

## **An Analysis of the Effects of Speed Limit Enforcement Cameras on Accident Rates**

Stephane Hess and John Polak  
Centre for Transport Studies  
Imperial College London

Stephane Hess (corresponding author)  
Centre for Transport Studies  
Department of Civil and Environmental Engineering  
Imperial College London  
Exhibition Road  
London SW7 2AZ  
United Kingdom  
Tel: +44(0)20 7594-6105  
Fax: +44(0)20 7594-6102  
[stephane.hess@imperial.ac.uk](mailto:stephane.hess@imperial.ac.uk)

John Polak  
Centre for Transport Studies  
Department of Civil and Environmental Engineering  
Imperial College London  
Exhibition Road  
London SW7 2AZ  
United Kingdom  
Tel: +44(0)20 7594-6089  
Fax: +44(0)20 7594-6102  
[j.polak@ic.ac.uk](mailto:j.polak@ic.ac.uk)

Word Count (including bibliography): 6,075 + 5 figures

### **ABSTRACT**

Speed Limit Enforcement Cameras (SLEC) have been in operation in Great Britain since 1991. However, there is still considerable dispute regarding their effectiveness in reducing accident rates. The aim of this research was to analyse the effects of SLECs on accident rates in Cambridgeshire, UK, using time series data collected over an 11 year period. A time series analysis of the accident data revealed the presence of both trend and seasonality components. A method was developed to remove the influence of these two components from the data and compare mean accident levels before and after installation of the camera. The method was also constructed in such a way that it would be able to distinguish between the actual effects of the camera installation and the effects of regression to mean. The initial investigation into the effects of SLECs showed an average decrease over sites in the monthly accident frequency by around 18%, a more detailed analysis suggested that the best approximation of the effect of the introduction of a SLEC is a decrease in injury accidents by 31.26%, thus giving clear evidence that SLECs do indeed contribute to a significant decrease in accident numbers.

## 1. INTRODUCTION

One of the most controversial recent road safety innovations in the UK has been the introduction of speed limit enforcement cameras (SLECs). These were introduced in 1991 as a means of assisting in the enforcement of speed limits, especially on rural roads. Motorists recorded by the cameras as exceeding the posted speed limit incur a fine. Although SLECs were popular with highway authorities, they are sometimes criticised by motoring organisations and the public, who see them as a covert means of raising revenue (1).

Initially the number of cameras was limited due to the high cost of installation, which had to be met in full by the highway authorities. However, a cost benefit analysis undertaken in 1996 (2) concluded that, if consideration is given to the full benefits to society, including the health service, SLECs generate benefits equivalent to roughly 5 times their cost in the first year of operation alone.

Following this report, in 1998, the UK Government decided to allow local police forces to use the fine income from speed and red light cameras to fund additional camera installations. The scheme was first introduced in April 2000 in eight police force areas known as partnerships. Certain criteria were set out for a site to be eligible for a camera in this "hypothecation" scheme. There is also a requirement that the authority monitors the effects of the camera on speed and accidents. For a camera to be used in this scheme it has to be installed at an accident hotspot. The stated aim of the camera installations is to reduce accident rates, not to punish drivers. To this end, camera locations are widely publicised, the camera housing is painted bright yellow and signing of cameras is provided within a radius of 1km of the camera locations.

A recent study (3) of the first year of operation of the new scheme conducted by the UK Department for Transport (DFT) concluded that SLECs reduced collisions at accident hotspots by 35% and the number of persons killed and seriously injured (KSI) by 47% with corresponding reductions in the wider partnership area of 6% and 18% respectively. However, this research was largely descriptive and did not attempt to explicitly identify the potential impacts of trend, seasonal and other factors and consequently there remains considerable controversy regarding both the existence and the magnitude of the effects of SLECs.

## 2. OBJECTIVES, DATA SOURCES AND PRELIMINARY ANALYSIS

### 2.1 Study Objectives

The aim of this study was to investigate the effects of the introduction of SLECs in one particular shire county of the UK (Cambridgeshire), and to compare these effects to the results produced by the DFT research. A particular concern was to control for seasonal and trend effects and (given the selection criteria of sites) for the possible effects of regression to mean. The research was carried out in collaboration with Cambridgeshire County Council and the Cambridgeshire Safety Camera Partnership (Cambridgeshire Constabulary).

### 2.2 Data Sources

The main data set used in the analysis contains information on the 31,042 injury accidents that were recorded in Cambridgeshire during the period from 1990 to 2001, with a detailed listing for every single accident. This data is collected by police officers completing a 'Stats 19' form at the scene of the accident. Accidents are classified into three degrees of severity (slight, serious and fatal injury) and the database also contains information on time and date, the accident site location and the site reference number (a site is generally defined as a stretch of road with a length of 1000 meters). The site reference number enables a division of the data into two classes, non-camera sites and camera sites. The reliability of the information contained in the Stats 19 data is increased by adjustment for under-reporting through comparison with other data sources, like hospital records for example. The number of camera sites used in this analysis was 43, this is significantly higher than the number used in most previous studies.

Speed survey data was made available by Cambridgeshire Police. This included results from speed surveys at camera sites before and after the installation of the camera, as well as some limited data on speed surveys in the surrounding areas during the same time periods.

This research looked at the number of injury accidents rather than the total number of accidents or the number of injuries (except where otherwise stated), thus treating an accident with one fatality in the same way as an accident with two or more fatalities and also ignoring non-injury accidents. This makes the division of the data into three main categories (and hence also the modelling) more straightforward and arguably is a more appropriate way of representing the data than a method that treats multiple fatalities as multiple fatal accidents. This is also the method used by Cambridgeshire County Council in its analysis of accident data.

### 2.3 Preliminary Analysis

A preliminary analysis of the data from speed surveys at camera sites before and after the installation of the camera showed a significant decrease in average vehicle speed as well as in the percentage of vehicles exceeding the posted speed limit, the percentage of vehicles exceeding the camera threshold and the 85%tile of vehicle speeds. These results are consistent with the hypothesis that the SLECs are associated with a reduction of speed and hence, given the well established link between speed and road safety, a reduction in accident incidence and severity, but do not allow us to discriminate between the effects of the SLECs *per se*, and such factors as trend and seasonality.

In addition to the decreases in speed at the camera sites, the preliminary analysis also showed that there were similar, but less significant decreases in speed in the surrounding areas. Speed in the rest of the network however seems to have stayed approximately constant. If this had not been the case it would have meant that the decreases observed at camera sites and in their neighbourhood were not necessarily linked to the installation of the camera, but could have been linked to the overall changes.

