	2AM ThuSat Sober
9	PM11SD	11PM SunWed Drunk
10	AM12SD	12AM SunWed Drunk
11	AM1SD	1AM SunWed Drunk
12	AM2SD	2AM SunWed Drunk
13	PM11TD	11PM ThuSat Drunk
14	AM12TD	12AM ThuSat Drunk
15	AM1TD	1AM ThuSat Drunk
16	AM2TD	2AM ThuSat Drunk

The time series are comprised of very small counts mostly 0's and some 1's, 2's and 3's. Such time series are better modelled using a Poisson distribution or negative binomial but as we say previously in the TIRF late-night time series analysis, the sensitivity was only slightly improved. As a first step we will use usual intervention analysis method.

May 1996 corresponds to observation #53 in this series. A step intervention model defined by,

$$z_t = \mu + \delta_1 \xi_t + N_t \tag{1}$$

where N_t is the error term. Based on the pre-intervention data we assume initially that N_t is normal and independent, so ordinary multiple linear regression can be used. The intervention series are defined by,

$$\xi_t = \begin{cases} 0 & t < 53 \\ 1 & t \geq 53 \end{cases}$$

The pre-intervention series is fairly short so the following two-step approach to the ARIMA identification of the error term N_t will be used. In the first stage the model in eqn. (1) is fit using standard regression. Then the residuals, the estimated values, $\hat{N}_t$, are obtained and the residual autocorrelation of these residuals is examined. This approach should be expected to work well and is indeed theoretically superior to the alternative approach of basing the model identification on the pre-intervention residuals. The reason for this is that in the estimates of the autocorrelations of $\hat{N}_t$ are first-order efficient as are the estimates in the pre-intervention approach. However since the sample size is larger, this approach provides better estimates.

Regression Results

There was no detectable autocorrelation.

		Estimate	SE	TStat	PValue
PM11SS	1	0.538462	0.101349	5.31295	9.07993×10^{-7}
	ξ	−0.225962	0.164204	−1.3761	0.172536

		Estimate	SE	TStat	PValue
AM12SS	1	0.615385	0.110758	5.55613	3.33407×10^{-7}
	ξ	−0.115385	0.179448	−0.642997	0.52202

		Estimate	SE	TStat	PValue
AM1SS	1	0.480769	0.0921389	5.21788	1.33641×10^{-6}
	ξ	−0.137019	0.149282	−0.917855	0.361387

		Estimate	SE	TStat	PValue
AM2SS	1	0.538462	0.0928975	5.7963	1.21829×10^{-7}
	ξ	−0.100962	0.150511	−0.670791	0.504239

		Estimate	SE	TStat	PValue
PM11TS	1	0.692308	0.115189	6.01019	4.90556×10^{-8}
	ξ	**-0.317308**	**0.186628**	**-1.70022**	**0.0928799**

		Estimate	SE	TStat	PValue
AM12TS	1	1.03846	0.138186	7.51494	6.33127×10^{-11}
	ξ	**-0.413462**	**0.223887**	**-1.84674**	**0.0683925**

		Estimate	SE	TStat	PValue
AM1TS	1	0.461538	0.0930551	4.95984	3.75512×10^{-6}
	ξ	−0.211538	0.150767	−1.40309	0.164366

		Estimate	SE	TStat	PValue
AM2TS	1	0.942308	0.131471	7.1674	3.03204×10^{-10}
	ξ	−0.129808	0.213008	−0.609404	0.543942

		Estimate	SE	TStat	PValue
PM11SD	1	0.538462	0.101349	5.31295	9.07993×10^{-7}
	ξ	−0.225962	0.164204	−1.3761	0.172536

		Estimate	SE	TStat	PValue
AM12SD	1	0.615385	0.110758	5.55613	3.33407×10^{-7}
	ξ	−0.115385	0.179448	−0.642997	0.52202

		Estimate	SE	TStat	PValue
AM1SD	1	0.480769	0.0921389	5.21788	1.33641×10^{-6}
	ξ	−0.137019	0.149282	−0.917855	0.361387

		Estimate	SE	TStat	PValue
AM2SD	1	0.538462	0.0928975	5.7963	1.21829×10^{-7}
	ξ	−0.100962	0.150511	−0.670791	0.504239

		Estimate	SE	TStat	PValue
PM11TD	1	0.692308	0.115189	6.01019	4.90556×10^{-8}
	ξ	**-0.317308**	**0.186628**	**-1.70022**	**0.0928799**

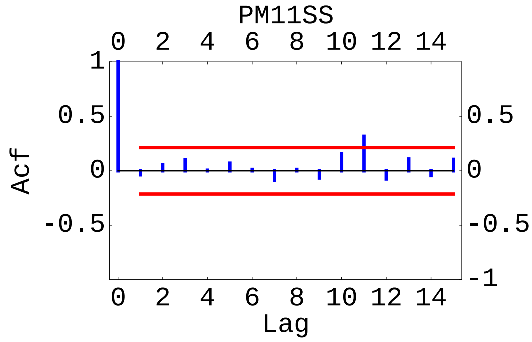
		Estimate	SE	TStat	PValue
AM12TD	1	1.03846	0.138186	7.51494	6.33127×10^{-11}
	ξ	**-0.413462**	**0.223887**	**-1.84674**	**0.0683925**

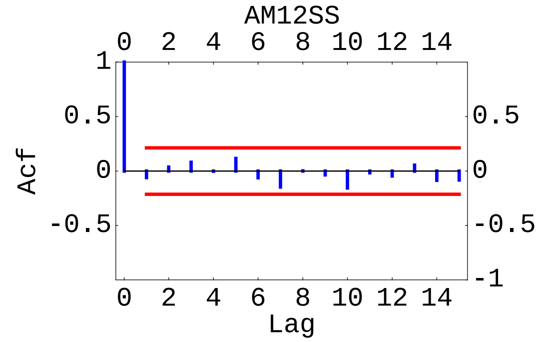
		Estimate	SE	TStat	PValue
AM1TD	1	0.461538	0.0930551	4.95984	3.75512×10^{-6}
	ξ	−0.211538	0.150767	−1.40309	0.164366

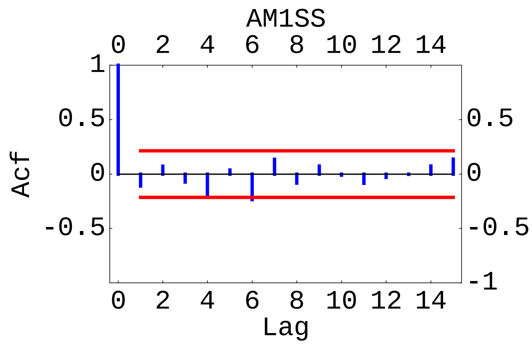
		Estimate	SE	TStat	PValue
AM2TD	1	0.942308	0.131471	7.1674	3.03204×10^{-10}
	ξ	−0.129808	0.213008	−0.609404	0.543942

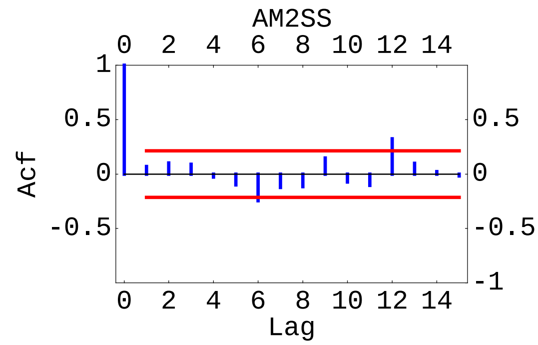
Residual Autocorrelation Analysis

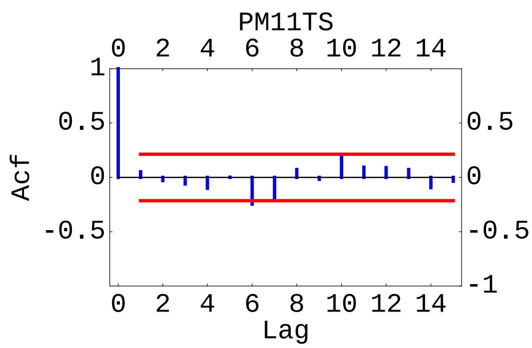
The residual autocorrelation from the regression model are examined and used to select tentative ARIMA models.

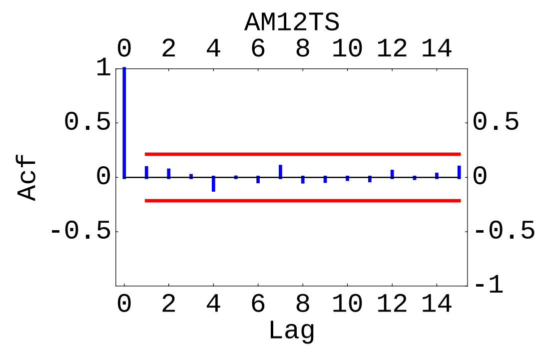

PM11SS
Acf
Lag

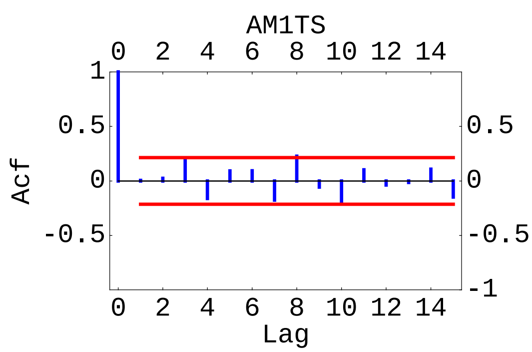

AM12SS
Acf
Lag

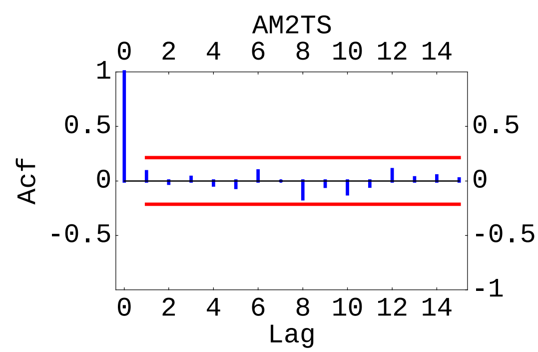

AM1SS
Acf
Lag

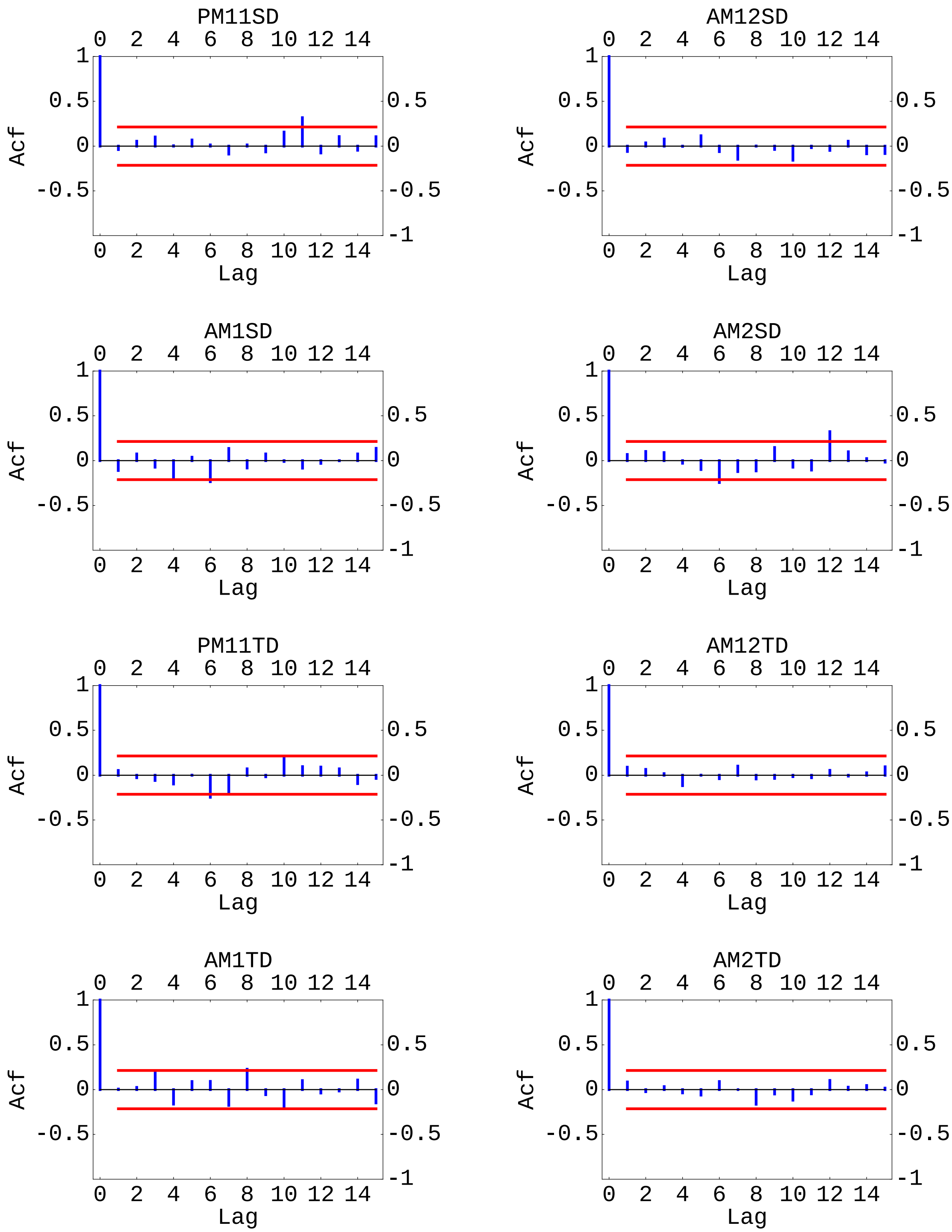

AM2SS
Acf
Lag


PM11TS
Acf
Lag

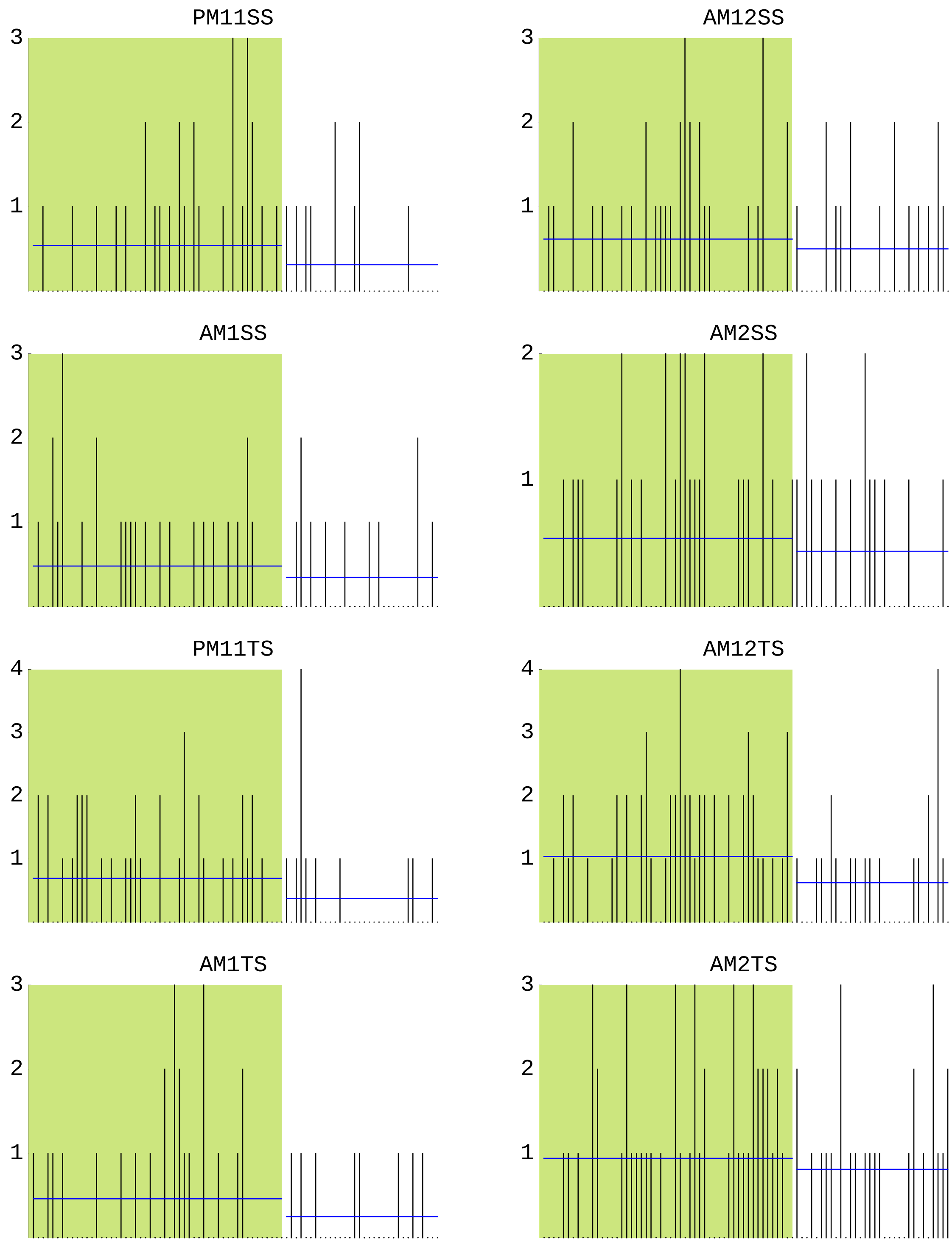

AM12TS
Acf
Lag

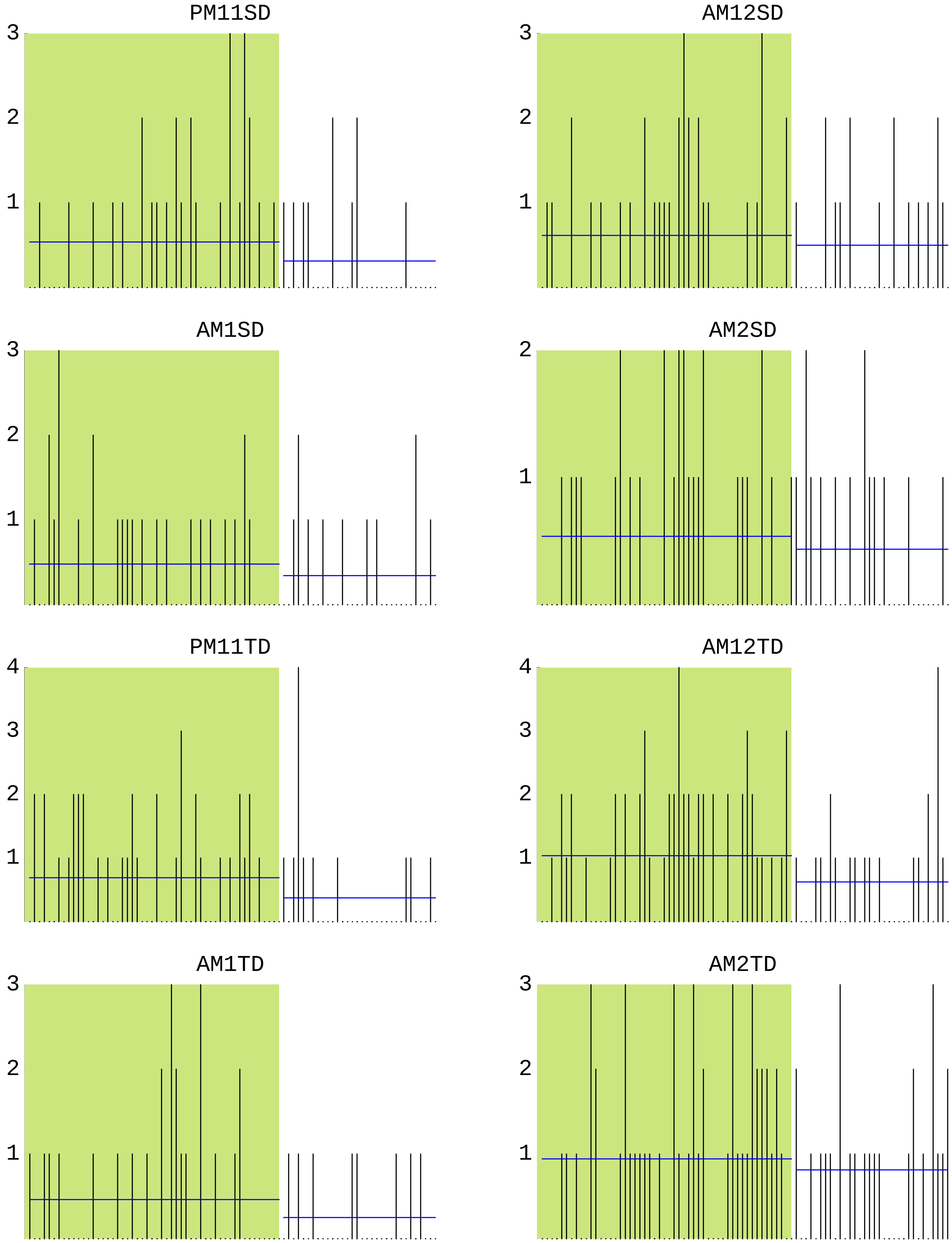

AM1TS
Acf
Lag


AM2TS
Acf
Lag


PM11SD
AM12SD
AM1SD
AM2SD
PM11TD
AM12TD
AM1TD
AM2TD
Acf
Lag

Intervention Visualization




PM11SD
3
2
1
AM12SD
3
2
1
AM1SD
3
2
1
AM2SD
2
1
PM11TD
4
3
2
1
AM12TD
4
3
2
1
AM1TD
3
2
1
AM2TD
3
2
1

MTO Major, DRCOND

The time series comprises all accidents which resulted in major injury. The data were aggregated to monthly time series from January 1992 to December 1998. So there are $n = 84$ observations. May 1996 corresponds to observation #53 in this series. There are 16 time series corresponding to the variables for hour window (11PM, 12AM, 1AM, 2AM), weekgroup (SunWed, ThuSat) and driver condition (sober, drunk). The codes used and their defintions are given in the table below.

#	Code	Definition
1	PM11SS	11PM SunWed Sober
2	AM12SS	12AM SunWed Sober
3	AM1SS	1AM SunWed Sober
4	AM2SS	2AM SunWed Sober
5	PM11TS	11PM ThuSat Sober
6	AM12TS	12AM ThuSat Sober
7	AM1TS	1AM ThuSat Sober
8	AM2TS	2AM ThuSat Sober
9	PM11SD	11PM SunWed Drunk
10	AM12SD	12AM SunWed Drunk
11	AM1SD	1AM SunWed Drunk
12	AM2SD	2AM SunWed Drunk
13	PM11TD	11PM ThuSat Drunk
14	AM12TD	12AM ThuSat Drunk
15	AM1TD	1AM ThuSat Drunk
16	AM2TD	2AM ThuSat Drunk

May 1996 corresponds to observation #53 in this series. A step intervention model defined by,

$$z_t = \mu + \delta_1 \xi_t + N_t$$

where N_t is the error term. Based on the pre-intervention data we assume initially that N_t is normal and independent, so ordinary multiple linear regression can be used. The intervention series are defined by,

$$\xi_t = \begin{cases} 0 & t < 53 \\ 1 & t \geq 53 \end{cases}$$

The pre-intervention series is fairly short so the following two-step approach to the ARIMA identification of the error term N_t will be used. In the first stage the model in eqn. (1) is fit using standard regression. Then the residuals, the estimated values, $\hat{N}_t$, are obtained and the residual autocorrelation of these residuals is examined. This approach should be expected to work well and is indeed theoretically superior to the alternative approach of basing the model identification on the pre-intervention residuals. The reason for this is that in the estimates of the autocorrelations of $\hat{N}_t$ are first-order efficient as are the estimates in the pre-intervention approach. However since the sample size is larger, this approach provides better estimates.

■ Regression Results

In the residual autocorrelation, a small amount of positive autocorrelation is detected. Hence these initial regression estimates of the intervention are statistically efficient but due to the small amount of positive autocorrelation, their significance is a little overstated. An ARIMA intervention will be below and it will be shown that there are only minor changes in the results below.

