

# **ELECTRICITY CASE: STATISTICAL ANALYSIS OF ELECTRIC POWER OUTAGES**

**Simonoff, J. (NYU-Stern); Zimmerman, R., Restrepo, C.,  
Dooskin, N., Hartwell, R., Miller, J.,  
Remington, W. (NYU-Wagner); Lave, L. (Carnegie Mellon);  
Schuler, R. (Cornell)**

CREATE REPORT  
Under FEMA Grant EMW-2004-GR-0112

**May 31 , 2005**



**Center for Risk and Economic Analysis of Terrorism Events  
University of Southern California  
Los Angeles, California**



**Electricity Case:  
Statistical Analysis of Electric Power Outages**

**CREATE Report**

**May 31, 2005**

**Jeffrey S. Simonoff (NYU-Stern);  
Rae Zimmerman, Carlos E. Restrepo, Nicole J.  
Dooskin, Ray V. Hartwell, Justin I. Miller and Wendy  
E. Remington (NYU-Wagner);  
Lester B. Lave (Carnegie Mellon);  
Richard E. Schuler (Cornell)**

**New York University-Wagner Graduate School,  
Institute for Civil Infrastructure Systems**

## Abstract

This report analyses electricity outages over the period January 1990-August 2004. A database was constructed using U.S. data from the DAWG database, which is maintained by the North American Electric Reliability Council (NERC). The data includes information about the date of the outage, geographical location, utilities affected, customers lost, duration of the outage in hours, and megawatts lost. Information found the DAWG database was also used to code the primary cause of the outage. Categories that included weather, equipment failure, human error, fires, and others were added to the database. In addition, information about the total number of customers served by the affected utilities, as well as total population and population density of the state affected in each incident, was also incorporated to the database. The resulting database included information about 400 incidents over this period.

The database was used to carry out two sets of analyses. The first is a set of analyses over time using three-, six-, or twelve-month averages for number of incidents, average outage duration, customers lost and megawatts lost. Negative binomial regression models, which account for overdispersion in the data, were used. For the number of incidents over time a seasonal analysis suggests there is a 9.3% annual increase in incident rate given season over this period. Given the year, summer is estimated to have 65-85% more incidents than the other seasons. The duration data suggest a more complicated trend; an analysis of duration per incident over time using a loess nonparametric regression “scatterplot smoother” suggests that between 1990-93 durations were getting shorter on average but this trend changed in the mid-1990s when average duration started to increase, and this increase became more pronounced after 2002. When looking at average customer losses by season there is weak evidence of an upward trend in the average customer loss per incident, with an estimated increase of a bit less than 10,000 customers per incident per year. Similar analyses of MW lost per incident over time showed no evidence of any time or seasonal patterns for this variable.

The second part of the report includes a number of event-level analyses. The data in these analyses are modeled in two parts. First, the different characteristics related to whether an incident has zero or nonzero customers lost are determined. Then, given that the number lost is nonzero, the characteristics that help to predict the customers lost are analyzed. Unlike the first set of models described, in this section a number of predictors such as primary cause of the outage (including variables such as weather, equipment failure, system protection, human error and others), total number of customers served by the affected utilities, and the population density of the states where the outages were used in the analyses to gain a better understanding of the three key variables: customers lost, megawatts lost and duration of electric outages. Logistic regression was used in these analyses. For logged customers lost, the only predictor showing much of a relationship was logged MW lost. The total number of customers served by the utility was found to be a marginally significant predictor of customers lost per incident. Customer losses are higher for natural weather related events, crime, unknown causes, and third party, and lower for capacity shortage, demand reduction, and equipment failure, holding all else in the model fixed.

The analyses for duration at the event level find that the two most common causes of outages, equipment failure and weather, are very different, with the former associated with shorter events and the latter associated with longer ones. When the primary cause of the events is included in the regression models, the time trend for the average duration per incident found in earlier analyses disappears. According to the data, weather related incidents are becoming more common in later years and equipment failures less common, and this change in the relative frequency of primary cause of the events accounts for much of the overall pattern of increasing average durations by season. Holding all else in the model constant, these analyses also suggest that winter events have an expected duration that is 2.25 times the duration of summer events, with autumn and spring in between.

## Acknowledgements

This research was supported by the United States Department of Homeland Security through the Center for Risk and Economic Analysis of Terrorism Events (CREATE), grant number EMW-2004-GR-0112. However, any opinions, findings, and conclusions or recommendations in this document are those of the author (s) and do not necessarily reflect views of the U.S. Department of Homeland Security.

## Table of Contents

Abstract and Acknowledgements	3
SUMMARY ANALYSES OVER TIME (pg. 3)	
Analysis of the number of incidents over time	3
Analysis of the number of incidents that were associated with nonzero MW loss or nonzero customer loss over time	8
Analysis of duration over time	14
Analysis of MW loss over time	26
Analysis of customer loss over time	31
EVENT-LEVEL ANALYSES (pg. 36)	
Analysis of customer loss at the event level	36
Analysis of duration at the event level	51
REFERENCES (pg. 62)	

## **Electricity Case: Statistical Analysis of Electric Power Outages**

This report presents the detailed results of the statistical analysis of electric power outage data summarized in the Electricity Case – Main Report. It contributes to the sections on the evaluation of risks and consequences of electric power outages.

### ***I. Summary analyses over time***

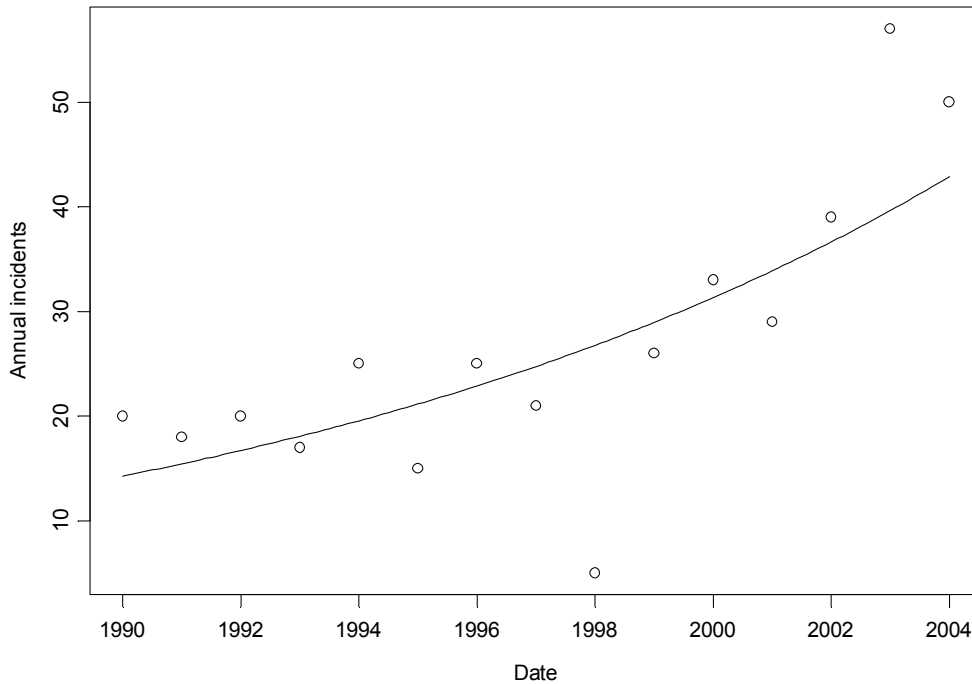
#### **A. Analysis of the number of incidents over time**

This report summarizes the analyses of incident counts over time. Such count data are typically analyzed using special count regression models based on the Poisson and negative binomial distributions; see Simonoff (2003, chapter 5), for extensive discussion of these models. The standard count regression model is based on the Poisson distribution. The Poisson distribution has the property that its mean equals its variance, which can account for the observed pattern in count data that variability increases with level.