The initial analysis also showed the strong influence of seasonality on accident data, calling for extra care in the analysis, something that been ignored in virtually all prior research projects.

## 3. MODELLING OF ACCIDENT DATA

### 3.1 ARIMA/SARIMA Model

To better understand the time dependent components of the accident data, a time series model was fitted to the combined data set of all injury accidents from all locations, meaning that no distinction could be made between accidents of different injury severities. The reason for this decision was the relatively low monthly frequency of fatal injury accidents at single accident sites, leading to a lot of random variation. The use of the combined accident data gave a very stable time series, with trend and seasonality factors that were relatively easy to estimate. Experiments using different weights for different levels of severity produced inconsistent modelling results, equal weights were thus used for the different types of accidents.

A number of time series approaches were compared including ARIMA (autoregressive, integrative, moving average) and SARIMA (seasonal autoregressive, integrative, moving average) processes, as well as exponential smoothing and the Holt-Winters methods. ARIMA/SARIMA models emerged as the preferred approach. Not only does ARIMA/SARIMA tend to outperform the Holt-Winters method in the precision of its seasonal indices, but it is also a better tool for forecasting; this becomes important in the validation process. Figure 1 shows a plot of the monthly injury accident numbers for Cambridgeshire from 1990 to 2000, summing over the three degrees of severity. The data for the year 2001 had only become available after the initial analysis had been completed and was used mainly for model validation by comparing forecasts to observations.

The implementation of the ARIMA/SARIMA model proceeds as follows. We define  $\{Z_t\}$  to be a time series, where  $Z_t$  is the observation at time  $t$  and define  $a_t$  (error term) to be a white noise process,  $a_t \sim N(0, \sigma^2)$ . The backward shift operator  $B$  is defined such that  $B^k Z_t = Z_{t-k}$  and the difference operator  $\nabla$  is defined so that  $\nabla_k = 1 - B^k$ .

As a tool for identifying a suitable model, the ACF (auto-correlation function) and PACF (partial auto-correlation function) for the combined accident data were plotted at different degrees of differencing (figure 2). The plots in the first row of the figure show the ACF and PACF before differencing the data. The ACF follows a cosine curve and fails to die out, suggesting that we should difference at least once non-seasonally. The results of the differencing are clearly visible in the two plots in the second row. However, the ACF is still significant mainly at multiples of the seasonal period (12 months), suggesting we should try differencing seasonally as well as non-seasonally. The plots in the final row show an ACF that is significant at lags 1, 12 and 13 (and less importantly at lags 26 and 27). This is consistent with the ACF of the model  $W_t = (1 - \theta B)(1 - \Theta B^{12})a_t$ , which has significant autocorrelations at lags 1, 12 and 13. The PACF is not consistent with any low order SARIMA model (applying the Principle of Parsimony we would usually look for an initial model with  $p + q \leq 2$ ). Our initial model will thus be SARIMA (0,1,1)x(0,1,1)<sub>12</sub>.

The data has been differenced non-seasonally as well as seasonally, we have thus applied  $\nabla$  and  $\nabla_{12}$  to the data, and if  $W_t$  is our model at time  $t$ , we get:

$$W_t = \nabla \nabla_{12} Z_t$$

From  $W_t = (1 - \theta B)(1 - \Theta B^{12})a_t$  (by definition of the SARIMA (0,1,1) $\times$ (0,1,1)<sub>12</sub> model), we get:

$$\begin{aligned}\nabla\nabla_{12}Z_t &= (1 - \theta B)(1 - \Theta B^{12})a_t \\ (1 - B)(1 - B^{12})Z_t &= (1 - \theta B)(1 - \Theta B^{12})a_t \\ Z_t - Z_{t-1} - Z_{t-12} + Z_{t-13} &= a_t - \theta a_{t-1} - \Theta a_{t-12} + \theta \Theta a_{t-13} \\ Z_t &= Z_{t-1} + Z_{t-12} - Z_{t-13} + a_t - \theta a_{t-1} - \Theta a_{t-12} + \theta \Theta a_{t-13}\end{aligned}$$

Here  $\theta$  and  $\Theta$  are the only two parameters to be estimated, there is no need to include a constant in this model. Indeed the mean after differencing will generally be sufficiently close to zero to make any constant insignificant.

This model renders the following values for the parameters, with standard errors shown in brackets:  $\theta = 0.8273(0.0508)$  and  $\Theta = 0.8708(0.0707)$ . The mean absolute percentage error (MAPE) for this model is 8.14284, showing an average deviation of the fits from the observations by 8.14%. Attempts were made to overfit the model by adding one of the possible extra parameters. None of the models resulting from this procedure leads to an improvement in the fit. The final model will thus be the SARIMA (0,1,1) $\times$ (0,1,1)<sub>12</sub>.

### 3.2 Residual Analysis

A detailed residual analysis was carried out on the fits of the model, beginning with an analysis of the residual plots that are shown in figure 3.

The first plot checks for autocorrelation of the residual errors. The results fall within the acceptable region of no more than one significant observation in every 20 (aside from the ACF at lag 0, which is 1 by definition). The quantile-quantile (QQ) plot shows that the residuals are approximately normally distributed, the greater spread in the top end shows that the residuals are right skewed, this is likely to have been caused by outliers which are months with a higher than expected number of accidents. The time series plot of the residuals shows that, except for some outliers, they can be regarded as being roughly white noise (where the outlier for February 1996 accounts for most of the skewedness observed in the second plot, there were 218 accidents in that month, the mean number for February is 185). Although the histogram also shows that the data is right skewed, it can be assumed that the residuals are normally distributed.

The Box-Ljung statistic is defined as:

$$Q(K) = n(n+2) \sum_{k=1}^K \frac{r_k^2}{n-k},$$

where  $r_k$  is the autocorrelation of the errors at lag  $k$ , and  $K$  is the number of lags used, which should be around  $\frac{1}{4}$  of the sample size, in this case 33. If the model has been correctly specified, then  $Q(K) \sim \chi_{\nu=K}^2$  as  $n \rightarrow \infty$ .

The calculation of the Box-Ljung statistic showed that that none of the autocorrelations is significantly different from zero, the autocorrelation at lag 20 lies outside the limits, but, coming at such a high lag, this outlier can be disregarded.

### 3.3 Model Validation by Forecasting

The next step was concerned with model validation, which compared the forecasts produced by the model to the actual observations. This process showed that the forecasts produced by the model are very accurate, the MAPE for the forecasts for the year 2001 is 7.8648, thus even lower than the MAPE comparing fits to observations in the fitting period (1990-2000).