		Estimate	SE	TStat	PValue
PM11SS	1	1.38462	0.153925	8.99539	7.34968×10^{-14}
	ξ	**-0.478365**	**0.249387**	**-1.91817**	**0.0585706**

		Estimate	SE	TStat	PValue
AM12SS	1	1.88462	0.188355	10.0057	8.88178×10^{-16}
	ξ	-0.353365	0.30517	-1.15793	0.250254

		Estimate	SE	TStat	PValue
AM1SS	1	1.28846	0.138821	9.28144	1.9984×10^{-14}
	ξ	**-0.413462**	**0.224916**	**-1.83829**	**0.0696404**

		Estimate	SE	TStat	PValue
AM2SS	1	1.92308	0.179189	10.7321	0.
	ξ	**-0.579327**	**0.290319**	**-1.99548**	**0.0493104**

		Estimate	SE	TStat	PValue
PM11TS	1	2.59615	0.220541	11.7717	0.
	ξ	**-1.40865**	**0.357318**	**-3.9423**	**0.000169088**

		Estimate	SE	TStat	PValue
AM12TS	1	3.42308	0.300949	11.3743	0.
	ξ	**-1.79808**	**0.487593**	**-3.68766**	**0.000405503**

		Estimate	SE	TStat	PValue
AM1TS	1	1.61538	0.167885	9.62196	4.21885×10^{-15}
	ξ	-0.0528846	0.272005	-0.194425	0.846324

		Estimate	SE	TStat	PValue
AM2TS	1	2.40385	0.226379	10.6187	0.
	ξ	0.0961538	0.366776	0.26216	0.793856

		Estimate	SE	TStat	PValue
PM11SD	1	1.38462	0.153925	8.99539	7.34968×10^{-14}
	ξ	**-0.478365**	**0.249387**	**-1.91817**	**0.0585706**

		Estimate	SE	TStat	PValue
AM12SD	1	1.88462	0.188355	10.0057	8.88178×10^{-16}
	ξ	-0.353365	0.30517	-1.15793	0.250254

		Estimate	SE	TStat	PValue
AM1SD	1	1.28846	0.138821	9.28144	1.9984×10^{-14}
	ξ	**-0.413462**	**0.224916**	**-1.83829**	**0.0696404**

		Estimate	SE	TStat	PValue
AM2SD	1	1.92308	0.179189	10.7321	0.
	ξ	**-0.579327**	**0.290319**	**-1.99548**	**0.0493104**

		Estimate	SE	TStat	PValue
PM11TD	1	2.59615	0.220541	11.7717	0.
	ξ	**-1.40865**	**0.357318**	**-3.9423**	**0.000169088**

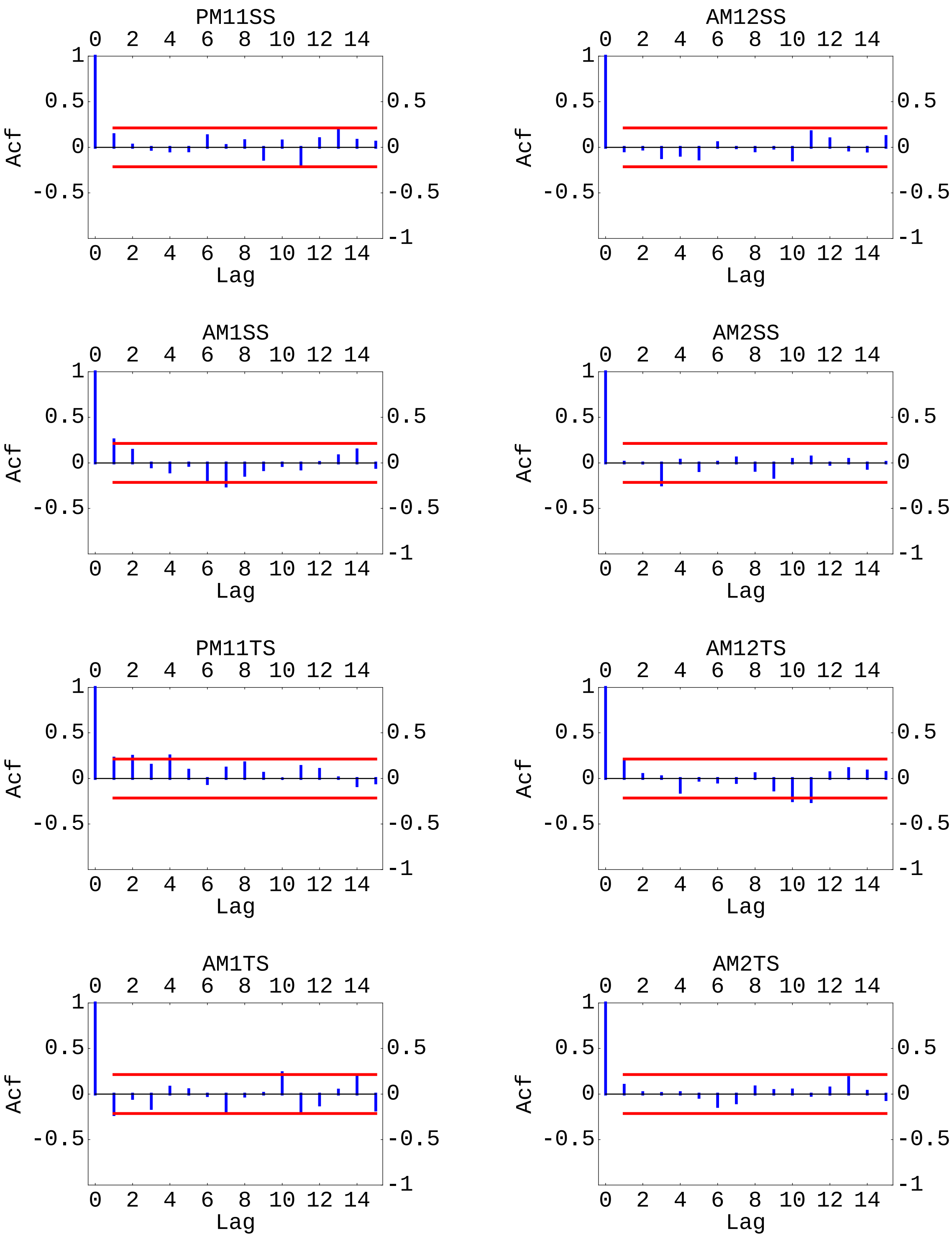
		Estimate	SE	TStat	PValue
AM12TD	1	3.42308	0.300949	11.3743	0.
	ξ	**-1.79808**	**0.487593**	**-3.68766**	**0.000405503**

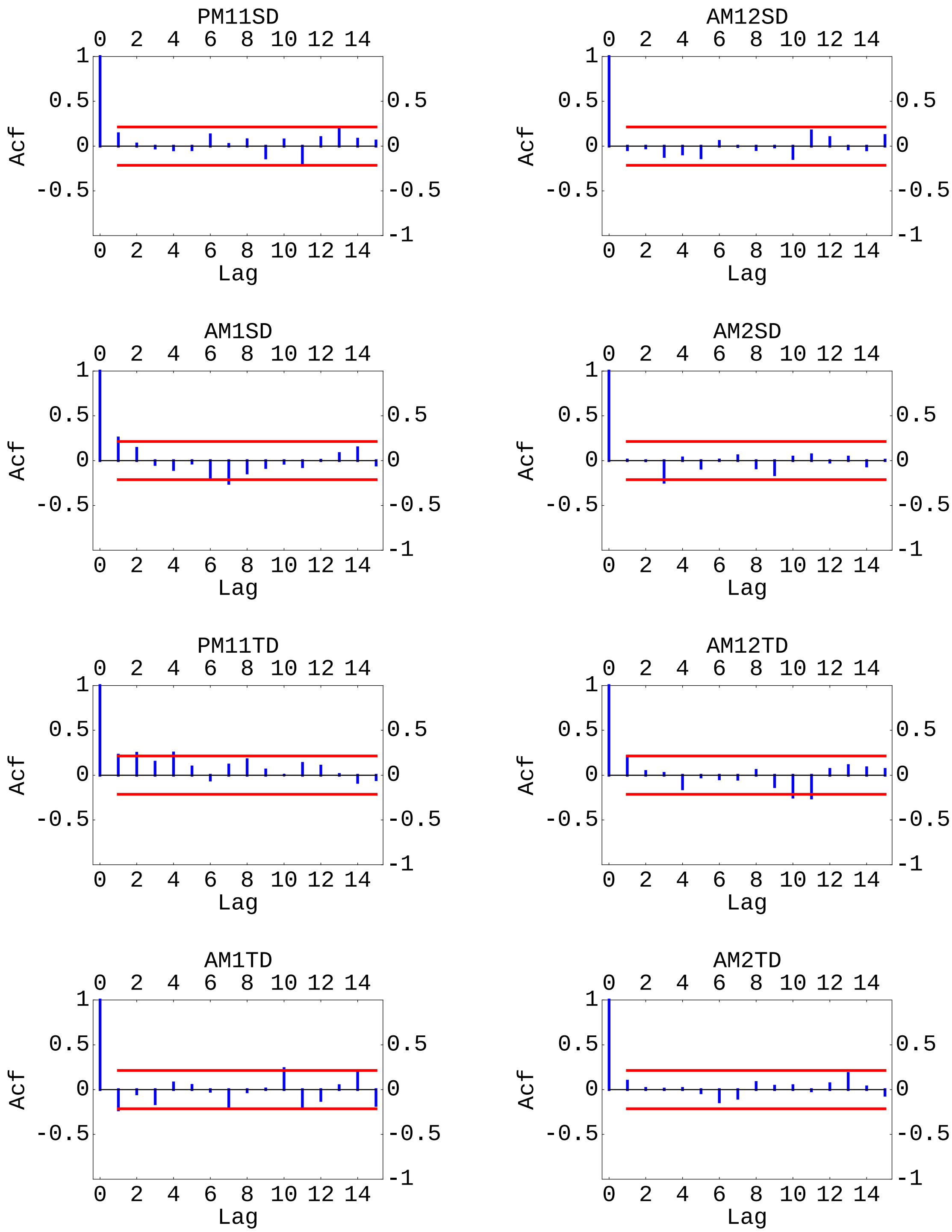
		Estimate	SE	TStat	PValue
AM1TD	1	1.61538	0.167885	9.62196	4.21885×10^{-15}
	ξ	-0.0528846	0.272005	-0.194425	0.846324

		Estimate	SE	TStat	PValue
AM2TD	1	2.40385	0.226379	10.6187	0.
	ξ	0.0961538	0.366776	0.26216	0.793856

Residual Autocorrelation Analysis

The residual autocorrelation from the regression model are examined and used to select tentative ARIMA models.


PM11SS
AM12SS
AM1SS
AM2SS
PM11TS
AM12TS
AM1TS
AM2TS
Acf
Lag
0 2 4 6 8 10 12 14
1
0.5
0
-0.5
-1



ARIMA Modelling

Based on the residual autocorrelations of the regression, the following ARIMA models are suggested.

■ Model Identification

#	Code	ARIMA(p,d,q)
1	PM11SS	$(0,0,1)(0,0,0)_{12}$
2	AM12SS	$(0,0,0)(0,0,0)_{12}$
3	AM1SS	$(1,0,0)(0,0,0)_{12}$
4	AM2SS	$(0,0,0)(0,0,0)_{12}$
5	PM11TS	$(0,0,2)(0,0,0)_{12}$
6	AM12TS	$(0,0,1)(0,0,0)_{12}$
7	AM1TS	$(0,0,1)(0,0,0)_{12}$
8	AM2TS	$(0,0,0)(0,0,0)_{12}$
9	PM11SD	$(0,0,0)(0,0,0)_{12}$
10	AM12SD	$(0,0,0)(0,0,0)_{12}$
11	AM1SD	$(1,0,0)(0,0,0)_{12}$
12	AM2SD	$(0,0,0)(0,0,0)_{12}$
13	PM11TD	$(0,0,2)(0,0,0)_{12}$
14	AM12TD	$(0,0,1)(0,0,0)_{12}$
15	AM1TD	$(0,0,1)(0,0,0)_{12}$
16	AM2TD	$(0,0,0)(0,0,0)_{12}$

■ 1. PM11SS (0, 0, 1)

	Estimate	SE	P-Value
μ	1.384	0.134	0
δ	-0.481	0.275	0.08
θ_1	-0.133	0.108	0.217

Since the θ_1 estimate is not significantly different from zero, we will go the the previous model for this data.

■ 3. AM1SS (0, 0, 1) $(0, 0, 0)_{12}$

	Estimate	SE	P-Value
μ	1.277	0.14	0
δ	-0.395	0.284	0.164
ϕ_1	0.255	0.106	0.016

Only minor change even though the AR parameter is significant. We keep this model.

■ 5. PM11TS (0, 0, 2) $(0, 0, 0)_{12}$

	Estimate	SE	P-Value
μ	2.586	0.223	0
δ	-1.426	0.451	0.002
θ_1	-0.189	0.108	0.08
θ_2	-0.162	0.108	0.132

The MA parameters are not significant at 5% so we keep the previous model.

■ 6. AM12TS (0, 0, 1) (0, 0, 0)$_{12}$

	Estimate	SE	P-Value
μ	3.803	0.279	0
δ	-1.828	0.57	0.001
θ_1	-0.204	0.107	0.056

The autocorrelation is small but significant, we keep this model.

■ 7. AM1TS (0, 0, 1) (0, 0, 0)$_{12}$

	Estimate	SE	P-Value
μ	1.816	0.089	0
δ	-0.078	0.186	0.675
θ_1	0.315	0.104	0.002

Very minor impact of the size or significance of the intervention coefficient but we keep this model.

■ 11. AM1SD (1, 0, 0) (0, 0, 0)$_{12}$

	Estimate	SE	P-Value
μ	1.277	0.14	0
δ	-0.395	0.284	0.164
ϕ_1	0.255	0.106	0.016

Only minor change in the intervention coefficient. AR parameter signficant so we keep this model.

■ 13. PM11TD (0, 0, 2) (0, 0, 0)$_{12}$

	Estimate	SE	P-Value
μ	2.586	0.223	0
δ	-1.426	0.451	0.002
Θ_1	-0.189	0.108	0.08
Θ_2	-0.162	0.108	0.132

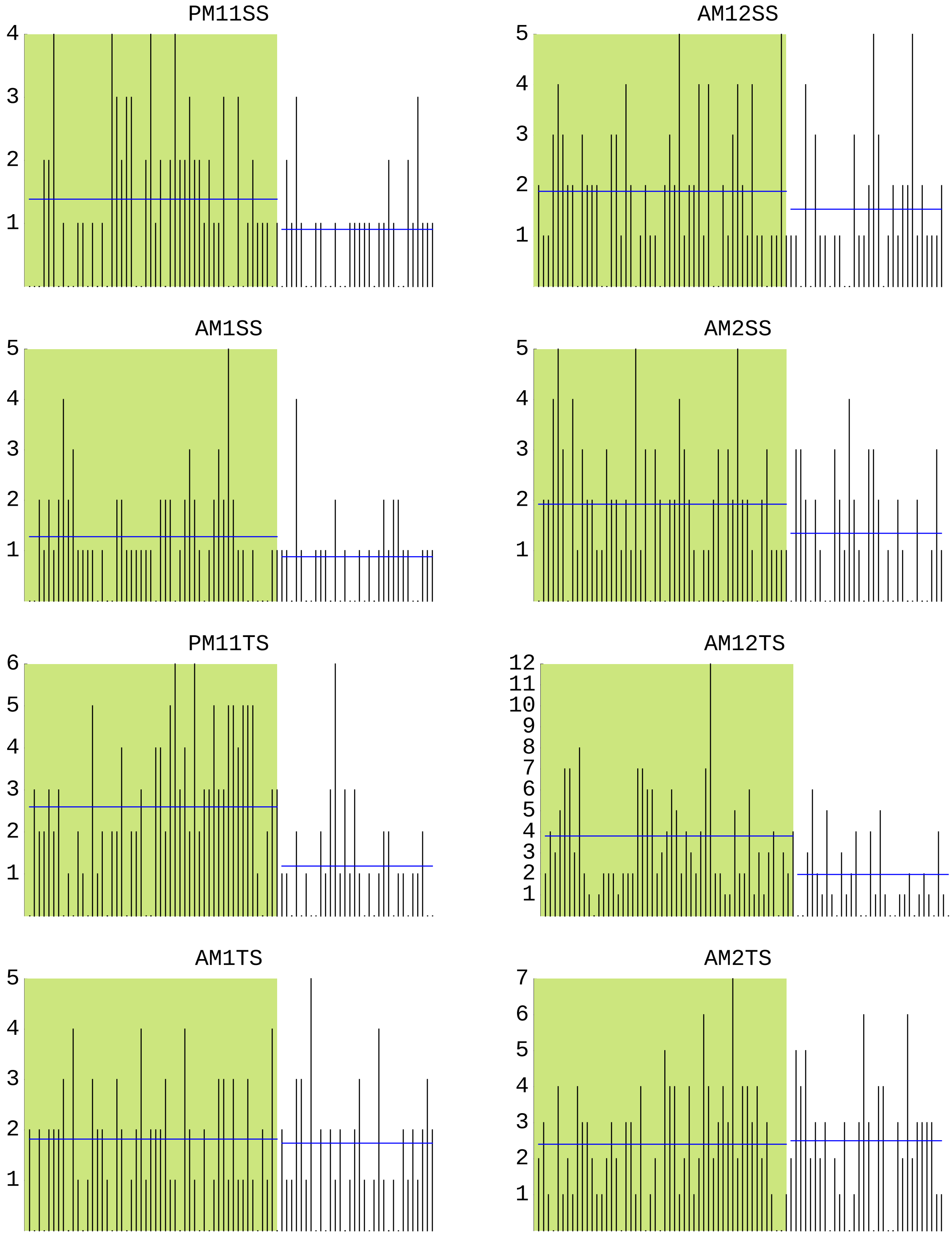
The ARIMA parameter is not significant, so the original model will be used.

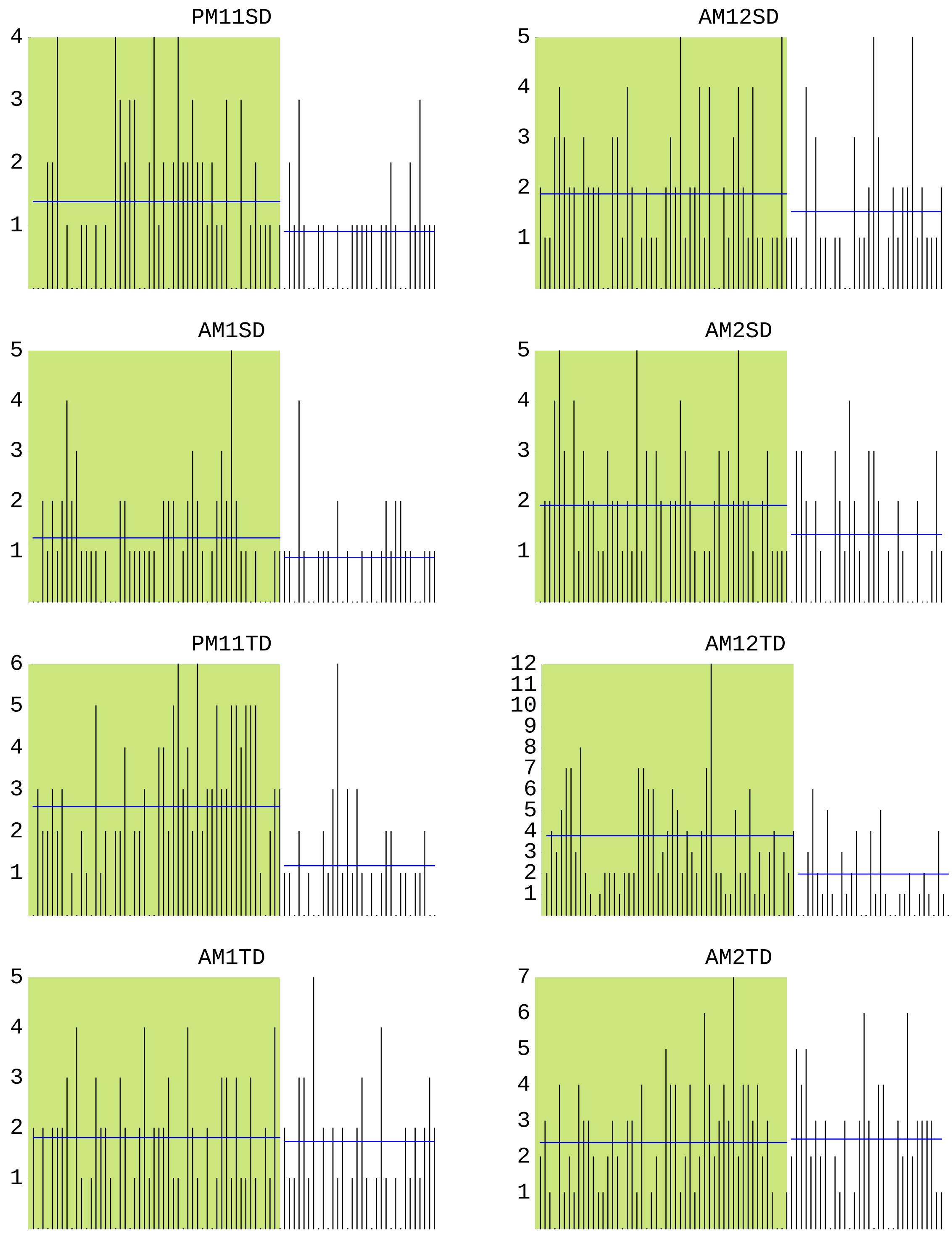
Final Intervention Parameter Estimates

The table below gives the estimates of the overall mean μ and the step-intevention parameter δ for the final selected model for each series. For each series μ is the upper entry and δ is the lower entry.

	Estimate	SD
PM11SS	1.38462 −0.478365	0.153925 0.249387
AM12SS	1.88462 −0.353365	0.188355 0.30517
AM1SS	1.27658 −0.394649	0.140046 0.283723
AM2SS	1.92308 −0.579327	0.179189 0.290319
PM11TS	2.59615 −1.40865	0.220541 0.357318
AM12TS	3.80263 −1.82818	0.278819 0.569666
AM1TS	1.81581 −0.0779229	0.0888371 0.185618
AM2TS	2.40385 0.0961538	0.226379 0.366776
PM11SD	1.38462 −0.478365	0.153925 0.249387
AM12SD	1.88462 −0.353365	0.188355 0.30517
AM1SD	1.27658 −0.394649	0.140046 0.283723
AM2SD	1.92308 −0.579327	0.179189 0.290319
PM11TD	2.59615 −1.40865	0.220541 0.357318
AM12TD	3.80263 −1.82818	0.278819 0.569666
AM1TD	1.81581 −0.0779229	0.0888371 0.185618
AM2TD	2.40385 0.0961538	0.226379 0.366776

Intervention Visualization




PM11SD
AM12SD
AM1SD
AM2SD
PM11TD
AM12TD
AM1TD
AM2TD

MTO Minor, DRCOND

The time series comprises all accidents which resulted in minor injury. The data were aggregated to monthly time series from January 1992 to December 1998. So there are $n = 84$ observations. May 1996 corresponds to observation #53 in this series. There are 16 time series corresponding to the variables for hour window (11PM, 12AM, 1AM, 2AM), weekgroup (SunWed, ThuSat) and driver condition (sober, drunk). The codes used and their defintions are given in the table below.

#	Code	Definition
1	PM11SS	11PM SunWed Sober
2	AM12SS	12AM SunWed Sober
3	AM1SS	1AM SunWed Sober
4	AM2SS	2AM SunWed Sober
5	PM11TS	11PM ThuSat Sober
6	AM12TS	12AM ThuSat Sober
7	AM1TS	1AM ThuSat Sober
8	AM2TS	2AM ThuSat Sober
9	PM11SD	11PM SunWed Drunk
10	AM12SD	12AM SunWed Drunk
11	AM1SD	1AM SunWed Drunk
12	AM2SD	2AM SunWed Drunk
13	PM11TD	11PM ThuSat Drunk
14	AM12TD	12AM ThuSat Drunk
15	AM1TD	1AM ThuSat Drunk
16	AM2TD	2AM ThuSat Drunk

May 1996 corresponds to observation #53 in this series. A step intervention model defined by,

$$z_t = \mu + \delta_1 \xi_t + N_t$$

where N_t is the error term. Based on the pre-intervention data we assume initially that N_t is normal and independent, so ordinary multiple linear regression can be used. The intervention series are defined by,

$$\xi_t = \begin{cases} 0 & t < 53 \\ 1 & t \geq 53 \end{cases}$$

The pre-intervention series is fairly short so the following two-step approach to the ARIMA identification of the error term N_t will be used. In the first stage the model in eqn. (1) is fit using standard regression. Then the residuals, the estimated values, $\hat{N}_t$, are obtained and the residual autocorrelation of these residuals is examined. This approach should be expected to work well and is indeed theoretically superior to the alternative approach of basing the model identification on the pre-intervention residuals. The reason for this is that in the estimates of the autocorrelations of $\hat{N}_t$ are first-order efficient as are the estimates in the pre-intervention approach. However since the sample size is larger, this approach provides better estimates.

■ Regression Results

In the residual autocorrelation, a small amount of positive autocorrelation is detected. Hence these initial regression estimates of the intervention are statistically efficient but due to the small amount of positive autocorrelation, their significance is a little overstated. An ARIMA intervention will be below and it will be shown that there are only minor changes in the results below.

