Exploring the Method to Assess the Effectiveness of Countermeasures for Proactive Traffic Safety Improvement

A PROJECT
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Fang Chen

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE IN MATHEMATICS AND STATISTICS

Advisers: Dr. Zhuangyi Liu and Dr. Hongyi Chen

July 2010
Acknowledgements

First of all, I would like to take this opportunity to sincerely thank my advisor and co-advisor, Dr. Zhuangyi Liu and Dr. Hongyi Chen, for their encouragement, supervision and support which motivated me to overcome the difficulties so that I could develop an understanding for the subject and finish it in a timely manner. It is also a pleasure to thank Dr. Bruce Peckham for being on my degree committee and making a great effort to revising this project report. Special thanks also go to Junyu Guo, who always support me and offer me with an insight in Transportation area. I am also indebted to many of my colleagues and friends, who encouraged me to think and offered me help when it was needed. The last but not least, I owe my deepest gratitude to my family, especially my parents for always being there and giving kind advice for my development.
Abstract

Transportation safety is always an important issue. In this report, the reactive and the proactive methods of eliminating the hazardous locations are reviewed. The proactive method is considered superior since it makes predictive analysis and helps to choose the most effective safety strategy for implementation to prevent accident from happening. In order to assess the effectiveness of traffic safety strategies, the before and after study methods are also reviewed. A few inaccurate and incomplete proofs in the literature are revised. Moreover, a methodology of forecasting accidents for road sections is proposed. Combining Hauer’s coherent method with proper extrapolation method will provide a possible way to forecast accident counts and solve the regression to the mean issue. This helps to give an accurate input for the optimization model proposed in the proactive method to improve traffic safety systematically.
# Table of contents

1. Introduction .............................................................................................................................................. 1

2. Literature review ...................................................................................................................................... 2

   2.1 General procedure .............................................................................................................................. 2

      2.1.1 The Six-step process ..................................................................................................................... 2

      2.1.2 The proactive process .................................................................................................................. 3

2.2 Assessment of treatment ...................................................................................................................... 6

   2.2.1 The Four-Step for before-and-after studies .................................................................................... 6

      2.2.1.1 Simple before-and-after study method ....................................................................................... 9

      2.2.1.2 The before-and-after study with Comparison Group method ............................................... 12

      2.2.1.3 The before-and-after study with the empirical bayes (EB) method .................................... 14

      2.2.1.4 Comparisons of the three methods .......................................................................................... 26

   2.2.2 A more coherent method .............................................................................................................. 27

3. Methodology of forecasting accidents for road sections ..................................................................... 34

4. Conclusion ............................................................................................................................................... 51

REFERENCES ................................................................................................................................................. 52
1. Introduction

Every year, many lives were killed by traffic accidents. Transportation safety has been a very important issue. Traffic engineers and college scholars are making their effort to improve traffic safety conditions. It is important to identify and eliminate the hazardous locations or the “hot spots” to prevent accidents from happening. The goal of this paper is to provide a proactive decision support system to improve transportation safety. By reviewing the current reactive procedure and the proposed proactive procedure proposed by Chen in [1], the system uses the optimizing model with the constraints to systematically assign countermeasures to the proper location so that the transportation safety will be improved to the maximum. It’s a proactive system because it provides an algorithm to forecast the happen of the accidents beforehand. In order to provide adequate inputs for the optimization model, this paper reviews several methods including:

- The simple before-and-after study method;
- The before-and-after study with Comparison Group (CG) method;
- The before-and-after study with the Empirical Bayes (EB) method;
- Hauer’s more coherent method.

It is found that the extrapolation combined with Hauer’s more coherent method will provide a proper way to forecast the expected accident count in the coming years and later will serve as an input to the system.

The rest of the paper is organized as follows: in section 2, we review the general procedure of eliminating hazardous locations. The three types of standard before and after study methods are discussed in detail and a summary is given. Then the Huaer’s coherent method is introduced with examples and case studies to demonstrate the whole process. In section 3, the methodology of forecasting the accident counts is proposed.
2. Literature review

Traffic safety has always been an important concern along with the development of traffic facilities. Every year hundreds of people were injured or killed in traffic accidents. According to a report of Minnesota Department of Transportation, there were 73,498 traffic crashes reported to Public Safety in 2009. Among them 31,074 were injured accidents, and 1,271 of these were server injuries [2]. A hazardous location, or a “hot spot” is a location with relatively high crash rate. Countermeasures have been made in order to reduce the crash rate. Professionalism requires that the effect of the treatment be known.

2.1 General procedure

2.1.1 The Six-Step process

The standard reactive procedure for identifying and eliminating hazardous locations consists of the following steps[3].

- Identify the highly hazardous locations according to crash reports;
- Analyze the potential design problems for these locations;
- Identify feasible countermeasures to deal with the design problems;
- Predict the effect of the potential countermeasures according to the crash reduction number;
- Implement the countermeasures with the highest cost effectiveness ratio;
- Estimate the effect of the countermeasure after implementing the countermeasures.

In the fourth step, we will choose to implement the countermeasures with the highest cost effectiveness ratio, which equals to the crash reduction factor divided by the implementation cost. The last step is to estimate the effect after implementing the countermeasures. The Six-Step is viewed as a reactive decision making process. This process of improving traffic
safety is the most pervasive process in real practice. It is straightforward and easy to use. However, this current process is found to be ineffective. The reasons are as follows:

This decision support system is a reactive system, which searches for remedy after traffic accident happens. An effective system should make predictive analysis beforehand and choose to implement the best possible countermeasure(s) to prevent accidents from happening.

The current system treats the study locations as isolated individuals therefore ignores the dependencies among safety improvements in related locations. Also, the “cost effectiveness ratio” is the only criterion to select countermeasures. This represents a lack of systematic optimization and long term planning in the resource allocation decision.

2.1.2 The proactive process

By reviewing the shortcomings of the Six-Step method, Chen proposed to develop a decision support system to address the problems [1]. The system was constructed by treating all the hazardous locations as an entirety. This comes in handy when we want to improve the traffic safety by regions. Also, this model is based on a proactive point of view in which the inputs are the forecast of what the accident numbers would be in the coming year(s). It uses an optimization model that combines the need of reducing accident counts with the restrictions on the budget.

Next, the underlying theory for the decision support system is introduced. Let \( P(\overline{A}|L_i, c_j) \) denote the conditional probability of having no accident in location \( i \) after the implementation of countermeasure \( j \). \( \overline{A} \) represents the event of having no accidents, \( L_i \) represents the location \( i \) and \( c_j \) represents the implementation of countermeasure \( j \). When more than one countermeasure is implemented in a certain location, then we define this combining implementation of different countermeasures in one location to be a “scenario”. For example, the probability of observing no accident in location \( L_1 \) after the implementation of \( c_1 \) and \( c_2 \) can be expressed as \( P(\overline{A}|L_1, c_1, c_2) \). In this case, we can define the jointly implementation of these two countermeasures to be scenario 1, namely \( S_1 = c_1 \cap c_2 \). Then for each road location
\( L \), suppose the number of the possible strategies is \( p \), then there should be \( (p^1) \) + \( (p^2) \) + \( \ldots + (p^{p-1}) \) + \( (p^p) \) = \( J \) scenarios to individually or jointly implement \( p \) strategies. In the same fashion, the conditional probability of having no accident in location \( i \) after the implementation of scenario \( j \) is \( P(\bar{A} | L_i, S_j) \).

The highly hazardous locations in this region are collected and named to be \( L_1, \ldots, L_I \) respectively. Since the goal is to improve the transportation safety in a region by reducing the crash rate. The objective function is chosen to be the probability of having no accident in this region.

Maximum \( P(\bar{A}) \) \[1\]

where \( I \) is the total number of selected locations in a certain region, \( J \) is the number of scenarios available for implementing for each of these locations. Then by Bayes formula

\[ P(\bar{A}) = \sum_{i=1}^I \sum_{j=1}^J P(\bar{A}, L_i, S_j) = \sum_{i=1}^I \sum_{j=1}^J P(\bar{A} | L_i, S_j) P(L_i, S_j) \] \[1\]

Let \( Y_{L_j} = P(L_i, S_j) \), a binary variable to represent whether a scenario is implemented in a location or not. If \( Y_{L_j} = 1 \), then scenario j is implemented in location i, and when \( Y_{L_j} = 0 \), then scenario j is not implemented in location i.

The types of model we will use depend on the budget condition. If the budget is predetermined, then the goal is to maximize total safety improvement while satisfying the budget constraints. Let \( C_{L_j} \) be the cost to implementing scenario j in location i. Also notice that no more than one scenario is implemented in each location, then

\[ \text{Maximum } P(\bar{A}) = \text{Maximum } \sum_{i=1}^I \sum_{j=1}^J P(\bar{A} | L_i, S_j) Y_{L_j} \]

Subject to \( \sum_{j=1}^J Y_{L_j} \leq 1 \) for \( i=1, \ldots, I \)

\[ \sum_{i=1}^I \sum_{j=1}^J C_{L_j} Y_{L_j} \leq \text{Budget} \]
\[ y_{S_j}^{L_i} \text{'s are binaries for } i=1, \ldots, I \text{ and } j=1, \ldots, J \text{ [1]} \]

If the budget is not fixed, then the model would be modified to

\[
\text{Maximum } w_1 P(\bar{A}) - w_2 C = \text{Maximum } w_1 \sum_{i=1}^{I} \sum_{j=1}^{J} P(\bar{A}|L_{ij}, S_j) y_{S_j}^{L_i} - w_2 C
\]

Subject to \( \sum_{j=1}^{J} y_{S_j}^{L_i} \leq 1 \text{ for } i=1, \ldots, I \)

\( \sum_{i=1}^{I} L_{ij} c_{S_j} y_{S_j}^{L_i} \leq \text{Budget estimate} \)

\[ y_{S_j}^{L_i} \text{'s are binaries for } i=1, \ldots, I \text{ and } j=1, \ldots, J \text{ [1]} \]

This is a multi-objective optimization problem where \( w_1 \) is the weight of reducing the crash rate, \( w_2 \) is the weight of minimizing cost. The problem subject to the same group of constraints except the right hand side of the budget constraint is replaced by the estimated budget upper bound.

At this moment, if we were given the inputs to the optimization model, including the future budget information, the future candidate location information, and the future conditional probability for each scenario in each location, we will be able to conduct an optimization calculation with the given model and provide suggestions about how to efficiently implement transportation facilities to prevent traffic accident from happening. The key is how to calculate \( P(\bar{A}|L_{ij}, S_j) \). Intuitively, this conditional probability is the forecasted traffic accident counts in the coming year divided by the traffic flow of that year. We will discuss later in section 2.2.1.1 about how regression to the mean will affect such estimation. To avoid the influence of regression to the mean, define

\[
P(\bar{A}|L_{ij}, S_j) = 1 - \frac{\text{expected traffic accident counts/year}}{\text{AADT} \times 365} \quad (2.3)
\]

Here AADT refers to annual average daily traffic. Now the challenge would be how to estimate the expected accident counts for a certain entity with or without a particular
treatment.

2.2 Assessment of treatment

2.2.1 The Four-Step for before-and-after studies

To evaluate the effectiveness of the roadway safety improvements, several methods are proposed. A measure of the accident prior to the treatment(s) is obtained and compared with a similar measure obtained after the implementation of the countermeasure to estimate the effectiveness of this countermeasure. This is called the before-and-after study method. In general, there are three types of before-and-after study methods [4]:

- The simple before-and-after study method;
- The before-and-after study with Comparison Group (CG) method;
- The before-and-after study with the Empirical Bayes (EB) method.