## 4. ANALYSIS OF THE EFFECTS OF REGRESSION TO MEAN

This section looks at the role of regression to mean in the analysis of time specific data. Regression to mean, the statistical phenomenon of a time series returning to its mean level, can partially or fully explain the fluctuations in the level of a time series and should thus always be given some consideration in the analysis.

#### 4.1 Regression to Mean in the Context of Accident Data Analysis

In the context of this analysis the notion of regression to mean means that if we observe a higher than expected number of accidents in one month, it is likely that the observation for the next month will be closer to the expected level. Hence, if such an outlying observation results in the installation of a SLEC at the specific site, a drop in accident numbers in the months following the installation is a likely event as the level of the time series returns to its mean. This drop may thus not be due to the installation of the camera but rather to regression to mean.

#### 4.2 Existing Methods of Detection

There are numerous methods that can be used to detect the effects of regression to mean and to see whether a certain change in the level of a time series is due to an intervention, or simply due to regression to mean. The best-known method is probably the extended version of the empirical Bayes model (4), which is however inapplicable in this case as there is insufficient data to estimate the true underlying accident rate needed for this method to work.

An approach that is simpler to apply for single sites was described by Abbess *et al* (5). This method uses accident data from sites similar to the treated site over the same period, and calculates the biggest explainable regression to mean effect for the treated site for a one-year period. This method does not take into account seasonality and trend, but as it looks at annual data, the effects should be almost completely independent of seasonality. This method was used for the individual camera sites and the effects were averaged over all sites. Although it was sometimes very difficult to find sites similar to the camera site under investigation, the method still produced a result, namely a maximum possible regression to mean effect of 9.69%. The mean change observed over a one-year period is a decrease by around 19% in injury accident frequency. This reduction in frequency can thus not be fully explained by regression to mean. However, due to the limitations of the data set, a new method had to be developed to further analyse the influence of regression to mean on the changes in accident numbers at camera sites.

#### 4.3 New Method

The nature of the phenomenon means that regression to mean can only really be observed over a long time horizon. Similarly, we can also only disregard the effects of regression to mean by considering a time series over a longer time period.

One way to do this is to consider the mean level of the data before the intervention, in this case the installation of the camera, and to compare this to the mean level of the data following the intervention. These should ideally be long term prior and posterior mean levels, but the length of the prior interval especially is important as this will provide us with a stable estimate of the natural mean level of the time series. However, this approach does not take into account variation in the data caused by time specific effects, mainly trend and seasonality. This method should thus be used on a data set from which these effects have been removed. The method used to do this will be introduced in the section concerned with the estimation of the effects of the SLECs.

In order to apply this idea we need to assume that (A1) we have sufficient prior and posterior data to enable us to produce good estimates of the corresponding long term means and that, (A2) after controlling for trend and seasonality, there are no additional external factors acting on the data.

Figure 4 illustrates three possible scenarios. The plots show a long (infinite) horizon time series plot of the data, with an intervention at some point in time (roughly in the middle of the series), and the prior and posterior mean levels. In the first two plots, the intervention takes place at the peak of the pulse in the data, whereas in the third plot the intervention takes place at an arbitrarily chosen point in time.

In the first plot, the data returns to the level it was at before the pulse in the data. The (long term) mean levels before and after the intervention are exactly equal, meaning that in this case there is no support for the hypothesis that the decrease is due to the intervention.

In the second plot, the data decreases to a level below the stable level before the intervention. The (long term) prior and posterior mean levels are clearly different, and in this case, the decrease can thus not be fully explained by the effects of regression to mean. Roughly we could assume that the decrease of the data to the old stable state is due to regression to mean and that any additional decrease is due to the effects of the intervention.

The final plot shows the situation where we observe a decrease after an intervention, where the level of the data prior to the intervention was constant. In this case the decrease of the mean level cannot be explained by regression to mean. Indeed, assuming an infinite time horizon, the data prior to the intervention was in a stable state and the time series reaches a new stable state after the intervention.

In each case, the difference between the prior and posterior long term mean provides a suitable estimate of the effect of the intervention, net of the effect of regression to the mean.

It will not always be possible to find appropriate long term mean levels by direct analysis of the data. However, if we assume that trend and seasonality account for almost all of the variation in the data, the removal of the effects of these components will give us estimates of the stable levels of the time series before and after the intervention. These stable levels can then be used as estimators of the long term mean levels. Special care should still be taken to guarantee that the prior and posterior periods are long enough (at least 12 months of data).

## 5. ESTIMATING THE EFFECTS OF THE CAMERA INSTALLATION

The aim of this section is to analyse the effects that the installation of SLECs has on the safety record at camera sites and in the surrounding areas. This section puts into use the knowledge we have acquired about the data in the first half of the project, especially the characteristics of the underlying time series.

It was recognised at an early stage that it would not be possible to use any SARIMA based intervention analysis methods on the accident data for single camera sites, where the average number of accidents per site over a twelve year period (1990-2001) was less than 40 accidents, thus less than 1 (0.28) accident per month. The models fitted in section 3 looked at the total combined accident count for Cambridgeshire, some 3000 injury accidents per year, the total count over all camera sites was an average annual number of around 140 accidents per year. Clearly there was no way to fit an ARIMA/SARIMA model for single camera sites, as there would have been too much fluctuation in the data to fit a useful model. The same problem applied to the data set combining accidents over all camera sites. Another method had to be found.

### 5.1 Requirements of new method

The new method had to meet the following requirements. It should:

- look at monthly accident data, indeed, when using annual data, it would be very difficult to specify the start of the effects after introduction
- use sufficiently large data sets to be able to discard random variation
- be able to distinguish between the effects of the SLEC and other effects like seasonality and trend.

The first two points contradict each other, since in the present context it is not possible to identify monthly data that are large enough to discard random variation. The solution to this problem is to combine the single sites into a group of sites, however, due to different installation dates, time-shifting has to be used to make the dates coincide, and any time specific effects will thus have to be removed beforehand.