		Estimate	SE	TStat	PValue
PM11SS	1	3.71154	0.274609	13.5157	0.
	ξ	-0.524038	0.444917	-1.17783	0.24227

		Estimate	SE	TStat	PValue
AM12SS	1	5.57692	0.345906	16.1226	0.
	ξ	**-1.45192**	**0.560432**	**-2.59072**	**0.0113332**

		Estimate	SE	TStat	PValue
AM1SS	1	3.5	0.305851	11.4435	0.
	ξ	-0.5625	0.495535	-1.13514	0.259625

		Estimate	SE	TStat	PValue
AM2SS	1	5.73077	0.354199	16.1795	0.
	ξ	**-1.82452**	**0.573869**	**-3.17933**	**0.00208331**

		Estimate	SE	TStat	PValue
PM11TS	1	6.59615	0.370879	17.7852	0.
	ξ	**-3.34615**	**0.600893**	**-5.56863**	**3.16511×10^{-7}**

		Estimate	SE	TStat	PValue
AM12TS	1	9.75	0.49226	19.8066	0.
	ξ	**-4.875**	**0.797552**	**-6.11246**	**3.16273×10^{-8}**

		Estimate	SE	TStat	PValue
AM1TS	1	3.98077	0.274993	14.4759	0.
	ξ	0.237981	0.44554	0.53414	0.59469

		Estimate	SE	TStat	PValue
AM2TS	1	6.98077	0.392441	17.7881	0.
	ξ	-0.574519	0.635828	-0.903577	0.368867

		Estimate	SE	TStat	PValue
PM11SD	1	3.71154	0.274609	13.5157	0.
	ξ	−0.524038	0.444917	−1.17783	0.24227

		Estimate	SE	TStat	PValue
AM12SD	1	5.57692	0.345906	16.1226	0.
	ξ	**-1.45192**	**0.560432**	**-2.59072**	**0.0113332**

		Estimate	SE	TStat	PValue
AM1SD	1	3.5	0.305851	11.4435	0.
	ξ	−0.5625	0.495535	−1.13514	0.259625

		Estimate	SE	TStat	PValue
AM2SD	1	5.73077	0.354199	16.1795	0.
	ξ	**-1.82452**	**0.573869**	**-3.17933**	**0.00208331**

		Estimate	SE	TStat	PValue
PM11TD	1	6.59615	0.370879	17.7852	0.
	ξ	**-3.34615**	**0.600893**	**-5.56863**	**3.16511×10^{-7}**

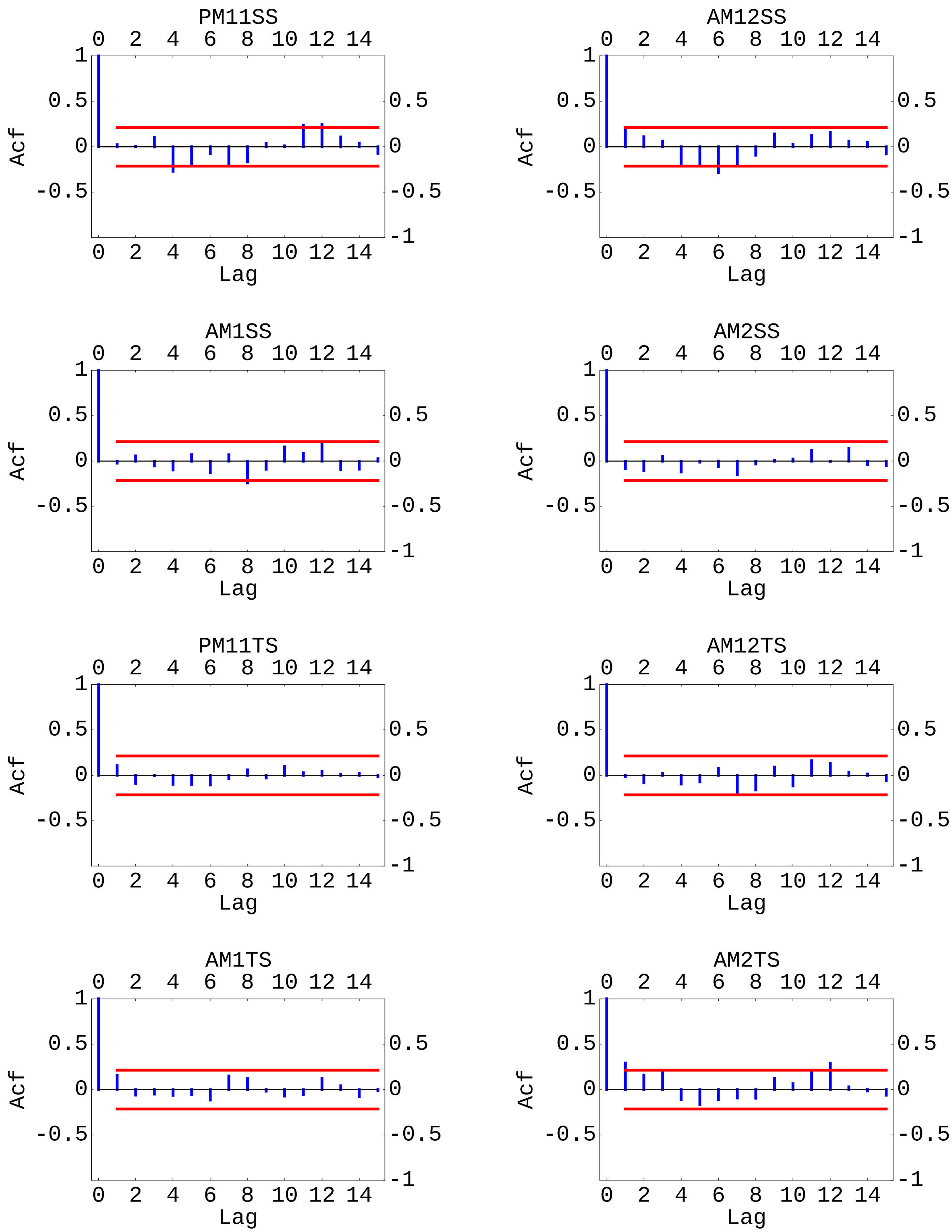
		Estimate	SE	TStat	PValue
AM12TD	1	9.75	0.49226	19.8066	0.
	ξ	**-4.875**	**0.797552**	**-6.11246**	**3.16273×10^{-8}**

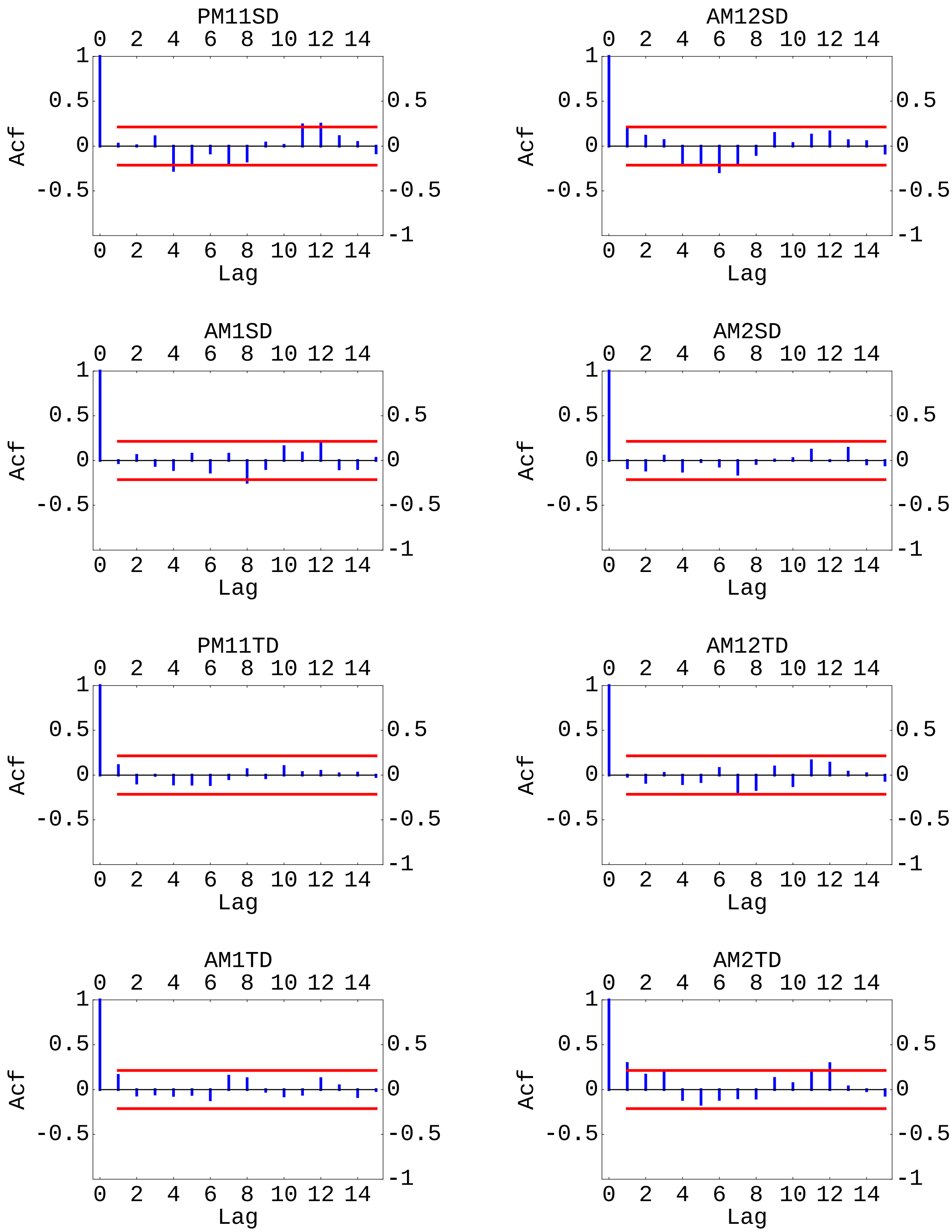
		Estimate	SE	TStat	PValue
AM1TD	1	3.98077	0.274993	14.4759	0.
	ξ	0.237981	0.44554	0.53414	0.59469

		Estimate	SE	TStat	PValue
AM2TD	1	6.98077	0.392441	17.7881	0.
	ξ	−0.574519	0.635828	−0.903577	0.368867

Residual Autocorrelation Analysis

The residual autocorrelation from the regression model are examined and used to select tentative ARIMA models.


PM11SS
AM12SS
AM1SS
AM2SS
PM11TS
AM12TS
AM1TS
AM2TS
Acf
Lag
0 2 4 6 8 10 12 14
1
0.5
0
-0.5
-1



ARIMA Modelling

Based on the residual autocorrelations of the regression, the following ARIMA models are suggested.

■ Model Identification

#	Code	ARIMA(p,d,q)
1	PM11SS	$(0,0,0)(0,0,1)_{12}$
2	AM12SS	$(0,0,0)(0,0,1)_{12}$
3	AM1SS	$(0,0,0)(0,0,1)_{12}$
4	AM2SS	$(0,0,0)(0,0,1)_{12}$
5	PM11TS	$(0,0,0)(0,0,0)_{12}$
6	AM12TS	$(0,0,0)(0,0,0)_{12}$
7	AM1TS	$(0,0,1)(0,0,1)_{12}$
8	AM2TS	$(0,0,3)(0,0,1)_{12}$
9	PM11SD	$(0,0,0)(0,0,1)_{12}$
10	AM12SD	$(1,0,0)(0,0,1)_{12}$
11	AM1SD	$(0,0,0)(0,0,1)_{12}$
12	AM2SD	$(0,0,0)(0,0,0)_{12}$
13	PM11TD	$(0,0,0)(0,0,1)_{12}$
14	AM12TD	$(0,0,0)(0,0,1)_{12}$
15	AM1TD	$(0,0,1)(0,0,1)_{12}$
16	AM2TD	$(0,0,3)(0,0,1)_{12}$

■ 1. PM11SS $(0, 0, 0)(0, 0, 1)_{12}$

	Estimate	SE	P-Value
μ	3.728	0.261	0
δ	-0.592	0.479	0.217
Θ_1	-0.269	0.105	0.01

There is almost no change is the regression parameter and it is still not significant at 10% on a one-sided test.

■ 2. AM12SS $(0, 0, 0)(0, 0, 1)_{12}$

	Estimate	SE	P-Value
μ	5.543	0.309	0
δ	**-1.377**	**0.592**	**0.02**
Θ_1	-0.167	0.108	0.121

Due to the small positive autocorrelation at the seasonal lag, the significance level has increased from 1% to 2%. But since the seasonal coefficient is not significant, the previous model will be used.

■ 3. AM1SS $(0, 0, 0)(0, 0, 1)_{12}$

	Estimate	SE	P-Value
μ	3.534	0.285	0
δ	-0.592	0.532	0.266
Θ_1	-0.232	0.106	0.029

Only minor change from coefficient of -0.592 from previous estimate of -0.5625. Still not significant at 10%.

■ 4. AM2SS (0, 0, 0) (0, 0, 1)$_{12}$

	Estimate	SE	P-Value
μ	5.731	0.274	0
δ	-1.825	0.566	0.001
Θ_1	0.004	0.109	0.969

The seasonal autocorrelation parameter is not significant, so the original model will be used.

■ 5. AM1TS (0, 0, 1) (0, 0, 1)$_{12}$

	Estimate	SE	P-Value
μ	3.993	0.295	0
δ	0.141	0.554	0.799
θ_1	-0.186	0.107	0.083
Θ_1	-0.2	0.107	0.061

Due to the positive autocorrelation the significance level has increased to 80% from 60% and the intervention coefficient has been reduced to 0.141 from 0.238. The ARIMA parameters are close to significance at 5% so we keep this model.

■ 8. AM2TS (0, 0, 3) (0, 0, 1)$_{12}$

	Estimate	SE	P-Value
μ	3.993	0.295	0
δ	0.141	0.554	0.799
θ_1	-0.186	0.107	0.083
Θ_1	-0.2	0.107	0.061

Due to the positive autocorrelation the significance level has increased to 64% from 37% and the intervention coefficient has been slightly reduced.

■ 9. PM11SD (0, 0, 0) (0, 0, 1)$_{12}$

	Estimate	SE	P-Value
μ	3.851	0.261	0
δ	-0.592	0.48	0.217
Θ_1	-0.269	0.105	0.01

Very minor impact of the size or significance of the intervention coefficient.

■ 10. AM12SD (1, 0, 0) (0, 0, 1)$_{12}$

	Estimate	SE	P–Value
μ	5.665	0.367	0
δ	–1.356	0.701	0.053
ϕ_1	0.183	0.107	0.088
Θ_1	–0.151	0.108	0.161

Due to the small positive autocorrelation, the signficance level has increased from 0.01% to 0.053% on a two-sided test. The coefficient in the intervention has decreased in absolute magnitude form -1.45 to -1.35. Since the ARIMA parameters are not significant at the 5% level the original model will be used in our visualization below.

■ 13. PM11TD (0, 0, 0) (0, 0, 1)$_{12}$

	Estimate	SE	P–Value
μ	7.059	0.303	0
δ	–3.344	0.614	0
Θ_1	–0.037	0.109	0.733

The ARIMA parameter is not significant, so the original model will be used.

■ 14. AM12TD (0, 0, 0) (0, 0, 1)$_{12}$

	Estimate	SE	P–Value
μ	10.397	0.43	0
δ	–4.908	0.841	0
Θ_1	–0.119	0.108	0.27

The ARIMA parameter is not significant, so the original model will be used.

■ 15. AM1TD (0, 0, 1) (0, 0, 1)$_{12}$

	Estimate	SE	P–Value
μ	4.082	0.295	0
δ	0.141	0.554	0.8
Θ_1	–0.186	0.107	0.083
Θ_1	–0.2	0.107	0.061

The coefficient of the intervention has changed from 0.238 to 0.141 has its significance level is now 80%. The ARIMA coefficients are both close to being significant at 5%. We keep this model.

■ 16. AM2TD (0, 0, 3) $(0, 0, 1)_{12}$

	Estimate	SE	P-Value
μ	6.867	0.587	0
δ	-0.479	1.02	0.639
θ_1	-0.253	0.104	0.015
θ_2	-0.183	0.105	0.083
θ_3	-0.311	0.104	0.003
Θ_1	-0.248	0.106	0.019

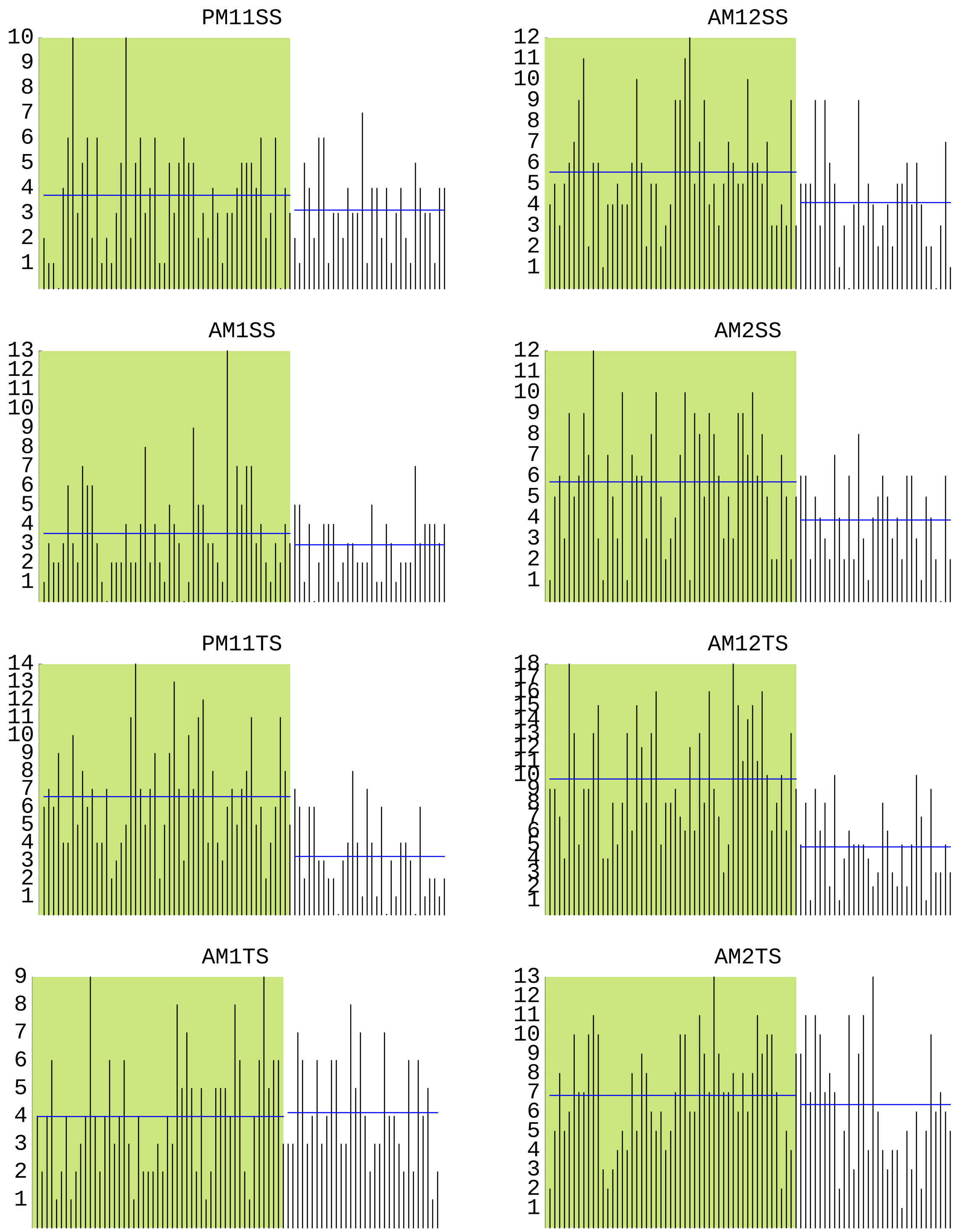
The coefficient of the intervention has changed from -0.575 to -0.479 has its significance level is now 64%. The ARIMA coefficients are jointly significant at 5%. We keep this model.

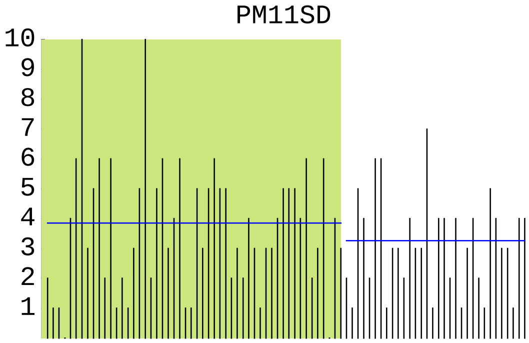
Final Intervention Parameter Estimates

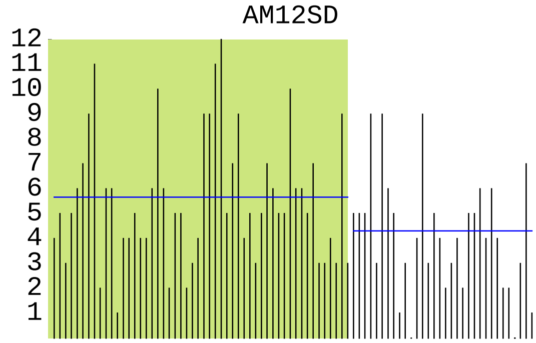
The table below gives the estimates of the overall mean μ and the step-intevention parameter δ for the final selected model for each series. For each series μ is the upper entry and δ is the lower entry. Intervention coefficients which are significant at 1% are indicated by bold font for the row label..

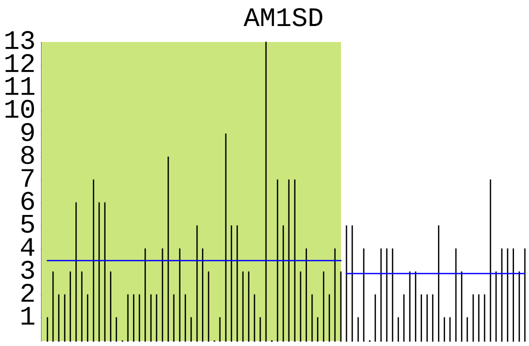
	Estimate	SD
PM11SS	3.72755 -0.591876	0.260549 0.47898
AM12SS	5.57692 -1.45192	0.345906 0.560432
AM1SS	3.53423 -0.591831	0.285165 0.532366
AM2SS	5.73077 -1.82452	0.354199 0.573869
PM11TS	3.99335 0.140668	0.295139 0.553715
AM12TS	9.75 -4.875	0.49226 0.797552
AM1TS	3.98077 0.237981	0.274993 0.44554
AM2TS	6.86714 -0.478591	0.586985 1.01992
PM11SD	3.85133 -0.591876	0.26092 0.479662
AM12SD	5.66476 -1.35557	0.367241 0.700847
AM1SD	3.5 -0.5625	0.305851 0.495535
AM2SD	5.73077 -1.82452	0.354199 0.573869
PM11TD	6.59615 -3.34615	0.370879 0.600893
AM12TD	9.75 -4.875	0.49226 0.797552
AM1TD	4.08239 0.140668	0.295307 0.553906
AM2TD	6.86714 -0.478591	0.586985 1.01992

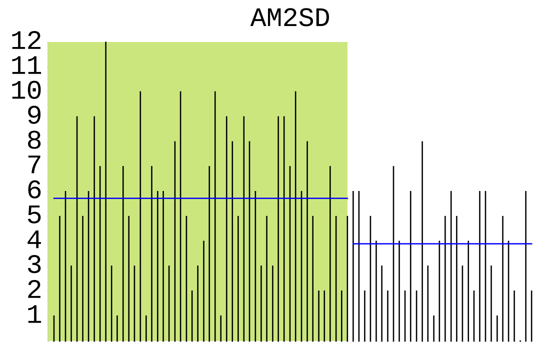
Intervention Visualization

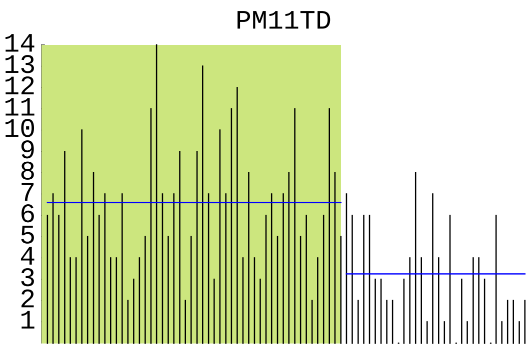


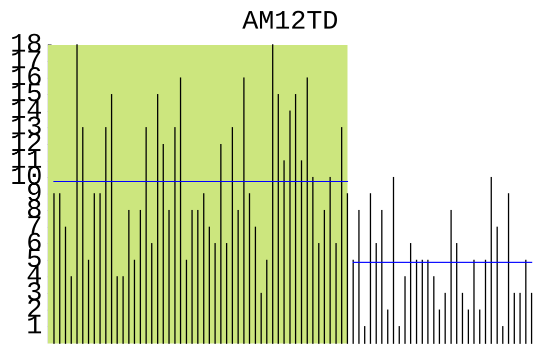

PM11SD

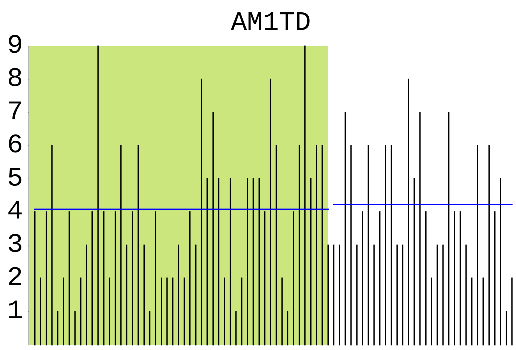

AM12SD

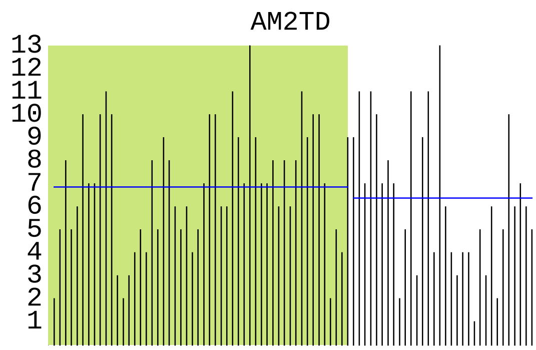

AM1SD


AM2SD


PM11TD


AM12TD


AM1TD


AM2TD

MTO Minimal, DRCOND

The time series consists of all minimal injury accidents per month from January 1992 to December 1998. So there are $n = 84$ observations. May 1996 corresponds to observation #53 in this series. There are 16 time series corresponding to the variables for hour window (11PM, 12AM, 1AM, 2AM), weekgroup (SunWed, ThuSat) and driver condition (sober, drunk). The codes used and their defintions are given in the table below.

Code	Definition
PM11SS	11PM SunWed Sober
AM12SS	12AM SunWed Sober
AM1SS	1AM SunWed Sober
AM2SS	2AM SunWed Sober
PM11TS	11PM ThuSat Sober
AM12TS	12AM ThuSat Sober
AM1TS	1AM ThuSat Sober
AM2TS	2AM ThuSat Sober
PM11SD	11PM SunWed Drunk
AM12SD	12AM SunWed Drunk
AM1SD	1AM SunWed Drunk
AM2SD	2AM SunWed Drunk
PM11TD	11PM ThuSat Drunk
AM12TD	12AM ThuSat Drunk
AM1TD	1AM ThuSat Drunk
AM2TD	2AM ThuSat Drunk

A step intervention model defined by,

$$z_t = \mu + \delta_1 \xi_t + N_t \qquad (1)$$

where N_t is the error term. Based on the pre-intervention data we assume initially that N_t is normal and independent, so ordinary multiple linear regression can be used. The intervention series are defined by,

$$\xi_t = \begin{cases} 0 & t < 53 \\ 1 & t \geq 53 \end{cases}$$

The pre-intervention series is fairly short so the following two-step approach to the ARIMA identification of the error term N_t will be used. In the first stage the model in eqn. (1) is fit using standard regression. Then the residuals, the estimated values, $\hat{N}_t$, are obtained and the residual autocorrelation of these residuals is examined. This approach should be expected to work well and is indeed theoretically superior to the alternative approach of basing the model identification on the pre-intervention residuals. The reason for this is that in the estimates of the autocorrelations of $\hat{N}_t$ are first-order efficient as are the estimates in the pre-intervention approach. However since the sample size is larger, this approach provides better estimates. Residual diagnostic checking did not reveal any significant autocorrelation for these time series.

There is a statistical significant decrease everywhere except at the ThuSat-1AM and ThuSat-2AM windows for both sober and drunk drivers. The results are tabulated and visualized below.