Two tasks need to be carried out in the before and after study. First predict what would have been the safety of an entity in the after period had the treatment not been implemented, then estimate the safety of the treated entity in the after period after treatment. The following four steps need to be followed to calculate the necessary values [5] (a “~” is a symbol represents the estimated value):

(a) Estimate the expected reduction number of accident count of a specific entity, denoted as \( \hat{\delta} \), and calculated as

\[
\hat{\delta} = \hat{\eta} - \hat{\lambda} \quad [5],
\]

where \( \hat{\eta} \) is the estimated expected number of accident of this entity in the after period with no treatment, and \( \hat{\lambda} \) is the estimated expected number of accident of this entity in the after period after treatment.

(b) Estimate the variance of \( \hat{\delta} \), calculated as

\[
\text{vår}(\hat{\delta}) = \text{vår}(\hat{\eta}) + \text{vår}(\hat{\lambda}) \quad [5],
\]

since \( \hat{\eta} \) and \( \hat{\lambda} \) are statistically independent.
(c) Estimate the crash reduction factor, $\hat{\theta}$.

The crash reduction factor is defined as the expected accident count in the after period after treatment divide by the expected accident count in the after period without treatment. Namely, $\theta = \frac{\lambda}{\pi}$. Intuitively, it can be calculated by $\hat{\theta}^* = \frac{\hat{\lambda}}{\hat{\pi}}$. However, although $\hat{\pi}$ and $\hat{\lambda}$ are unbiased estimate of $\pi$ and $\lambda$, respectively, the faction $\frac{\hat{\lambda}}{\hat{\pi}}$ is biased estimation of $\theta$. An approximately unbiased estimate is given by

$$\hat{\theta} = \frac{\hat{\lambda}}{\hat{\pi}} [1 + \text{var}(\hat{\pi})/\hat{\pi}^2]$$

where $[1 + \text{var}(\hat{\pi})/\hat{\pi}^2]$ serves as a correction factor to remove the bias. [5] provides an incomplete proof, a more accurate proof is given below.

Proof of equation (2.6) as an unbiased estimate of $\theta$:

Since $\hat{\theta}^*$ is a function of $\hat{\pi}$ and $\hat{\lambda}$, denoted by $\hat{\theta}^* = f(\hat{\pi}, \hat{\lambda}) = \frac{\hat{\lambda}}{\hat{\pi}}$. Using the Taylor Expansion, we have:

$$f(\hat{\pi}, \hat{\lambda}) \equiv f(\pi, \lambda) + \frac{(\hat{\pi} - \pi)}{1!} \frac{\partial f(\hat{\pi}, \hat{\lambda})}{\partial \hat{\pi}} \bigg|_{\hat{\pi} = \pi} \frac{(\hat{\lambda} - \lambda)}{1!} \frac{\partial f(\hat{\pi}, \hat{\lambda})}{\partial \hat{\lambda}} \bigg|_{\hat{\lambda} = \lambda}$$

$$= \pi + \frac{(\hat{\lambda} - \lambda)}{1!} \frac{\partial f(\hat{\pi}, \hat{\lambda})}{\partial \hat{\lambda}} \bigg|_{\hat{\lambda} = \lambda}$$

$$= \pi + \frac{(\hat{\lambda} - \lambda)}{2!} \frac{\partial^2 f(\hat{\pi}, \hat{\lambda})}{\partial \hat{\pi}^2} \bigg|_{\hat{\pi} = \pi} \frac{\partial^2 f(\hat{\pi}, \hat{\lambda})}{\partial \hat{\lambda}^2} \bigg|_{\hat{\lambda} = \lambda}$$

(2.7)

Then

$$\text{E}(\hat{\theta}^*) = \text{E}(f(\hat{\pi}, \hat{\lambda})) \equiv f(\pi, \lambda) + \text{E}(\hat{\pi} - \pi) \frac{\partial f(\hat{\pi}, \hat{\lambda})}{\partial \hat{\pi}} \bigg|_{\hat{\pi} = \pi} \frac{\partial f(\hat{\pi}, \hat{\lambda})}{\partial \hat{\lambda}} \bigg|_{\hat{\lambda} = \lambda} + \frac{1}{2} \text{E}(\hat{\lambda} - \lambda)^2 \frac{\partial^2 f(\hat{\pi}, \hat{\lambda})}{\partial \hat{\pi}^2} \bigg|_{\hat{\pi} = \pi} \frac{\partial^2 f(\hat{\pi}, \hat{\lambda})}{\partial \hat{\lambda}^2} \bigg|_{\hat{\lambda} = \lambda}$$

$$= \pi + \frac{1}{2} \text{E}(\hat{\lambda} - \lambda)^2 \frac{\partial^2 f(\hat{\pi}, \hat{\lambda})}{\partial \hat{\pi}^2} \bigg|_{\hat{\pi} = \pi} \frac{\partial^2 f(\hat{\pi}, \hat{\lambda})}{\partial \hat{\lambda}^2} \bigg|_{\hat{\lambda} = \lambda}$$

Since $\frac{\partial^2 f(\hat{\pi}, \hat{\lambda})}{\partial \hat{\pi}^2} \bigg|_{\hat{\pi} = \pi} = \frac{\partial^2 f(\hat{\pi}, \hat{\lambda})}{\partial \hat{\lambda}^2} \bigg|_{\hat{\lambda} = \lambda} = 0$, $\text{E}(\hat{\pi} - \pi)^2 = \text{var}(\hat{\pi}) - (\text{E}(\hat{\pi} - \pi)^2 = \text{var}(\pi)$, and $\text{E}(\hat{\pi} - \lambda)^2 = \text{var}(\lambda)$, therefore
\[ E(\tilde{\theta}^*) = \frac{\lambda}{\pi} \left[ 1 + \frac{\text{var}(\hat{\pi})}{\pi^2} \right] \]  
(2.8)

Equation (2.8) indicates that if we estimate \( \theta \) by \( \tilde{\theta}^* = \frac{\hat{\pi}}{\pi} \) repeatedly, the result would be greater than the actual value by a factor of \( 1 + \frac{\text{var}(\hat{\pi})}{\pi^2} \). Therefore the correction factor need to be introduced into the expression which \( \tilde{\theta} = \frac{\hat{\pi}}{\pi} / [1 + \text{var}(\hat{\pi})/\pi^2] \).

(d) Estimate the variance of \( \tilde{\theta} \), which is calculated as

\[ \text{var}(\tilde{\theta}) = \left( \frac{\hat{\pi}}{\pi} \right)^2 \left[ \frac{\text{var}(\hat{\lambda})}{\lambda^2} + \frac{\text{var}(\hat{\pi})}{\pi^2} \right] / \left[ 1 + \frac{\text{var}(\hat{\pi})}{\pi^2} \right]^2 \]  
(2.9)

Note in [5], (2.9) is written as

\[ \text{var}(\hat{\theta}) = (\hat{\theta})^2 \left[ \frac{\text{var}(\hat{\lambda})}{\lambda^2} + \frac{\text{var}(\hat{\pi})}{\pi^2} \right] / \left[ 1 + \frac{\text{var}(\hat{\pi})}{\pi^2} \right]^2, \]  
but it is not accurate.

Proof of equation (2.9):

According to the previous deduction of (2.8), the question now will become that what will happen when we estimate \( \theta \) by \( \tilde{\theta} = \frac{\hat{\pi}}{\pi} / [1 + \text{var}(\hat{\pi})/\pi^2] \)?

\( \tilde{\theta} \) is a function of \( \hat{\pi} \) and \( \hat{\lambda} \), denoted by \( \tilde{\theta} = g(\hat{\pi}, \hat{\lambda}) = \frac{\hat{\pi}}{\pi} / [1 + \text{var}(\hat{\pi})/\pi^2] \). The value of \( [1 + \text{var}(\hat{\pi})/\pi^2] \) is usually close to 1. Representing this value as a constant “a”, then \( \tilde{\theta} = \frac{\hat{\pi}}{\pi} / a \). By equation (2.7)

\[ \text{Var}(\tilde{\theta}) = 0 + \text{Var}(\hat{\pi} - \pi) * \left( \frac{\partial g(\hat{\pi}, \hat{\lambda})}{\partial \pi} \right)^2 \left| \begin{array}{c} \hat{\pi} = \pi \\ \hat{\lambda} = \lambda \end{array} \right. + \text{Var}(\hat{\lambda} - \lambda) * \left( \frac{\partial g(\hat{\pi}, \hat{\lambda})}{\partial \lambda} \right)^2 \left| \begin{array}{c} \hat{\pi} = \pi \\ \hat{\lambda} = \lambda \end{array} \right. 
\]

Since \( \frac{\partial g(\hat{\pi}, \hat{\lambda})}{\partial \pi} = -\frac{\hat{\lambda}}{\pi^2 a} \), \( \frac{\partial g(\hat{\pi}, \hat{\lambda})}{\partial \lambda} = \frac{1}{\pi a} \), then

\[ \text{Var}(\tilde{\theta}) = \left( -\frac{\hat{\lambda}}{\pi^2 a} \right)^2 * \text{Var}(\hat{\pi}) + \left( \frac{1}{\pi a} \right)^2 * \text{Var}(\hat{\lambda}) = \left( \frac{\hat{\lambda}}{\pi} \right)^2 \left[ \frac{\text{var}(\hat{\pi})}{\pi^2} + \frac{\text{var}(\hat{\lambda})}{\lambda^2} \right] / a^2 \]

Therefore \( \text{var}(\tilde{\theta}) = \left( \frac{\hat{\lambda}}{\pi} \right)^2 \left[ \frac{\text{var}(\hat{\lambda})}{\lambda^2} + \frac{\text{var}(\hat{\pi})}{\pi^2} \right] / \left[ 1 + \frac{\text{var}(\hat{\pi})}{\pi^2} \right]^2 \).
2.2.1.1 Simple before-and-after study method

The essence of the simple before-and-after study is based on the assumption that the number of the accidents in the after period is expected to be the same as the before period if no improvement has been made [5]. To apply this method, we use the accident count before implementation to estimate what would have happened during the after period had the treatment not been implemented. For example, consider a simple before and after study with 100 accidents count in the before year and 66 accident count in the after year after treatment. By the assumption, if no treatment has been implemented, we would expect the same number of accident count in the after period. Hence $\hat{\pi} = 100$. Using the Four-Step, the effect of the treatment would be estimated to be $\hat{\delta} = \hat{\pi} - \lambda = 100 - 66 = 34$ accidents. Also assume the happen of the accident is Poisson distributed, therefore, $\text{var}(\hat{\delta}) = \text{var}(\hat{\pi}) + \text{var}(\lambda) = 100 + 66 = 166$. Also,

$$\hat{\theta} = \frac{\hat{\delta}}{\sqrt{\text{var}(\hat{\theta})} + \frac{\text{var}(\lambda)}{\pi^2}} = \frac{66}{100} \left[ \frac{1}{1 + \frac{100}{100^2}} \right] = 0.6666,$$

$$\text{var}(\hat{\theta}) = \left( \frac{\hat{\delta}}{\sqrt{\text{var}(\hat{\delta})} + \frac{\text{var}(\lambda)}{\pi^2}} \right)^2 = 0.66^2 \left[ \frac{66}{66^2} + \frac{100}{100^2} \right] = 0.0110.$$ 

The logic of the simple before-and-after study method is straight forward and easy to use. However, this method ignores several factors including regression to the mean, crash migration, maturation, as well as external causal factors [3] that can distort the estimates and lead to an inaccurate estimate.

Regression to the mean (RTM) is a statistical phenomenon that can make natural variation in repeated data looks like real change. It happens when unusual large or small measurements tend to follow measurements that are close to the mean [6]. This problem is the most
frequently cited problem in before-and-after studies. Usually the location with the large number of crashes will be selected for treatment. Because of the existence of RTM, the extreme crash frequencies would likely be followed by less extreme values even when the countermeasure is completely ineffective. In this case, we may overestimate the effectiveness of the improvement. The following examples will show how the regression to the mean can affect the result of the analysis.