### 5.2 Detrending and Deseasonalising

Apart from an irregular component, the two main components of the accident data are seasonality and trend. It was recognised straightaway that smoothing could not be used to remove seasonality and trend from the data as this could introduce artificial cycles as well as put us at risk of losing some of the information contained in the data. Also, the new method should be reversible, such that after the analysis, one should be able to apply the inverse transformation to the data and the forecasts so as to get back to the original data and obtain forecasts that reflect the trend and seasonality observed in the original data. Two assumptions were made in the calculation of the seasonality and trend indices:

- The seasonal indices are assumed to be constant from year to year, however, as they are multiplicative indices, the variation for years with higher data will be more important.
- The trend is assumed to be a constant decrease in accident numbers. This is a strong assumption as it ignores changes in trend over the 12 years of observation, but this is the only way to include a trend component in forecasting, and the assumption is also justified by the fact that as the decrease in the total number of accidents from 1990 to 2001 is only around 2%, the effect of trend (and also random fluctuation) is not as important as are for example the effects of seasonality.

An analysis using a combination of Holt-Winters and SARIMA modelling was carried out to calculate the (multiplicative) seasonal indices ( $s_i, i=1, \dots, 12$ ), with  $\sum_{i=1}^{12} s_i = 12$ . This analysis showed that the months with the highest accident frequencies are January, November and December, those with the smallest frequencies are February, April and May. The analysis showed a decrease in deseasonalised accident frequency from January 1990 to Dec 2001 by 2.2% (0.015% per month), found by estimating the trend from 1991 to 2001 and backforecasting for the year 1990, the inclusion of which would have biased the results (see figure 1, 1990 was

the wettest year in England and Wales since 1872). We let  $Y_i$  be the trend level at time  $i$ , and  $Y_1$  be the initial trend level, and we set  $p_i = \frac{Y_i}{Y_1}$ .

Then, by assuming trend to be linear and normalizing by setting the trend for January 1990 to 1, the trend effect can be transformed into a multiplicative effect.

The data at time point  $i$ ,  $X_i$  can now be detrended and deseasonalised, by:

$$\overline{X}_i = \frac{X_i}{s \cdot \text{MOD}_{12}^i \cdot p_i} \text{ where } \text{MOD}_{12}^i \text{ is the remainder of the division } \frac{i}{12}, \text{ and } s_0 = s_{12}.$$

If we now subtract from each  $\overline{X}_i$  the mean of the detrended and deseasonalised data,  $\frac{\sum_{i=1}^n \overline{X}_i}{n}$ , where  $n$  is the length of the data set, we get a series with mean equal to zero. This process will be referred to as differencing.

The resulting time series is, except for some outliers (and notably the year 1990) very similar to a white noise process, something that can be established by a runs test on the data, which shows a probability of 0.53 of the data being purely random, which is quite high considering the inclusion of the year 1990, with its clear decreasing trend (still present even after applying this procedure).

Figure 5 shows the effects of this procedure, on the fits of the model and on the actual data. The effects on the fits will necessarily produce a flat line at zero, as the coefficients were calculated from the model producing these very fits.

If the behaviour of the data for a specific site is similar to the overall behaviour of the data for all sites, the resulting series (after detrending, deseasonalising and differencing) should be very similar to a white noise process. Slight deviations from this do not necessarily mean that the series is different from other sites, some marginal difference may be explained by the fact that the indices have not explained all the variation in the original data and that the original models did not give a perfect fit to the data.

### 5.3 Time-Shifting

Time-shifting is used when combining data from single camera sites into one group. After detrending and deseasonalising the data, it could be assumed that any components other than the effect of the camera installation and a remaining irregular component had been removed from the data. This meant that the data for the single camera sites could be grouped together into one single data set by shifting the data so that the installation dates coincided. The shifting process did however involve the loss of some prior data for sites that were commissioned late and some posterior data for sites that were commissioned early. This was necessary as the grouped data could only cover a period of time where data was available from all sites in the group, but obviously the prior and posterior periods for sites with different installation dates are of different lengths.

A plot of the grouped and time-shifted data was then produced to check whether any decrease in the accident data was visible after introduction of the cameras. However, the posterior period was now very short (length equal to the length of the posterior period for the last camera to be installed), so that although the plot showed an apparent decrease, some other method had to be developed.

## 5.4 Calculation of Effects

### 5.4.1 Change in Total Detrended and Deseasonalised Accident Numbers

The change in the mean level of monthly prior and posterior detrended and deseasonalised (DD) accident frequency data was calculated for each camera site. Most of these changes showed a decrease in the monthly mean level of accidents. In order to compute the overall effect across all sites, the prior and posterior long-term mean accident frequency levels were initially simply summed across all sites and the difference of these two totals was calculated. This resulted in an estimate of a reduction in injury accident frequency by 31.26%. Although most individual sites showed reductions in accident frequency, the proportionate change at individual sites showed considerable variation ranging from -100% to +336% (where the site with the 336% increase was a site with a very low accident frequency of only around 1 accident per year prior to the installation of the SLEC, which was installed not because of an accident blackspot but because of persistent speeding offences). The average proportionate effect across all sites was -15.17% with a 95% confidence interval (truncated at -100%) of

[-100%,97%], which, although very wide, does suggest that the effect at specific sites is more likely to be a decrease than an increase.

#### 5.4.2 Changes for Individual Sites and Weighting Strategies

One problem with simply averaging effects across different sites is that equal weight is given to all sites. However, the data for the sites with longer periods of posterior runs and/or higher average monthly accidents levels will in general be more reliable. To illustrate this effect, the estimate of the effect derived from averaging across all sites (i.e., -15.17%) was used to forecast the effects at each one of the 43 camera sites. By comparing these 'forecasts' to the effects actually observed on the single sites, the sum of squared errors (SSE) was calculated for the forecasts over all sites. This was found to be equal to 714.44, with a total sum of squares (TSS) of 3298, and was especially high for some of the sites with the most credibility (relatively high prior monthly data levels and long posterior periods). Clearly more weight had to be given to these sites. A number of approaches were used to develop weights.

An initial set of weights were found by multiplying the prior monthly accident DD data level with the length of the posterior period. From this, the weighted average was calculated, giving an estimated reduction in accident numbers by 22.74% (standard deviation of 61.11). The use of this coefficient reduced the SSE from 714.44 to 599.67. The prior data level was deemed to be more important than the duration of the posterior period (sites with high data were not very prone to random fluctuation, making the conclusions more reliable, even over a shorter posterior period). An attempt to use the prior monthly accident DD data level for the weights resulted in an estimated reduction by 30.54% (standard deviation of 60.38), and the SSE was reduced to 552.06. Finally, a least squares method was used to find the coefficients that minimized the SSE, which was essentially equivalent to reducing the squared error for the sites with the highest credibility. This gave a mean effect of -31.79% (this method did not provide a standard error), and reduced the SSE only marginally further to 551.14.