		Estimate	SE	TStat	PValue
PM11SS	1	3.59615	0.241768	14.8744	0.
	ξ	**-1.6899**	**0.391709**	**-4.31419**	**0.0000444032**

		Estimate	SE	TStat	PValue
AM12SS	1	4.53846	0.278886	16.2736	0.
	ξ	**-1.44471**	**0.451846**	**-3.19735**	**0.00197109**

		Estimate	SE	TStat	PValue
AM1SS	1	3.98077	0.29014	13.7202	0.
	ξ	**-1.44952**	**0.470081**	**-3.08355**	**0.00278678**

		Estimate	SE	TStat	PValue
AM2SS	1	4.92308	0.286975	17.1551	0.
	ξ	**-1.70433**	**0.464953**	**-3.66559**	**0.000436712**

		Estimate	SE	TStat	PValue
PM11TS	1	5.30769	0.316432	16.7735	0.
	ξ	**-2.87019**	**0.512679**	**-5.59842**	$\mathbf{2.7958 \times 10^{-7}}$

		Estimate	SE	TStat	PValue
AM12TS	1	7.65385	0.388219	19.7153	0.
	ξ	**-2.96635**	**0.628987**	**-4.71607**	$\mathbf{9.73915 \times 10^{-6}}$

		Estimate	SE	TStat	PValue
AM1TS	1	3.23077	0.251977	12.8217	0.
	ξ	0.394231	0.408249	0.965663	0.337052

		Estimate	SE	TStat	PValue
AM2TS	1	6.25	0.375654	16.6376	0.
	ξ	0.21875	0.608629	0.359414	0.720209

		Estimate	SE	TStat	PValue
PM11SD	1	3.59615	0.241768	14.8744	0.
	ξ	**-1.6899**	**0.391709**	**-4.31419**	**0.0000444032**

		Estimate	SE	TStat	PValue
AM12SD	1	4.53846	0.278886	16.2736	0.
	ξ	**-1.44471**	**0.451846**	**-3.19735**	**0.00197109**

		Estimate	SE	TStat	PValue
AM1SD	1	3.98077	0.29014	13.7202	0.
	ξ	**-1.44952**	**0.470081**	**-3.08355**	**0.00278678**

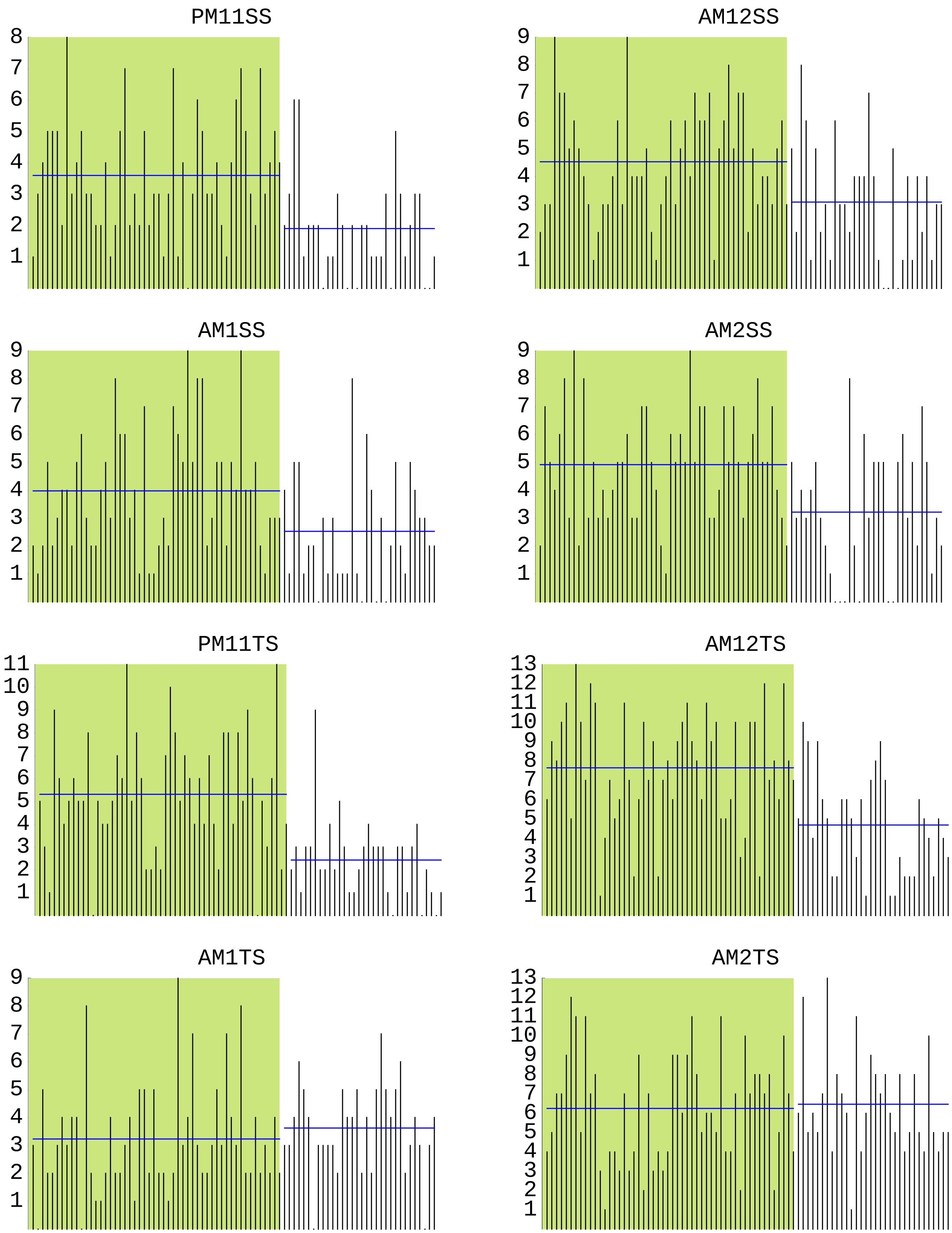
		Estimate	SE	TStat	PValue
AM2SD	1	4.92308	0.286975	17.1551	0.
	ξ	**-1.70433**	**0.464953**	**-3.66559**	**0.000436712**

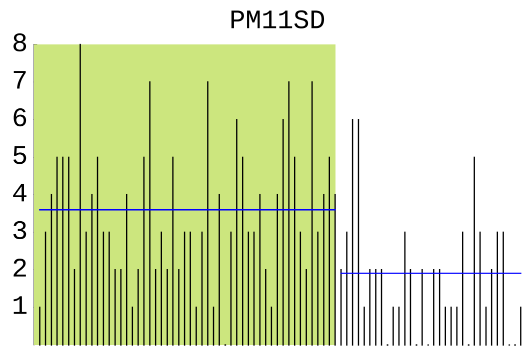
		Estimate	SE	TStat	PValue
PM11TD	1	5.30769	0.316432	16.7735	0.
	ξ	**-2.87019**	**0.512679**	**-5.59842**	**2.7958×10^{-7}**

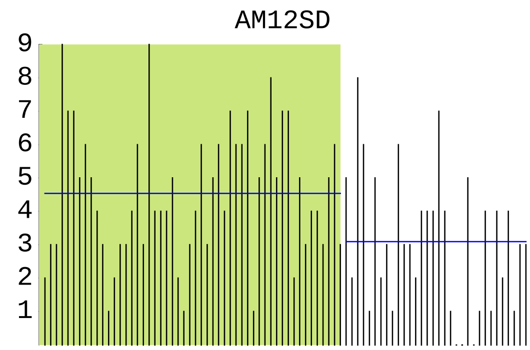
		Estimate	SE	TStat	PValue
AM12TD	1	7.65385	0.388219	19.7153	0.
	ξ	**-2.96635**	**0.628987**	**-4.71607**	**9.73915×10^{-6}**

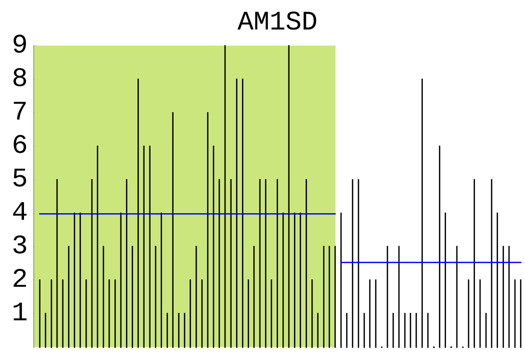
		Estimate	SE	TStat	PValue
AM1TD	1	3.23077	0.251977	12.8217	0.
	ξ	0.394231	0.408249	0.965663	0.337052

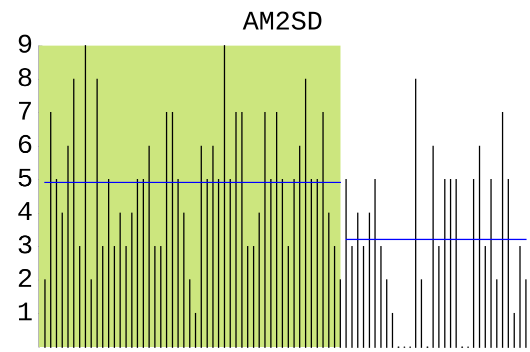
		Estimate	SE	TStat	PValue
AM2TD	1	6.25	0.375654	16.6376	0.
	ξ	0.21875	0.608629	0.359414	0.720209

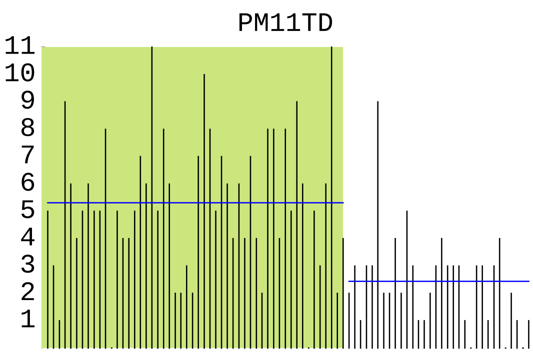

PM11SS
AM12SS
AM1SS
AM2SS
PM11TS
AM12TS
AM1TS
AM2TS

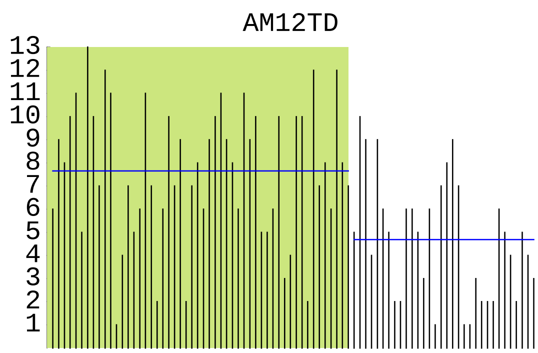

PM11SD

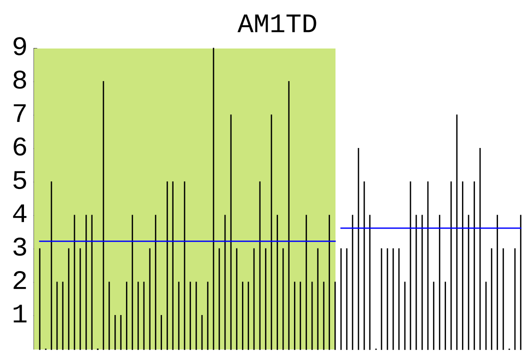

AM12SD

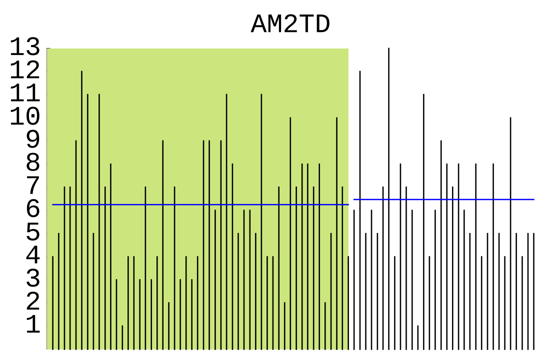

AM1SD


AM2SD


PM11TD


AM12TD


AM1TD


AM2TD

FARS

FARS time series are comprised on the monthly number of deaths in the hourly windows 11PM, 12AM, 1AM, 2AM, 3AM and week groups SunWed, ThuSat. There are ten series in all and their code names are given in the table below.

#	Code	Definition
1	PM11S	11PM SunWed
2	AM12S	12AM SunWed
3	AM1S	1AM SunWed
4	AM2S	2AM SunWed
5	AM3S	3AM ThuSat
6	PM11T	11PM ThuSat
7	AM12T	12AM ThuSat
8	AM1T	1AM ThuSat
9	AM2T	2AM ThuSat
10	AM3T	3AM ThuSat

The fars database includes a blood alcohol variable but in 38% of the records this was missing or not available. These records comprised 45% of the deaths. Until more is known about why so large a portion of the data is incomplete, it was decided that for the purpose of modelling we will use the entire fars dataset, ie. we will ignore the BAC variable.

The purpose of the intervention analysis is to compare our findings with those of the TIRF and MTO datasets.

May 1996 corresponds to observation #49 in this series. A step intervention model defined by,

$$z_t = \mu + \delta_1 \xi_t + N_t \tag{1}$$

where N_t is the error term. Based on the pre-intervention data we assume initially that N_t is normal and independent, so ordinary multiple linear regression can be used. The intervention series are defined by,

$$\xi_t = \begin{cases} 0 & t < 49 \\ 1 & t \geq 49 \end{cases}$$

The pre-intervention series is fairly short so the following two-step approach to the ARIMA identification of the error term N_t will be used. In the first stage the model in eqn. (1) is fit using standard regression. Then the residuals, the estimated values, $\hat{N}_t$, are obtained and the residual autocorrelation of these residuals is examined. This approach should be expected to work well and is indeed theoretically superior to the alternative approach of basing the model identification on the pre-intervention residuals. The reason for this is that in the estimates of the autocorrelations of $\hat{N}_t$ are first-order efficient as are the estimates in the pre-intervention approach. However since the sample size is larger, this approach provides better estimates.

Regression Results

There are no statistical significant interventions. None of the lag one autocorrelations is significant at 5%.

		Estimate	SE	TStat	PValue
PM11S	1	1.8125	0.316267	5.73091	1.31234×10^{-7}
	ξ	−0.244318	0.457322	−0.534237	0.594495

		Estimate	SE	TStat	PValue
AM12S	1	1.6875	0.479426	3.51983	0.000679735
	ξ	0.903409	0.693249	1.30315	0.195847

		Estimate	SE	TStat	PValue
AM1S	1	1.91667	0.349493	5.48414	3.77136×10^{-7}
	ξ	0.0833333	0.505365	0.164897	0.869395

		Estimate	SE	TStat	PValue
AM2S	1	1.85417	0.358629	5.17015	1.40167×10^{-6}
	ξ	0.0549242	0.518577	0.105913	0.915887

		Estimate	SE	TStat	PValue
AM3S	1	1.16667	0.272091	4.28777	0.0000452363
	ξ	0.0378788	0.393444	0.096275	0.923516

		Estimate	SE	TStat	PValue
PM11T	1	11.875	0.832699	14.2609	0.
	ξ	−1.23864	1.20408	−1.0287	0.306378

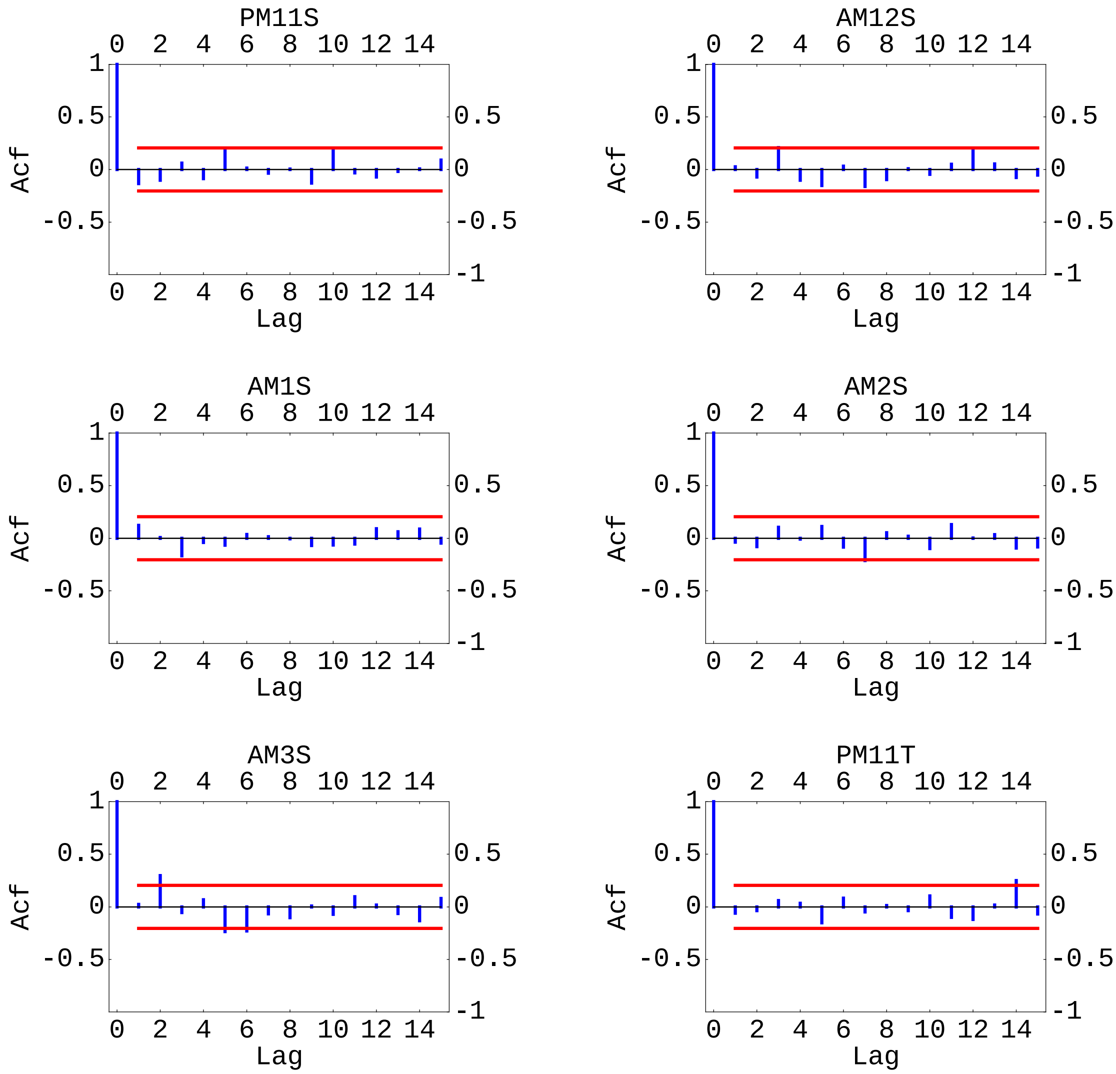
		Estimate	SE	TStat	PValue
AM12T	1	10.3125	0.845908	12.191	0.
	ξ	0.232955	1.22318	0.19045	0.849386

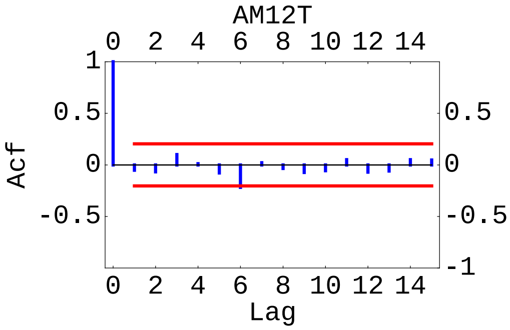
		Estimate	SE	TStat	PValue
AM1T	1	10.5625	0.964922	10.9465	0.
	ξ	−0.448864	1.39528	−0.321703	0.748424

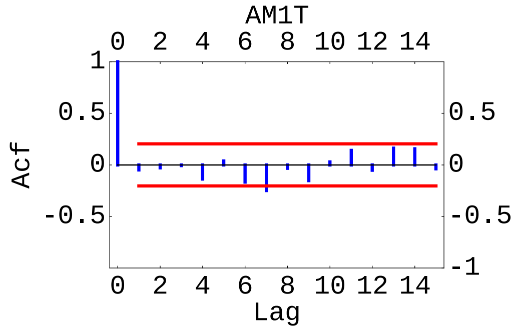
		Estimate	SE	TStat	PValue
AM2T	1	14.125	1.09225	12.932	0.
	ξ	−1.30682	1.57939	−0.82742	0.410188

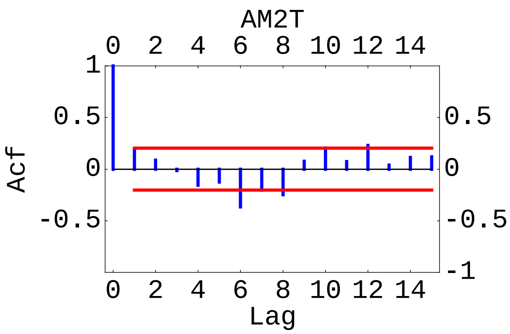
		Estimate	SE	TStat	PValue
AM3T	1	8.89583	0.791846	11.2343	0.
	ξ	−0.645833	1.14501	−0.564043	0.574128

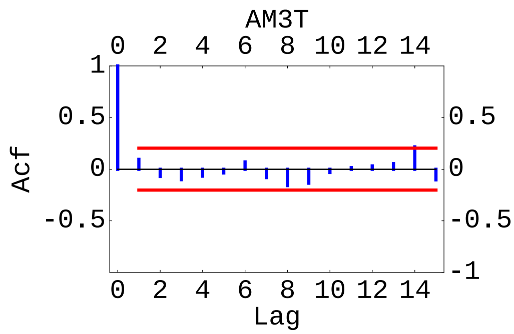
Residual Autocorrelation Analysis



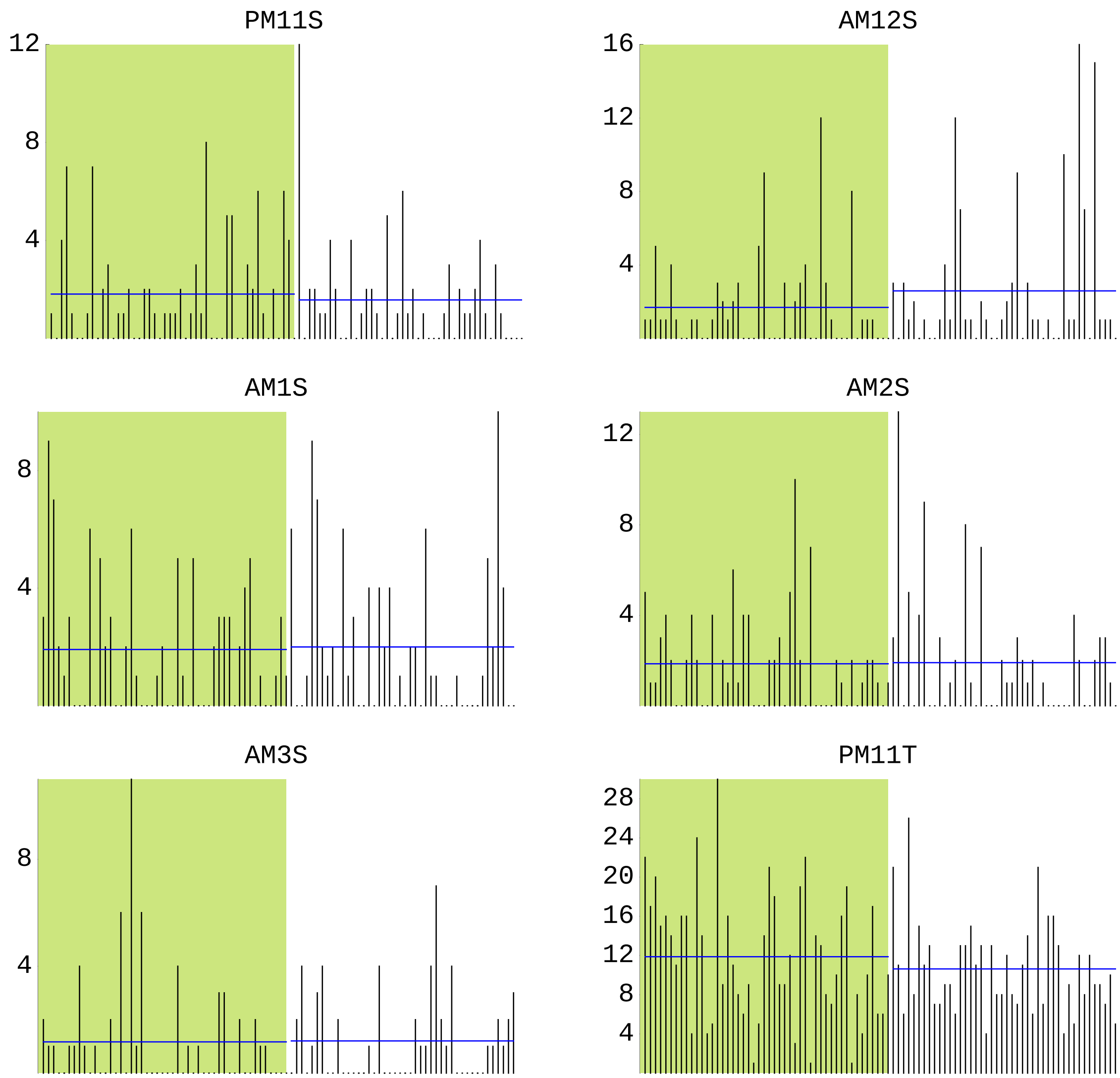

AM12T
0 2 4 6 8 10 12 14
1
0.5
0
-0.5
-1
Acf
Lag

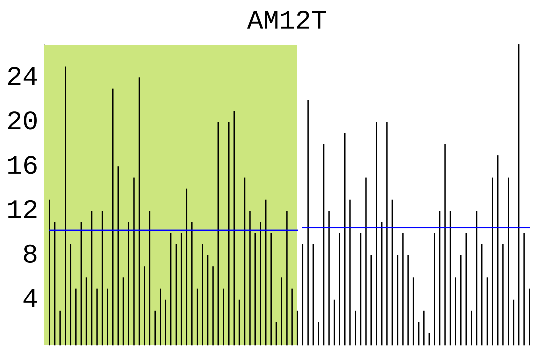

AM1T
0 2 4 6 8 10 12 14
1
0.5
0
-0.5
-1
Acf
Lag

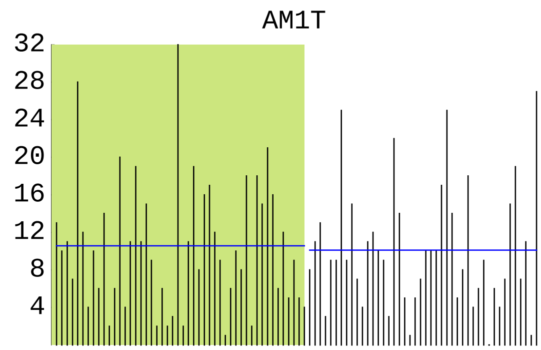

AM2T
0 2 4 6 8 10 12 14
1
0.5
0
-0.5
-1
Acf
Lag

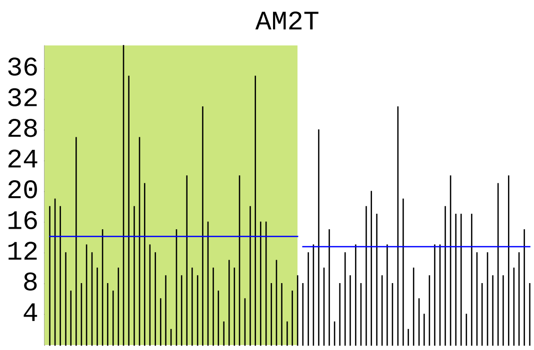

AM3T
0 2 4 6 8 10 12 14
1
0.5
0
-0.5
-1
Acf
Lag

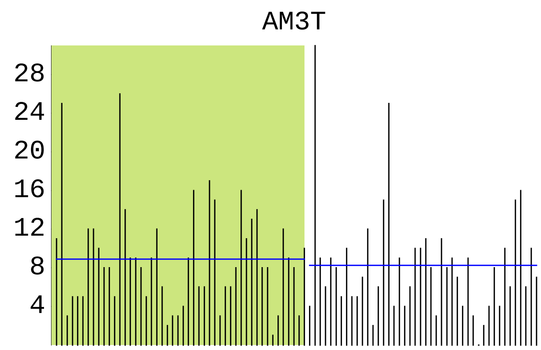
Data Visualization




AM12T
24
20
16
12
8
4


AM1T
32
28
24
20
16
12
8
4


AM2T
36
32
28
24
20
16
12
8
4


AM3T
28
24
20
16
12
8
4

MISTPOL

Michigan State Police data on car accidents for three border counties. The time series used start in May 1992 and run through to April 1999. There are 84 observations in all and the extended bar hours intervention starts May 1996 corresponding to the 49th observation in the series. The specific time series selected for intervention analysis is the number of fatalities per month in which alcohol was involved. This is compared also with the corresponding time series of monthly fatalities not involving drinking.