Count regression models are generally fit as *loglinear* models; that is, it is the logarithm of the mean that is modeled as a linear function of predictors, or equivalently, the mean is modeled as an exponential function of the predictors. This implies, for example, a *proportional* relationship with a variable, rather than an *additive* one. Loglinear models are natural for count data because the true mean of the response cannot possibly be negative; a linear model on predictors can lead to estimated negative means, but a loglinear model cannot.

#### *Annual data*

We start with data measured at the annual level. The following is a plot of the annual incident figures for the U.S. data, along with the estimate of the time trend based on a Poisson regression model. Note that the estimated time trend is based only on years 1990 through 2003, since the 2004 data are incomplete (the data only run through August).



There are several noteworthy points here:

1. The fitted curve is consistent with an estimated annual increase in incidents of 8.2%. Note, by the way, that it is apparent from the plot that a loglinear model is more reasonable than a linear model here, as the increase in incidents is slower in the 1990s than in the 2000s.
2. The estimated increase is highly significantly different from zero, with a Wald statistic (the analogue of a t-statistic for Poisson regression models) of 5.8.

Here is output from the model detailing the significance testing based on the Poisson model:

Coefficients:

	Value	Std. Error	Wald
(Intercept)	-153.6724123	27.2931301	-5.630443
Date	0.0785583	0.0136619	5.750174

The significance of the time trend can be assessed by calculating a tail probability for the 5.75 based on a normal distribution; in this case it is zero to 8 digits. The estimated annual increase in incidents comes from exponentiating the slope coefficient in this model, as  $\exp(.0786) = 1.082$ , implying an estimated 8.2% annual increase.

3. The year 1998 was obviously a very unusual one, with a very small number of incidents (5, where the model implies an estimate of 26.0).

4. There is evidence that the incident rate is increasing in recent years. The model implies an estimated 39.6 incidents in 2003, when there were, in fact, 57 (this is roughly  $2\frac{3}{4}$  standard deviations above the expected number), and an estimated 42.9 in 2004, when there were 50 in only the first 8 months of the year. The 2003 number is apparently not because of the August 14, 2003 blackout, since that event only accounts for 8 incidents.

There is a flaw in this analysis, in that the Poisson regression model does not fit the data well, because of *overdispersion*. Overdispersion occurs when there is unmodeled heterogeneity in the data. The Poisson model treats each year as identical, other than the actual difference in year. This is unlikely to be true, as the chances are very good that there have been many changes to the structure of power generation over the years (new power plants come on line, old ones go off, new drains on power generation occur, political situations change, and so on). The Poisson model does not account for this possibility, and as a result the observed variability in the response is larger than that implied by the Poisson model (recall that the Poisson distribution has the property that the mean equals the variance). An important result of overdispersion is that the statistical significance of effects in the model are overstated.

Overdispersion has occurred here, as both the Pearson ( $X^2=34.1$ ) and deviance ( $G^2=42.0$ , both on 12 degrees of freedom) goodness-of-fit statistics indicate that the Poisson model does not fit the data.

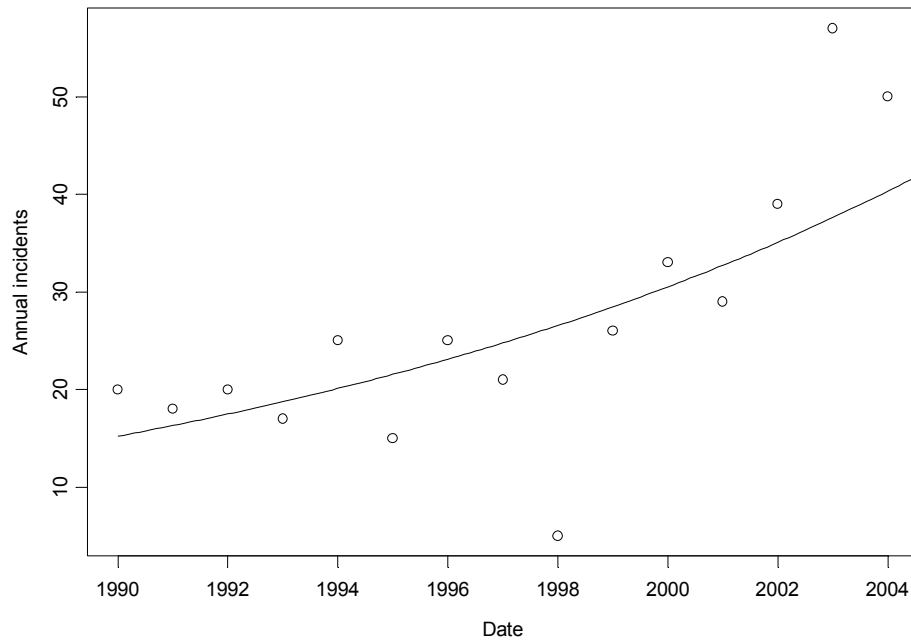
A way of addressing overdispersion is to fit a count regression model that allows for the variance being larger than the mean. The standard model of this type is the negative binomial regression model. Here is output for this model:

```

Coefficients:
                Value  Std. Error    Wald
(Intercept) -135.76307835  47.91257415  -2.833558
          Date   0.06959082  0.02399384   2.900362

```

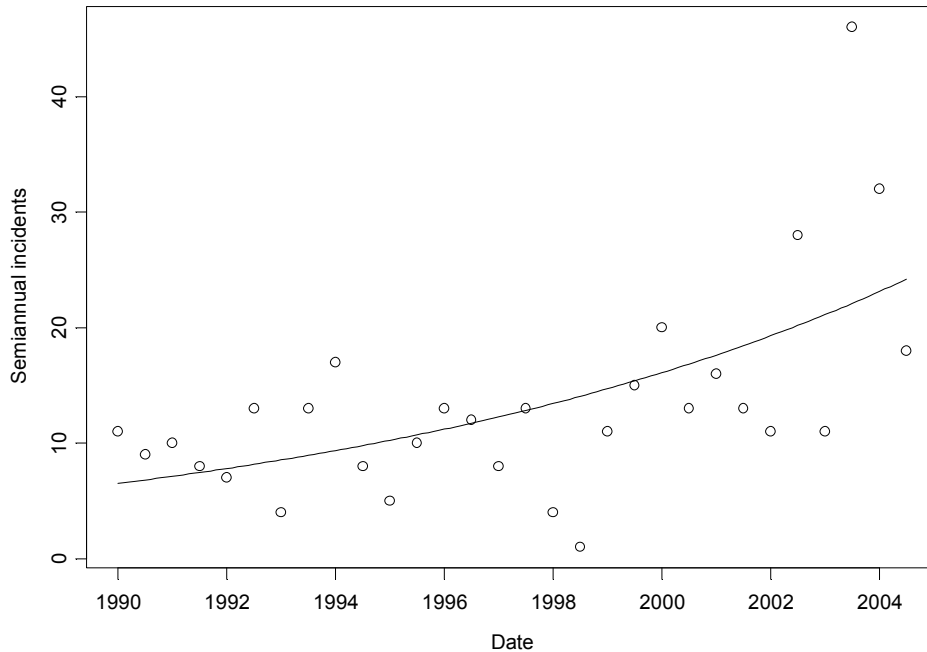
The Wald statistic for this model is smaller than in the Poisson model, but it is still highly significant ( $p=.002$ ). The model fits the data well, as the deviance equals 15.3 on 12 degrees of freedom ( $p=.23$ , not rejecting the fit of the model). The estimated annual increase in incidents based on this model is slightly lower than before, implying a 7.2% annual increase in incidents. Here is a graphical representation of the estimated trend:



### *Semiannual data*

One potential problem with the analysis on annual data is that there are only 14 data points. The following analysis is based on semiannual incident counts, resulting in roughly twice as many data points. Once again the last data point (corresponding to the second half of 2004) has been omitted, since it is incomplete.





This analysis reinforces and refines some of the earlier impressions.