This iterative process has thus decreased the SSE from 714.44 to 551.14. Clearly this final coefficient explains the overall effect far better than the simple average taken earlier on. This estimate is also very close to the initial estimate of -31.26% found in the calculation of the change in total detrended and deseasonalised accident numbers, thus increasing confidence in that earlier result.

The above discussion shows that sites with higher initial data levels should be given more weight in the calculation of coefficients (the weighted mean coefficient explains the overall changes far better than the simple average over all sites). An explanation for the difference between the different estimators is that of the ratio of means versus mean of ratios issue. The initial estimate of -31.26% derives from taking the ratio of two totals, namely the change in the total monthly DD accident count over all sites following the installation of the cameras. Basic algebra shows that this is equivalent to the ratio of the two means (prior and posterior DD). The second estimate of -15.17% is found by taking the mean of ratios, namely the average of the changes on individual sites. The difference between the two estimates is inevitable, when the units over which the averaging is being done are of significantly different size.

By weighting sites according to prior DD means in the averaging of changes (leading to the estimate of -30.54%), this difference is simply cancelled out, hence the similarity between the two estimates (-31.26% and -30.54%). The mean of ratios versus ratio of means issue is essentially a question of alternative estimators of change or effect. The relative size of the two estimators will depend amongst other things on the correlation structure in the data (e.g., if bigger sites have bigger improvements, then ratio of means will always be bigger than mean of ratios, which is the situation we have here).

The results from the SSE analysis which produced an estimate very close to the ratio of means estimate suggest that we should prefer this method to the mean of ratios approach. The fact that reliability of the data comes into play in this calculation, and the fact that the data from larger sites is deemed more reliable gives additional support to this decision. Indeed, the ratio of means approach does, by looking at the ratio of the averages over all sites (rather than the average of the ratios over all sites) indirectly assign different weights to sites with different levels of data (as higher values play a bigger role in the calculation of a mean). This minimizes the bias introduced by the inclusion in the calculation of data from sites with relatively low accident frequency and resulting high random fluctuation. The final estimate of the reduction in detrended and deseasonalised average monthly injury accident numbers will thus be a decrease by 31.26%.

## 6. COMPARISON WITH OTHER RESEARCH RESULTS

The principal studies against which comparisons can be made are the cost benefit analysis undertaken for the DFT in 1996 (2) and the more recent assessment of the first year of operation of the new SLEC regime, also carried out for the DFT (3). One of the problems in comparing the results of the current analysis to these earlier results is that little or no consideration was given to seasonality and trend in the earlier work.

The cost benefit study showed that across all sites, SLECs decreased casualties by an average 28%. The analysis of the effects on the new sites, which, unlike some older sites, were selected only because of their long-term accident record, showed an average decrease in collisions by 35%, compared to 28% on the old camera sites. More importantly the number of KSI accidents was found to have decreased by 47% (3). The total number of injury accidents has decreased by some 32%. These figures are broadly consistent with the estimates presented in section 5.

## **7. CONCLUSIONS AND RECOMMENDATIONS**

The research has shown that, after removing the effects of seasonality and trend, and independent of the influence of regression to the mean, there is a significant reduction in the mean level of monthly injury accident frequency following the installation of a SLEC. The analysis has shown that the actual effects observed on the camera sites in Cambridgeshire are best explained by a reduction in the detrended and deseasonalised monthly accident mean level by around 31%, giving clear evidence that the installation of a SLEC does indeed have a positive effect on the accident rates at the camera site.

## **ACKNOWLEDGEMENTS**

This paper is based on work undertaken by the first author in the Statistical Laboratory at the University of Cambridge. The authors would like to acknowledge the contribution of: Mr Bob Menzies, Mr Graham Amis and Mr Malcolm Mugridge, Cambridgeshire County Council; Sergeant Andy Chatfield, Safety Camera Project Manager, Cambridgeshire Constabulary; Mr Anthony Taylorson, Public Relations Manager, Cambridgeshire Safety Camera Partnership and Dr Alastair Young, Statistical Laboratory, Cambridge University.

**REFERENCES**

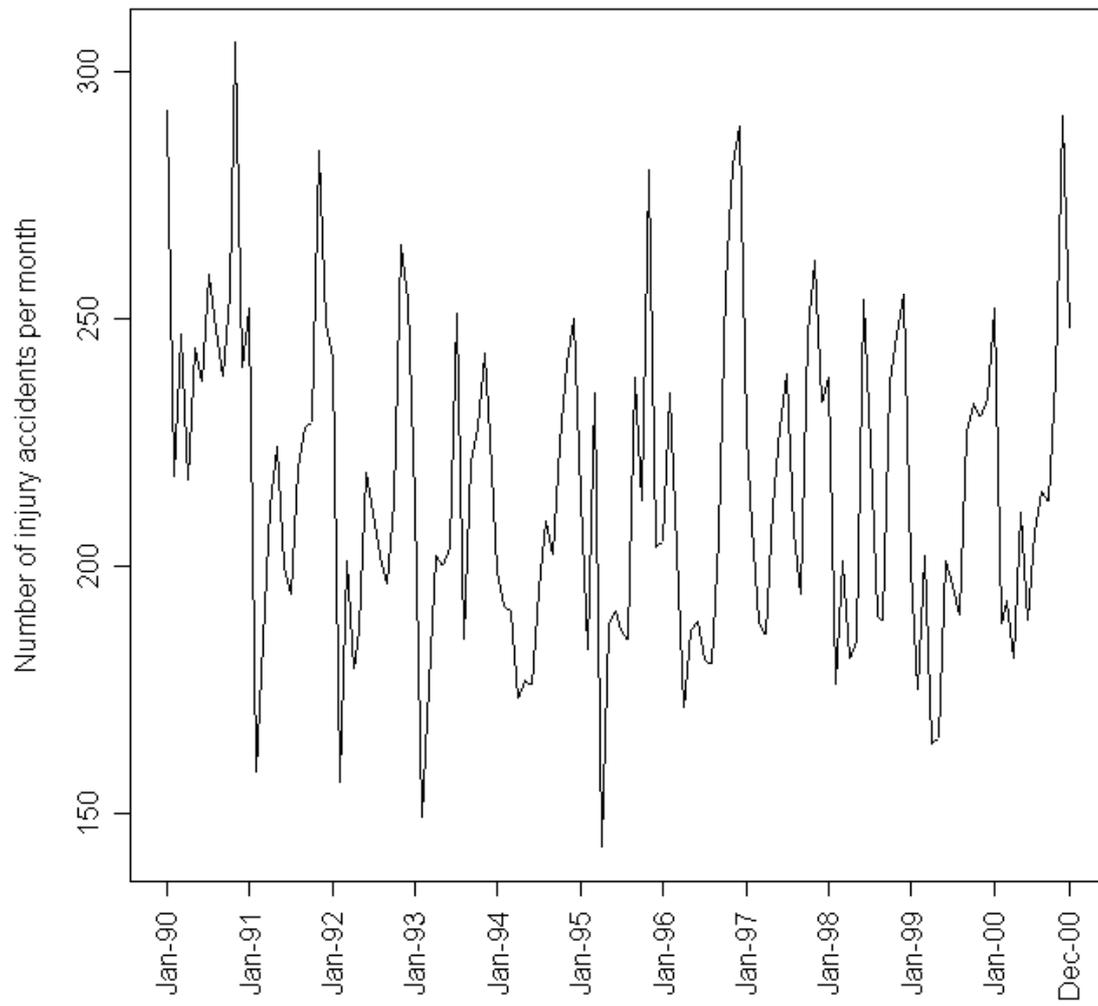
1. Corbett, C. and Simon, F. The effects of speed cameras: How drivers respond, DTLR Road Safety Division, 1999.
2. Hooke A., Knox J., Portas D., Cost benefit analysis of Speed Cameras, Police Research Series Paper 20, Police Research Group, Home Office, London, 1996.
3. PA Consulting, Cost Recovery System for Traffic Safety Cameras – First Year Report, DTLR (Department for Transport, Local Government and the Regions – now DFT) Road Safety Division, London, 2001.
4. Hauer, E., Observational before-after studies in road safety : estimating the effect of highway and traffic engineering measures on road safety. Elsevier Science, Oxford, 2001.
5. Abbess C., Jarrett J. and Wright C.C., Accidents at blackspots: estimating the effectiveness of remedial treatment, with special reference to the `regression-to-mean' effect. Traffic Engineering and Control, 22(10), 1981.