Intervention Analysis

May 1996 corresponds to observation #49 in this series. A step intervention model defined by,

$$z_t = \mu + \delta \ \xi_t + N_t \tag{1}$$

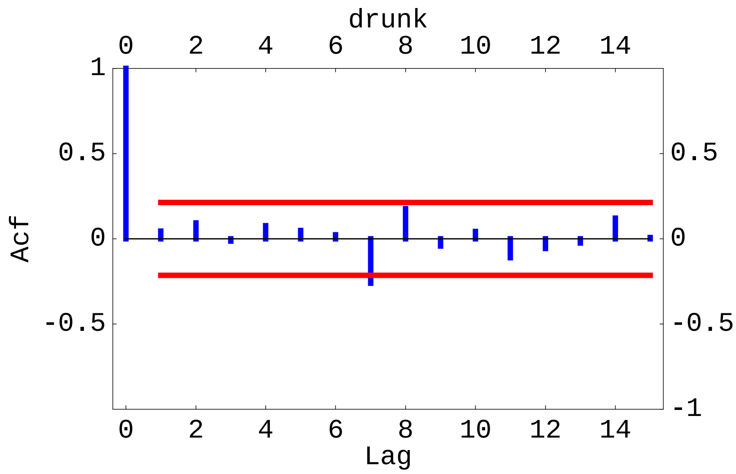
where N_t is the error term. Based on the pre-intervention data we assume initially that N_t is normal and independent, so ordinary multiple linear regression can be used. The intervention series are defined by,

$$\xi_t = \begin{cases} 0 & t < 49 \\ 1 & t \geq 49 \end{cases}$$

The pre-intervention series is fairly short so the following two-step approach to the ARIMA identification of the error term N_t will be used. In the first stage the model in eqn. (1) is fit using standard regression. Then the residuals, the estimated values, $\hat{N}_t$, are obtained and the residual autocorrelation of these residuals is examined. This approach should be expected to work well and is indeed theoretically superior to the alternative approach of basing the model identification on the pre-intervention residuals. The reason for this is that in the estimates of the autocorrelations of $\hat{N}_t$ are first-order efficient as are the estimates in the pre-intervention approach. However since the sample size is larger, this approach provides better estimates.

Residual Autocorrelation of Fitted Regression

Since there are no large autocorrelations evident at lags one and tweleve, the initial model $N_t \sim$ ARIMA(0, 0, 0) is accepted.



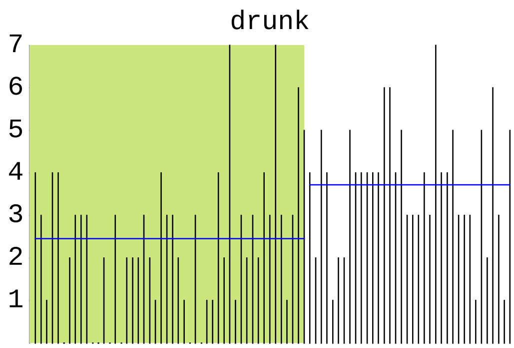
■ Regression Coefficients

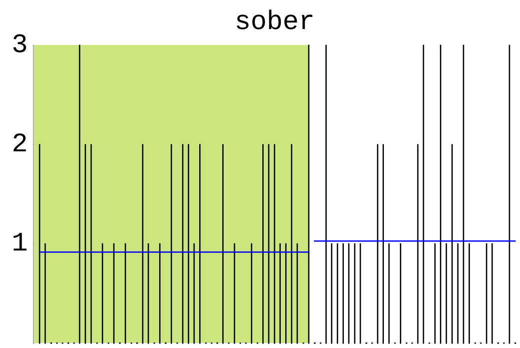
		Estimate	SE	TStat	PValue
drunk	1	2.45833	0.232704	10.5642	0.
	ξ	**1.26389**	**0.355461**	**3.55563**	**0.000629329**

		Estimate	SE	TStat	PValue
sober	1	0.916667	0.141348	6.48515	6.26466×10^{-9}
	ξ	0.111111	0.215913	0.51461	0.60821

■ Visualization

There has clearly been an increase in deaths involving drinking since May 1996 and there is no corresponding change in the death rate for sober drivers.


drunk
7
6
5
4
3
2
1

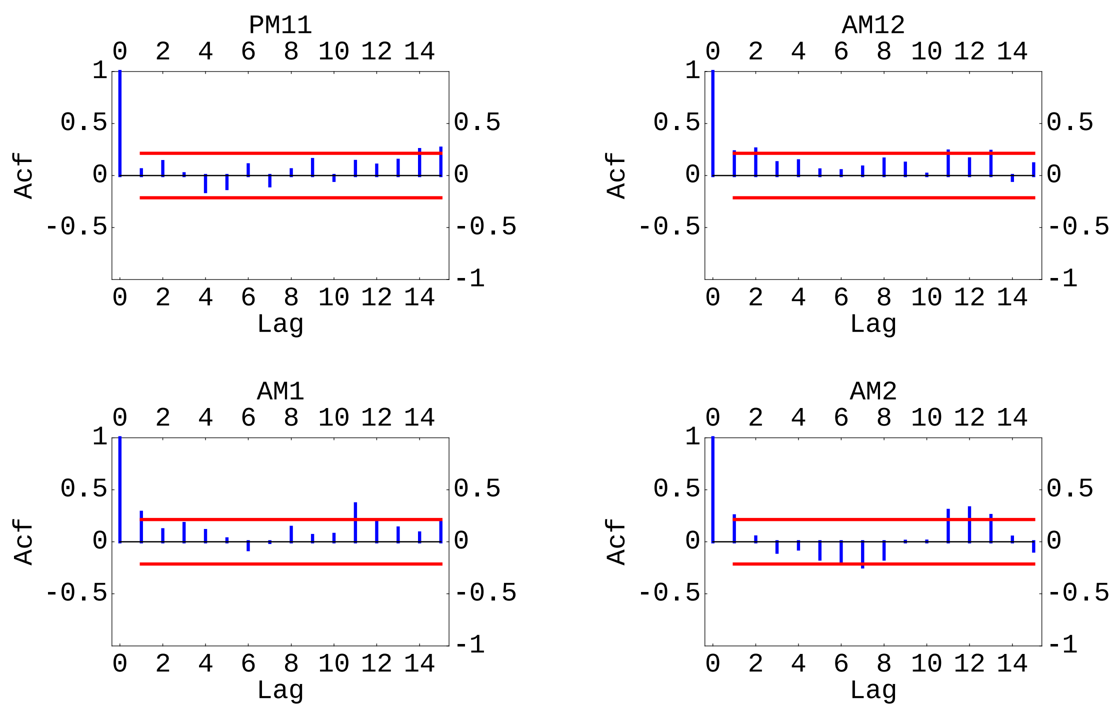

sober
3
2
1

MVA

Autocorrelation Function

Based on the autocorrelation function, the following models were tenatively identified.

Data	ARMA Model
PM11	(0, 0, 0)
AM12	$(0, 0, 2)\,(0, 0, 1)_{12}$
AM1	$(0, 0, 1)\,(0, 0, 1)_{12}$
AM2	$(0, 0, 1)\,(0, 0, 1)_{12}$



Intervention Analysis

May 1996 corresponds to observation #49 in this series. A step intervention model defined by,

$$z_t = \mu + \delta\ \xi_t + N_t \tag{1}$$

where N_t is the error term. Based on the pre-intervention data we assume initially that N_t is normal and independent, so ordinary multiple linear regression can be used. The intervention series are defined by,

$$\xi_t^{(1)} = \begin{cases} 0 & t < 49 \\ 1 & t \geq 49 \end{cases}$$

For each series the ARMA component has been tentatively identifed as follows. For PM11,

$$N_t = a_t \tag{2}$$

For AM12,

$$N_t = (1 - \theta_1 \mathfrak{B} - \theta_2 \mathfrak{B}^2)(1 - \Theta_1 \mathfrak{B}^{12}) \tag{3}$$

For AM1,

$$N_t = (1 - \theta_1 \mathfrak{B})(1 - \Theta_1 \mathfrak{B}^{12}) \tag{4}$$

and for AM2

$$N_t = (1 - \theta_1 \mathfrak{B})(1 - \Theta_1 \mathfrak{B}^{12}) \tag{5}$$

where $a_t \sim \text{NID}(0, \sigma_a^2)$ and $\mathfrak{B}$ denotes the backshift operator on t, $\mathfrak{B}\, a_t = a_{t-1}$. After fitting my exact maximum likelihood the parameter estimates are shown in the tables below. The intervention parameter δ is statistically significant on a two-sided test for PM11 at 9.4% and for AM1 at 4.7%. No departures from model assumptions were indicated in the residual analysis.

- **PM11**

	Estimate	SE	P-Value
μ	3.833	0.21	0
δ	**-1.028**	**0.424**	**0.015**

- **AM12**

	Estimate	SE	P-Value
μ	2.632	0.315	0
δ	-0.23	0.577	0.691
θ_1	-0.103	0.106	0.333
θ_2	-0.233	0.106	0.028
Θ_1	-0.185	0.107	0.085

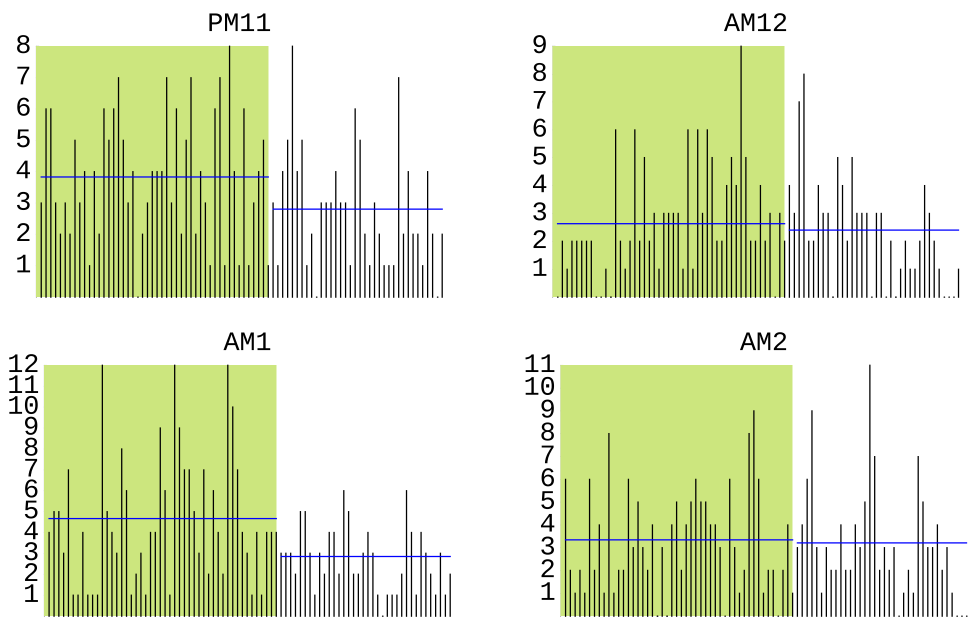
- **AM1**

	Estimate	SE	P-Value
μ	3.341	0.352	0
δ	-0.127	0.622	0.839
θ_1	-0.148	0.108	0.171
Θ_1	-0.305	0.104	0.003

■ AM2

	Estimate	SE	P–Value
μ	–0.341	0.352	0
δ	–0.127	0.622	0.839
Θ_1	–0.148	0.108	0.171
Θ_1	–0.305	0.104	0.003

■ Intervention Visualizations

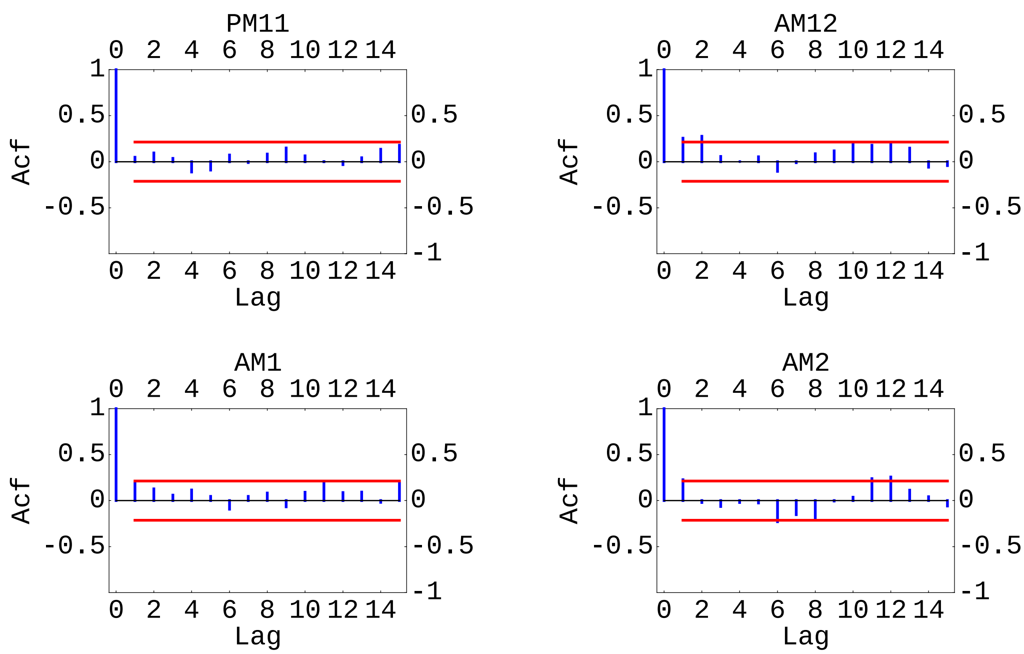


OTR

Autocorrelation Function

Based on the autocorrelation function, the following models were tenatively identified.

Data	**ARMA Model**
PM11	(0, 0, 0)
AM12	$(0, 0, 2)\,(0, 0, 1)_{12}$
AM1	(0, 0, 1)
AM2	$(0, 0, 1)\,(0, 0, 1)_{12}$



Intervention Analysis

May 1996 corresponds to observation #49 in this series. A step intervention model defined by,

$$z_t = \mu + \delta\, \xi_t + N_t \tag{1}$$

where N_t is the error term. Based on the pre-intervention data we assume initially that N_t is normal and independent, so ordinary multiple linear regression can be used. The intervention series are defined by,

$$\xi_t^{(1)} = \begin{cases} 0 & t < 49 \\ 1 & t \geq 49 \end{cases}$$

For each series the ARMA component has been tentatively identifed as follows. For PM11,

$$N_t = a_t \quad (2)$$

For AM12,

$$N_t = (1 - \theta_1 \mathfrak{B} - \theta_2 \mathfrak{B}^2)(1 - \Theta_1 \mathfrak{B}^{12}) \quad (3)$$

For AM1,

$$N_t = a_t - \theta_1 a_{t-1} \quad (4)$$

and for AM2

$$N_t = (1 - \theta_1 \mathfrak{B})(1 - \Theta_1 \mathfrak{B}^{12}) \quad (5)$$

where $a_t \sim \text{NID}(0, \sigma_a^2)$ and $\mathfrak{B}$ denotes the backshift operator on t, $\mathfrak{B}\, a_t = a_{t-1}$. After fitting my exact maximum likelihood the parameter estimates are shown in the tables below. The intervention parameter δ is statistically significant on a two-sided test for PM11 at 9.4% and for AM1 at 4.7%. No departures from model assumptions were indicated in the residual analysis.

- **PM11**

	Estimate	SE	P-Value
μ	6.312	0.314	0
δ	**-1.062**	**0.635**	**0.094**

- **AM12**

	Estimate	SE	P-Value
μ	4.455	0.443	0
"δ	0.271	0.783	0.729
θ_1	-0.175	0.103	0.091
θ_2	-0.323	0.103	0.002
Θ_1	-0.238	0.106	0.025

- **AM1**

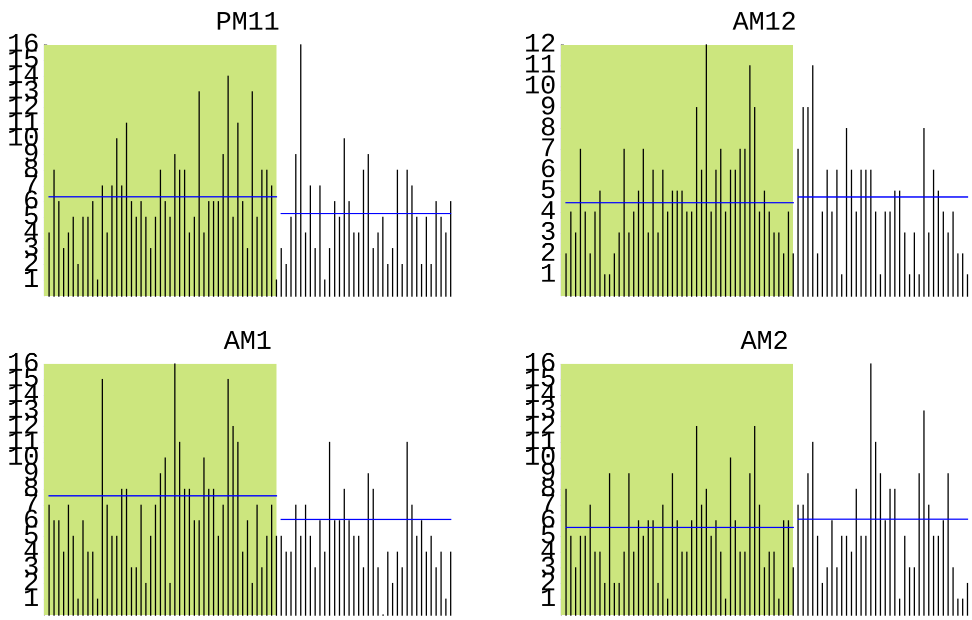
	Estimate	SE	P-Value
μ	7.599	0.379	0
"δ	**-1.513**	**0.762**	**0.047**
θ_1	-0.12	0.108	0.266

- **AM2**

	Estimate	SE	P-Value
μ	5.574	0.464	0
"δ	0.532	0.828	0.521
θ_1	-0.179	0.107	0.096
Θ_1	-0.282	0.105	0.007

■ Intervention Visualizations

The effect of the intervention is visualized graphically.



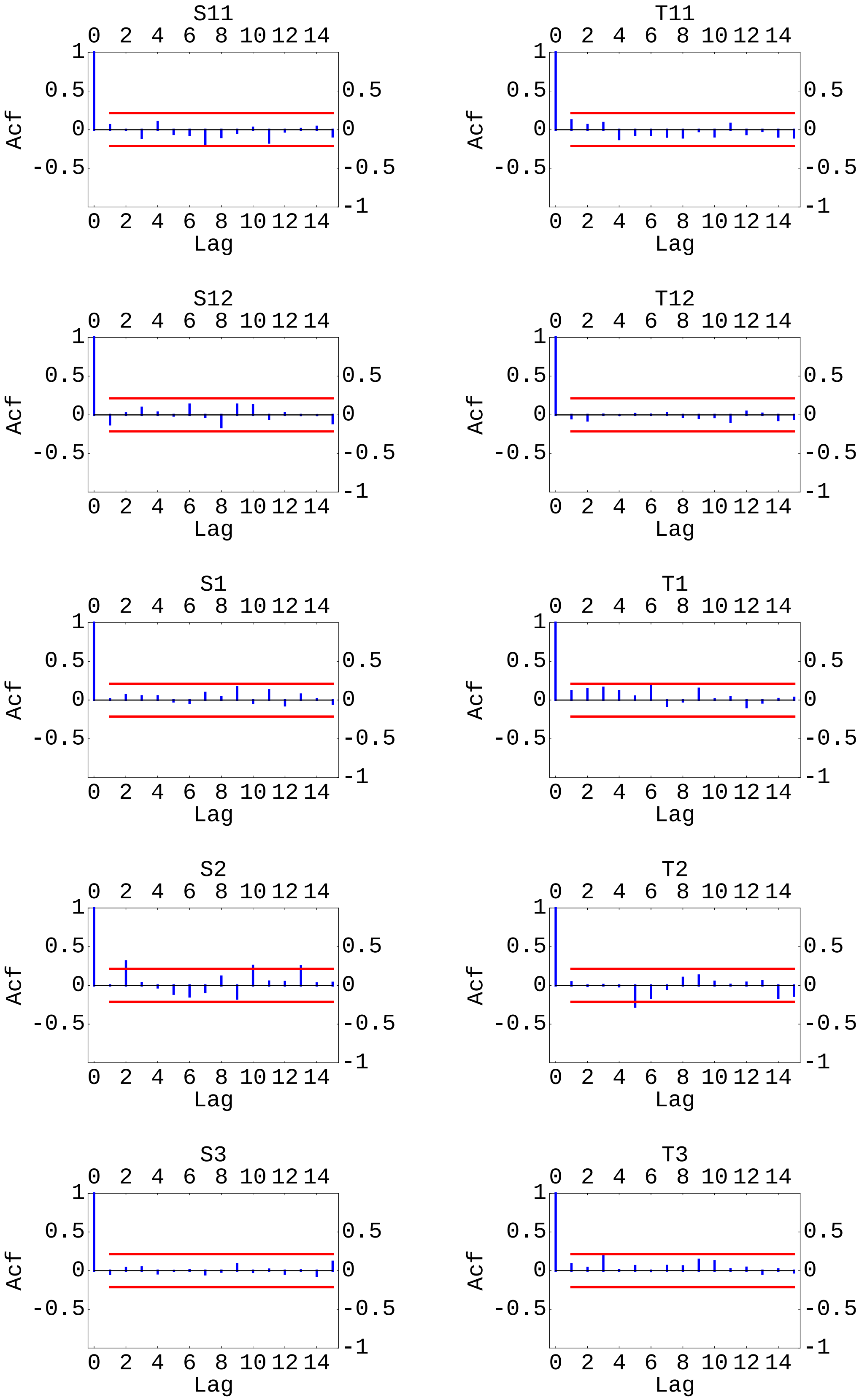
LPS Assaults

Introduction

The LPS assault time series are comprised of $n = 85$ monthly observations beginning May 1992 and ending May 1999. Ten time series were created corresponding the two weekgroup variables SunWed and ThuSat and the five hour one hour periods beginning at 11PM, 12AM, 1AM, 2AM and 3AM. For brevity we will refer to these time series using the codes S11, S12, S1, S2, S3, T11, T12, T1, T2 and T3. We are primarily interested in testing to see if a change occurred starting effective with May 1996, the 49th observation.

Autocorrelation Function

There does not seem to be any strong autocorrelations at either lag one or lag twelve across the entire series, $n = 85$. So we can model the intervention using regression and check that there is no autocorrelation remaining in the residuals.


S11
T11
S12
T12
S1
T1
S2
T2
S3
T3
Acf
Lag

Intervention Analysis

May 1996 corresponds to observation #49 in this series. A step intervention model defined by,

$$z_t = \mu + \delta \, \xi_t + N_t \tag{1}$$

where N_t is the error term. Based on the pre-intervention data we assume initially that N_t is normal and independent, so ordinary multiple linear regression can be used. The intervention series are defined by,

$$\xi_t = \begin{cases} 0 & t < 49 \\ 1 & t \geq 49 \end{cases}$$

The term N_t represents the disturbance or error term and it has been tentatively identified as Gaussian white noise, that is $N_t = a_t$, where $a_t \sim \text{NID}(0, \ \sigma^2)$.