1. The implications regarding the increase in incidents are similar for these semiannual data to those for the annual data. The estimated rate of increase is 9.5% annually, similar to what was seen before.
2. The estimated increase is even more significantly different from zero, with a Wald statistic of 7.1. Here is output for the model:

Coefficients:

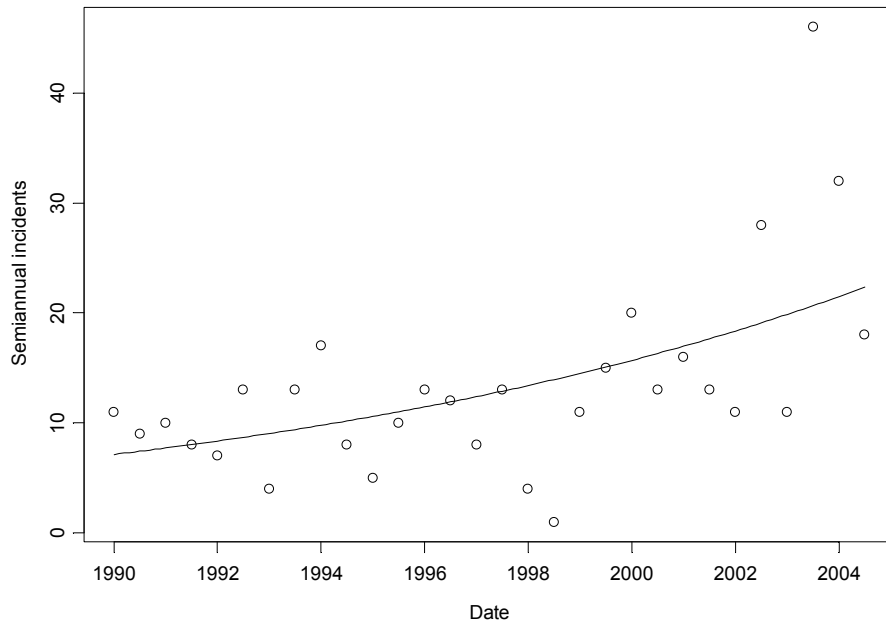
	Value	Std. Error	Wald
(Intercept)	-178.12611859	25.49168622	-6.987616
Date	0.09045251	0.01275513	7.091459

3. We can see from the plot that it was the second half of 1998 that was so unusual, with only 1 incident (and 14.1 predicted by the model).
4. The first part of 2003 actually had fewer incidents than expected; it was the 46 incidents in the second half of 2003 (more than twice the expected number) that was so unusual. Again, only eight of these were from the August 14 blackout. The first half of 2004 was a bit above normal, but not overwhelmingly so, but the 18 incidents in the first two months of the second half of 2004 is noticeably high. Thus, in addition to the relatively stable annual increase in incidents, there is still (limited) evidence of an increasing rate recently.
5. There is evidence of overdispersion in these data as well, as the Pearson ( $X^2=87.2$ ) and deviance ( $G^2=92.4$ , both on 27 degrees of freedom) indicate lack of fit. A negative binomial fit to these data is as follows:

Coefficients:

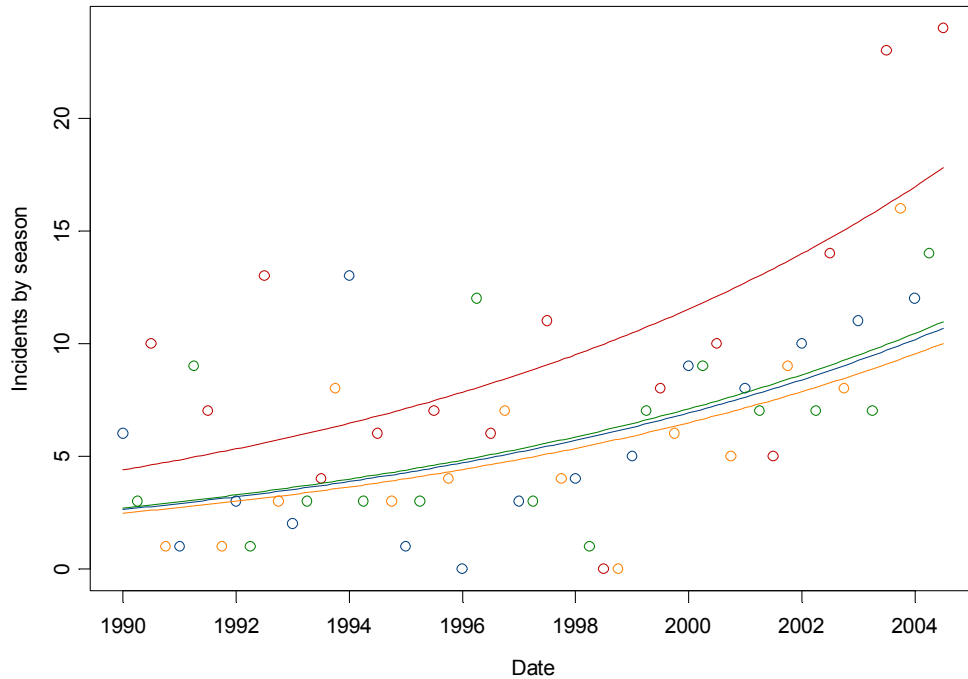
	Value	Std. Error	Wald
(Intercept)	-154.57911519	44.34697394	-3.485674
Date	0.07866603	0.02220172	3.543240

The time trend is highly statistically significant ( $p=.0004$ ). The model fits the data well (the deviance is 30.4 on 27 degrees of freedom,  $p=.29$ ), and it implies an estimated 8.2% annual increase in incidents. Here is a plot:



### *Seasonal data*

Examining the data at a seasonal level allows for the inclusion of different levels for different seasons. Winter is defined as December through February, spring as March through May, summer as June through August, and autumn as September through November. In this plot, the winter points and line are in blue, the spring points and line are in green, the summer points and line are in red, and the autumn points and line are in orange. Note that all of the data points other than the first one are used, since the data go through August 2004 (that is, summer 2004), but the first data point only includes two months instead of three.



1. Taking the season into account, the estimated annual increase in incident rate is 10.1%. All of the estimates obtained thus far are within the estimated standard errors of each other, so from a statistical point of view, all are equally reasonable. That is, what is most reasonable is to say is that the estimated increase in incidents is roughly 7-10% annually.
2. This increase is highly statistically significantly different from zero (Wald statistic 7.7). Here is output; the tests for the seasons take Autumn as a baseline category.

Coefficients:

	Value	Std. Error	Wald
(Intercept)	-191.08893534	25.12676351	-7.6049960
Date	0.09647909	0.01257097	7.6747538
SeasonSpring	0.09125890	0.15677009	0.5821193
SeasonSummer	0.57571503	0.14186204	4.0582740
SeasonWinter	0.06511136	0.15980717	0.4074371

3. While the winter, spring, and autumn estimated rates are similar to each other (with autumn having a rate that is slightly lower), summer has a noticeably higher rate of incidents. This is presumably from weather effects: snow and ice in the winter, thunderstorms in parts of the US in spring, and most importantly thunderstorms and intense heat (with corresponding air conditioner use) in the summer (and the lack of all of these factors in the autumn; we might have expected evidence of a hurricane effect in autumn, but only Hurricane Floyd in 1999 and Hurricane Isabel in 2003 show up as noteworthy). The difference between the summer rate and that of the other seasons is highly statistically significant, but more importantly, corresponds to an important effect