**TABLE OF CONTENTS**

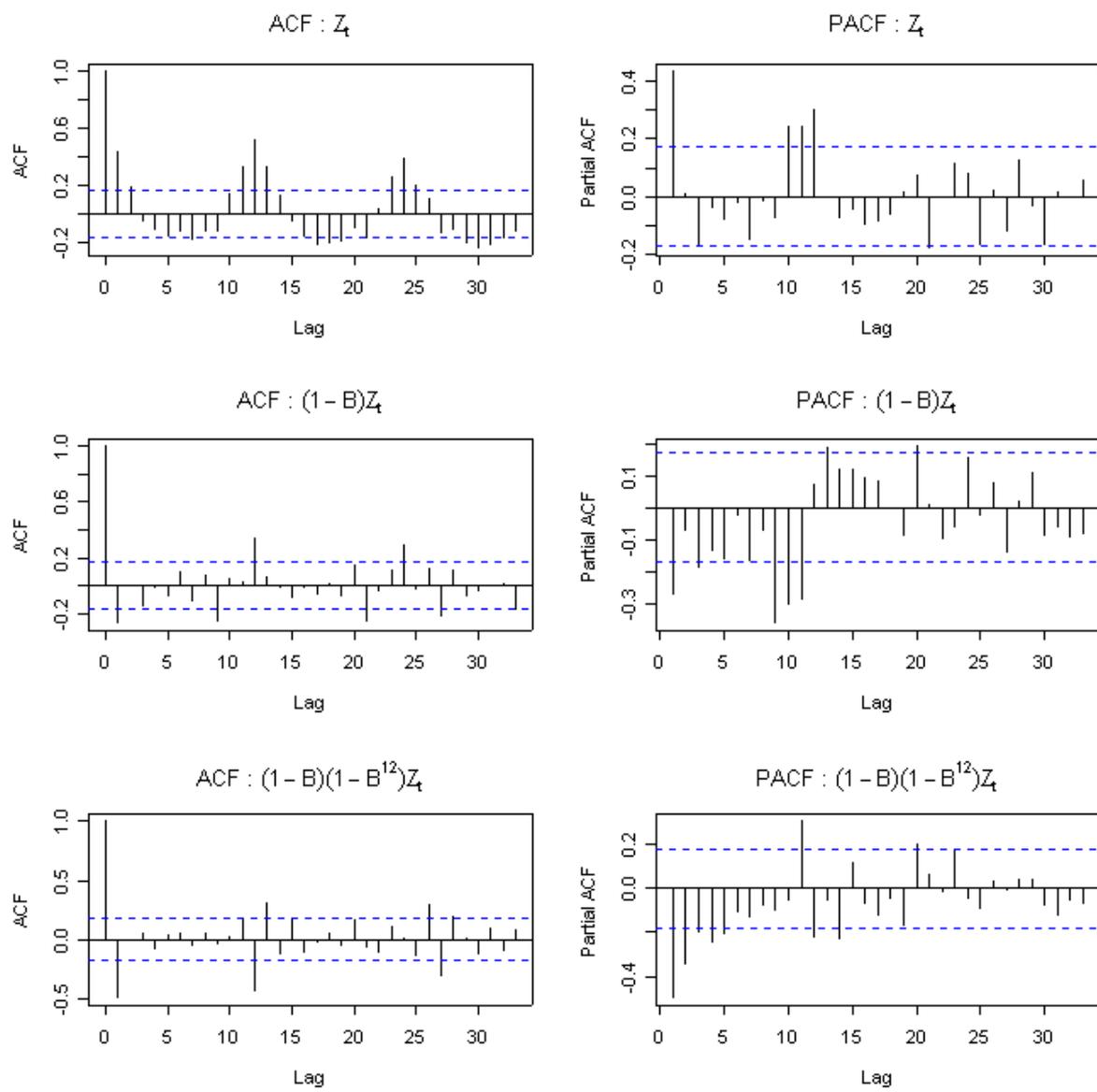
ABSTRACT .....	1
1. INTRODUCTION .....	2
2. OBJECTIVES, DATA SOURCES AND PRELIMINARY ANALYSIS .....	2
2.1 Study Objectives .....	2
2.2 Data Sources .....	2
2.3 Preliminary Analysis .....	3
3. MODELLING OF ACCIDENT DATA .....	3
3.1 ARIMA/SARIMA Model .....	3
3.2 Residual Analysis .....	4
3.3 Model Validation by Forecasting .....	4
4. ANALYSIS OF THE EFFECTS OF REGRESSION TO MEAN.....	4
4.1 Regression to Mean in the Context of Accident Data Analysis.....	5
4.2 Existing Methods of Detection .....	5
4.3 New Method .....	5
5. ESTIMATING THE EFFECTS OF THE CAMERA INSTALLATION .....	6
5.1 Requirements of new method .....	6
5.2 Detrending and Deseasonalising.....	6
5.3 Time-Shifting.....	7
5.4 Calculation of Effects .....	7
5.4.1 Change in Total Detrended and Deseasonalised Accident Numbers .....	7
5.4.2 Changes for Individual Sites and Weighting Strategies.....	8
6. COMPARISON WITH OTHER RESEARCH RESULTS .....	8
7. CONCLUSIONS AND RECOMMENDATIONS .....	9
ACKNOWLEDGEMENTS.....	9
REFERENCES .....	10
TABLE OF CONTENTS .....	11
TABLE OF FIGURES.....	11