Figure 2.1 shows a pattern called the regression to the mean effect in transportation studies. Suppose in a certain location i, we manually divide the before period and the after period by the beginning of year 3, and each period a year long. (Before period- beginning of year 2 to beginning of year 3; after period-beginning of year 3 to beginning of year 4) The variation of the accident frequencies in years represents the natural change of the crash number around the mean. Notice that no treatment has been implemented in this location throughout the study period. At the end of year 2, we observe accident count a. At the end of year 3, the observation is b and b<a. By the logic of simple before and after method, we would estimate the expected accident count to be “a” in the after period if there was no treatment. However, the actual value is b. This difference indicates that even if the treatment is totally ineffective, there could still be reduction in accident number. As a result, we may overestimate (or underestimate) the effect of the treatment and produce an inaccurate estimate result.

This issue is illustrated with a specific example: Imagine a city with 100 two lane highways
segments that were equipped with rumble strips at the end of year 2005. Assume that for each of these road segments, the expected number of accidents is 3 at year 2005. In practice, the observations would vary from site to site. Suppose that the accident count was Poisson distributed with mean 3, then \( f(K|\lambda = 3) = \frac{e^{-3\lambda}3^K}{K!} \). The observation values can be estimated. For example, the probability that a site would expect to have 0 accidents at year 2005 would be \( f(K = 0|\lambda = 3) = \frac{e^{-3\lambda}3^0}{0!} = 0.0498 \) and the expected number of sites with 0 accidents can be estimated by \( 0.0498 \times 100 = 5 \). Any other probabilities with certain accident counts can be calculated in the same fashion. Below is a list of results:

**Table 2.1. Expected number of sites with K accidents**

<table>
<thead>
<tr>
<th>Accident Count(K)</th>
<th>Prob. That site has K accidents</th>
<th>Expected number of sites with K accidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0498</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>0.1494</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>0.2240</td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>0.2240</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td>0.1680</td>
<td>17</td>
</tr>
<tr>
<td>5</td>
<td>0.1008</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>0.0504</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>0.0216</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>0.0081</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>0.0027</td>
<td>1</td>
</tr>
<tr>
<td>10 and above</td>
<td>0.0008</td>
<td>-</td>
</tr>
</tbody>
</table>

At the end of year 2005, the transportation planners decided to equip the rumble strips on the road segments with accident count greater than 5. Table 2.1 shows that there are \((10+5+2+1+1=19) \) road segments in total that recorded more than 5 accidents in year 2005. Suppose this strategy reduces the correctable accident count by 10%. Then one would estimate there are \(0.9 \times 3 = 2.7 \) accidents expected to occur on each of these 19 road segments. However, have we estimated by the simple before and after study method, the expected number of before accident counts would be \( 5 \times 10+6 \times 5+7 \times 2+8 \times 1+9 \times 1=111 \), and the expected number of after accident counts would be \( 2.7 \times 19=51.3 \). Hence, the estimated crash reduction would be \( \hat{\theta} = (51.3/111)/(1+1/111)=0.46 \). The crash reduction factor appears to be \( 1-0.46=54\% \). However, there is only 10\% of reduction. The difference comes from the fact that this 19 road segments had recorded an unusually high number of accidents. Actually, this
is exactly why they were selected to get improved. In this example, the existence of RTM tends to be overestimate the effect of the traffic safety strategies.

The pervious example shows how the RTM Phenomenon will become a disturbing factor. In fact, it is the most pervasive problem in before-and-after studies and needs to be settled properly. Other than the RTM effect, the potential existence of other disturbing factors can also be destructive and attention needs to be paid to them.

Crash migration is the phenomenon that the crash rate or crash severity apparently rises at the untreated sites but adjacent to treated sites as a result of the treatment [3]. When crash migration occurs, crash rates in the treated sites may decline whereas they may increase in the surrounding area. Boyle and Wright (1984) first pointed out the potential existence of the crash migration [7]. Many researchers tried to demonstrate the existence of this phenomenon; some find no evidence to support it. Elvik (1997), reviewing United Kingdom and other accident studies, found that very little of accident reduction could be directly attributed to this factor [8]. If this was a genuine effect, attention should be paid to individual links within a roadwork, which can help avoid missing potential system-wide effects.

Maturation refers to the effect of collision trends over time[9]. For example, in a treated site, the crash frequency is reduced between the before and the after period, this change could fully or partially be due to an extension of a continuing decreasing trend which has been occurring for years. Maturation could be another disturbing factor in the study that it overestimates the effectiveness of the improvement. The difference between maturation and regression to the mean is that maturation occurs due to change in external factors such as traffic flow, economy and weather conditions. While regression to the mean is merely a statistical phenomenon that occurs whenever you have a nonrandom sample from a population and two measures that are imperfectly correlated [10].

2.2.1.2 The before-and-after study with Comparison Group method

The simple before and after study cannot distinguish between what is caused by the treatment
and what is caused by many other influences such as regression to the mean, maturation and crash migration. The comparison group method was developed to solve the maturation and external causal factors. This method can potentially provide more accurate estimates than the simple before and after method.

The comparison group is a group of sites that have similar traffic or geometric conditions as the treated sites [5]. Conceptually, the comparison group method estimates the number of crashes that would have been occurred if no improvements have been made at a treatment site in the after period. Hauer claims that this method is based on two assumptions [5]: First, that the sundry factors that affect safety have changed from the before to the after period in the same manner on both the treatment and the comparison group. Second, this change in the sundry factors influences the safety of both groups in the same way.

The C-G method is based on the hope that, without the implementation of the treatment, the ratio of the accident count of the before and after period in the treatment site should be the same as in the comparison group. In the table below K, M, L, N denote the observed accident count in different periods, and κ, μ, λ, ν denote the corresponding expectation values [5].

Table 2.2 Observed and expected accident counts in comparison group method

<table>
<thead>
<tr>
<th>Accident count and expected values</th>
<th>Treatment Group</th>
<th>Comparison Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>K, κ</td>
<td>M, μ</td>
</tr>
<tr>
<td>After w/o treatment</td>
<td>π</td>
<td>N, ν</td>
</tr>
<tr>
<td>After w treatment</td>
<td>L, λ</td>
<td>--</td>
</tr>
</tbody>
</table>

Define: \( r_c \equiv \nu/\mu \) to be the ratio of the expected accident count for the comparison group. \( r_t \equiv \pi/\kappa \) to be the ratio of the expected accident count for the treatment group.

Our hope is \( r_c = r_t \). Hence, \( \pi = r_t \kappa = r_c \kappa \). The estimate of \( \pi \) now requires information about \( r_c \) and \( \kappa \). The algorithm for comparison before and after study method is listed below. The proof of the before and after study with comparison group method is similar to the simple
before and after study method.

\[
\hat{\lambda} = L \quad [5],
\]

\[
\hat{\tau}_t = \hat{\tau}_c = \frac{N}{1 + \frac{M}{N}} \quad [5],
\]

\[
\hat{\pi} = \hat{\tau}_c \cdot K \quad [5],
\]

\[
\text{vår}(\hat{\lambda}) = L \quad [5],
\]

\[
\frac{\text{vår}(\hat{\tau}_t)}{\hat{\tau}_t^2} = \frac{1}{M} + \frac{1}{N} + \text{vår}\left(\frac{\hat{\tau}_c}{\hat{\tau}_t}\right) \quad [5],
\]

\[
\text{vår}(\hat{\pi}) = \hat{\pi}^2 \left[\frac{1}{K} + \frac{\text{vår}(\hat{\tau}_t)}{\hat{\tau}_t^2}\right] \quad [5].
\]

Equations 2.10-2.15 serve as building blocks for the Four-Step. From them we get the estimate of \( \lambda \) and \( \pi \) and their variance. To finish the comparison before-and-after study, we then need to follow the Four Step which listed in equation (2.4)-(2.9). The detailed proof for equation (2.10)-(2.15) can be found in book [5], from page 125 to page 127.

The before-and-after study with comparison group method is considered a better approach than the simple method because it accounts the effect of maturation. However, the accuracy of this method highly depends on the availability of comparison sites and the similarity between the comparison and the treatment site [3].

2.2.1.3 The before-and-after study with the Empirical Bayes (EB) method

*Bayes and Empirical Bayes*

First, the fundamental concepts of the Bayes and Empirical Bayes Theorem will be introduced. In probability theorem, Bayes Theorem shows the relationship between the conditional probability and its inverse. Bayesian analysis depends on a prior distribution. The Empirical Bayes approach uses the observed data to estimate the prior and then proceeds as though the prior was known [11]. Below is the definitions and Bayes theorem which serves as a core concept of the case study are listed [12].
Definition 1

If $T$, is a statistic, $T = t(x_1, x_2, \ldots x_n)$, is an estimator of $\tau(\theta)$, then the loss function $L(T; \theta) \geq 0$ for all $t$, and $L(T; \theta) = 0$ when $t = \tau(\theta)$.

Definition 2

The risk function is defined as the expected loss $R_T(\theta) = E_{x|\theta}[L(T; \theta)]$.

Definition 3

For a random sample from $f(x; \theta)$, the Bayes Risk of an estimator $T$ related to a risk function $R_T(\theta)$ and pdf $p(\theta)$ is the average risk with respect to $p(\theta)$

$$A_T = E[R_T(\theta)] = \int R_T(\theta) p(\theta) d\theta.$$

Definition 4

For a random sample from $f(x; \theta)$, the Bayes Estimator $T^*$ relative to the risk function $R_T(\theta)$ and pdf $p(\theta)$ is the estimator with respect to the minimum expected risk

$$A_{T^*} = E[R_{T^*}(\theta)] \leq A_T.$$

Definition 5

The conditional density $\theta$ given the sample observations $X = (x_1, x_2, \ldots x_n)$ is called the posterior density/pdf, is given by:

$$f(\theta|x_1, x_2, \ldots x_n) = \frac{f(\theta, x_1, x_2, \ldots x_n)}{f(x_1, x_2, \ldots x_n)} = \frac{f(x_1, x_2, \ldots x_n|\theta)p(\theta)}{\int f(x_1, x_2, \ldots x_n|\theta)p(\theta)d\theta}.$$

For a single observation

$$f(\theta|x) = \frac{f(x|\theta)p(\theta)}{\int f(x|\theta)p(\theta)d\theta}.$$

Bayes Theorem

The Bayes estimator $T$, of $\tau(\theta)$ under the squares error loss function, $L(t; \theta) = [t - \tau(\theta)]^2$ is the conditional mean of $\tau(\theta)$ relative to the posterior distribution.
\[ T^* = \mathbb{E}_{\theta|x} [\tau(\theta)] = \int \tau(\theta)f(\theta|x)d\theta. \]

Proof: \[ A_T = \int R_T(\theta)p(\theta)d\theta = \int \int [T - \tau(\theta)]^2 f(x|\theta)p(\theta)dx d\theta \]

\[ = \int \left[ \int f(\theta|x)p(\theta)d\theta - 2T\int \tau(\theta)f(x|\theta)p(\theta)d\theta + \int \tau^2(\theta)f(x|\theta)p(\theta)d\theta \right] dx \]

\[ = \int \left[ \int f(x|\theta)p(\theta)d\theta \left[ T - \frac{\int \tau(\theta)f(x|\theta)p(\theta)d\theta}{\int f(x|\theta)p(\theta)d\theta} \right]^2 - \frac{\left( \int \tau(\theta)f(x|\theta)p(\theta)d\theta \right)^2}{\int f(x|\theta)p(\theta)d\theta} + \int \tau^2(\theta)f(x|\theta)p(\theta)d\theta \right] dx \]

Then \( A_T \) is minimized when

\[ T^* = \frac{\int \tau(\theta)f(x|\theta)p(\theta)d\theta}{\int f(x|\theta)p(\theta)d\theta} = \int \tau(\theta)f(\theta|x)d\theta. \]

**Before and after study with Empirical Bayes method**

The most critical part in the before-and-after study is to estimate what would have been the crash frequency if there were no treatments implemented in the after period, which is denoted by \( \bar{\tau} \) in the Four Step. Both of the simple and the CG before and after approach are based on the assumption that for any treated entity, the before accident count \( K \) is a sensible estimate for the expected after accident count \( \kappa \) with no treatment. This is not necessarily the case. If an entity is treated for its unusually high accident count, then this accident count would not be a good estimate of its expected accident count \( \kappa \) in the after period. The reason is statistically straightforward, since an unusual accident count cannot be a good estimate for the usual case.