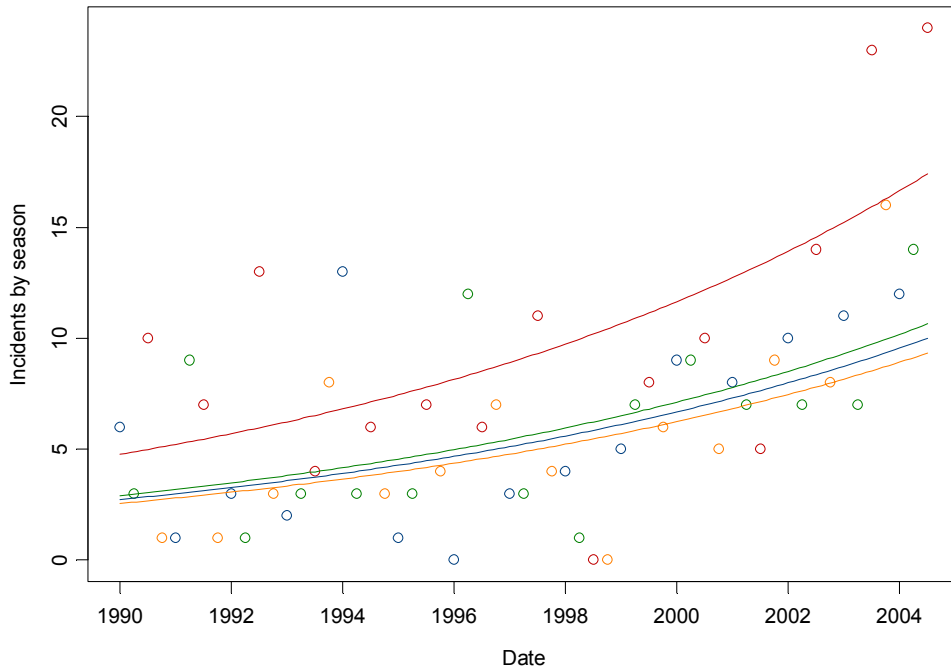
in practical terms, since the estimated number of incidents is 60% to 80% higher in summer than in the other seasons, given the year.

4. The unusually high rates of the last two years noted earlier come from the summers of 2003 and 2004, which have incident counts that are unusually high. Note, however, that once season is taken into account, these observations are no longer alarmingly high, only being between 1 ½ and 2 standard deviations above the expected value.
5. There is evidence of overdispersion here, as the Pearson ( $X^2=127.1$ ) and deviance ( $G^2=132.8$ , both on 53 degrees of freedom) tests indicate lack of fit. A negative binomial fit to these data is as follows:

Coefficients:

	Value	Std. Error	Wald
(Intercept)	-176.83433714	38.06541369	-4.645538
Date	0.08933316	0.01905175	4.688973
SeasonSpring	0.13104417	0.23103941	0.567194
SeasonSummer	0.62354889	0.22063885	2.826107
SeasonWinter	0.06813003	0.23598164	0.288709

The model fits the data adequately, but not as well as the earlier negative binomial models fit (the deviance is 68.4 on 53 degrees of freedom,  $p=.08$ ). The model implies an estimated 9.3% annual increase in incidents given season, and an estimated 65-85% higher rate for summer than for the other seasons given year. Here is a plot summarizing the model; it is clearly broadly similar to the one based on the Poisson model:



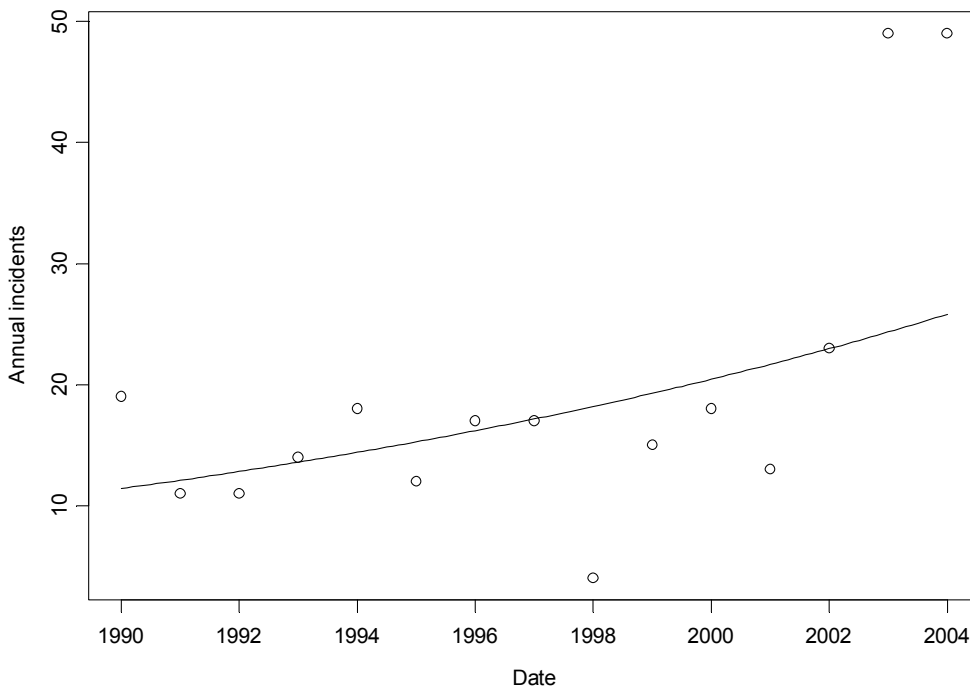
## B. Analysis of the number of incidents that were associated with nonzero MW loss or nonzero customer loss over time

The earlier analyses included incidents where there was no effect on the customer base, either in terms of customers affected or power loss. It could be argued that it is incidents that affect customers that are most interesting, so this portion of the report focuses on those incidents. Overdispersion occurs in all of the models, so all analyses are based on the negative binomial model.

### 1. INCIDENTS WITH NONZERO MW LOST

#### *Annual data*

Here is a plot with the negative binomial fit superimposed.



Here is output for this model:

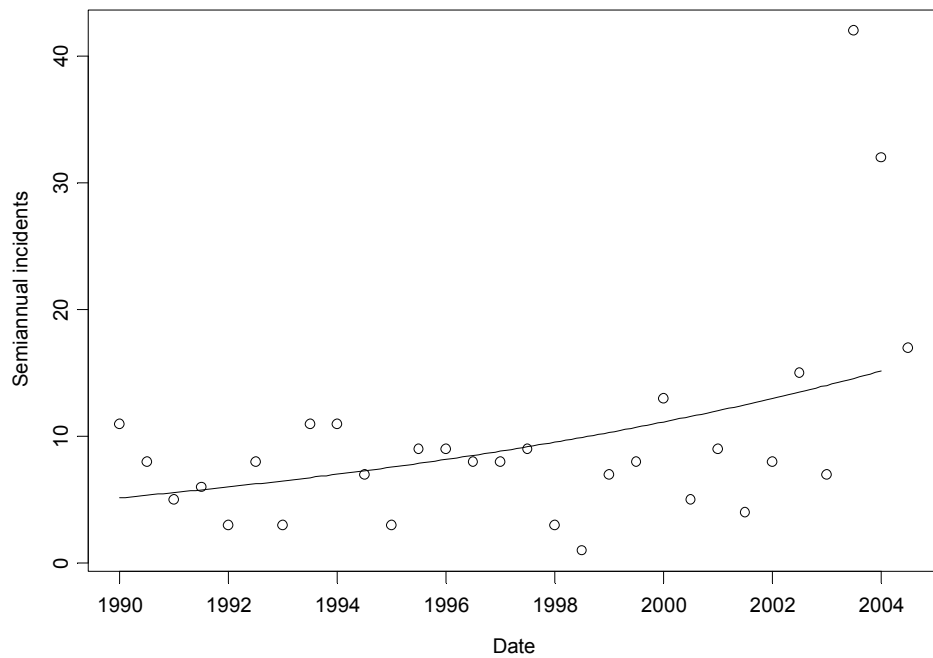
Coefficients:

	Value	Std. Error	Wald
(Intercept)	-113.5721447	57.91502205	-1.961014
Date	0.0582949	0.02900386	2.009901

The strength of the time trend is weaker than for the complete data, having a tail probability of .044. The estimated annual increase in incidents with nonzero MW is  $\exp(.058295)-1=6.0\%$ , so apparently the incidents with zero MW lost inflated the rate slightly (since this rate, with those incidents omitted, is smaller than the rate estimated based on all incidents). Note that 1998 is still unusually low, and 2003 and 2004 are unusually high.