■ Parameter Estimates

S11		Estimate	SE	TStat	PValue
	1	3.0625	0.308224	9.93595	8.88178×10^{-16}
	ξ	0.126689	0.46717	0.271184	0.786923

S12		Estimate	SE	TStat	PValue
	1	4.14583	0.373477	11.1006	0.
	ξ	**-0.848536**	**0.566073**	**-1.49899**	**0.137671**

S1		Estimate	SE	TStat	PValue
	1	3.72917	0.334674	11.1427	0.
	ξ	-0.810248	0.507261	-1.5973	0.113999

S2		Estimate	SE	TStat	PValue
	1	2.35417	0.293271	8.02728	5.74141×10^{-12}
	ξ	**0.835023**	**0.444506**	**1.87854**	**0.0638179**

S3		Estimate	SE	TStat	PValue
	1	1.66667	0.190058	8.76926	1.89182×10^{-13}
	ξ	-0.153153	0.288068	-0.531657	0.596384

T11		Estimate	SE	TStat	PValue
	1	3.79167	0.392325	9.6646	3.10862×10^{-15}
	ξ	0.478604	0.594641	0.804861	0.423199

T12		Estimate	SE	TStat	PValue
	1	5.70833	0.840903	6.78833	1.57354×10^{-9}
	ξ	-0.221847	1.27454	-0.17406	0.862242

T1		Estimate	SE	TStat	PValue
	1	6.54167	0.38298	17.0809	0.
	ξ	**-1.89302**	**0.580477**	**-3.26114**	**0.00161083**

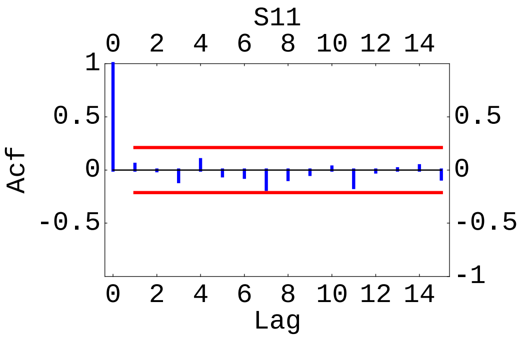
T2		Estimate	SE	TStat	PValue
	1	4.97917	0.41123	12.108	0.
	ξ	**0.939752**	**0.623295**	**1.50772**	**0.135424**

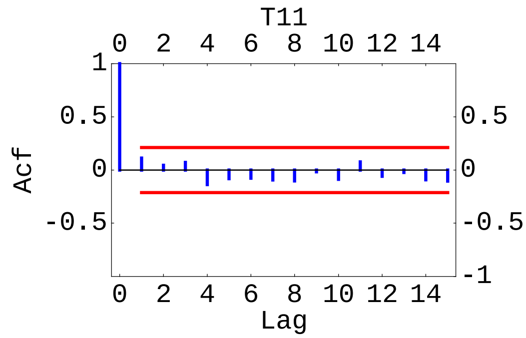
T3		Estimate	SE	TStat	PValue
	1	3.04167	0.314055	9.68514	2.66454×10^{-15}
	ξ	**1.1205**	**0.476008**	**2.35394**	**0.0209387**

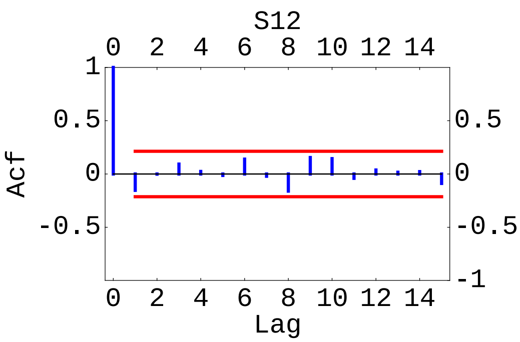
Using a one-sided test, the intervention parameter is significant at 10% for S12 and T2 and it is significant at 5% for S2, T1 and T3. The intervention is especially strong for T1 and T3.

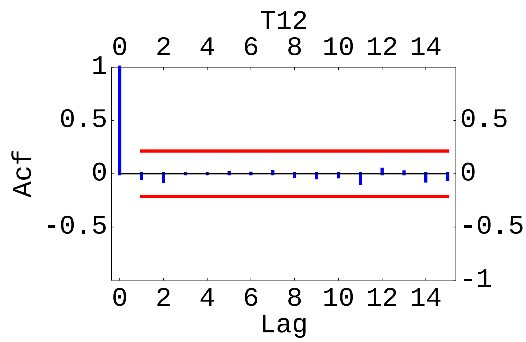
■ Residual Autocorrelation Analysis

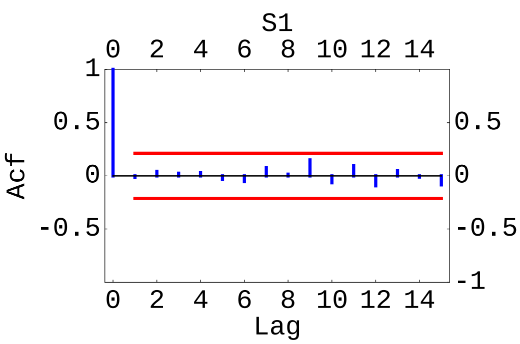
The autocorrelation in the residuals is examined particularly at lag one and lag twelve. Everything looks reasonable.

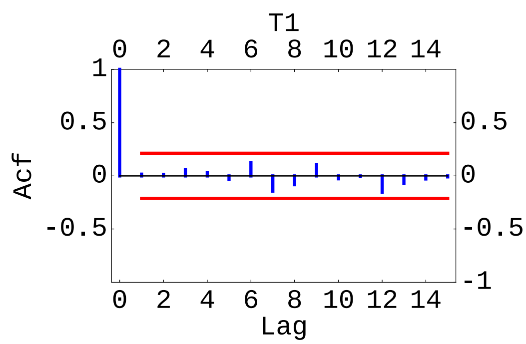


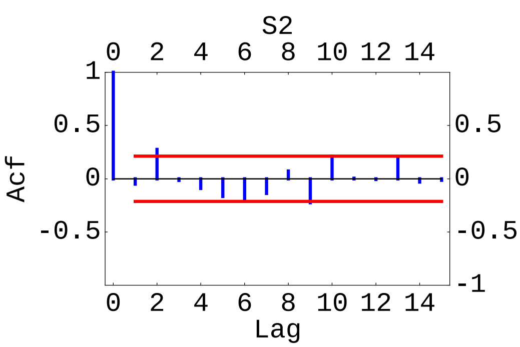


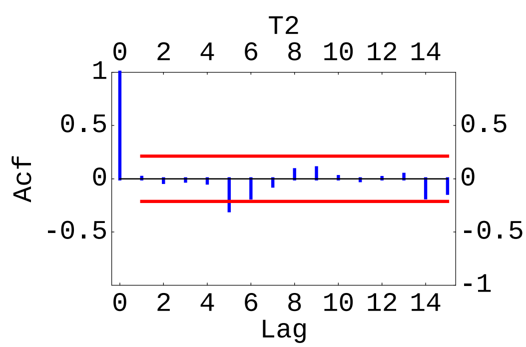


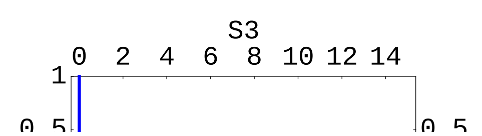


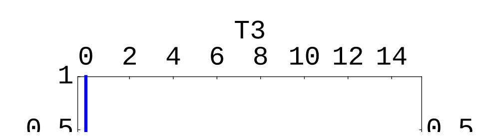


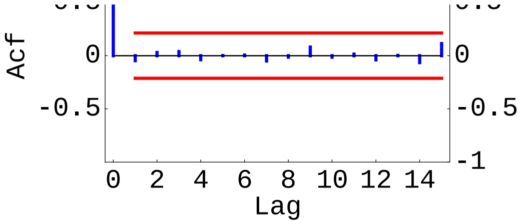


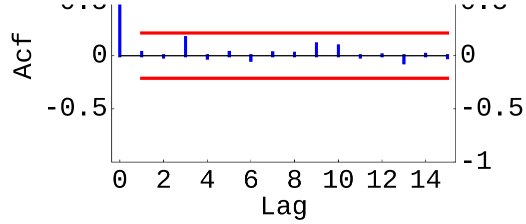




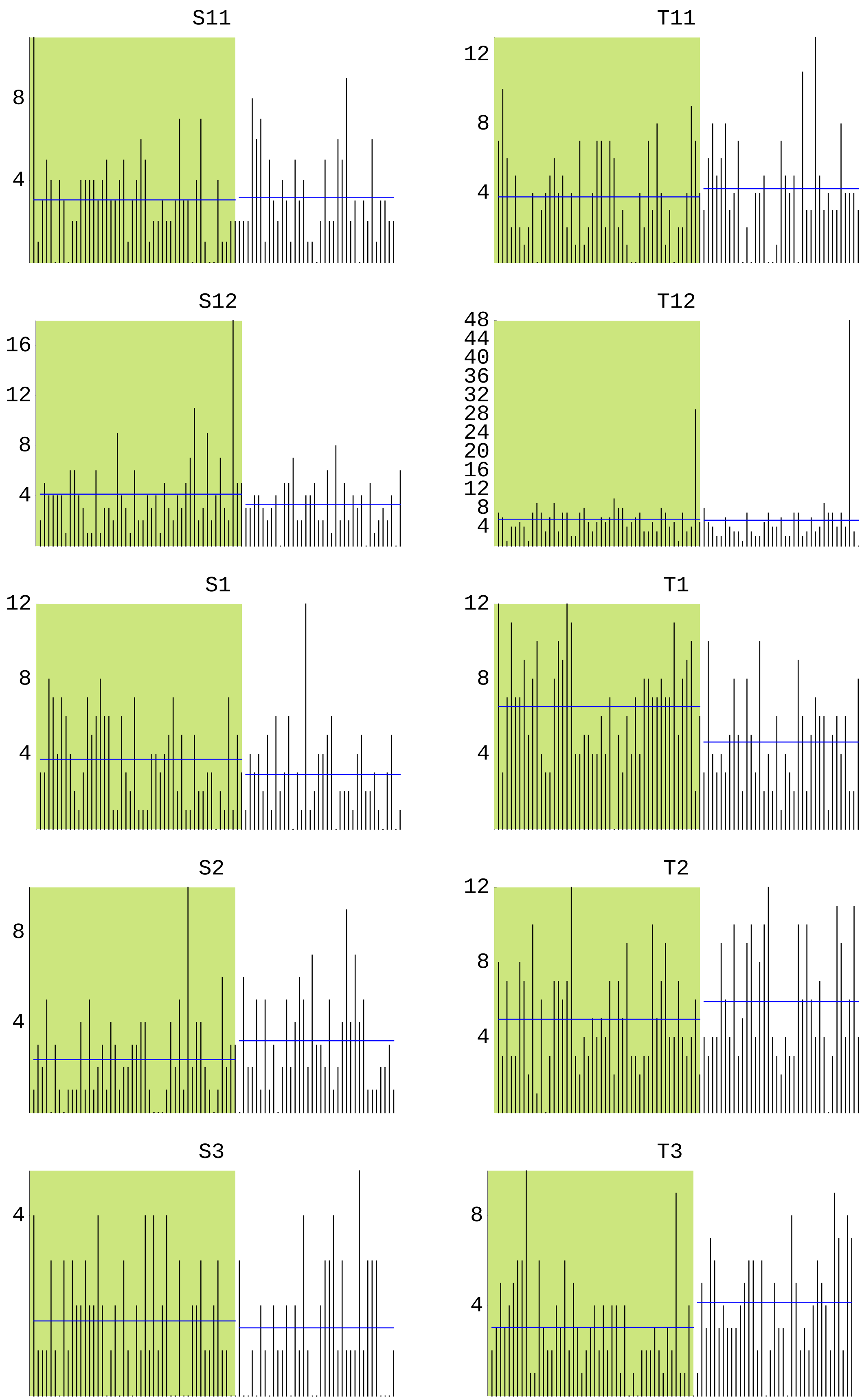








■ Visualization of Intervention


S11
8
4
T11
12
8
4
S12
16
12
8
4
T12
48
44
40
36
32
28
24
20
16
12
8
4
S1
12
8
4
T1
12
8
4
S2
8
4
T2
12
8
4
S3
4
T3
8
4

Logged Data

The IA graphics indicated some strong outliers present especially for T12 and S12. The models were refit using a logarithmic transformation, $Y = \log_2(X + 1)$.

		Estimate	SE	TStat	PValue
S11	1	1.81152	0.117815	15.376	0.
	ξ	0.0623211	0.17857	0.349001	0.727972

		Estimate	SE	TStat	PValue
S12	1	2.17771	0.10731	20.2936	0.
	ξ	**-0.243936**	**0.162648**	**-1.49978**	**0.137465**

		Estimate	SE	TStat	PValue
S1	1	2.05043	0.120043	17.0809	0.
	ξ	**-0.327416**	**0.181947**	**-1.79951**	**0.0755716**

		Estimate	SE	TStat	PValue
S2	1	1.51989	0.118814	12.7923	0.
	ξ	**0.344939**	**0.180084**	**1.91544**	**0.0588807**

		Estimate	SE	TStat	PValue
S3	1	1.23009	0.114479	10.7451	0.
	ξ	-0.117828	0.173514	-0.679068	0.498984

		Estimate	SE	TStat	PValue
T11	1	2.01823	0.137384	14.6904	0.
	ξ	0.106145	0.208231	0.509745	0.611583

		Estimate	SE	TStat	PValue
T12	1	2.57308	0.113033	22.764	0.
	ξ	**-0.220464**	**0.171322**	**-1.28684**	**0.201728**

		Estimate	SE	TStat	PValue
T1	1	2.79222	0.0956368	29.1961	0.
	ξ	**-0.43482**	**0.144955**	**-2.99968**	**0.0035669**

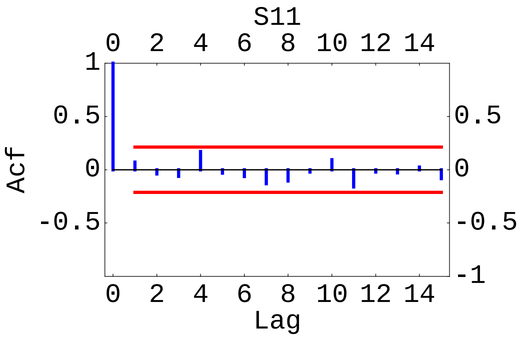
		Estimate	SE	TStat	PValue
T2	1	2.42798	0.104299	23.2791	0.
	ξ	**0.200226**	**0.158084**	**1.26658**	**0.208848**

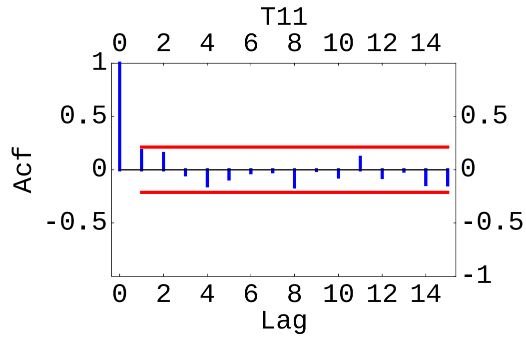
		Estimate	SE	TStat	PValue
T3	1	1.82001	0.112327	16.2029	0.
	ξ	**0.38055**	**0.170252**	**2.23522**	**0.0280885**

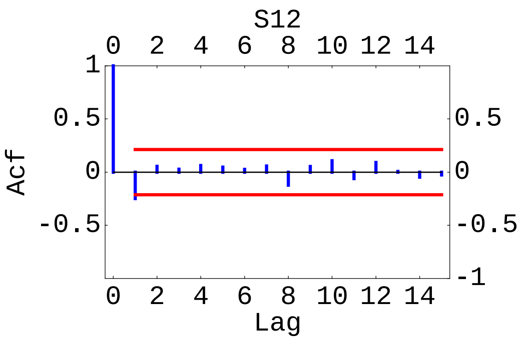
Using a one-sided test, the intervention parameter is significant at 10% for S12, T12 and T2 and it is significant at 5% for S2, T1 and T3. The intervention is especially strong for T1 and T3. The logarithmic transformation has improved the sensitivity of the analysis.

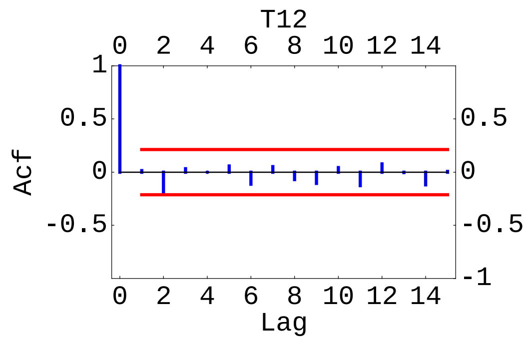
Residual Autocorrelation Analysis

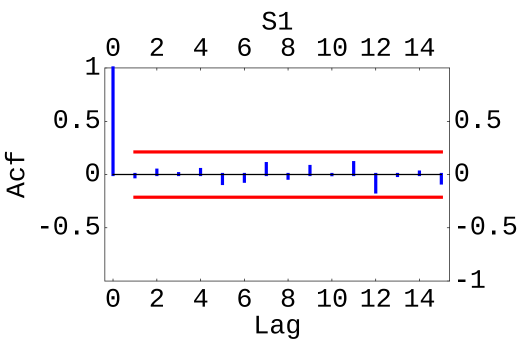
The autocorrelation in the residuals is examined particularly at lag one and lag twelve. There appears to be significant positive autocorrelation at lag one for T11 and negative lag one autocorrelation for S12 and negative lag 12 autocorrelation for T1. As a rule positive autocorrelation implies that the regression significance levels are too low, ie. the significance is overstated and negative autocorrelation implies the reverse. Thus we may expect that after fitting models with autocorrelated error the already statistically significant results for S12 and T1 would be increased. The result for T11 would remain statistically not significant.

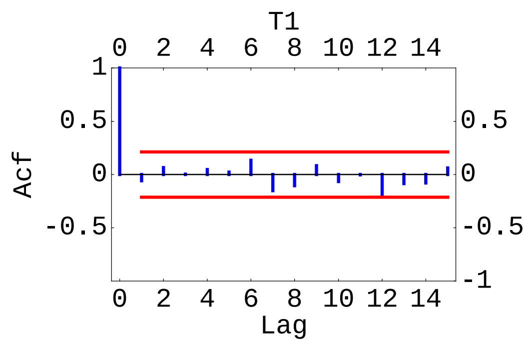


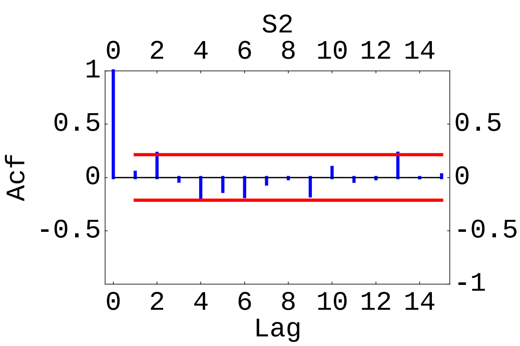


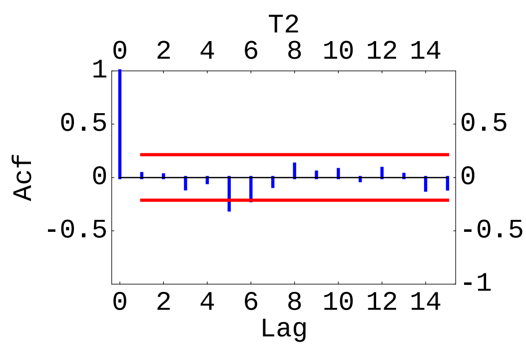






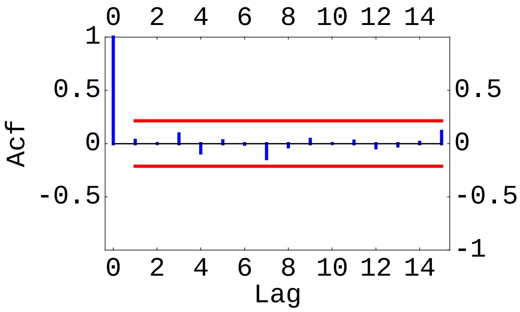


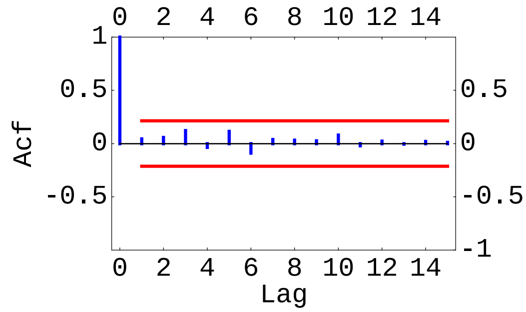




S3

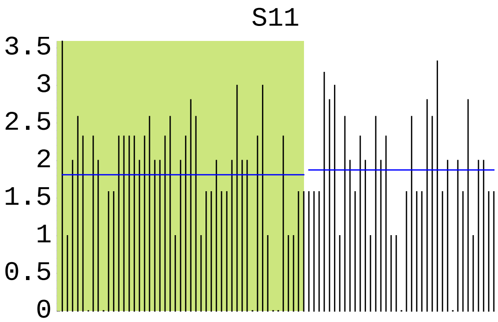
T3

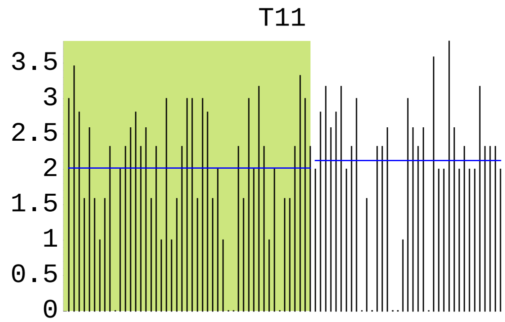


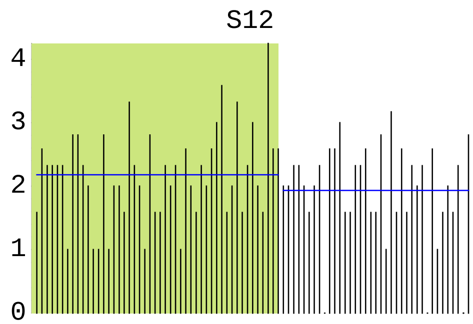


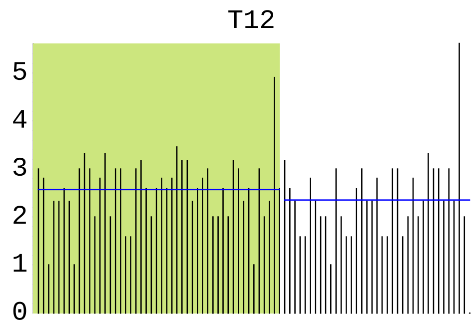
■ Visualization of Logged Data Intervention Analysis

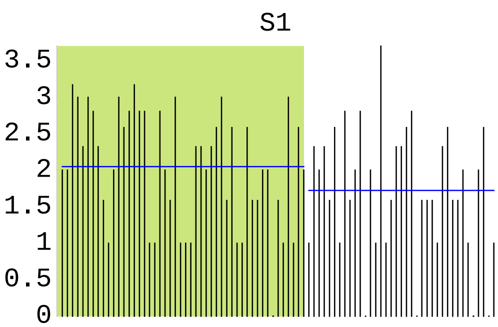
We examine the logarithmically transformed series and the intervention effect on a log scale.

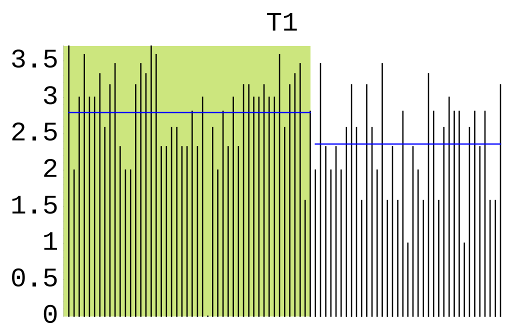

S11
3.5 3 2.5 2 1.5 1 0.5 0

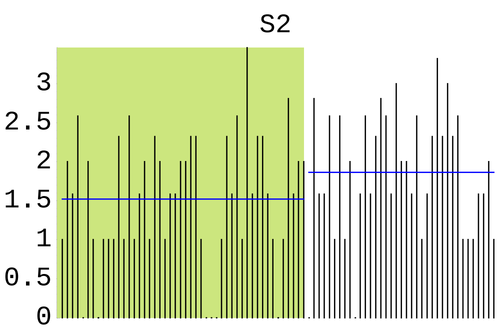

T11
3.5 3 2.5 2 1.5 1 0.5 0

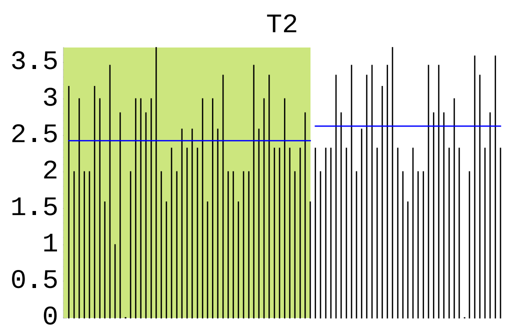

S12
4 3 2 1 0

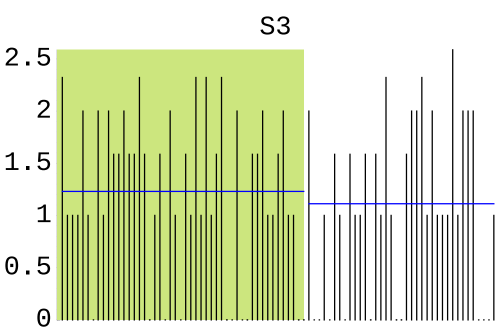

T12
5 4 3 2 1 0

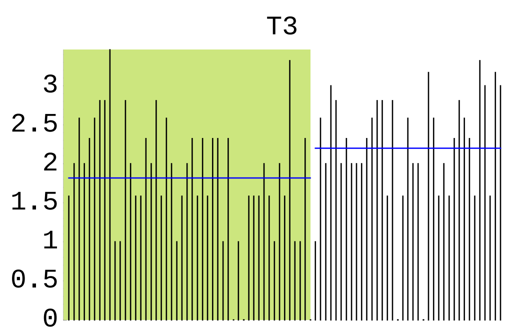

S1
3.5 3 2.5 2 1.5 1 0.5 0


T1
3.5 3 2.5 2 1.5 1 0.5 0


S2
3 2.5 2 1.5 1 0.5 0


T2
3.5 3 2.5 2 1.5 1 0.5 0


S3
2.5 2 1.5 1 0.5 0

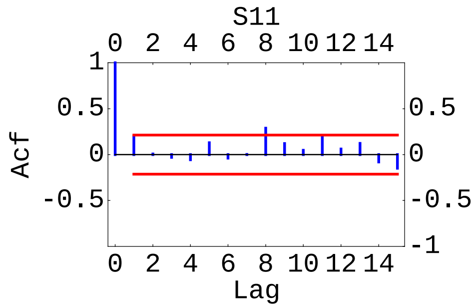

T3
3 2.5 2 1.5 1 0.5 0

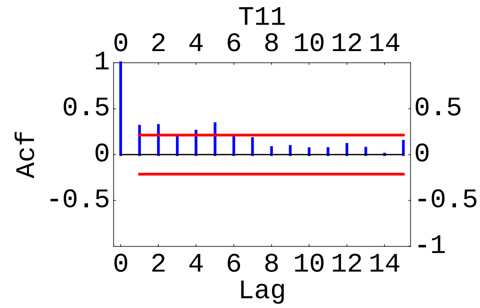
LPS Impaired

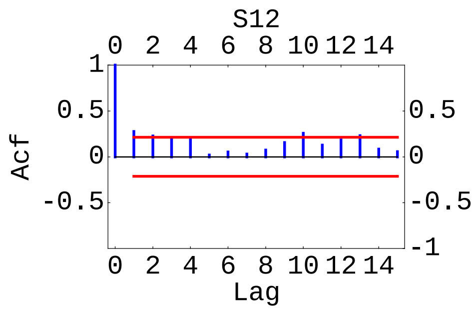
The LPS impaired driving charges time series are comprised of $n = 85$ monthly observations beginning May 1992 and ending May 1999. Ten time series were created corresponding the two weekgroup variables SunWed and ThuSat and the five hour one hour periods beginning at 11PM, 12AM, 1AM, 2AM and 3AM. For brevity we will refer to these time series using the codes S11, S12, S1, S2, S3, T11, T12, T1, T2 and T3. We are primarily interested in testing to see if a change occurred starting effective with May 1996, the 49th observation.