In the transportation safety studies, a traffic entity is more likely to be treated due to unusually high accident count[5]. This causes the so called “selection bias” or “regression to the mean” bias.

The Empirical Bayes approach is designed to eliminate the regression to the mean bias. The essence of the EB approach is that it uses two different kinds of clues to estimate the safety of an entity [5]. Clues of the first kind contain in the traits of the safety entity. A few are the traffic flow, road condition, weather condition. Clues of the second kind are derived from the history of accident occurrence, including the number of accidents.
To use both clues to estimate $\kappa$ for a certain entity by EB method, first identify which reference population that the entity belongs to—the entities that have expected number of accident count $\kappa$ in the after period with mean $E(\kappa)$ and variance $\text{Var}(\kappa)$; Second, select the entities from the reference population that record $K$ accidents in the before period. Let $E(\kappa|K)$ and $\text{Var}(\kappa|K)$ denote the mean and the variance in this “sub-population.” [5]

The steps to apply EB method is listed as follow. Intuitively, $E(\kappa|K)$ will be decided by both $E(\kappa)$ and $K$. Actually, the value of $E(\kappa|K)$ is a combination of $E(\kappa)$ and $K$, which will be proved later that

$$E(\kappa|K) = \alpha E(\kappa) + (1 - \alpha)K \ [5].$$

(2.16)

In this expression $\alpha$ is a number between 0 and 1. To estimate the $\kappa$ of the entity with maximum precision

$$\alpha = \frac{1}{1 + \frac{\text{Var}(\kappa)}{E(\kappa)}} \ [5].$$

(2.17)

Two assumptions are listed below in order to get the parameter $\alpha$ in our case.

1. The expectation of crash counts is gamma distributed, which is $g(\kappa) = \frac{b^b}{\Gamma(b)} \kappa^{b-1} e^{-a\kappa}$, where $E(\kappa) = \frac{b}{a}$, $\text{Var}(\kappa) = \frac{b}{a^2}$ [5].

2. The observation of the crash counts given its expectation is Possion distributed, denote by $\pi(K|\kappa) = \frac{\kappa^K e^{-\kappa}}{K!}$ [5].

Hauer did not give a complete deduction in his book. A theoretical deduction about the formula $E(\kappa|K) = \alpha E(\kappa) + (1 - \alpha)K$[5] is shown below:

By the Definition 5,

$$f(\kappa|K) = \frac{P(K|\kappa) \cdot P(\kappa)}{P(K)} = \frac{\pi(K|\kappa) \cdot g(\kappa)}{\int \pi(K|u) \cdot g(u) \, du}$$

and according to the assumptions,

$$f(\kappa|K) = \frac{\pi(K|\kappa) \cdot g(\kappa)}{\int \pi(K|u) \cdot g(u) \, du} =$$
Therefore, the expected accident count given its observation is \( \text{GAM}(\frac{K+b}{1+a}, \frac{K+b}{(1+a)^2}) \).

Let \( \alpha = \frac{1}{1 + \frac{\text{Var}(\kappa)}{\mu(\kappa)}} = \frac{a}{1+a} \), then \( E(\kappa|K) = \frac{K+b}{1+a} = \alpha E(\kappa) + (1 - \alpha)K \), and \( \text{Var}(\kappa|K) = (1 - \alpha)E(\kappa|K) \).

Had we estimated \( \kappa \) in usual way, using only the history of its accident occurrence \( K \), the value of \( \alpha \) should be 0. And by the formula \( \text{Var}(\kappa|K) = (1 - \alpha)E(\kappa|K) \), the variance should be \( E(\kappa|K) \). If we use both clues of the safety, since \( 0 < \alpha = \frac{1}{1 + \frac{\text{Var}(\kappa)}{\mu(\kappa)}} < 1 \), variance \( \text{Var}(\kappa|K) \) never exceed \( E(\kappa|K) \) and is always smaller.

The same result can be produced by using the Bayes Theorem with a single observation.

The Bayes Estimator

\[
\hat{\kappa} = E(\kappa|K) = \frac{\int K \pi(\kappa|K)g(\kappa) \, d\kappa}{\int \pi(\kappa|K)g(\kappa) \, d\kappa} = \frac{\int K^{a+b} e^{-\kappa a} e^{-\kappa(K+b)} \, d\kappa}{\int K^{a+b} e^{-\kappa a} e^{-\kappa(K+b)} \, d\kappa} = \frac{\Gamma(K+b+1)}{\Gamma(K+b)} \frac{1}{1+a} = \frac{K+b}{1+a},
\]

then \( E(\kappa|K) = \frac{K+b}{1+a} = \alpha E(\kappa) + (1 - \alpha)K \).

Example [2]: There are 2 accident counts in a 5 year period in a certain location. The multivariate method suggests that \( \bar{E}(\kappa) = 0.0239/\text{year}, \text{Var}(\kappa) = 0.0011/\text{year} \). What’s the estimate of \( \kappa \) ?

For a 5 year period, \( \bar{E}(\kappa) = 0.0239 \times 5 = 0.1195, \text{ Var}(\kappa) = 0.0011 \times 5 = 0.0075 \)

\[
\hat{\alpha} = \frac{1}{1 + \frac{\text{Var}(\kappa)}{\bar{E}(\kappa)}} = \frac{1}{1 + \frac{0.0275}{0.1195}} = 0.81, \]

\[
\hat{\kappa} = \bar{E}(\kappa|K) = \hat{\alpha} \bar{E}(\kappa) + (1 - \hat{\alpha})K = 0.81 \times 0.1195 + 0.19 \times 2 = 0.48. \]

\[
\text{Var}(\kappa|K) = (1 - \hat{\alpha})\bar{E}(\kappa|K) = 0.19 \times 0.48 = 0.09. \]

Notice that had only the accident count been used, the estimate of \( \kappa \) would have been 2.
accidents in 5 years and the standard deviation of that estimate would be estimated at \( \sqrt{2} = 1.4 \) accidents in 5 years. The EB approach makes use of both of these clues to produce a more accurate, location-specific safety estimate.

Now the focus will on be how to estimate \( E(\kappa) \) and \( \text{Var}(\kappa) \). By Adam’s formula in probability theory, \( E(K) = E(E(K|\kappa)) \). By definition \( E(K|\kappa) = \kappa \), therefore

\[
E(K) = E(\kappa) \tag{2.18}
\]

By Eve’s formula in probability theory, \( \text{Var}(K) = E(\text{Var}(K|\kappa)) + \text{Var}(E(K|\kappa)) \). Since observation of the crash counts given a its expectation is poison distributed, \( \text{Var}(K|\kappa) = E(K|\kappa) = \kappa \), therefore

\[
\text{Var}(K) = E(\kappa) + \text{Var}(\kappa) \tag{2.19}
\]

An example:
Consider a reference population of two lane highway in rural area with speed limits 70 m/h. To do an accurate calculation, we need the sample size to be large.

Let \( \bar{K} \) be the average accident count among the reference population. \( \bar{K} \) is the sample mean and \( \bar{K} = \sum K \times n(K)/n \) where \( n(K) \) is the number of locations that records \( K \) accidents in this year. \( n \) is the total number of accident among the reference group. The sample variance \( S^2 \) is defined as \( S^2 = \sum (K - \bar{K})^2 \times n(K)/n \). As \( n \) becomes large, \( \bar{K} \rightarrow E(K) \) and \( S^2 \rightarrow \text{Var}(K) \). Therefore, if we know the occurrence of the accident in a certain year we are able to do the estimation for \( E(\kappa) \) and \( \text{Var}(\kappa) \). Below is a table shows the observation and the calculation.

<table>
<thead>
<tr>
<th></th>
<th>( K )</th>
<th>( n(\kappa) )</th>
<th>( K \times n(\kappa) )</th>
<th>( (K - \bar{K})^2\times n(\kappa) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5930</td>
<td>0</td>
<td>0</td>
<td>3.416</td>
</tr>
<tr>
<td>1</td>
<td>120</td>
<td>120</td>
<td>144</td>
<td>114.309</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>8</td>
<td>32</td>
<td>15.618</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>15</td>
<td>75</td>
<td>44.283</td>
</tr>
<tr>
<td>Total</td>
<td>6059</td>
<td>143</td>
<td>858</td>
<td>177.626</td>
</tr>
</tbody>
</table>
Therefore by using relations 2.18 and 2.19, \( \hat{E}(\kappa) = \bar{E}(K) = 0.024 \) and \( \text{Var}(\kappa) = \text{Var}(K) - \hat{E}(\kappa) = 0.029 - 0.024 = 0.005 \). Therefore \( \hat{\alpha} = \frac{1}{1 + \frac{\text{Var}(\kappa)}{\hat{E}(\kappa)}} = \frac{1}{1 + \frac{0.005}{0.024}} = 0.8276 \). Now the estimated conditional mean can be calculated. For instance, for all the locations in the reference group which recorded \( K = 2 \) in a certain year, then we would estimate \( \hat{E}(\kappa|K) = \hat{\alpha}\bar{E}(\kappa) + (1 - \hat{\alpha})K = 0.8276 \times 0.024 + (1 - 0.8276) \times 2 = 0.3647 \) in that year. We should also pay attention to the variance of \( \kappa \). It is estimated to be 0.005, which is relatively small when compare to \( \hat{E}(\kappa) \). This is partly because that the size of the reference group is large. In some cases, the reference group is not large enough to make an accurate estimation by the method of the sample moments. To address this problem the multivariate regression method is introduced here [3].

Let \( X_1, X_2, ..., X_n \) be the independent variables of the reference sites, such as AADT, road section length, or number of lanes, which are believed to be the most important factors for the occurrence of accidents. Assume that \( K \) mostly depends on these independent variables, and their relationships with \( \kappa \) is exponential.

\[
K = \beta_0 + X_1^{\beta_1} + X_2^{\beta_2} + \ldots + X_n^{\beta_n} + \varepsilon \quad \text{where} \quad \beta_0, \beta_1, ..., \beta_n \quad \text{are parameters of the independent variables, and} \quad E(\varepsilon) = 0
\]

Therefore:

\[
E(K) = \beta_0 + X_1^{\beta_1} + X_2^{\beta_2} + \ldots + X_n^{\beta_n} \quad [1]
\]

\( \text{Var}(K) \) can be estimated by the maximum likelihood estimate.

The multivariate method is better than the method of sample moments in the following two aspects. First, a large number of reference sites are not needed for any particular combination of characteristics. And second, it provides estimates of \( \text{Var}(\kappa) \) as well as \( E(\kappa) \) for the reference sites.

In summary, had we got the historical information from the treatment site and the reference...
site in the before period, we could do the EB estimate for value $\hat{\kappa} = \hat{E}(\kappa|K)$. Note that this $\hat{\kappa}$ is also for the before period. The next thing is to estimate what would happen if the treatment site remains untreated in the after period. Let $\kappa_b$ be the expected accident count in the before year, usually a year before implementation take place. Also, let $\kappa_a$ be the expected accident count in the after year without treatment. Since the observation in the after period with no treatment is no longer available, then there is no way to distinguish $\kappa_a$ with $\kappa_b$, except the performance of their reference group during different periods. Then it is reasonable to introduce the formula here:

$$\hat{\kappa}_a = \frac{E(\kappa_a)}{E(\kappa_b)} \times \hat{\kappa}_b \quad [5]$$

(2.21)

The loop is now closed. Had we got the estimate of $\kappa_a$, the expected accident count in the after year without treatment for the treatment site, and the observation of accident frequency in the after year, we would be able to conduct a before-and-after study using the Four-Step to measure the effectiveness of the treatment.