*Semiannual data*

Here is a plot with the fitted model on the semiannual data.



Here is output for the model:

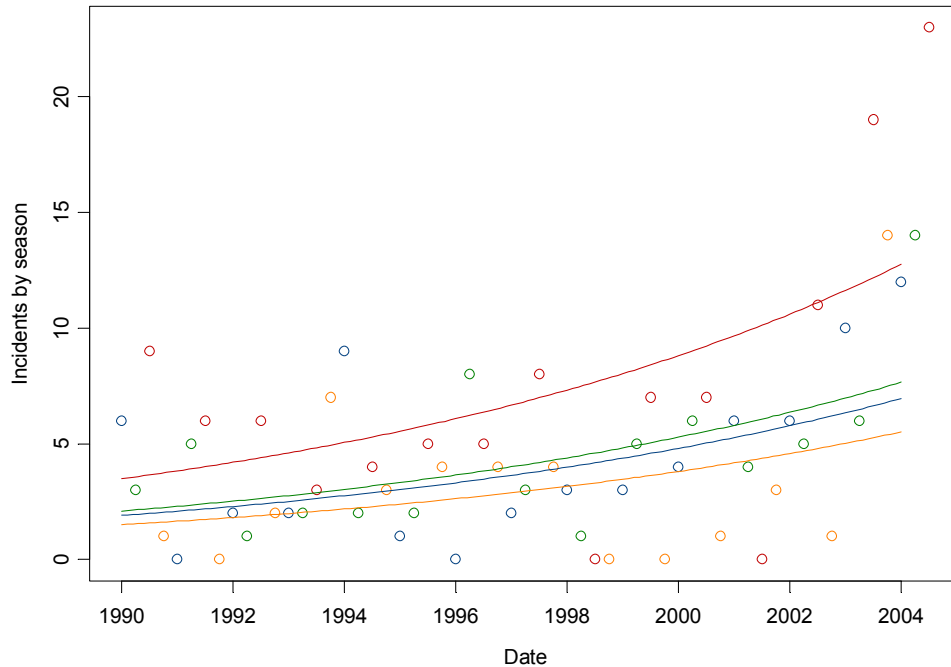
Coefficients:

	Value	Std. Error	Wald
(Intercept)	-151.48640808	54.82831060	-2.762923
Date	0.07694855	0.02744971	2.803256

The estimated annual increase in incidents with nonzero MW loss is 8.0%, and is highly significant ( $p=.005$ ). Note that while the second half of 2003 and of 2004 (only two months) are still high, now the first half of 2003 is also very high (this is because all of the incidents in the first half of 2003 were nonzero MW loss incidents).

## Seasonal data

Here is a plot by season.



The fit to these data is as follows:

Coefficients:

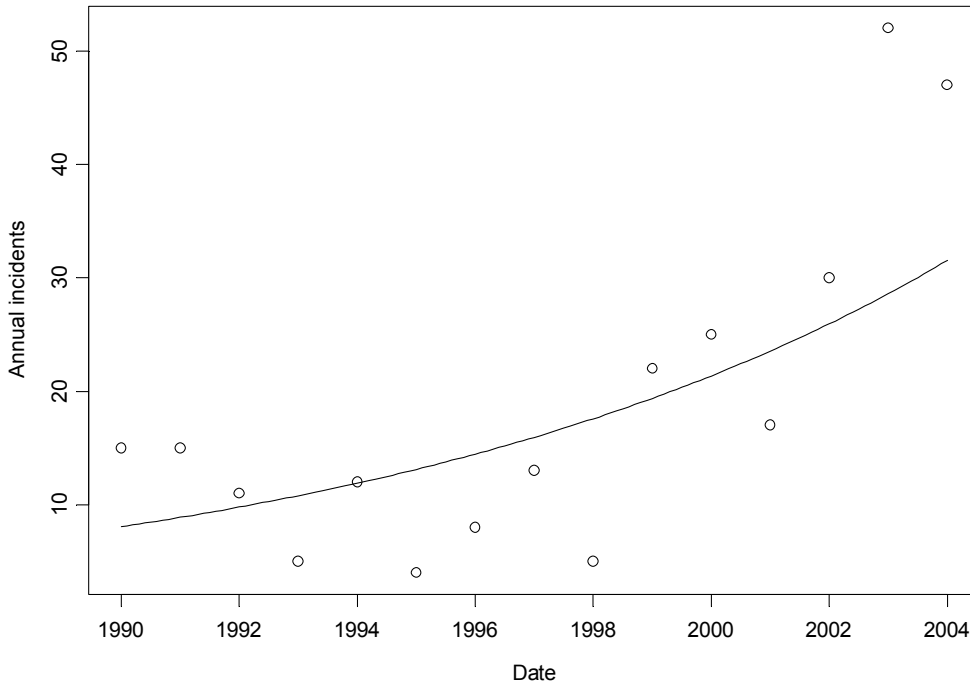
	Value	Std. Error	Wald
(Intercept)	-183.80666893	47.40955640	-3.876996
Date	0.09257176	0.02372817	3.901344
SeasonSpring	0.32906080	0.29114767	1.130220
SeasonSummer	0.83965494	0.27956975	3.003383
SeasonWinter	0.23304525	0.29788955	0.782321

The model implies an estimated 9.7% annual increase in incident rate given season, which is highly statistically significant ( $p < .0001$ ), and an estimated 65-130% higher rate for summer than for the other seasons. The summer effect is stronger than before, which is easy to understand: while more than 90% of the summer incidents had nonzero MW loss, roughly  $\frac{1}{4}$  of the incidents in the autumn had zero MW loss. That is, nonzero MW incidents are more likely in the summer, thereby strengthening the “summer effect” here. In terms of the time trend, we see a similar pattern to before, of a 6-10% annual increase in incidents from the analyses based on the three different time aggregations.

## 2. INCIDENTS WITH NONZERO CUSTOMERS LOST

### *Annual data*

Here is a plot with the negative binomial fit superimposed.



Here is output for this model:

Coefficients:

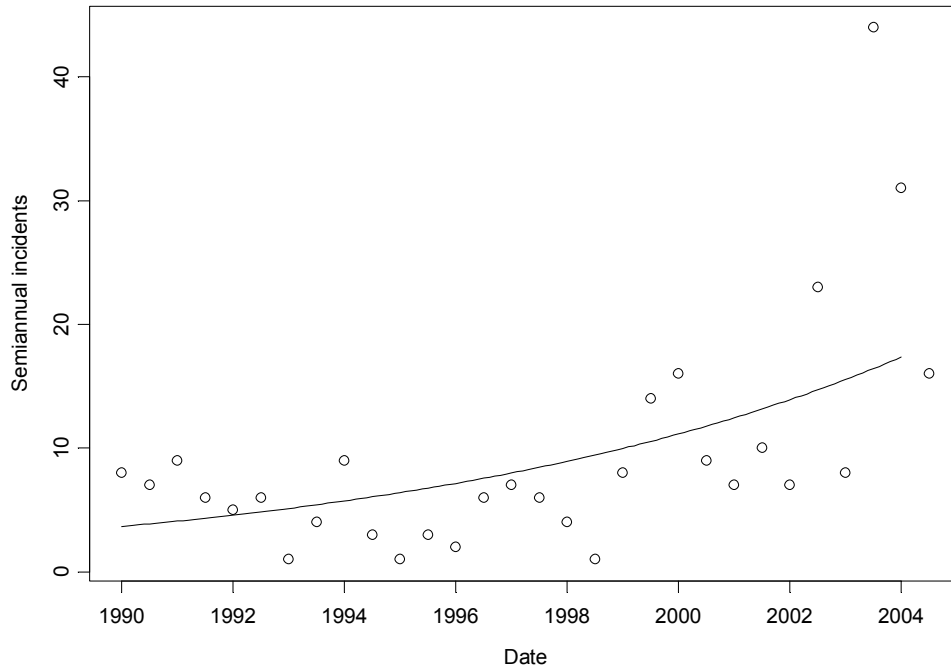
	Value	Std. Error	Wald
(Intercept)	-191.83850183	69.27401101	-2.769271
Date	0.09745008	0.03469084	2.809101

The strength of the time trend is stronger than for the analysis based only on nonzero MW incidents, having a tail probability of .005. The estimated annual increase in incidents with nonzero customer loss is  $\exp(.09745)-1=10.2\%$ , so apparently the incidents with zero customers lost deflated the rate earlier (the rate is higher once the zero customer loss events are omitted). This makes sense: the rate of incidents that had no customer loss was more than 35% from 1990-1997, but has been only 7.5% since then. Note that 1998 is not unusually low now, but 2003 and 2004 are still unusually high.