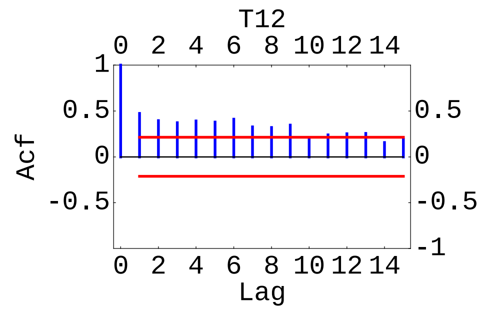
Autocorrelation Function

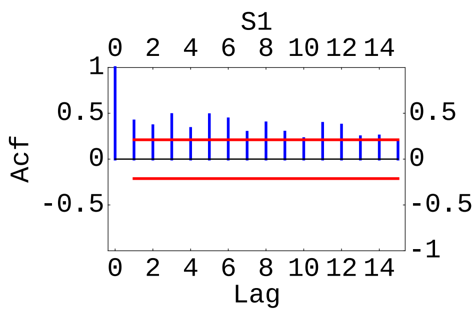
Many of the series have significant autocorrelations. A two-step approach will be used to develop the intervention analysis model. First the intervention model will be fit using ordinary regression. Then the residuals from this model will be used to identify the ARIMA noise component. Finally the model will be refit using exact maximum likelihood.

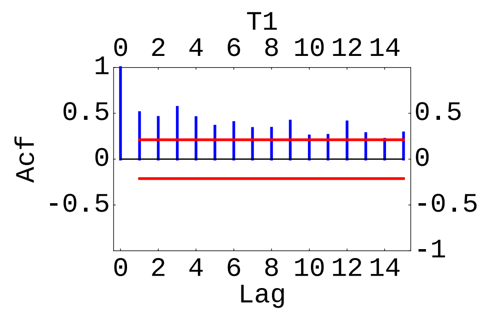

S11
Acf
Lag

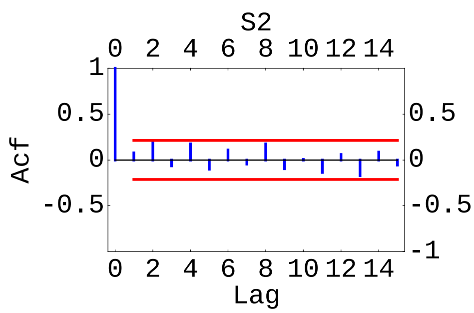

T11
Acf
Lag

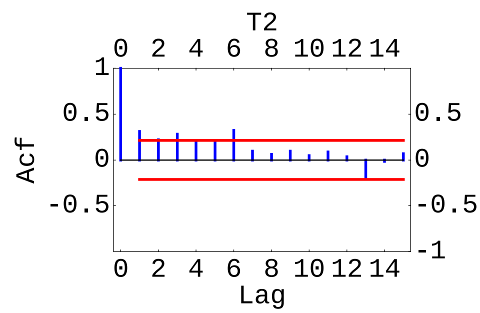

S12
Acf
Lag

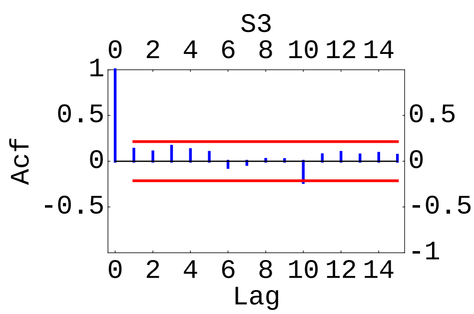

T12
Acf
Lag

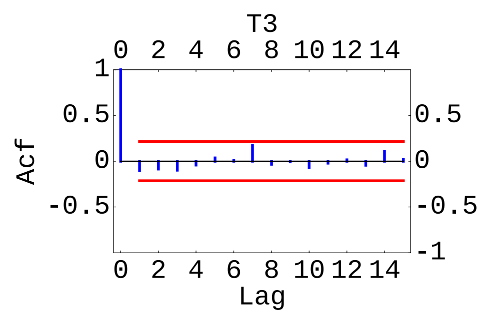

S1
Acf
Lag


T1
Acf
Lag


S2
Acf
Lag


T2
Acf
Lag


S3
Acf
Lag


T3
Acf
Lag

Intervention Analysis

May 1996 corresponds to observation #49 in this series. A step intervention model defined by,

$$z_t = \mu + \delta\,\xi_t + N_t \tag{1}$$

where N_t is the error term. Based on the pre-intervention data we assume initially that N_t is normal and independent, so ordinary multiple linear regression can be used. The intervention series are defined by,

$$\xi_t = \begin{cases} 0 & t < 49 \\ 1 & t \geq 49 \end{cases}$$

The term N_t represents the disturbance or error term and it has been tentatively identified as Gaussian white noise, that is $N_t = a_t$, where $a_t \sim \text{NID}(0,\ \sigma^2)$.

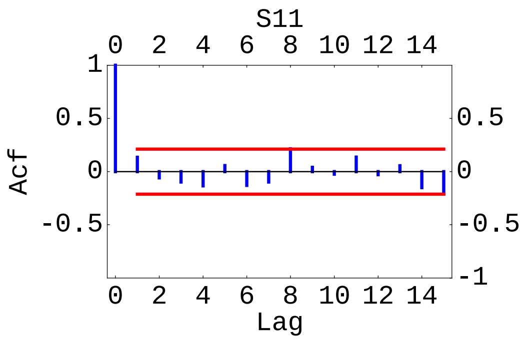
■ Parameter Estimates

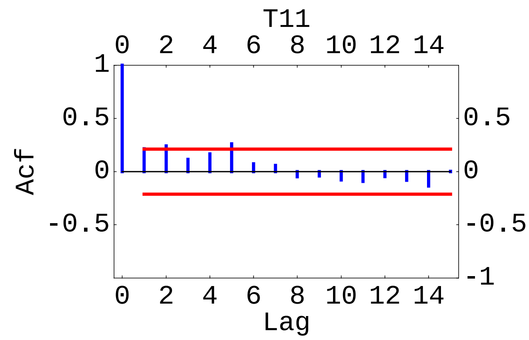
Using a one-sided test, the intervention parameter is significant at <5% for S12, S1, S2, S3, T1 and T12.

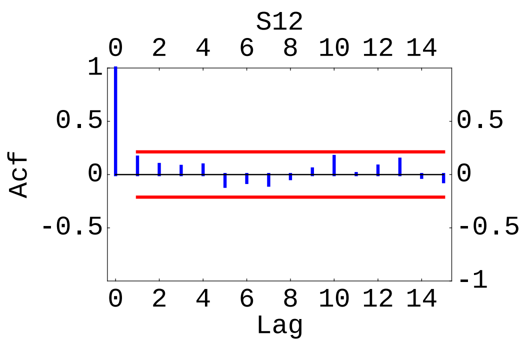
Series		Estimate	SE	TStat	PValue
S11	1	3.70833	0.313341	11.8348	0.
	ξ	**-1.19482**	**0.474926**	**-2.5158**	**0.0138033**
S12	1	5.58333	0.396049	14.0976	0.
	ξ	**-2.09685**	**0.600285**	**-3.49308**	**0.000768253**
S1	1	9.89583	0.553203	17.8883	0.
	ξ	**-5.97691**	**0.83848**	**-7.12827**	**3.43187×10^{-10}**
S2	1	6.0625	0.405999	14.9323	0.
	ξ	**-1.00845**	**0.615367**	**-1.63877**	**0.105046**
S3	1	1.79167	0.318729	5.62128	2.47785×10^{-7}
	ξ	**1.45158**	**0.483092**	**3.00476**	**0.00351374**
T11	1	7.04167	0.484278	14.5405	0.
	ξ	**-2.39302**	**0.734013**	**-3.26019**	**0.00161564**
T12	1	11.2708	0.595437	18.9287	0.
	ξ	**-5.10867**	**0.902494**	**-5.66062**	**2.10084×10^{-7}**
T1	1	19.6042	0.952457	20.5827	0.
	ξ	**-11.2258**	**1.44362**	**-7.77612**	**1.81282×10^{-11}**
T2	1	10.2708	0.607125	16.9172	0.
	ξ	-1.10867	0.92021	-1.2048	0.231704
T3	1	4.08333	0.403061	10.1308	4.44089×10^{-16}
	ξ	-0.191441	0.610913	-0.313369	0.754786

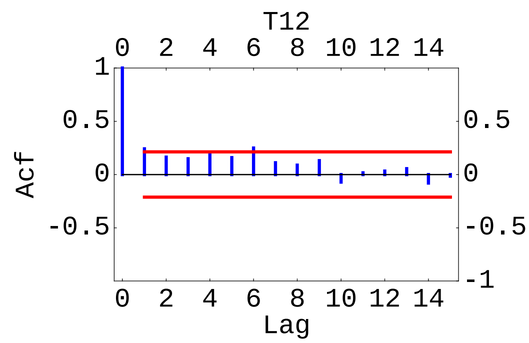
■ Residual Autocorrelation Analysis and ARIMA Identification

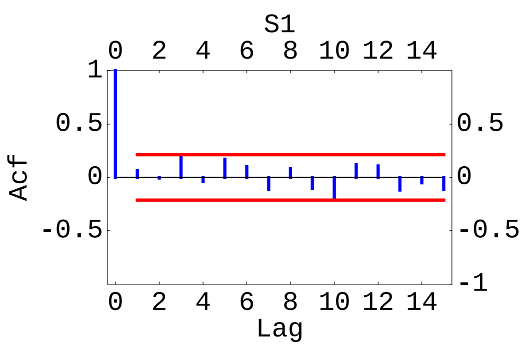
The autocorrelation in the residuals is examined particularly at lag one and lag twelve. Everything looks reasonable except perhaps for T3 has a large negative autocorrelation at lag one.

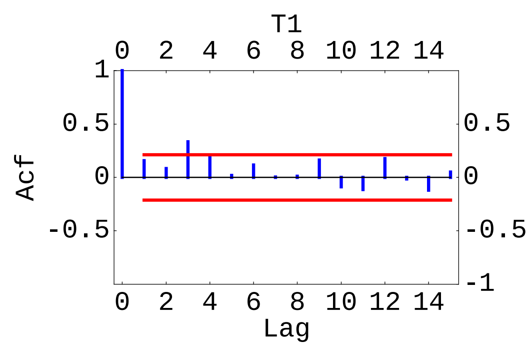


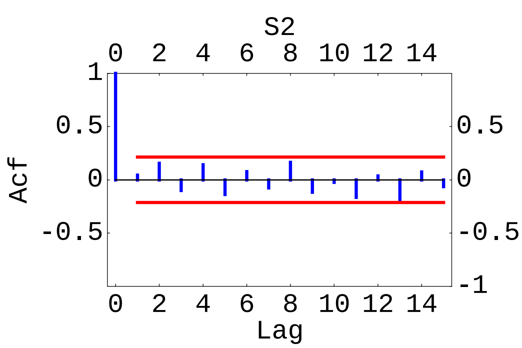


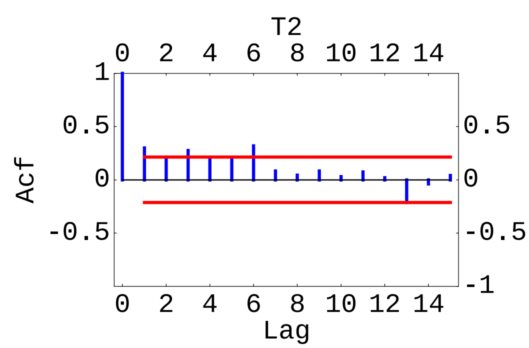


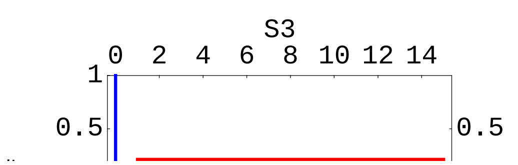


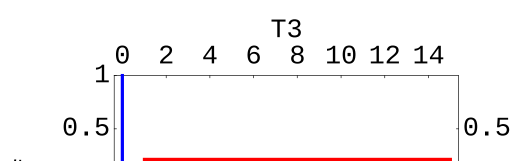


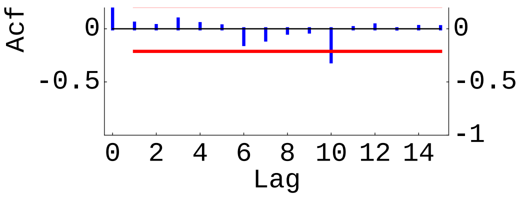


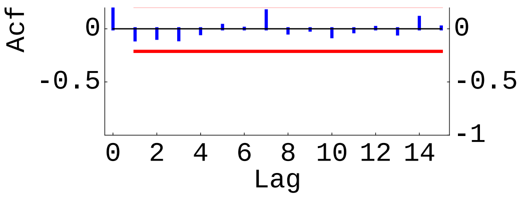




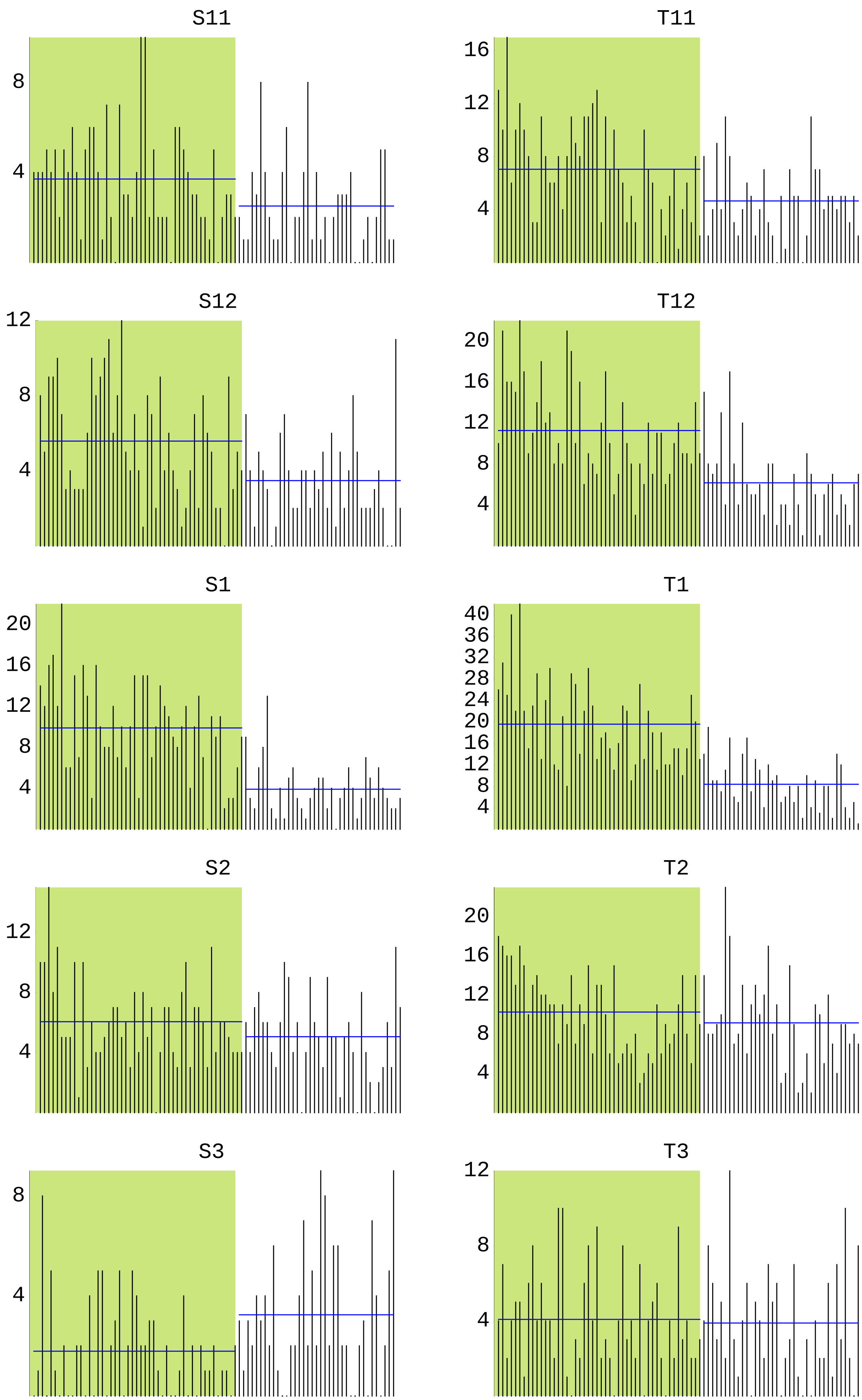








■ Visualization


S11
8
4
T11
16
12
8
4
S12
12
8
4
T12
20
16
12
8
4
S1
20
16
12
8
4
T1
40
36
32
28
24
20
16
12
8
4
S2
12
8
4
T2
20
16
12
8
4
S3
8
4
T3
12
8
4

ARIMA Modelling

Based on the residual autocorrelations of the regression, the following ARIMA models are suggested for the error term N_t.

Model Identification

Dataset	**ARIMA(p, d, q)**
S11	(0, 0, 1)
S12	(0, 0, 1)
S1	(0, 0, 0)
S2	(0, 0, 0)
S3	(0, 0, 0)
T11	(0, 0, 2)
T12	(0, 0, 1)
T1	$(0, 0, 1)(0, 0, 1)_{12}$
T2	(0, 0, 3)
T3	(0, 0, 0)

These ARIMA model specifications are then used in a full joint exact maximum likelihood estimation of the intervention models. The results are summarized below. As expected there are only some minor changes in the significance level. The intervention parameter δ only changes slightly.

S11 (0, 0, 1)

The intervention is still statistically significant at 5% on a one-sided test. The magnitude of the intervention parameter has not changed.

	Estimate	SE	P-Value
μ	3.704	0.265	0
δ	**-1.19**	**0.531**	**0.025**
θ_1	-0.15	0.107	0.162

S12 (0, 0, 1)

The intervention is still statistically significant at 5% on a one-sided test. The magnitude of the intervention parameter has changed very slightly.

	Estimate	SE	P-Value
μ	5.574	0.333	0
δ	**-2.075**	**0.668**	**0.002**
θ_1	-0.146	0.107	0.174

T11 (0, 0, 2)

The intervention is still statistically significant at about 1% on a one-sided test. The magnitude of the intervention parameter has changed very slightly.

	Estimate	SE	P-Value
μ	7.014	0.482	0
δ	**-2.304**	**0.952**	**0.016**
θ_1	-0.183	0.106	0.083
θ_2	-0.214	0.106	0.043

- **T12 (0, 0, 1)**

The intervention is still statistically highly significant. The magnitude of the intervention parameter has changed very slightly.

	Estimate	SE	P-Value
μ	12.227	0.53	0
δ	**-4.958**	**1.06**	**0**
θ_1	-0.2	0.106	0.06

- **T1 (0, 0, 1) (0, 0, 1)$_{12}$**

The intervention is still statistically highly significant. The magnitude of the intervention parameter has changed slightly.

	Estimate	SE	P-Value
μ	19.396	1.024	0
δ	**-10.432**	**1.833**	**0**
θ_1	-0.191	0.106	0.072
Θ_1	-0.275	0.104	0.008

- **T2 (0, 0, 3)**

The intervention parameter has only slightly changed and it remains not significant at the 10% level.

	Estimate	SE	P-Value
μ	10.293	0.652	0
δ	-1.067	1.274	0.402
θ_1	-0.242	0.107	0.024
θ_2	-0.145	0.109	0.183
θ_3	-0.15	0.107	0.161

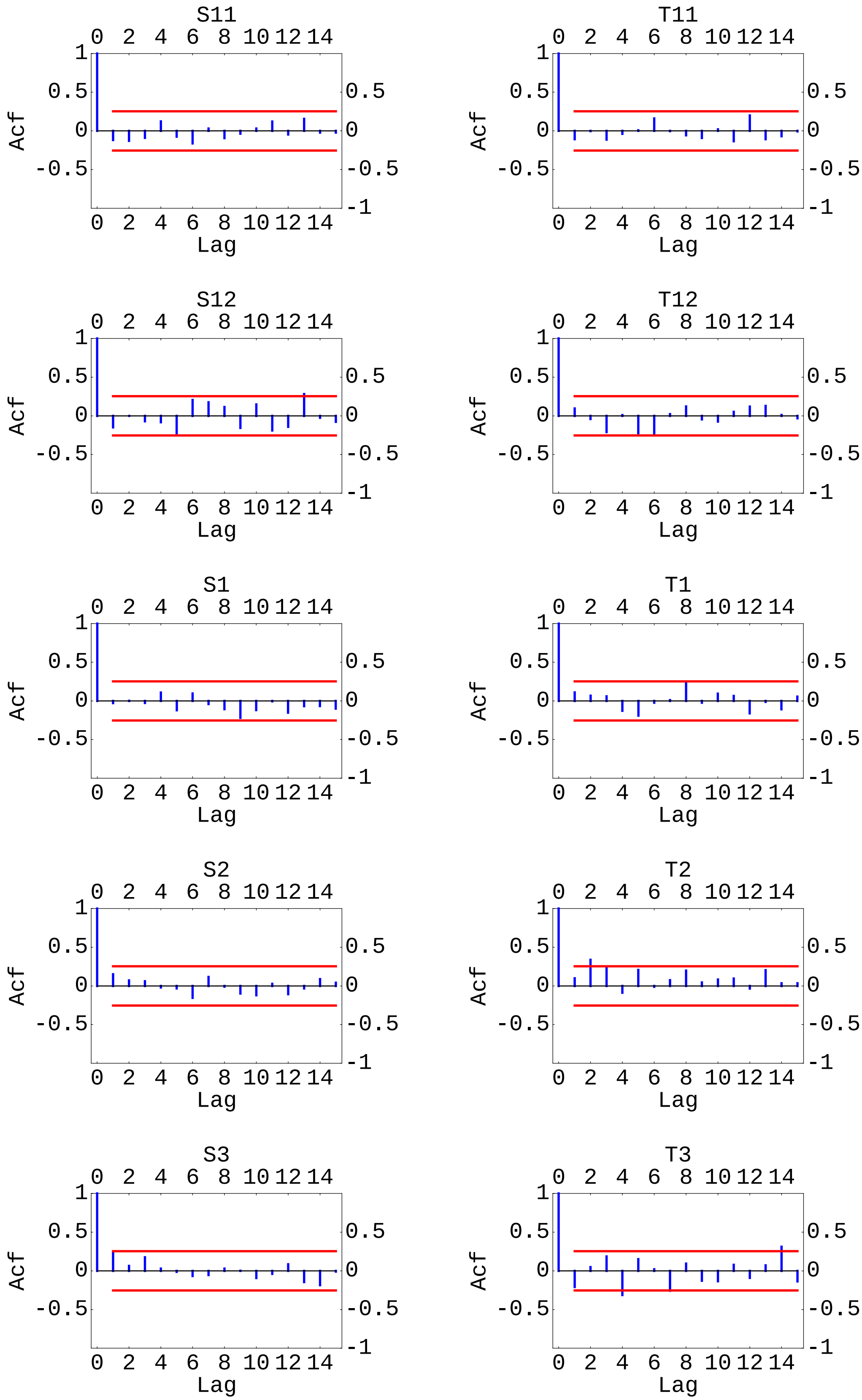
WPS Assaults

Introduction

The WPS assault time series are comprised of $n = 60$ monthly observations beginning January 1994 and ending December 1998. Ten time series were created corresponding the two weekgroup variables SunWed and ThuSat and the five hour one hour periods beginning at 11PM, 12AM, 1AM, 2AM and 3AM. For brevity we will refer to these time series using the codes S11, S12, S1, S2, S3, T11, T12, T1, T2 and T3. We are primarily interested in testing to see if a change occurred starting effective with May 1996, the 29th observation.

Autocorrelation Function

There does not seem to be any strong autocorrelations at either lag one or lag twelve across the entire series, $n = 60$. So we can model the intervention using regression and check that there is no autocorrelation remaining in the residuals.


S11
T11
S12
T12
S1
T1
S2
T2
S3
T3
Acf
Lag

Intervention Analysis

May 1996 corresponds to observation #29 in this series. A step intervention model defined by,

$$z_t = \mu + \delta \xi_t + N_t \tag{1}$$

where N_t is the error term. Based on the pre-intervention data we assume initially that N_t is normal and independent, so ordinary multiple linear regression can be used. The intervention series are defined by,

$$\xi_t = \begin{cases} 0 & t < 29 \\ 1 & t \geq 29 \end{cases}$$

The term N_t represents the disturbance or error term and it has been tentatively identified as Gaussian white noise, that is $N_t = a_t$, where $a_t \sim \text{NID}(0,\ \sigma^2)$.