The EB before-and-after study method cures the RTM problem. The reasons are as following. First, it is said that the conceptual frame of the EB method fits the reality of observational study[2]. We estimate $\kappa$ by calculating $E(\kappa|K)$, the mean of the $\kappa$'s in the subpopulation. When using EB method, there is a two-stage selection process. The first stage is to identify the group of locations with similar traits and remain untreated throughout the study period—the reference population. Next from the reference population, select the subpopulation with $K$ accident counts. This way of estimation is based on the belief that locations that record $K$ accident counts have a mean that different from the locations with $L$ accidents. $K \neq L$, then $E(\kappa|K) \neq E(\kappa|L)$.

Second, since $E(\kappa|K) = \alpha E(\kappa) + (1 - \alpha)K$ and $0 < \alpha = \frac{1}{1+\frac{\text{var}(\kappa)}{E(\kappa)}} \leq 1$, then $E(\kappa|K)$ is always between $E(\kappa)$ and $K$. Thus at least qualitatively, $E(\kappa|K)$ does what the logic of RTM predicts. Namely it shifts the estimate of accident count in the direction of the population mean. Hauer used a real world problem to show that the reasons are logically
sound. The example appears to be incomplete in the book. Below is a more complete explanation.

The Table 2.4 is based on the accident report of 1139 intersections in San Francisco in year 1974 and 1975[5]. The total accident count in year 1974 is 1211, the same as in year 1975.

Table 2.4[5] Juxtaposition of EB estimates for 1974 and observed average in 1975

<table>
<thead>
<tr>
<th>n(K) of 1974</th>
<th>K of 1974</th>
<th>( \hat{K} ) of 1974</th>
<th>Avg(K) of 1975</th>
</tr>
</thead>
<tbody>
<tr>
<td>553</td>
<td>0</td>
<td>0.48</td>
<td>0.54</td>
</tr>
<tr>
<td>296</td>
<td>1</td>
<td>1.03</td>
<td>0.97</td>
</tr>
<tr>
<td>144</td>
<td>2</td>
<td>1.57</td>
<td>1.53</td>
</tr>
<tr>
<td>65</td>
<td>3</td>
<td>2.11</td>
<td>1.97</td>
</tr>
<tr>
<td>31</td>
<td>4</td>
<td>2.65</td>
<td>2.10</td>
</tr>
<tr>
<td>21</td>
<td>5</td>
<td>3.19</td>
<td>3.24</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>3.73</td>
<td>5.67</td>
</tr>
<tr>
<td>13</td>
<td>7</td>
<td>4.27</td>
<td>4.69</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>4.81</td>
<td>3.80</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>5.35</td>
<td>6.50</td>
</tr>
</tbody>
</table>

The first column in Table 2.4 is the number of the reference locations that recorded K accidents in 1974. The last column is the average accident counts in year 1975 of these n(K) locations. The third column in the table lists the estimate of \( \kappa \) of year 1974 using EB method.

To get the value of column 3, we need to do an EB estimate like the following:

Table 2.5 Juxtaposition of EB estimates for 1974

<table>
<thead>
<tr>
<th>n(K) (1974)</th>
<th>K(1974)</th>
<th>K*n(K)</th>
<th>( (K - \hat{K})^2 \cdot n(K) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>553</td>
<td>0</td>
<td>0</td>
<td>621.35</td>
</tr>
<tr>
<td>296</td>
<td>1</td>
<td>296</td>
<td>1.07</td>
</tr>
<tr>
<td>144</td>
<td>2</td>
<td>288</td>
<td>127.24</td>
</tr>
<tr>
<td>65</td>
<td>3</td>
<td>195</td>
<td>244.63</td>
</tr>
<tr>
<td>31</td>
<td>4</td>
<td>124</td>
<td>267.95</td>
</tr>
<tr>
<td>21</td>
<td>5</td>
<td>105</td>
<td>326.00</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>54</td>
<td>219.63</td>
</tr>
<tr>
<td>13</td>
<td>7</td>
<td>91</td>
<td>458.69</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>40</td>
<td>240.82</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>18</td>
<td>126.09</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td>1211</td>
<td>2633.64</td>
</tr>
</tbody>
</table>
\[ \hat{E}(\kappa) = \tilde{E}(K) = \tilde{K} = \frac{1211}{1139} = 1.06, \]

\[ \text{Var}(K) = S^2 = \frac{2633.64}{1139} = 2.31, \]

\[ \text{Var}(\kappa) = \text{Var}(K) - \hat{E}(K) = S^2 - \tilde{K} = 2.31 - 1.06 = 1.25, \]

\[ \hat{\alpha} = \frac{1}{1 + \frac{\text{Var}(\kappa)}{\hat{E}(\kappa)}} = \frac{1}{1 + \frac{1.25}{1.06}} = 0.46. \]

Therefore, if \( K = 4 \), then \( \hat{K} = \hat{E}(\kappa|K) = \hat{\alpha}\hat{E}(\kappa) + (1 - \hat{\alpha})K = 0.46 \times 1.06 + 0.54 \times 4 = 2.65. \)

The other values can be calculated similarly.

The next step is to predict what would happen had these 1124 locations remained untreated.

The formula (2.21) \( \hat{K}_a = \frac{\hat{E}(\kappa_a)}{\hat{E}(\kappa_b)} \hat{K}_b \) is used. In the formula, “a” refers to year 1974, “b” refers to year 1975. In this case, \( \frac{\hat{E}(\kappa_{1972})}{\hat{E}(\kappa_{1974})} = \frac{1211}{1139} = 1 \), therefore \( \hat{K}_{1975} = \hat{K}_{1974}. \)

The simple before-and-after study method would use only the first year accident count to predict the happen of accidents in the after period, while EB method uses both the accident history and the crash information from the reference population. It is necessary to compare these two types of estimates. Namely, column 2 compares to column 3 in table 2.4. The idea is, check which estimate is closer to the average accidents frequency in 1975. The result are shown in figure 2.2
As shown above, the ordinate of each diamond is the value of simple B+A estimate of what would happen in year 1975, the ordinate of triangle is the EB B+A estimate of what would happen in year 1975. Both these two estimates are plotted against \( \text{avg}(K) \) for 1975. The line is a standard line where \( K \) and estimate of kappa=Average of \( K \). Except one point where \( K=6 \), all the other points of EB B+A estimates are shown to be closer to the standard line than the simple B+A estimates. That means in this example, EB method tends to be more accurate than simple before-and–after method.

A case study [5]

The California office of traffic safety decided to provide a grant to the San Francisco Department of Public Works to conduct a before and after study for 49 intersections in 1973. These 49 intersections were converted from two-way to four-way stop control in the period 1969-1972. Using comparison group method, the reduction in total appears to be 71%. Since the original analysis did not account for the RTM, Hauer suspected that the original analysis
might exaggerate the result. So he decided to conduct an Empirical Bayes before-and-after study to these 49 intersections.

First, he collected the additional data needed to remove the bias. In this case, intersections with two-way stop control. The information needed is listed below:

Using the above information we can calculate $\hat{E}(\kappa|K)$ for any $K$ by using relations

$$\hat{R} = \hat{E}(\kappa|K) = \hat{\alpha}\hat{E}(\kappa) + (1 - \hat{\alpha})K.$$  
For example, if $K$ in 1969 was 4, then $\hat{E}(\kappa|K)$ for 1969 is $0.357 \times 0.966 + (1 - 0.357) \times 4 = 2.92$. So, if an intersection recorded 4 accidents in year 1969, we would estimate its $\kappa$ to be 2.92. Corresponding values for other years and $K$’s are given in Table 2.7

Among the 49 intersections, 3 intersections recorded 4 right angle accidents at year 1969. And they were converted to four way stop control at the end of the year. Now the question would be what would their $\kappa$ be in year 1970, had they remained untreated?

Since the sample mean was 0.996 in year 1969 and 0.925 in year 1970, our prediction for $\kappa$ in 1970 should be, by equation (2.21), $\hat{R}_a = \frac{\hat{E}(\kappa_a)}{\hat{E}(\kappa_b)} \times \hat{R}_b = 0.925/0.966 \times 2.92 = 2.79$. This value was listed in Table 2.6. The number in the parentheses is the observed accident number in the

Table 2.6 Sample mean variance and $\hat{\alpha}$ for right angle accidents

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.35</td>
<td>0.34</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td>1</td>
<td>0.99</td>
<td>0.97</td>
<td>0.96</td>
<td>0.94</td>
</tr>
<tr>
<td>2</td>
<td>1.63</td>
<td>1.60</td>
<td>1.57</td>
<td>1.54</td>
</tr>
<tr>
<td>3</td>
<td>2.27</td>
<td>2.23</td>
<td>2.19</td>
<td>2.14</td>
</tr>
<tr>
<td>4</td>
<td>2.92</td>
<td>2.86</td>
<td>2.81</td>
<td>2.74</td>
</tr>
<tr>
<td>5</td>
<td>3.56</td>
<td>3.49</td>
<td>3.42</td>
<td>3.34</td>
</tr>
<tr>
<td>6</td>
<td>4.20</td>
<td>4.12</td>
<td>4.04</td>
<td>3.94</td>
</tr>
<tr>
<td>7</td>
<td>4.84</td>
<td>4.75</td>
<td>4.65</td>
<td>4.54</td>
</tr>
<tr>
<td>8</td>
<td>5.49</td>
<td>5.39</td>
<td>5.27</td>
<td>5.14</td>
</tr>
</tbody>
</table>

Table 2.7 Estimates of $E(\kappa|K)$ in the before years

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.35</td>
<td>0.34</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td>1</td>
<td>0.99</td>
<td>0.97</td>
<td>0.96</td>
<td>0.94</td>
</tr>
<tr>
<td>2</td>
<td>1.63</td>
<td>1.60</td>
<td>1.57</td>
<td>1.54</td>
</tr>
<tr>
<td>3</td>
<td>2.27</td>
<td>2.23</td>
<td>2.19</td>
<td>2.14</td>
</tr>
<tr>
<td>4</td>
<td>2.92</td>
<td>2.86</td>
<td>2.81</td>
<td>2.74</td>
</tr>
<tr>
<td>5</td>
<td>3.56</td>
<td>3.49</td>
<td>3.42</td>
<td>3.34</td>
</tr>
<tr>
<td>6</td>
<td>4.20</td>
<td>4.12</td>
<td>4.04</td>
<td>3.94</td>
</tr>
<tr>
<td>7</td>
<td>4.84</td>
<td>4.75</td>
<td>4.65</td>
<td>4.54</td>
</tr>
<tr>
<td>8</td>
<td>5.49</td>
<td>5.39</td>
<td>5.27</td>
<td>5.14</td>
</tr>
</tbody>
</table>

25
last year 1967. Following the same fashion, we can fill in the table with many estimations in the after years and their observations in before year. The numbers in the parentheses add up to 49.

Table 2.8 Predictions of \( k \) for right-angle accidents in the after year.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.33;(2)</td>
<td>0.33;(2)</td>
<td>0.32;(5)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.56;(2)</td>
<td>0.93;(1)</td>
<td>0.91;(1)</td>
<td>0.89;(1)</td>
</tr>
<tr>
<td>2</td>
<td>2.18;(1)</td>
<td>1.53;(3)</td>
<td>1.50;(5)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2.79;(3)</td>
<td>2.13;(2)</td>
<td>2.09;(4)</td>
<td>2.03;(1)</td>
</tr>
<tr>
<td>4</td>
<td>3.34;(2)</td>
<td>3.26;(2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.94;(1)</td>
<td>3.85;(1)</td>
<td>3.74;(1)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4.54;(1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5.25;(2)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Therefore, had the intersections remain untreated, the expected accident counts in the after period will add up to 0.33*2+1.56*2+2.18*1+…+3.74*1=93.05 right-angle-accidents. The actual accident number in these 49 intersections in the after year is 16. Therefore, there seems to have 93-16=77 accidents. Notice that a count of the before period accidents is 0*2+2*2+3*1+…+6*1=129. If we estimated the crash reduction by simple before-and-after study method, then the expected accident count in the after period would be 129. That is, a reduction of 129-16=113. The difference between 113 and 77 was the estimate of the RTM bias in this case.