### *Semiannual data*

Here is a plot with the fitted model on the semiannual data.



Here is output for the model:

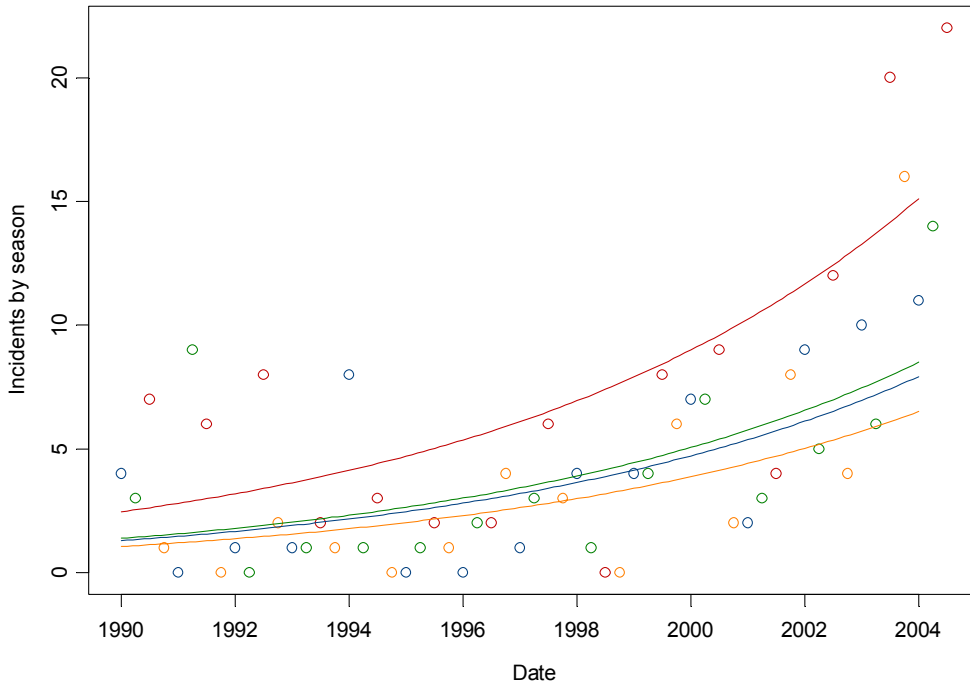
Coefficients:

	Value	Std. Error	Wald
(Intercept)	-219.3758662	59.57344642	-3.682444
Date	0.1108936	0.02982306	3.718385

The estimated annual increase in incidents with nonzero MW loss is 11.7%, and is highly statistically significant ( $p=.0002$ ). The second half of 2003 and first half of 2004 are unusually high.

## Seasonal data

Here is a plot of the seasonal data.



The fit to these data is as follows:

Coefficients:

	Value	Std. Error	Wald
(Intercept)	-257.7296583	50.12320622	-5.1419228
Date	0.1295428	0.02508268	5.1646296
SeasonSpring	0.2660959	0.30367008	0.8762666
SeasonSummer	0.8417521	0.28977870	2.9048102
SeasonWinter	0.1965276	0.30967757	0.6346199

The model implies an estimated 13.8% annual increase in incident rate (p zero to six digits), and an estimated 75-135% higher rate for summer than for the other seasons. The summer effect is similar to that for the nonzero MW loss data, but the pattern is a little more complicated: both summer and winter have lower rates of incidents with zero customer loss compared to spring and autumn, so the estimated relative chances of incidents in those seasons compared to spring and autumn are now higher. Overall, while removing the zero MW loss incidents has relatively little

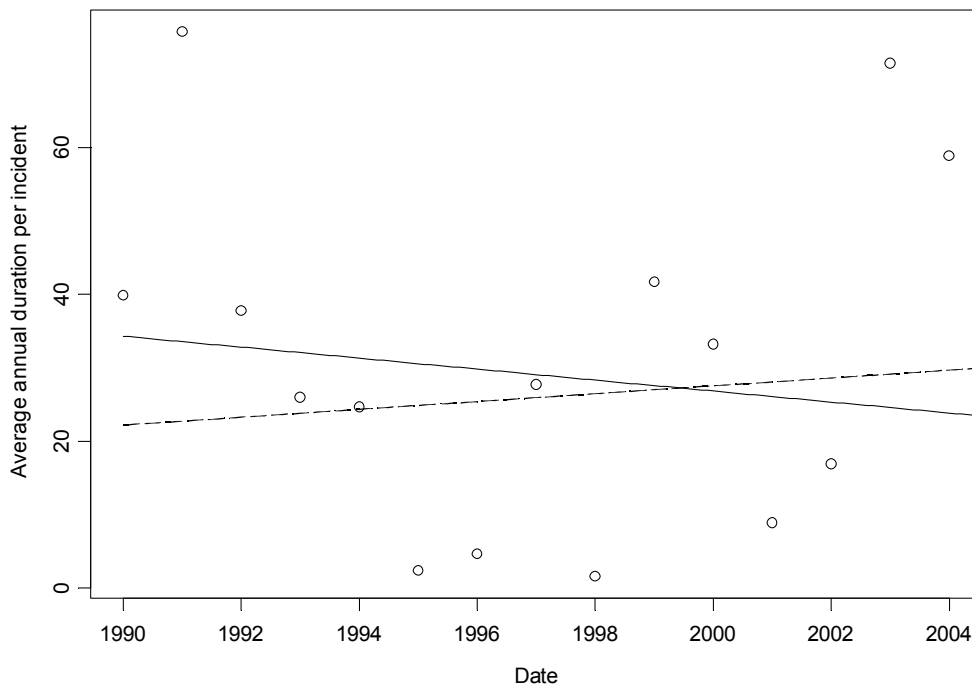
effect on the estimated annual increase in incident rate, removing the zero customer loss incidents has a stronger effect on the estimated annual increase of rates, increasing it to 12-14%.

### C. Analysis of duration over time

We now discuss the pattern of average duration of incidents over time. The response variable, whether measured annually, semiannually, or seasonally, is the average duration per incident over that time period. Note that zero-loss events are included, since they seem to be directly relevant to an analysis of duration. Obviously, events with missing duration are not included, which raises the issue of nonresponse bias. If the incidents for which duration is missing are different from those in which it was reported, that can bias the results in ways that are impossible to ascertain.

#### *Annual data*

We start with analyses based on a linear model for durations. Here is a plot of the average duration versus time, with two lines superimposed.