**TABLE OF FIGURES**

FIGURE 1 : MONTHLY INJURY ACCIDENT DATA FOR CAMBRIDGESHIRE - 1990 TO 2000 .....	12
FIGURE 2 : ACF AND PACF FOR THE COMBINED CAMBRIDGESHIRE ACCIDENT DATA .....	13
FIGURE 3 : RESIDUAL ANALYSIS FOR THE SARIMA MODEL FOR COMBINED ACCIDENT DATA .....	14
FIGURE 4 : DETECTION OF REGRESSION TO MEAN.....	15
FIGURE 5 : EFFECTS OF DETRENDING, DESEASONALISING AND DIFFERENCING.....	16



**Figure 1 : Monthly injury accident data for Cambridgeshire - 1990 to 2000**



**Figure 2 : ACF and PACF for the combined Cambridgeshire accident data**

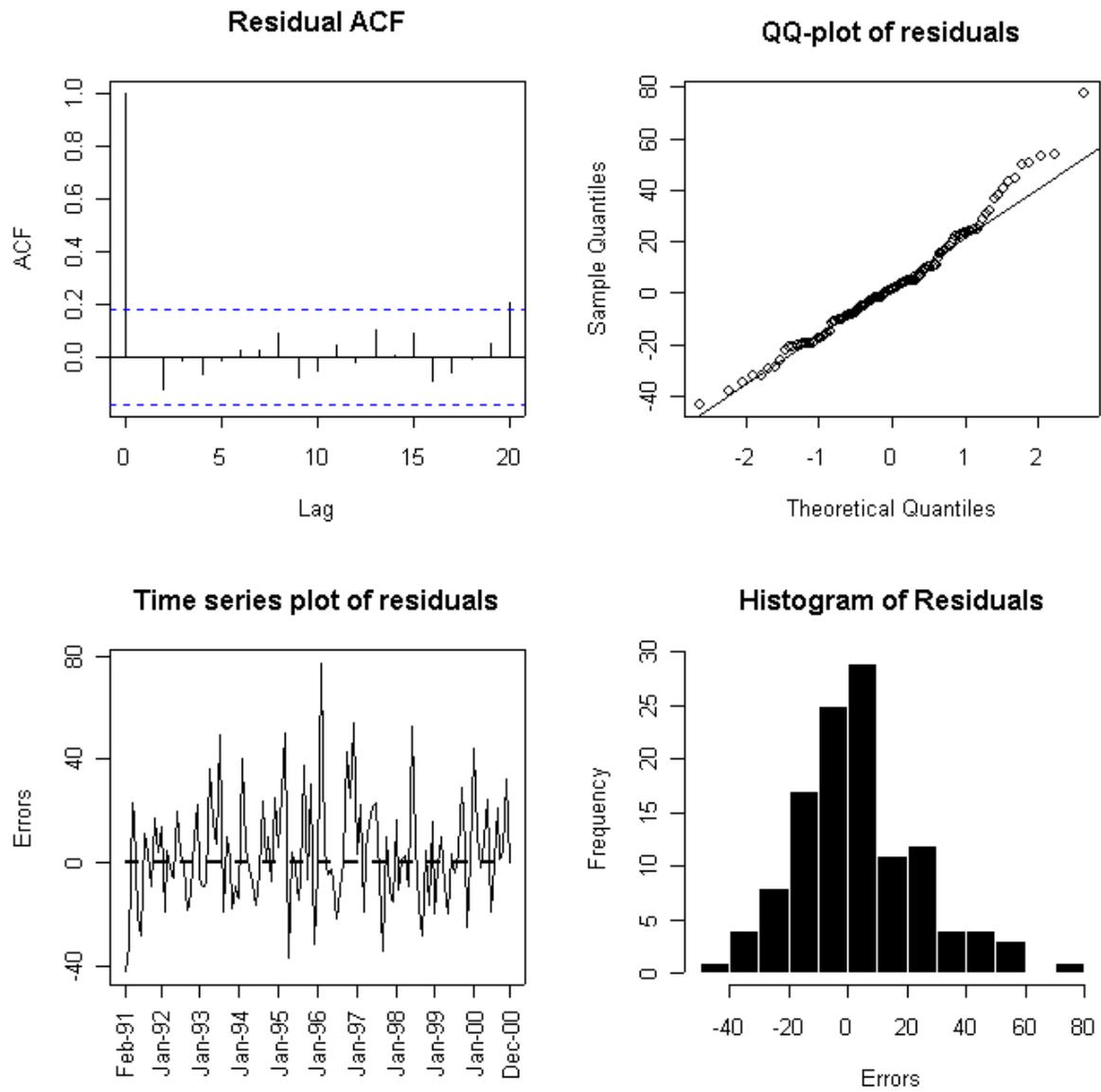
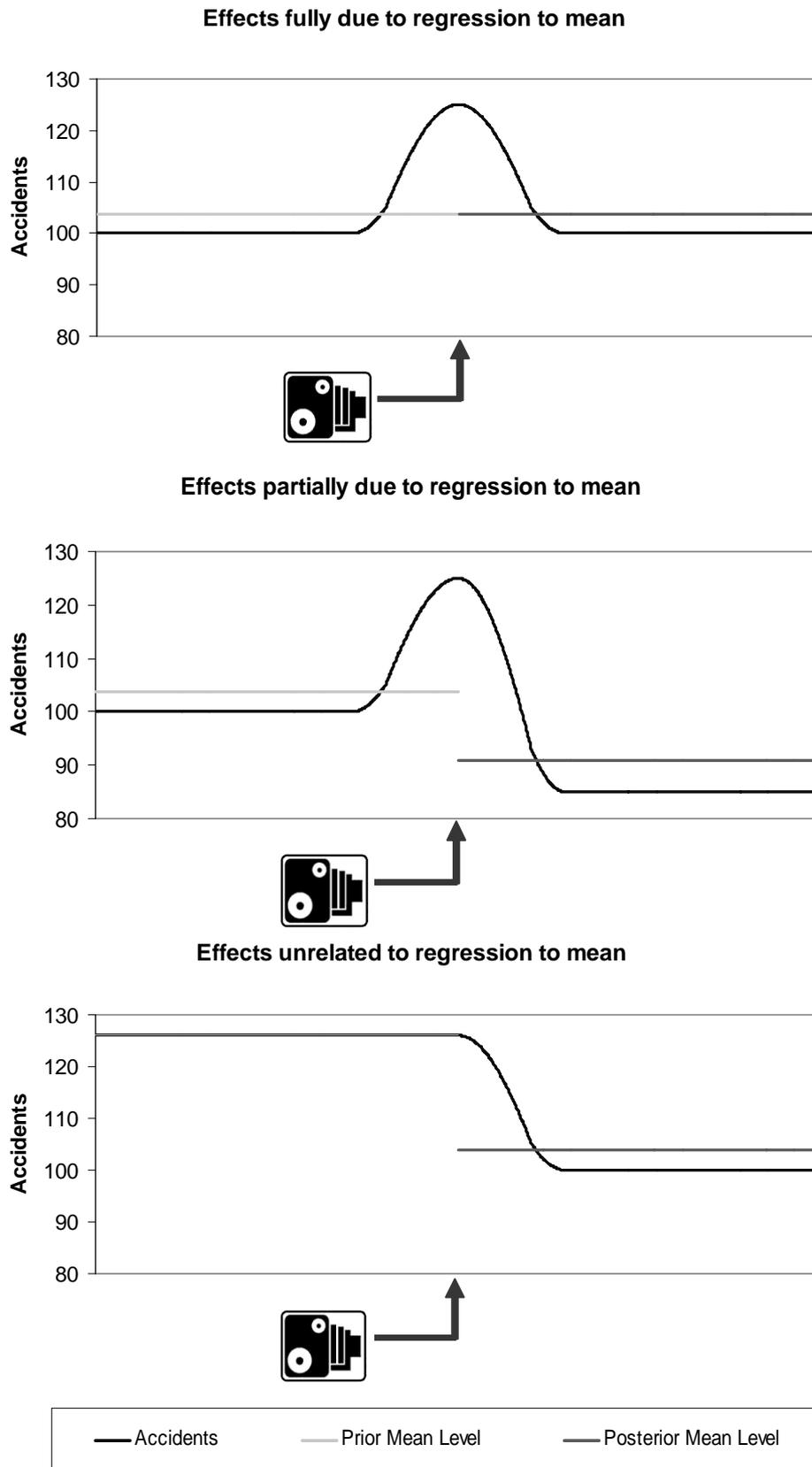
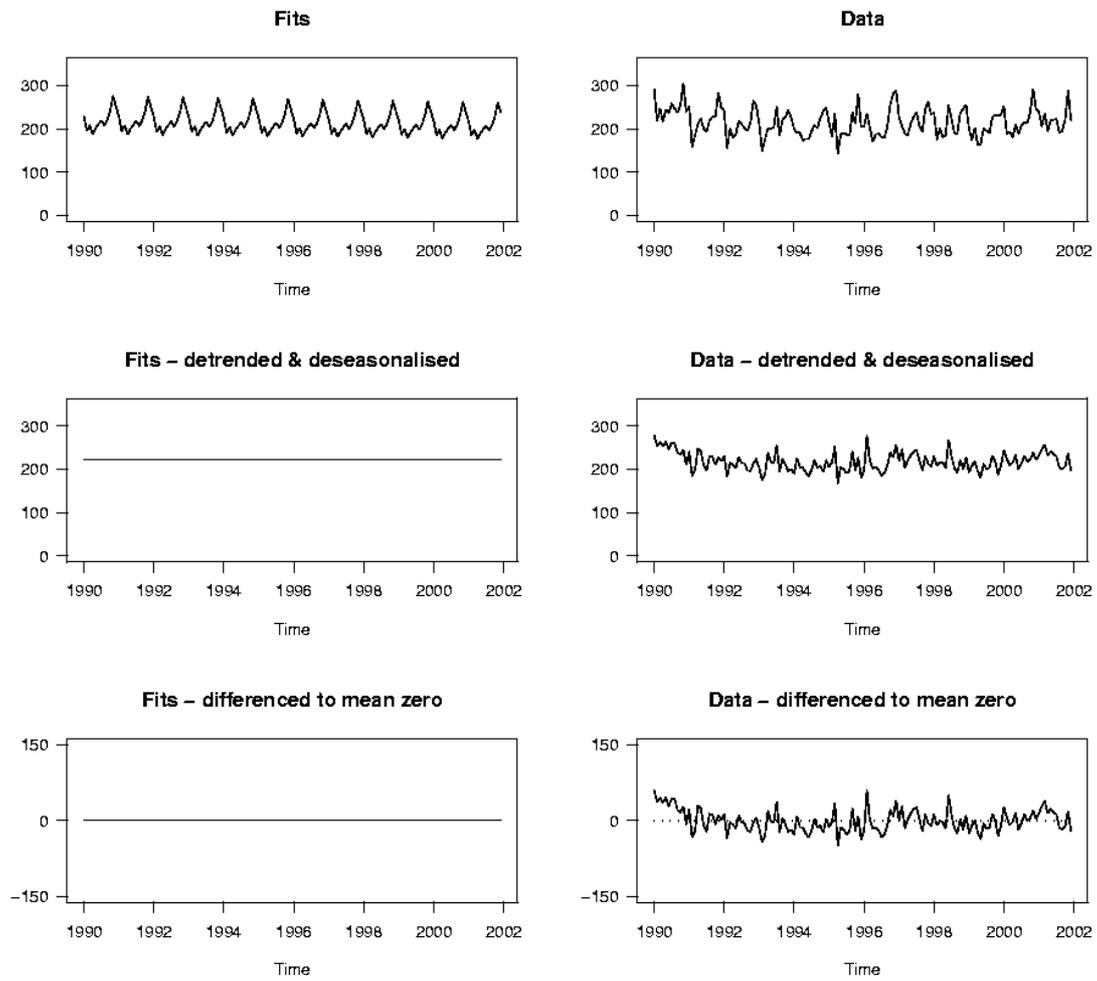


Figure 3 : Residual analysis for the SARIMA model for combined accident data



**Figure 4 : Detection of Regression to Mean**



**Figure 5 : Effects of Detrending, Deseasonalising and Differencing**