2.2.1.4 Comparisons of the three methods

So far we have discussed all the three types of before and after study method. They are all designed to estimate the effect of safety treatment. Putting aside their similarities, each method has its own way of managing data and has its advantages and downsides. Below is a table that summarizes the traits of the three methods.

The following table listed the comparison critiques and the results among the three methods.
Table 2.9 Comparisons of the three methods

<table>
<thead>
<tr>
<th></th>
<th>Existing method</th>
<th>The Simple B+A method</th>
<th>The B+A with Comparison Group method</th>
<th>The B+A with Empirical Bayes method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection</td>
<td>Crash history of only the treatment site</td>
<td>History of both the treatment site and the comparison group</td>
<td>History of both the treatment site and the reference group</td>
<td></td>
</tr>
<tr>
<td>Advantages</td>
<td>Only need the history data of the improvement location, calculation is straightforward</td>
<td>Eliminated the maturation.</td>
<td>Eliminated the RTM.</td>
<td></td>
</tr>
<tr>
<td>Shortcomings</td>
<td>Ignored the existence of regression to the mean, crash migration and maturation.</td>
<td>Usually requires a relatively large size of comparison groups. Still neglects RTM.</td>
<td>The calculation is relatively complicated. Still neglects maturation.</td>
<td></td>
</tr>
</tbody>
</table>

2.2.2 A more coherent method

This approach introduced by Hauer in [5] serve as an extension of the basic empirical before and after study method. It provides a possible way to estimate and predict the accident count all in one setting.

Let $K_{i,1}, K_{i,Y}, K_{i,Y+1}, \ldots, K_{i,Y+Z}$ be the observed accident count in the $i$th location from year 1 through year $Y+Z$, in which year 1 to year $Y$ are the $Y$ years before treatment, and year $Y+1$ to year $Y+Z$ are the $Z$ years after treatment. Let $\kappa_{i,1}, \kappa_{i,Y}, \kappa_{i,Y+1}, \ldots, \kappa_{i,Y+Z}$ be the expected
accident count corresponding to the observations. The task is to estimate \( \kappa_{i,1} \ldots \kappa_{i,Y} \) and to predict \( \kappa_{i,Y+1} \ldots \kappa_{i,Y+Z} \) if the site remains untreated.

Model selection

Hauer emphasizes that the selection of the model is more influential in determining the quality of the product than the methodology used to estimate the parameter values. The choice of the model should reflect the prior knowledge of the relationship between accident count and the factors that potentially influence traffic safety. In a road section study, Hauer suggests to use model [5]:

\[
\kappa_{i,y} = d_i \alpha_y F_{i,y}^\beta + \varepsilon_{i,y}
\]

where:

- \( d_i \) is the \( i \)th road section length;
- \( F_{i,y} \) is the annual average daily traffic (AADT);
- \( \alpha_y \) and \( \beta \) are parameters of the model;
- \( \varepsilon_{i,y} \) is the error of the model, where \( E(\varepsilon_{i,y}) = 0, \text{Var}(\varepsilon_{i,y}) = \sigma^2 \).

This model is based on the belief that the occurrence of accidents mainly depends on the traffic flow and the road section length. The use of \( \alpha_y \)'s in the model reflects that other than road section length and traffic flow, all the other factors that influence the road safety change from year to year, and these changes of each year affect the safety between locations in the same manner. The parameter \( \beta \) determines how the change of traffic flow (AADT) could affect the incidence of the accidents. This model also indicates the fact that when \( d_i = 0 \) or \( F_{i,y} = 0 \), \( \kappa_{i,y} = 0 \).

By the model,
Likelihood function for parameter estimation

One of the most widely used methods of statistical estimation is the Maximum likelihood Estimation (MLE) method. We introduce it in this transportation problem in order to get a sensible estimate for the parameters $\alpha_y$'s, $\beta$, and $b$ which will be introduced later in this section. All these parameters need to be estimated by MLE. Later they will be used to learn what would have been if there were no treatment in the treatment site. To do a MLE estimate we need the accident records of the reference group from the beginning of the before period though the after period.

Assume the occurrence of accidents at a certain entity and year is Poisson distributed. Then[5]:

$$P(K_{i,y} | k_{i,y}) = k_{i,y}^{K_{i,y}} e^{-k_{i,y}} / K_{i,y}! \quad [5]$$

(2.23)

Then for $R$ reference locations and $Y+Z$ years,

$$P(\text{accident count} | [K_{i,y}] | \text{parameters}([k_{i,y}])) \prod_{i=1}^{R} \prod_{y=1}^{Y+Z} k_{i,y}^{K_{i,y}} e^{k_{i,y}} / K_{i,y}! \quad [5].$$

since $K_{i,y}$'s are independent.

There are $R^*(Y+Z)$ unknowns in the formula. The next task would be to replace many unknowns by the parameters. Let [5]

$$\frac{E(k_{i,y})}{E(k_{i,1})} = C_{i,y} \quad \text{and} \quad \frac{k_{i,y}}{k_{i,1}} = C_{i,y} \quad [5]$$

(2.24)

By doing this, the many $k_{i,y}$'s can be expressed as a function of $k_{i,1}$ and $C_{i,y}$'s. Most of the time, $k_{i,y}$'s in different years are not equal. The reason is as follows: in this equation, people assume that over the years the $k_{i,y}$'s will remain similar in some aspects, but there will also be some change from year to year. And this change should not be totally unpredictable. The author assumed that this change have something to do with the traffic flow (AADT) that it can
be captured by the model as well. Also $\kappa_{i,y}$ will be different from $\text{E}(\kappa_{i,y})$. For a certain year $y$, the expected accident counts of the reference sites are similar for they share similar traits, but will still be different because other factors that are not been captured could also influence the incidence of accident.

Now for a certain location $i$, the many unknowns $\kappa_{i,y}$’s can be replaced by the combination of $\kappa_{i,1}$ and $C_{i,y}$’s, then

$$P(K_{i,1}, ..., K_{i,Y}, K_{i,Y+1}, ..., K_{i,Y+Z} | k_{i,1}, ..., k_{i,Y}, k_{i,Y+1}, ..., k_{i,Y+Z}) =$$

$$(\prod_{y=1}^{Y+Z} \frac{C_{i,y}^{K_{i,y}}}{K_{i,y}!}) (\prod_{y=1}^{Y+Z} \frac{\sum_{y=1}^{Z} K_{i,y} e^{-\sum_{y=1}^{Z} C_{i,y} K_{i,y}}}{\sum_{y=1}^{Z} C_{i,y}!}) [5].$$

Now the likelihood function becomes:

$$P(\text{accident count}|K_{i,Y}) | \text{parameters}(k_{i,Y}) = \left(\prod_{y=1}^{Y+Z} \frac{C_{i,y}^{K_{i,y}}}{K_{i,y}!}\right) (\prod_{y=1}^{Y+Z} \frac{\sum_{y=1}^{Z} K_{i,y} e^{-\sum_{y=1}^{Z} C_{i,y} K_{i,y}}}{\sum_{y=1}^{Z} C_{i,y}!})$$

$$= P(K_{i,1}, ..., K_{i,Y}, K_{i,Y+1}, ..., K_{i,Y+Z} | k_{i,1}) = \prod_{i=1}^{R} \left(\prod_{y=1}^{Y+Z} \frac{C_{i,y}^{K_{i,y}}}{K_{i,y}!}\right) (\prod_{y=1}^{Y+Z} \frac{\sum_{y=1}^{Z} K_{i,y} e^{-\sum_{y=1}^{Z} C_{i,y} K_{i,y}}}{\sum_{y=1}^{Z} C_{i,y}!}) [5].$$

By using the $C_{i,y}$’s, the dimension of unknowns has been reduced a lot. However, the expected accident counts for different sites of year $1$ ($k_{1,1}, k_{2,1}, ..., k_{R,1}$) still remain unknown.

To solve this problem, assume that $k_{1,1}, k_{2,1}, ..., k_{R,1}$ are Gamma distributed with parameters $a_i$ and $b$. Note that the reference locations have similar traits as location $i$, but their expected number of accident count are not necessary the same. This is because the reference locations were selected for they have similar traits, but the number of the traits we used to identify the reference location is limited. There still exist distinct characteristics among the reference locations, and this difference should not be unpredictable. Therefore we assume this change subject to the Gamma distribution

$$f(k_{i,1}) = \frac{a_i^b}{\Gamma(b)} k_{i,1}^{b-1} e^{-a_i k_{i,1}},$$

where $a_i = \frac{\text{E}(k_{i,1})}{\text{Var}(k_{i,1})}, b = \frac{\text{E}^2(k_{i,1})}{\text{Var}(k_{i,1})}, i = 1, ..., R \quad (2.26)$

Therefore, $k_{i,1}^{b-1} e^{-a_i k_{i,1}} = f(k_{i,1}) \frac{\Gamma(b)}{a_i^b} \quad (2.27)$

Also, following the same fashion,

$$k_{i,1}^{\sum_{y=1}^{Z} K_{i,y} + b-1} e^{-\sum_{y=1}^{Z} C_{i,y} K_{i,y}} = f(k_{i,1}) \frac{\Gamma\left(\sum_{y=1}^{Z} C_{i,y} + b\right)}{(a_i + \sum_{y=1}^{Z} C_{i,y})^{\sum_{y=1}^{Z} C_{i,y} + b}} \quad (2.28)$$
Divide (2.28) by (2.27), then
\[ \kappa_{i,1} \sum_{y=1}^{Y+Z} K_{i,y} e^{-\left( \sum_{y=1}^{Y+Z} C_{i,y} \right) \kappa_{i,1}} \]

Plugging the result into function (2.25),
\[
P_T(K_{i,1}, \ldots, K_{i,Y}, K_{i,Y+1}, \ldots, K_{i,Y+Z} | \kappa_{i,1}) = \prod_{i=1}^{R} \left( \prod_{y=1}^{Y+Z} \frac{C_{i,y}}{K_{i,y}} \right) \left( \frac{b}{E(K_{i,1})} \right)^b \left( \frac{\sum_{y=1}^{Y+Z} K_{i,y} + b - 1)!}{(b - 1)!} \right) \]

The likelihood \( L \) is given by the joint probability distribution evaluated at observed accident count \( K_{i,y} \), hence
\[
L = \prod_{i=1}^{R} \left( \prod_{y=1}^{Y+Z} \frac{C_{i,y}}{K_{i,y}} \right) \left( \frac{b}{E(K_{i,1})} \right)^b \left( \frac{\sum_{y=1}^{Y+Z} K_{i,y} + b - 1)!}{(b - 1)!} \right) \] (2.29)

So far we have introduced the parameter \( b \) into the likelihood function.