It is apparent that there is little evidence of any time trend in average duration. There is one early outlier, corresponding to 1991, although it is not that different from the values for 2003 and 2004. The solid line is the fitted time trend of average duration based on all of the data other than

2004 (since those data were incomplete), which has a negative slope that is far from statistically significant:

Coefficients:

	Value	Std. Error	t value	Pr(> t )
(Intercept)	1521.9638	3165.4638	0.4808	0.6393
Date	-0.7476	1.5855	-0.4715	0.6457

Residual standard error: 23.91 on 12 degrees of freedom

Multiple R-Squared: 0.01819

The dashed line in the plot gives the estimated time trend omitting 1991. The slope has shifted to be positive, but there is still no evidence of any trend:

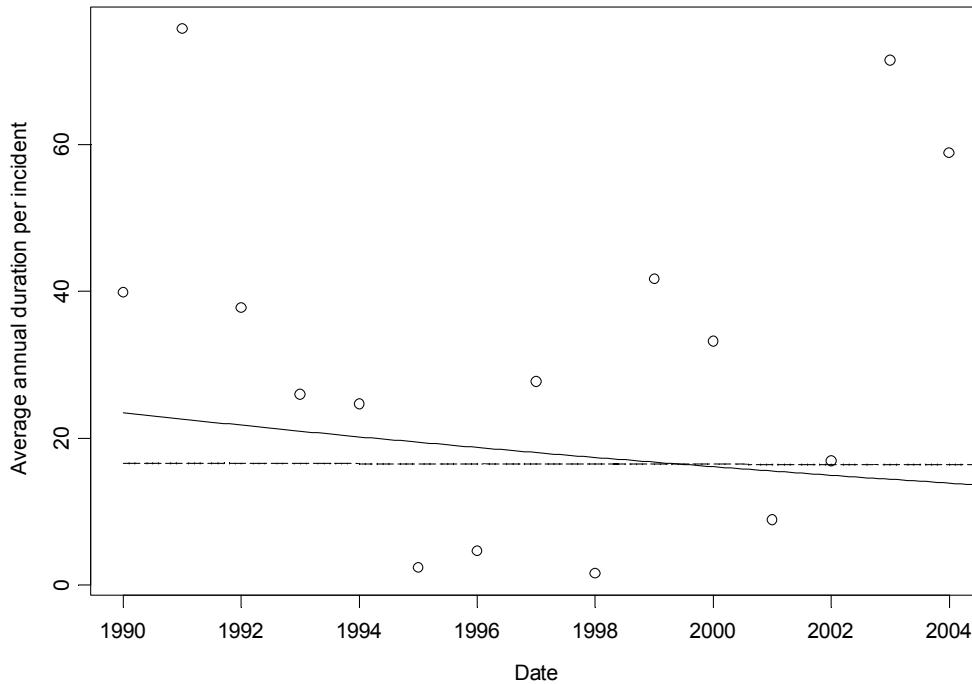
Coefficients:

	Value	Std. Error	t value	Pr(> t )
(Intercept)	-1039.5609	2933.9925	-0.3543	0.7298
Date	0.5335	1.4693	0.3631	0.7234

Residual standard error: 20.51 on 11 degrees of freedom

Multiple R-Squared: 0.01185

It is possible that multiplicative growth or decay of average duration might be sensible, which would imply the use of a model where logged duration is the response variable. In fact, the results are effectively the same:



*Full data set*

Coefficients:

	Value	Std. Error	t value	Pr(> t )
(Intercept)	77.8956	165.8329	0.4697	0.6470
Date	-0.0376	0.0831	-0.4522	0.6592

Residual standard error: 1.253 on 12 degrees of freedom  
Multiple R-Squared: 0.01675

*Data set omitting 1991*

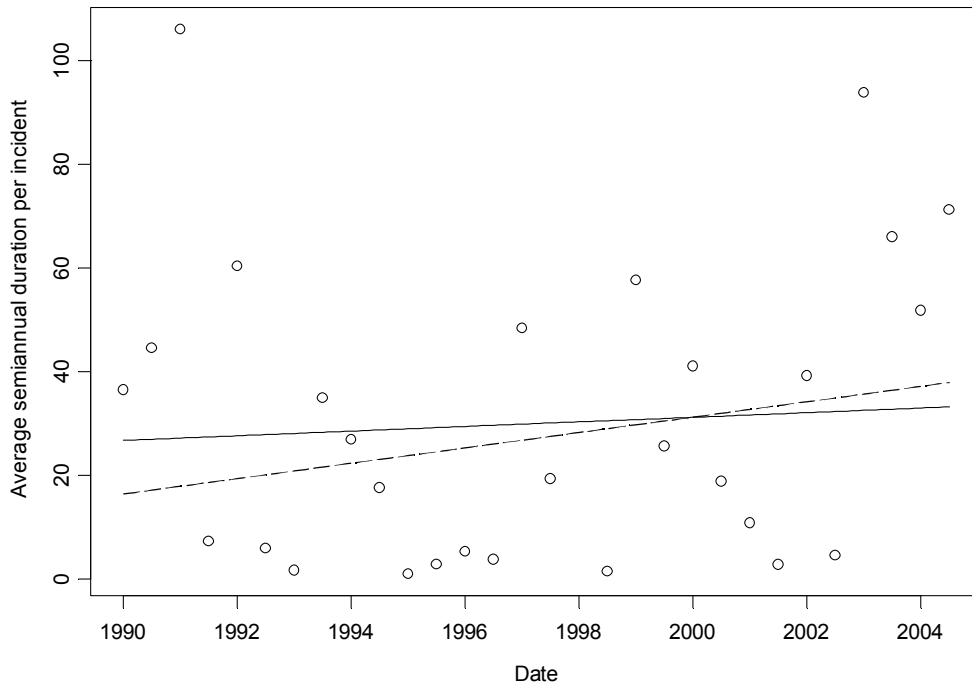
Coefficients:

	Value	Std. Error	t value	Pr(> t )
(Intercept)	4.3909	177.7861	0.0247	0.9807
Date	-0.0008	0.0890	-0.0089	0.9930

Residual standard error: 1.243 on 11 degrees of freedom  
Multiple R-Squared: 7.267e-006

*Semiannual data*

Here is a plot of the semiannual data, with linear trend lines superimposed.



The results are similar to those for the annual data. The estimated time trend is slightly positive when the 1991 time period is included, and slightly negative when it is not included, but in neither case is it close to statistical significance. Note that the value for the second half of 2004 is not included in either model, since the data are incomplete for that time period.

Here is computer output for the two models:

*Full data set*

Coefficients:

	Value	Std. Error	t value	Pr(> t )
(Intercept)	-857.6518	2569.1990	-0.3338	0.7412
Date	0.4445	1.2865	0.3455	0.7325

Residual standard error: 28.95 on 26 degrees of freedom  
Multiple R-Squared: 0.004569

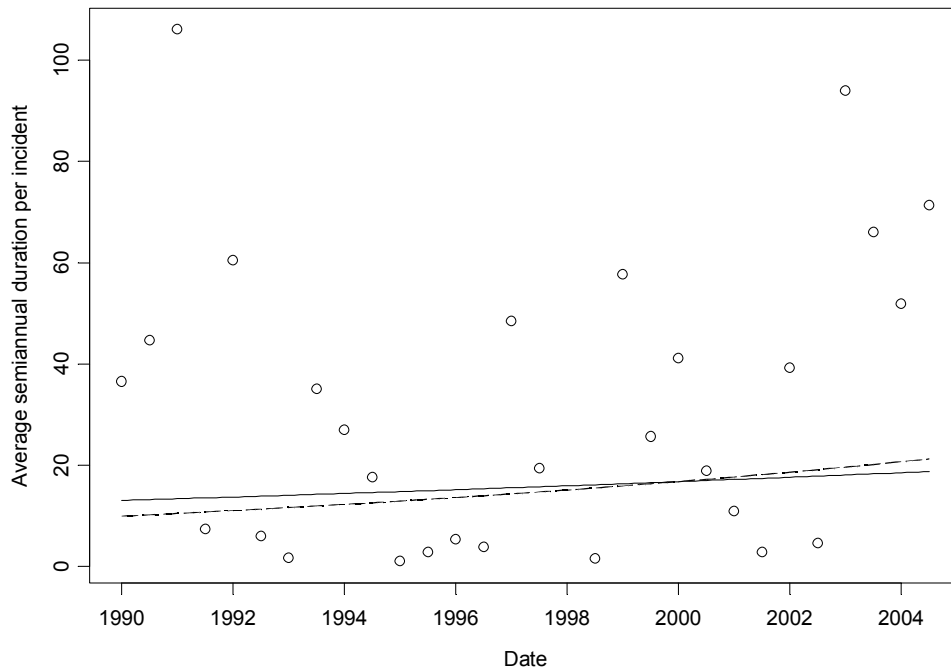
*Data set omitting first half of 1991*

Coefficients:

	Value	Std. Error	t value	Pr(> t )
(Intercept)	-2933.8052	2246.1745	-1.3061	0.2034
Date	1.4825	1.1247	1.3182	0.1994

Residual standard error: 24.37 on 25 degrees of freedom  
Multiple R-Squared: 0.06499

Models for logged duration also find no evidence of any effect:



*Full data set*

Coefficients:

	Value	Std. Error	t value	Pr(> t )
(Intercept)	-47.1819	122.4689	-0.3853	0.7032
Date	0.0250	0.0613	0.4077	0.6869

Residual standard error: 1.38 on 26 degrees of freedom  
Multiple R-Squared: 0.006351

*Data set omitting first half of 1991*

Coefficients:

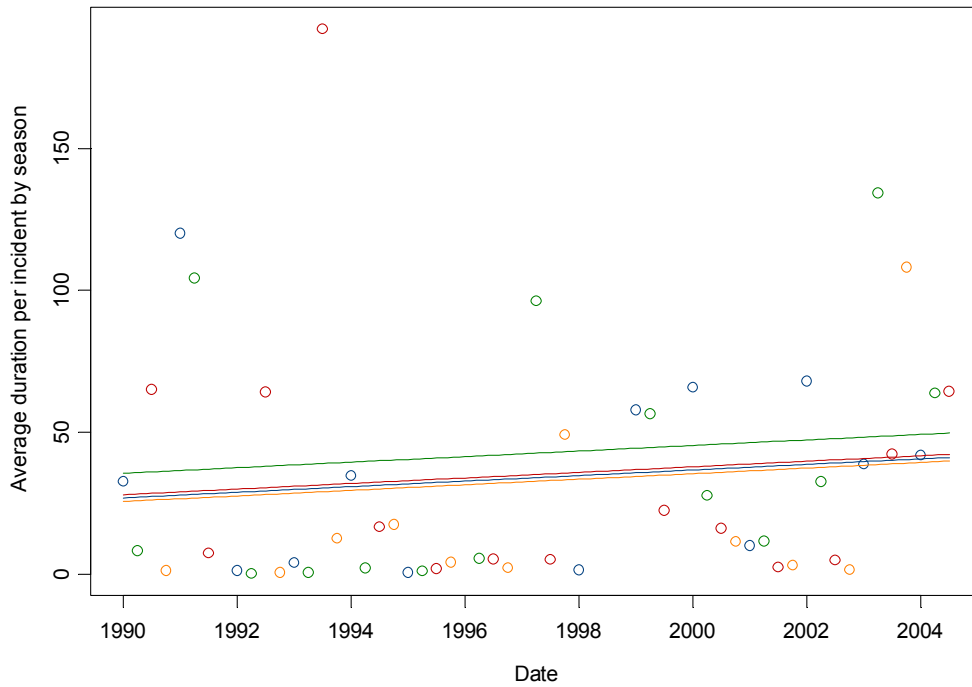
	Value	Std. Error	t value	Pr(> t )
(Intercept)	-101.6895	123.2912	-0.8248	0.4173
Date	0.0523	0.0617	0.8465	0.4053

Residual standard error: 1.338 on 25 degrees of freedom

Multiple R-Squared: 0.02786

### Seasonal data

Here is a plot based on all of the data other than the first data point, fitting a linear time trend.



There is a slight upward slope, but it is not statistically significant. There is also no evidence of a season effect; the spring line is marginally higher than the other lines, but this is not close to significance.

Coefficients:

	Value	Std. Error	t value	Pr(> t )
(Intercept)	-1924.2276	2748.3347	-0.7001	0.4874



Date	0.9798	1.3759	0.7121	0.4800
Season1	9.9299	8.6390	1.1494	0.2563
Season2	2.4131	4.7839	0.5044	0.6164
Season3	1.2941	3.5458	0.3650	0.7168

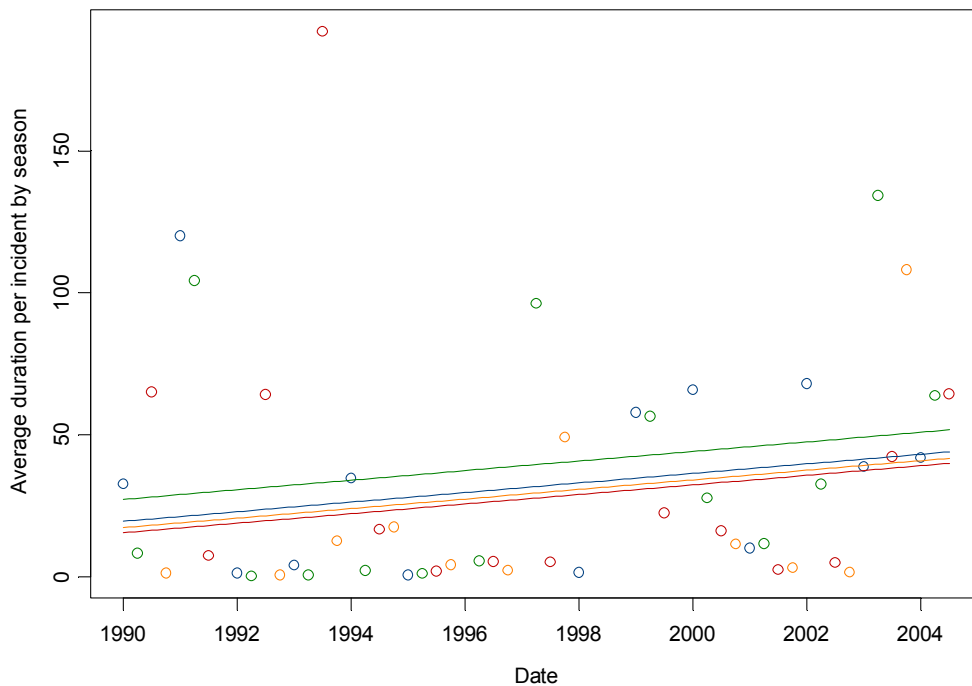
Residual standard error: 42.88 on 46 degrees of freedom  
Multiple R-Squared: 0.04353  
F-statistic: 0.5233 on 4 and 46 degrees of freedom, the p-value is 0.719

Anova Table

Response: Duration

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	901.21	1	0.4902004	0.4873640
Date	932.31	1	0.5071208	0.4799843
Season	2921.74	3	0.5297475	0.6641179
Residuals	84568.48	46		

The summer 1993 point is unusual, so here is a summary omitting that data point.



There is now (very) weak evidence of an upward slope, but no season effect. This is presumably coming from the last six seasons, wherein four had average durations of more than 60 hours.

Coefficients:

	Value	Std. Error	t value	Pr(> t )
(Intercept)	-3338.4889	2295.7900	-1.4542	0.1528
Date	1.6863	1.1493	1.4672	0.1493
Season1	10.0033	7.1542	1.3982	0.1689
Season2	-1.7791	4.0609	-0.4381	0.6634
Season3	2.2743	2.9438	0.7726	0.4438

Residual standard error: 35.51 on 45 degrees of freedom

Multiple R-Squared: 0.09675

F-statistic: 1.205 on 4 and 45 degrees of freedom, the p-value is 0.3218

Anova Table

Response: Duration

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	2666.13	1	2.114636	0.1528365
Date	2714.16	1	2.152735	0.1492709
Season	3466.54	3	0.916493	0.4405690
Residuals	56735.87	45		

The increasing trend in the last few data points suggests that a model for logged duration based on seasonal data might be appropriate, since in such a model while the proportional increase in duration is constant over time, the absolute level increases more quickly as time goes on (assuming that the slope is positive).



























































