■ Parameter Estimates

		Estimate	SE	TStat	PValue
S11	1	3.42857	0.567344	6.0432	1.15666×10^{-7}
	ξ	−0.491071	0.776868	−0.632117	0.529794

		Estimate	SE	TStat	PValue
S12	1	2.85714	0.400813	7.12837	1.78397×10^{-9}
	ξ	−0.638393	0.548836	−1.16318	0.249521

		Estimate	SE	TStat	PValue
S1	1	2.39286	0.361005	6.62832	1.23055×10^{-8}
	ξ	−0.361607	0.494327	−0.731515	0.46741

		Estimate	SE	TStat	PValue
S2	1	1.46429	0.294064	4.97948	6.05137×10^{-6}
	ξ	**0.754464**	**0.402664**	**1.87368**	**0.0660148**

		Estimate	SE	TStat	PValue
S3	1	0.714286	0.209407	3.411	0.00118467
	ξ	**0.410714**	**0.286742**	**1.43235**	**0.157412**

		Estimate	SE	TStat	PValue
T11	1	3.10714	0.42323	7.34149	7.81845×10^{-10}
	ξ	−0.263393	0.579532	−0.454492	0.651171

		Estimate	SE	TStat	PValue
T12	1	3.21429	0.439494	7.31361	8.70954×10^{-10}
	ξ	−0.245536	0.601801	−0.408001	0.684775

		Estimate	SE	TStat	PValue
T1	1	5.	0.555408	9.00239	1.31406×10^{-12}
	ξ	−0.96875	0.760524	−1.27379	0.207818

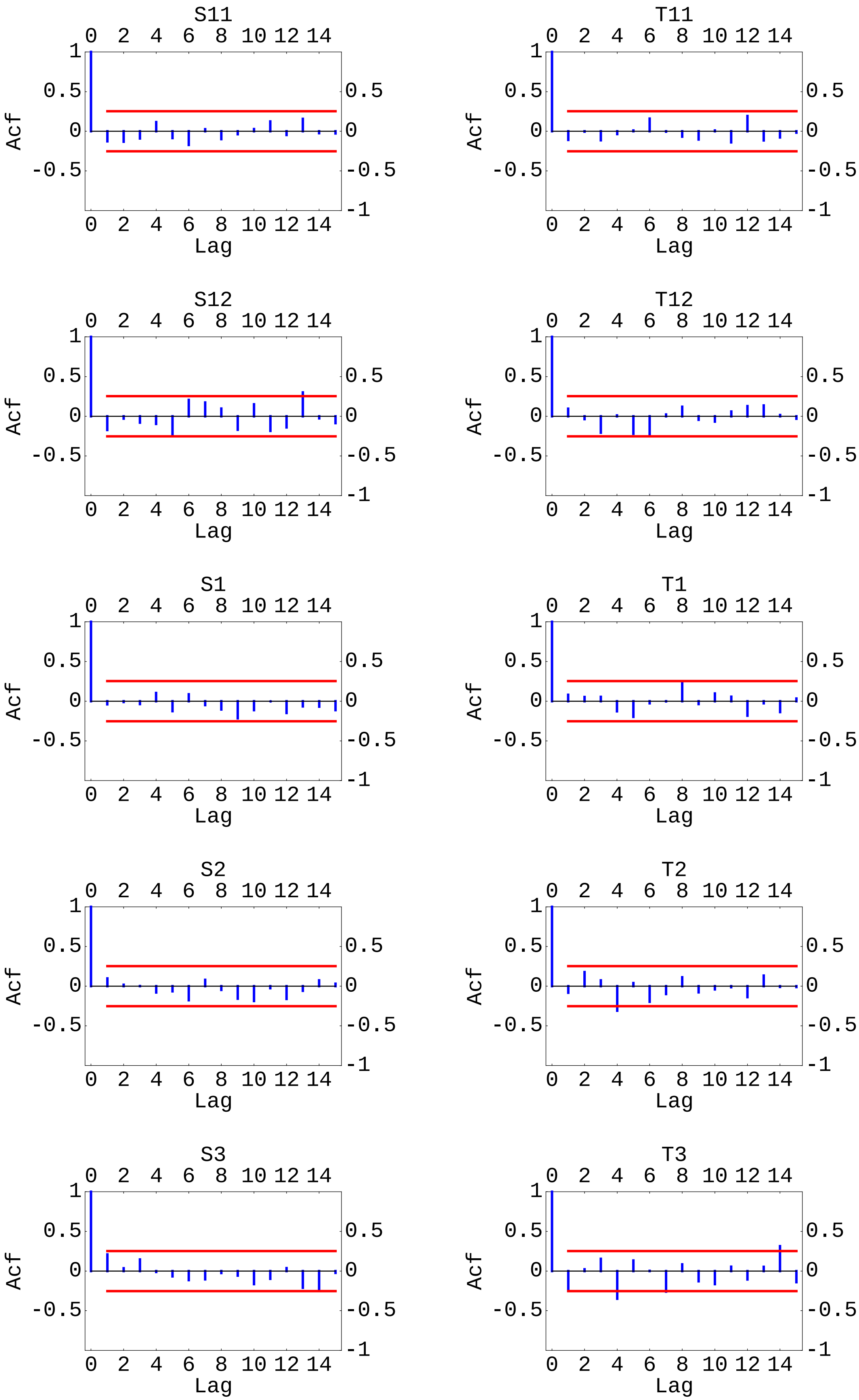
		Estimate	SE	TStat	PValue
T2	1	2.35714	0.47856	4.92549	7.35334×10^{-6}
	ξ	**2.26786**	**0.655296**	**3.46081**	**0.0010171**

		Estimate	SE	TStat	PValue
T3	1	1.75	0.283745	6.16751	7.20387×10^{-8}
	ξ	**0.625**	**0.388534**	**1.60861**	**0.113131**

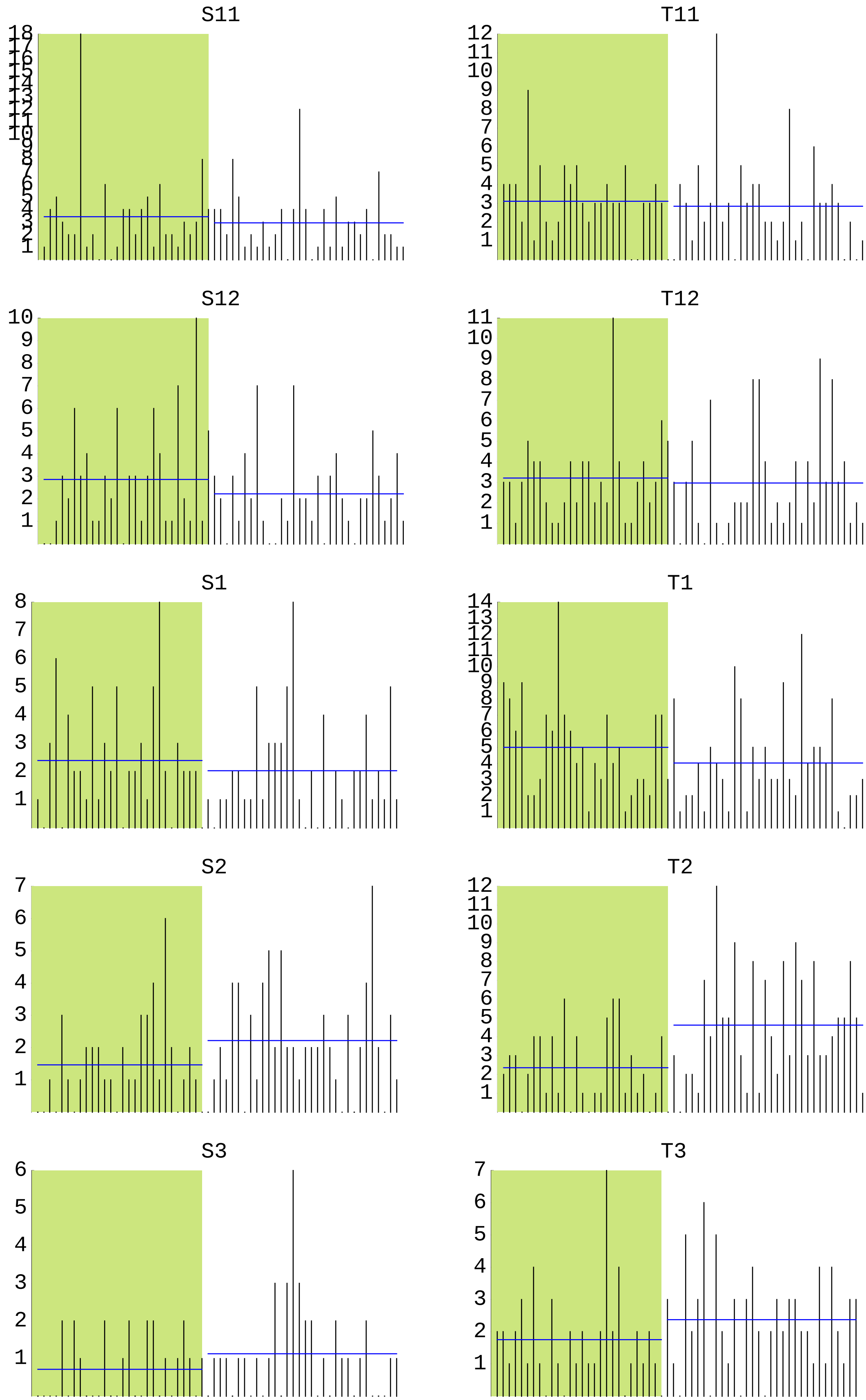
Using a one-sided test, the intervention parameter is significant at 10% for S3 and T3 and it is significant at 5% for S2 and T2. The intervention is especially strong for T2.

■ Residual Autocorrelation Analysis

The autocorrelation in the residuals is examined particularly at lag one and lag twelve. Everything looks reasonable except perhaps for S3 and T3 which have rather large autocorrelations at lag one.


S11
T11
S12
T12
S1
T1
S2
T2
S3
T3
Acf
Lag

■ Visualization of Intervention


S11
T11
S12
T12
S1
T1
S2
T2
S3
T3

Additional Fitting

The residual autocorrelations suggested that for the S3 and T3 series we should consider a model of the form $N_t \sim \text{ARIMA}(0, 0, 1)$ for the disturbance term.

S3 (0, 0, 1)

	Estimate	SE	P-Value
μ	0.725	0.168	0
δ	0.385	0.333	0.248
θ_1	-0.225	0.126	0.074

As expected the statistical significance of the intervention effect is reduced due to the positive autocorrelation.

T3 (0, 0, 1)

	Estimate	SE	P-Value
μ	1.763	0.14	0
δ	**0.594**	**0.285**	**0.037**
θ_1	0.24	0.125	0.055

As expected the statistical significance of the intervention effect is increased due to the negative autocorrelation.

WPS Impaired

The WPS impaired driving charges time series are comprised of $n = 60$ monthly observations beginning January 1994 and ending December 1998. Ten time series were created corresponding the two weekgroup variables SunWed and ThuSat and the five hour one hour periods beginning at 11PM, 12AM, 1AM, 2AM and 3AM. For brevity we will refer to these time series using the codes S11, S12, S1, S2, S3, T11, T12, T1, T2 and T3. We are primarily interested in testing to see if a change occurred starting effective with May 1996, the 29th observation.

May 1996 corresponds to observation #29 in this series. A step intervention model defined by,

$$z_t = \mu + \delta \xi_t + N_t \quad (1)$$

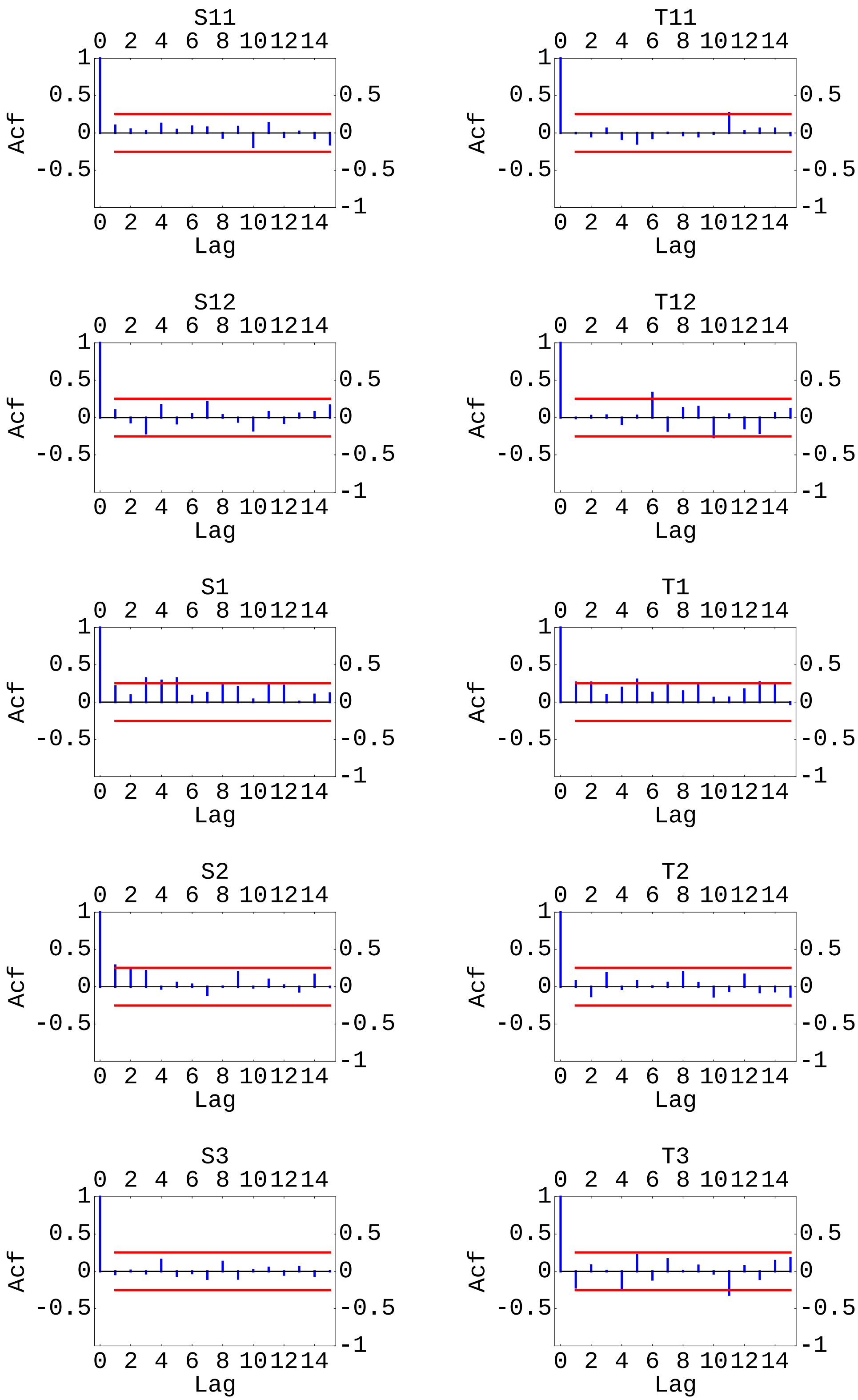
where N_t is the error term. Based on the pre-intervention data we assume initially that N_t is normal and independent, so ordinary multiple linear regression can be used. The intervention series are defined by,

$$\xi_t = \begin{cases} 0 & t < 29 \\ 1 & t \geq 29 \end{cases}$$

The term N_t represents the disturbance or error term and it has been tentatively identified as Gaussian white noise, that is $N_t = a_t$, where $a_t \sim \mathrm{NID}(0,\ \sigma^2)$.

Autocorrelation Function

There does not seem to be any strong autocorrelations at either lag one or lag twelve across the entire series, $n = 60$. So we can model the intervention using regression and check that there is no autocorrelation remaining in the residuals.


S11
T11
S12
T12
S1
T1
S2
T2
S3
T3
Acf
Lag
0 2 4 6 8 101214
1
0.5
0
-0.5
-1

Parameter Estimates

S11		Estimate	SE	TStat	PValue
	1	3.07143	0.425174	7.22393	1.23247×10^{-9}
	ξ	−0.665179	0.582194	−1.14254	0.257925

S12		Estimate	SE	TStat	PValue
	1	4.32143	0.57989	7.45215	5.09419×10^{-10}
	ξ	**−1.57143**	**0.794048**	**−1.97901**	**0.0525691**

S1		Estimate	SE	TStat	PValue
	1	10.0714	0.727315	13.8474	0.
	ξ	**−4.41518**	**0.995918**	**−4.43328**	**0.0000418896**

S2		Estimate	SE	TStat	PValue
	1	3.78571	0.622143	6.08495	9.86758×10^{-8}
	ξ	**2.15179**	**0.851905**	**2.52585**	**0.0142952**

S3		Estimate	SE	TStat	PValue
	1	1.42857	0.35514	4.02256	0.000168637
	ξ	**1.04018**	**0.486295**	**2.13899**	**0.0366585**

T11		Estimate	SE	TStat	PValue
	1	3.28571	0.347926	9.44372	2.47802×10^{-13}
	ξ	−0.598214	0.476417	−1.25565	0.214277

T12		Estimate	SE	TStat	PValue
	1	5.46429	0.499053	10.9493	8.88178×10^{-16}
	ξ	−0.839286	0.683357	−1.22818	0.22434

T1		Estimate	SE	TStat	PValue
	1	11.5	0.848245	13.5574	0.
	ξ	**−6.375**	**1.16151**	**−5.48856**	**9.33453×10^{-7}**

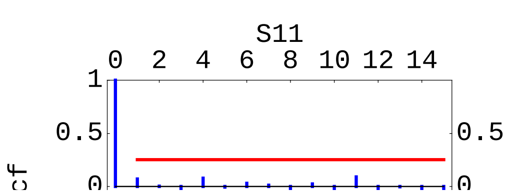
T2		Estimate	SE	TStat	PValue
	1	6.10714	0.7828	7.80167	1.3176×10^{-10}
	ξ	**3.17411**	**1.07189**	**2.96122**	**0.00443452**

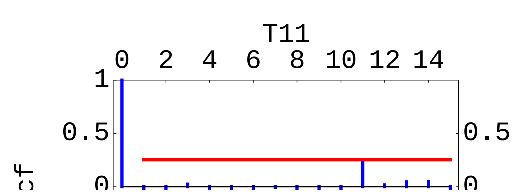
T3		Estimate	SE	TStat	PValue
	1	2.53571	0.375663	6.74997	7.69926×10^{-9}
	ξ	0.620536	0.514398	1.20633	0.232586

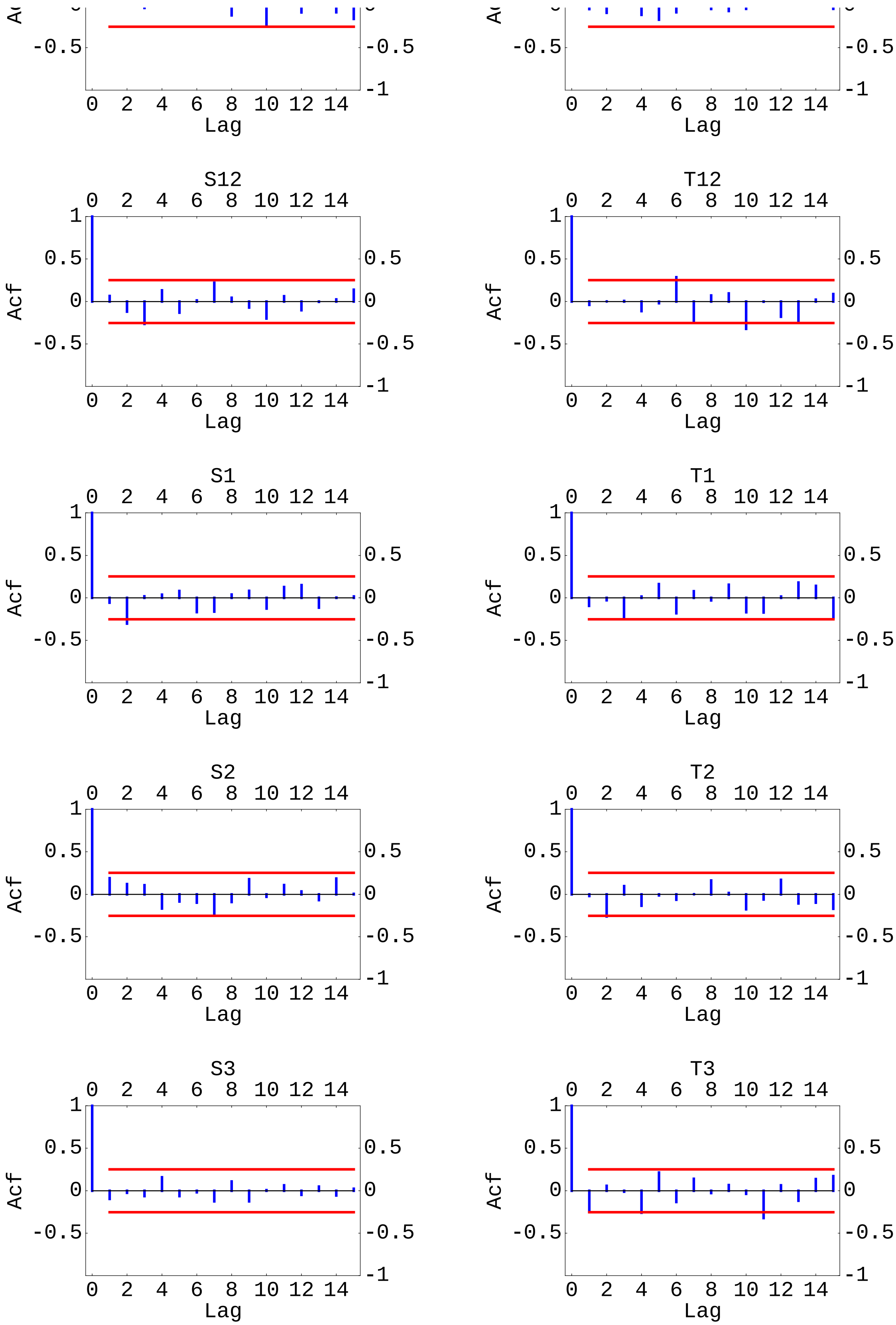
Using a one-sided test, the intervention parameter is significant at <5% for S12, S1, S2, S3, T1 and T12.

Residual Autocorrelation Analysis

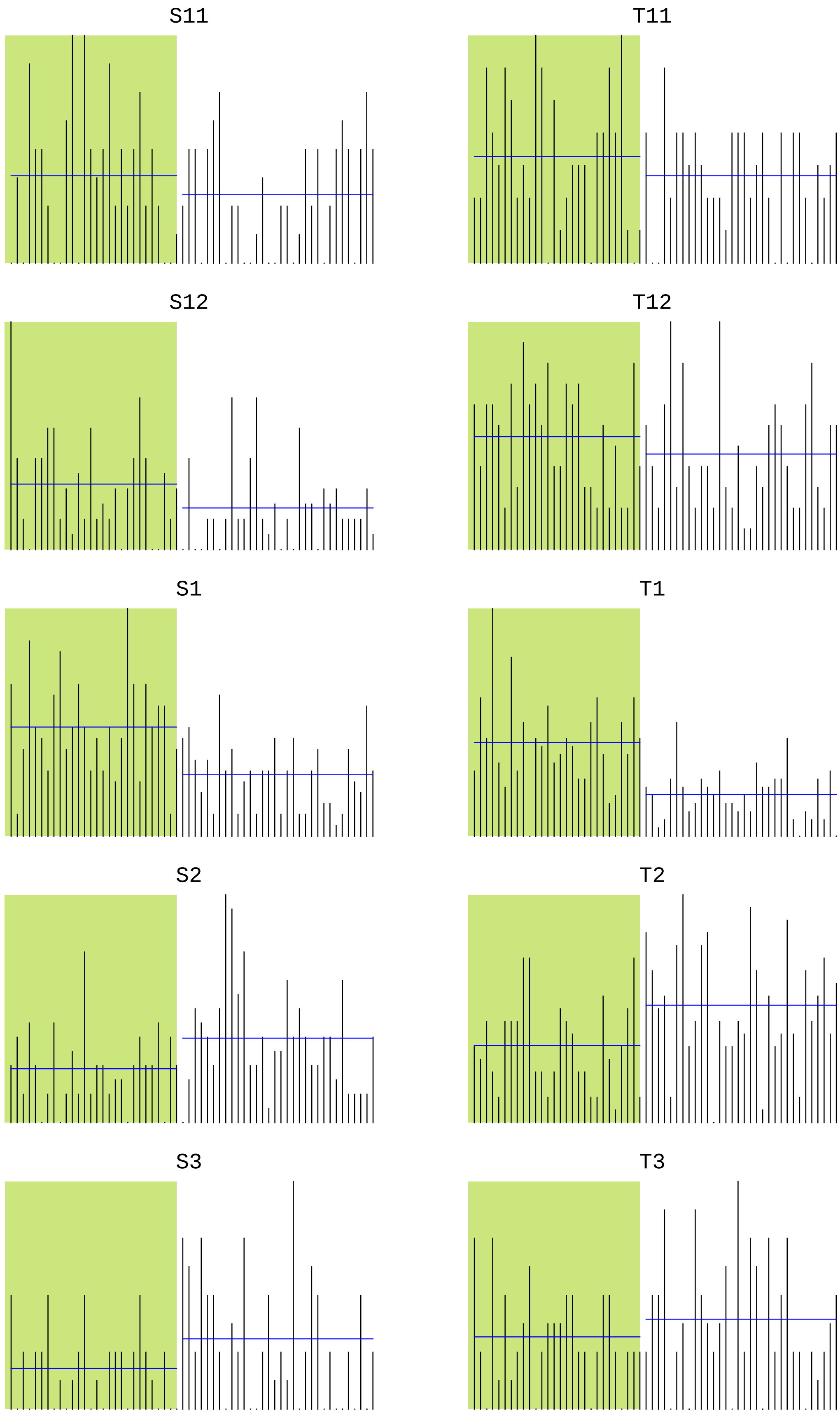
The autocorrelation in the residuals is examined particularly at lag one and lag twelve. Everything looks reasonable except perhaps for T3 has a large negative autocorrelation at lag one.







■ Visualization


S11
T11
S12
T12
S1
T1
S2
T2
S3
T3

Additional Fitting

T3 ARIMA (0, 0, 1)

The residual autocorrelation analysis indicated a negative lag one autocorrelation for the T3 series. Refitting with an ARIMA(0,0,1) error term, the intervention is now statistically significant at 5% on a one-sided test.

	Estimate	SE	P-Value
μ	2.505	0.18	0
δ	**0.647**	**0.366**	**0.078**
θ_1	0.262	0.125	0.036