Then
\[
\ln(L) = \sum_{i=1}^{R} \left( \sum_{y=1}^{Y+Z} K_{i,y} \ln \left( \frac{C_{i,y}}{E(K_{i,1})} \right) \right) + b \ln \left( \frac{b}{E(K_{i,1})} \right) - \left( \sum_{y=1}^{Y+Z} K_{i,y} + b \right) \ln \left( \frac{b}{E(K_{i,1})} + \sum_{y=1}^{Y+Z} C_{i,y} \right) + \ln \left( \frac{\sum_{y=1}^{Y+Z} K_{i,y} + b - 1)!}{(b - 1)!} \right) \] (2.30)

Once the parameter values \( \alpha_y \)'s, \( \beta \) and \( b \) are chosen, the values of \( E(K_{i,y}) \)'s can be calculated. Then by using equation (2.24), the parameter \( C_{i,y} \) for each location \( i=1, \ldots, R \) and year \( y=1, \ldots, Y+Z \) can be calculated. A sensible estimate of those parameters would be the parameters that maximize the log-likelihood function. The values of the parameters that maximize the log-likelihood cannot be expressed in a nice closed form solution. (Normally, the method to solve for the values of the parameter is taking the partial derivative.) Instead they must be determined numerically by starting with a set of initial values and iterating to the maximum of the log-likelihood function. Technically, this procedure is called an iteratively re-weighted least squares method [13]. However, for this case, the form is complicated and it may be difficult to solve. It turns out that the Excel software provides a “Solver Function”
that will do the job. More information about how to use solver function will be given in Section 3. From this step, we will get the estimate of $\alpha'_i$s, $\beta$ and $b$.

**Estimate $\kappa_{i,1}\kappa_{i,2}...\kappa_{i,Y}$ for a certain entity**

Let “$i$” be a treated entity. Suppose the treatment was taken at the end of year $Y$.

The $K_{i,1}, ..., K_{i,Y}$ for this treated site are available. Also to do the estimate people will need to know the value of independent variables of the model (road section length and traffic flows for $Y$ year).

There are two kinds of estimation, namely Maximum Likelihood estimation and Empirical Bayes estimation.

(a) Maximum Likelihood Estimation\[5\]

From year 1 to year $Y$ before treatment happens, the likelihood function can be written as:

$$L(k_{i,1}) = P_T(K_{i,1}, ..., K_{i,Y} | k_{i,1}) = \left(\prod_{y=1}^{Y} \frac{c_{iy}}{k_{iy}}\right)^{k_{i,1}} \frac{e^{-k_{i,1} \sum_{y=1}^{Y} k_{iy}}}{k_{iy}},$$

this is the likelihood function for $k_{i,Y}$ and we wish to find the value of $k_{i,1}$ that maximized the likelihood function. First take log on both sides,

$$\ln L(k_{i,1}) = \ln \left(\prod_{y=1}^{Y} \frac{c_{iy}}{k_{iy}}\right) + \sum_{y=1}^{Y} k_{i,y} \ln(k_{i,1}) - k_{i,1} \sum_{y=1}^{Y} c_{i,y},$$

then take derivative on both sides:

$$\frac{d[\ln L(k_{i,1})]}{dk_{i,1}} = \sum_{y=1}^{Y} k_{i,y} - \sum_{y=1}^{Y} c_{i,y}$$

If $\hat{k}_{i,1}$ is that value of $k_{i,1}$ at which the derivative equals 0, then

$$\hat{k}_{i,1} = \frac{\sum_{y=1}^{Y} k_{i,y}}{\sum_{y=1}^{Y} c_{i,y}} \quad (2.31)$$

For the remaining $\hat{k}_{i,y}$ where $y \neq 1$, $\hat{k}_{i,y} = \hat{k}_{i,1} c_{i,y}$. \quad (2.32)

(b) Empirical Bayes Estimation\[5\]

In the previous section we have demonstrated the existence of RTM and its influence on the estimation. The Maximum Likelihood estimation for $k_{i,Y}$’s is straightforward but still subject to RTM. The EB estimation is a remedy to that. With EB approach, the estimation of $k_{i,Y}$ is based on the joint use of two clues: those contained in accident counts of the treated entity.
and those contained in traits of this entity. $\kappa_{i,1}$ is the expected accident count for treated entity $i$ in year 1, which had been treated at the end of year $Y$. Now introduce its reference population. We have discussed in the first step, that the reference locations have similar traits as location $i$, but their expected number of accident count are not necessary the same. And we assume this change subject to the Gamma distribution, then $f(\kappa_{i,1}) = \frac{a_i \kappa_{i,1}^{b-1} e^{-a_i \kappa_{i,1}}}{\Gamma(b)}$ and 

$$a_i = \frac{E(\kappa_{i,1})}{\text{Var}(\kappa_{i,1})}, \quad b = \frac{(E(\kappa_{i,1}))^2}{\text{Var}(\kappa_{i,1})}$$

Under this condition, consider what is the probability density function of $f(\kappa_{i,1}, \ldots, \kappa_{i,Y}| K_{i,1}, \ldots, K_{i,Y})$.

$$f(\kappa_{i,1}, \ldots, \kappa_{i,Y}| K_{i,1}, \ldots, K_{i,Y}) = f(\kappa_{i,1}, \ldots, \kappa_{i,Y}| K_{i,1}, \ldots, K_{i,Y}) = \frac{f_T(\kappa_{i,1}| K_{i,1}, \ldots, K_{i,Y}) \cdots f_T(\kappa_{i,Y}| K_{i,1}, \ldots, K_{i,Y})}{\int f_T(\kappa_{i,1}| K_{i,1}, \ldots, K_{i,Y}) d\kappa_{i,1} \cdots \int f_T(\kappa_{i,Y}| K_{i,1}, \ldots, K_{i,Y}) d\kappa_{i,Y}}$$

Let $\int f_T(\kappa_{i,1}, \ldots, K_{i,Y}| K_{i,1}, \ldots, K_{i,Y}) d\kappa_{i,1} = 1$. is a constant, then

$$f(\kappa_{i,1}, \ldots, \kappa_{i,Y}| K_{i,1}, \ldots, K_{i,Y}) = m_1 \cdot f_T(\kappa_{i,1}| K_{i,1}, \ldots, K_{i,Y}) f_T(\kappa_{i,Y}| K_{i,1}, \ldots, K_{i,Y})$$

$$m_1 = \left( \prod_{y=1}^{Y} \frac{C_y^{K_y}}{K_y^!} \right) \left( \kappa_{i,1}^{\sum_{y=1}^{Y} K_{i,y}} e^{-\kappa_{i,1} \sum_{y=1}^{Y} C_y} \right) \left( \frac{(a_i \kappa_{i,1})^{b-1} e^{-a_i \kappa_{i,1}}}{\Gamma(b)} \right) = m_2 \kappa_{i,1}^{b+\sum_{y=1}^{Y} K_{i,y}-1} e^{-\kappa_{i,1}(a_i + \sum_{y=1}^{Y} C_y)}$$

where $m_2 = \frac{(a_i + \sum_{y=1}^{Y} C_y)^{b+\sum_{y=1}^{Y} K_{i,y}}}{\Gamma(b+\sum_{y=1}^{Y} K_{i,y})}$.

At this moment we can see the conditional probability is also gamma distributed with mean $E(\kappa_{i,1}|K_{i,1}, \ldots, K_{i,Y}) = \frac{b+\sum_{y=1}^{Y} K_{i,y}}{a_i + \sum_{y=1}^{Y} C_y}$ and variance $\text{Var}(\kappa_{i,1}|K_{i,1}, \ldots, K_{i,Y}) = \frac{(b+\sum_{y=1}^{Y} K_{i,y})^2}{(a_i + \sum_{y=1}^{Y} C_y)^2}$

By the Bayes Theorem, the Bayes estimator

$$\hat{\kappa}_{i,1} = \frac{b+\sum_{y=1}^{Y} K_{i,y}}{a_i + \sum_{y=1}^{Y} C_y} \cdot \frac{B(\frac{b+\sum_{y=1}^{Y} K_{i,y}}{a_i + \sum_{y=1}^{Y} C_y})}{\Gamma(\frac{b+\sum_{y=1}^{Y} K_{i,y}}{a_i + \sum_{y=1}^{Y} C_y})}$$

$$\text{V} \bar{\text{r}}(\kappa_{i,1}) = \frac{b+\sum_{y=1}^{Y} K_{i,y}}{\left( a_i + \sum_{y=1}^{Y} C_y \right)^2}$$

For the remaining $\hat{\kappa}_{i,y}$ where $y \neq 1$, $\hat{\kappa}_{i,y} = \hat{\kappa}_{i,1} \bar{C}_y$.

(c) The relationship between Maximum Likelihood estimation and the EB estimation
When the distribution of $\kappa_{i,\ell}$'s is such that the standard deviation $\sigma(\kappa_{i,\ell})$ is large compared with $E(\kappa_{i,\ell})$, then $b = \frac{\left[\frac{E(\kappa_{i,\ell})}{\sigma(\kappa_{i,\ell})}\right]^2}{E(\kappa_{i,\ell})}$ is small and $a_i = \frac{b}{E(\kappa_{i,\ell})}$ is also a small number, then the EB estimate will converge toward the maximum likelihood estimate. The Maximum likelihood estimator converges to the average accident counts (accident count divided by the number of years) when $\kappa_{i,\ell}$'s do not change from year to year [2] ($C_{i,Y} = 1$ for all $i=1,\ldots,Y$). The advantage of EB method is that it provides not only the estimator, but also the estimate of the variance.

Estimate $\kappa_{i,Y+1}\kappa_{i,Y+2} \ldots \kappa_{i,Y+Z}$ for a certain entity

We have stated that the entity $i$ had been treated at the end of year $Y$. Then the observation $K_{i,Y+1}K_{i,Y+2} \ldots K_{i,Y+Z}$ could no longer represent the historical data under untreated situation. Our task is to predict what would have been the expected accident frequencies $\kappa_{i,Y+1}\kappa_{i,Y+2} \ldots \kappa_{i,Y+Z}$ in the after year had the treatment not been applied. From the maximum likelihood function in (2.30) we get the estimation of necessary parameters $C_{i,Y+1}, \ldots, C_{i,Y+Z}$, and for the treated entity we have the estimate of $\kappa_{i,\ell}$, using either Maximum Likelihood estimation or Empirical Bayes estimation. What remains to be done is pretty straightforward:

For $y > Y$, $\hat{\kappa}_{i,Y} = \tilde{C}_{i,Y}\hat{\kappa}_{i,1}$ and $\text{Var}(\kappa_{i,Y}) = \tilde{C}_{i,Y}^2 \hat{\kappa}_{i,1}$. (2.35)

3. Methodology of forecasting accidents for road sections

The goal of this study is to forecast the expected accident counts in the coming year with or without treatment. To achieve this, we first divide the study period into two parts, the before period is from the beginning of the study till the current year, the after period is the coming year. With the methodology provided from the previous section 2.2.1, we are now able to forecast the expected accident count of treatment site in the after period with or without treatment. This value appears in the numerator of the function (2.3),

$P(A|I_4, S_t) = \frac{\text{expected traffic accident counts/year}}{\text{AADT} \times 365}$. And the function value will serve as an input to an optimization model. Notice that for each location under each scenario, if the conditional...
probability is believed to be different, then we need to do forecasts under each condition.

Let’s assume \( S_p \) is a scenario of safety strategy that we want to study.

Moreover, this study based on the belief that:

1. The RTM problem is the main problem that could affect the accuracy of the estimate;

2. The road section length and the traffic flow are two main factors that will affect traffic safety.

It is also important to emphasize the using of the expectation value in this probability function. We use the expected traffic accident counts for two reasons. First, the accident count in the coming year could not be observed in the current year, so it is not available to use. Second, in section 2.2.1.3, the RTM problem is shown to be disturbing, which will often exaggerate the effect of the treatment. The use of the expected value will remove the random effect, especially for the RTM, and represent the actual accident frequency from implementing the safety facilities. This will make our inputs to the previous optimization model more accurate to use. Typically, the method of forecasting the accident frequencies consists of 3 steps. The detailed discussion is listed below.

**Step 1: Data collection and preparation**

An adequate set of data need to be prepared before applying the algorithm. Basically, the following data will be needed for the study:

1. The accident report of the treatment site from the beginning of the study period till the current year.

2. The accident report of the first reference group G1, from the beginning of the study period till current year. G1 is the group of sites that share similar traits with the treatment site. The sites in G1 were not been treated by \( S_p \) throughout the study period.

3. The accident report of the second reference group G2, from the beginning of the study period till current year. G2 is the group of sites that share similar traits with the treatment site and all the sites had been treated by \( S_p \) at least from the beginning of the study period.
For missing data in the data set, we will interpolate to predict the data value. Interpolation is a method of constructing new data points within the range of a discrete set of known data points. A brief introduction of different types of interpolation method is given below [14]:

Linear interpolation is one of the simplest methods. Basically it tells that if a point \((x_i, y_i)\) is missing between two known points \((x_{i-1}, y_{i-1})\) and \((x_{i+1}, y_{i+1})\). The point to be interpolated is given by fitting \((x_i, y_i)\) into the line that created by \((x_{i-1}, y_{i-1})\) and \((x_{i+1}, y_{i+1})\)

\[
y_i = y_{i-1} + \frac{x_i-x_{i-1}}{x_{i+1}-x_{i-1}} (y_{i+1} - y_{i-1}) \text{ at the point } (x_i, y_i).\]

It is said that linear interpolation is easy to handle, but it is not very precise for data with random effects.

Polynomial interpolation is a generalization of linear interpolation. We replace the linear function with polynomial function when predict. Polynomial extrapolation can create a smoother curve.

The polynomial interpolation subjects to great error when its degree is large. Spline interpolation uses low-degree polynomials in each of the intervals, and chooses the polynomial pieces such that they fit smoothly together. The resulting function is called a spline.

In practice, the Mathematica software is available to use for interpolation. It uses minimum polynomial to fit the data. For example, suppose there is a sequence of accident counts from year 1 to 5, which the third year data is missing: 2, 3, *, 0, 5. In Mathematica, use the Interpolate statement:
It is shown in this example that when year is 3, the predict value is $\frac{5}{6}$. The function it used to fit the data is also showed in the output.

In addition to interpolation for the missing values, it is required that we forecast the occurrence of the accidents of the treatment site and the two reference groups of the coming year to prepare for the next step. The extrapolation method will do the job. We choose the extrapolation method because of three reasons. First, notice that the occurrence of the accidents varies in certain pattern, and this pattern should not be unpredictable. Second, the extrapolation method offers several ways to extrapolate according to the pattern of variation. The last but not least, the extrapolation is a handy tool which is found to be useful in many transportation projects. For example, Hauer used the linear extrapolation method to simulate the missing data in the study of measure the effect of implementation in California.

There are mainly five types of extrapolation, and a choice of which type to use depend on a prior knowledge of the data pattern. A brief introduction of different types of extrapolation
method is given below[15]:

Linear extrapolation means creating a tangent line using the two end points (or more than two points) and extending it beyond that limit. This is a sound choice when the data points are approximately linearly distributed. For example, suppose the two end points of the data are \((x_{1-1}, y_{1-1})\) and \((x_i, y_i)\), then the point to be extrapolated is \(y_{i+1} = y_{i-1} + \frac{x_{i+1} - x_{i-1}}{x_i - x_{i-1}}(y_i - y_{i-1})\).

Polynomial extrapolation is to generate a polynomial curve through the entire known data or just near the end. Polynomial extrapolation is typically done by means of Lagrange interpolation or using Newton's method of finite differences to create a Newton series that fits the data. The resulting polynomial may be used to extrapolate the data.

Conic extrapolation uses five points at the end of the data to create a conic section. If the section created is an ellipse or circle, it will loop back and rejoin itself. A parabolic or hyperbolic curve will not rejoin itself, but may curve back relative to the X-axis.

French curve extrapolation is suitable for any distribution with accelerating or decelerating factors.

It is suggested that we use the linear extrapolation method. Although the polynomial extrapolation can create a smooth result, it subjects to great uncertainty. The polynomial extrapolation provides a sound result only near the end point. The next example shows that polynomial is not suitable. Suppose the four year accident counts for a certain site is 5, 4, 9, 6. We want to forecast the accident frequency of the fifth year. By using Mathematica, the result is shown below:
We can see year 5 is out of the domain. As a result, a warning appears claiming that extrapolation is used. When the point is far apart from the end point, the error became large. -19 definitely is not a nice result since we expect a result that is greater or equal to 0.

Moreover, conic extrapolation is not suitable for predicting accident count since the set of points does not have any trend to “loop back”. French curve is good when the distribution has accelerating or decelerating factors. The occurrence of the accident reflects the combined effect of many sundry factors, such as road condition, traffic condition and drive’s behavior. In practice the trend of accelerating or decelerating is not obvious.

Above are the reasons why the other extrapolations are not suitable and we may use linear extrapolation. There are different types of linear extrapolation, depend on how we choose to use the data. The extrapolation with two end points is easy to use, but it cannot reflect the trend of accident frequency over the years. Here we introduce linear regression model to fit the data. This model use least square estimate (LSE) to fit the data and provide a regression
line through the two periods. The coming year accident frequency can be predicted by the linear regression model.

Model: \( y = \beta_0 + \beta_1 x + \varepsilon \), where \( \varepsilon \sim N(0, \sigma^2) \). \hfill (2.36)

In the model \( x \) represent the year, \( y \) represent the accident frequency. The statistical software SAS will provide a LSE estimate for the parameters \( \beta_0 \) and \( \beta_1 \) along with the variance.

An example:

A report from Saint Louis city reveals in year 2004-2008, the accident frequencies for the road section with number 0069009999 are 2, 3, 2, 4, 3. The data is missing in year 2009. We want to make a predictive analysis about how many accidents will be happen in 2009. Use simple SAS code will give the regression values.

**SAS Code:**

```sas
proc reg data=StLouis_Crash;
model Acc_Num =Year;
symbol value=circle INTERPOL=R;
proc gplot;
plot Acc_Num *Year;
Run;
```

Result is shown in table 2.10:
Table 2.10. SAS output for the Five-Year data

```
\textbf{The REG Procedure}
\textbf{Model: MODEL1}
\textbf{Dependent Variable: accnum accnum}

\begin{array}{|c|c|c|c|c|}
\hline
\textbf{Source} & \textbf{DF} & \textbf{Sum of Squares} & \textbf{Mean Square} & \textbf{F Value} & \textbf{Pr > F} \\
\hline
\text{Model} & 1 & 0.90000 & 0.90000 & 1.42 & 0.3189 \\
\text{Error} & 3 & 1.90000 & 0.63333 & & \\
\text{Corrected Total} & 4 & 2.80000 & & & \\
\hline
\end{array}
```

\textbf{Root MSE} 0.79582 \quad \textbf{R-Square} 0.3214
\textbf{Dependent Mean} 2.80000 \quad \textbf{Adj R-Sq} 0.0952
\textbf{Coeff Var} 28.42223

\textbf{Parameter Estimates}

```
\begin{array}{|c|c|c|c|c|c|c|}
\hline
\textbf{Variable} & \textbf{Label} & \textbf{DF} & \textbf{Parameter Estimate} & \textbf{Standard Error} & \textbf{t Value} & \textbf{Pr > |t|} \\
\hline
\text{Intercept} & \text{Intercept} & 1 & 1.90000 & 0.83467 & 2.28 & 0.1073 \\
\text{year} & \text{year} & 1 & 0.30000 & 0.25166 & 1.19 & 0.3189 \\
\hline
\end{array}
```

The scatter plot with the regression line (2004-2008) is also given:

Graph2.3. Scatter plot and regression line for the Five-Year data

We can see from the result the estimate of $\beta_0$ is 1.9 and the estimate of $\beta_1$ is 0.3. Therefore
for year 2009, the expected accident count is predicted as: $\beta_0 + \beta_1 X = 1.9 + 0.3 \times 6 = 3.7$.

When necessary, the result can be round up to 4.

By using the interpolation and extrapolation method, the forecasting values are prepared and ready for use.

**Step 2: Forecast the expected accident count of treatment site without treatment**

We will be using Hauer’s more coherent method to predict the expected accident count in the after period if no treatment has been made. The reason for choosing this method is because that it provide the estimate of the expected accident count for treatment site of each individual year. Also, the use of the Empirical Bayes estimation in Hauer’s method mitigates the influence of the RTM problem. Moreover, the coherent approach takes into account the road characteristics that will relate to the number of accidents. This is achieved by selecting the adequate model $E(K_{1Y}) = d_1 \alpha_y F_{1Y}^{\beta_1}$. This model reflects the concern of road section length and traffic flow as the main external factors that will affect the accident frequency.

However, this method cannot be used directly to forecast the expected accident frequencies in the coming year. From the previous step, we have got the extrapolation value ready to use. Hereby we introduce the extrapolation method combining with the Hauer’s more coherent method to do the forecast.

Suppose the before period of the treatment site t pertains L years and it’s actually accident numbers are $K_{t,1}K_{t,2} \ldots K_{t,L}$. Our goal is to forecast the expected accidents count for the next coming year—year $L+1$ under the condition that no improvement has been made. In this case, the after period is the year $L+1$. To start, we introduce the reference group $G_1$. As mentioned before, $G_1$ is a group with no treatment and has similar traits with the treatment site. Let’s assume $G_1$ consists of $m$ sites and has $L$ years of accident counts report. The accident counts of site i year j is denoted by $K_{i,j}$. Then for each site, the actual accident frequencies are:
If we were given the accident frequencies of the untreated reference group G1 in the coming year, and then use the EB method to mitigate the regression to the mean phenomenon, we then able to conduct relatively accurate forecast estimation. The extrapolation will be use here. Let’s denote the crash frequency of a certain site i in the coming year without treatment is $K_{i,L+1}$ for $i=1,\ldots,m$ then extrapolate the data we get the estimated accident frequency for the next year:

$$K_{1,1},K_{1,2},\ldots,K_{1,L} \xrightarrow{\text{extrapolation}} K_{1,L+1}$$

$$\vdots$$

$$K_{m,1},K_{m,2},\ldots,K_{m,L} \xrightarrow{\text{extrapolation}} K_{m,L+1}$$

The next step is to apply the maximum likelihood estimate. The idea is that after extrapolation, we proceed as the accident counts of the coming year were known. At this step, the problem becomes how to predict the expected accident counts of the treatment site in year $L+1$ without treatment. The model to be used is from Hauer’s coherence method $E(K_{i,y}) = d_i \alpha_y F_{i,y}^\beta$.

Also, by equation (2.30), the log likelihood function for this problem is

$$\ln(L) = \sum_{i=1}^{m} \left[ \sum_{y=1}^{L+1} K_{i,y} \ln \left( G_{i,y} \right) \right] + b \ln \left( \frac{b}{E(K_{i,1})} \right) -$$

$$\left( \sum_{y=1}^{L+1} K_{i,y} + b \right) \ln \left( \frac{b}{E(K_{i,1})} + \sum_{y=1}^{L+1} C_{i,y} \right) + \ln \left( \frac{\sum_{y=1}^{L+1} K_{i,y} + b - 1)!}{(b-1)!} \right).$$

As introduced in section 2.2.2, apply the solver function in Excel will give the estimate of $\alpha_1, \ldots, \alpha_L, \alpha_{L+1}, \beta, b$. Below is an example about how to use Excel to estimate the parameters. The four year of data of G1 is from [5]. In our case, if the last year is the current year, then we will extrapolate to get the estimate accident number of the coming year, then proceeds as we have five year of data. Due to the availability of a practical data source, we will just use the four year data provided by Hauer. However, the way to process the data is the same no matter how we prepare the data in the previous step.
An example:

Suppose there is four year of data of six rural two lane road section, their information including road section length, accident counts and AADT are shown in table:

Table 2.11 Data for six road section [5]

<table>
<thead>
<tr>
<th>Road section</th>
<th>Year</th>
<th>AADT</th>
<th>Length (km)</th>
<th>Accident counts</th>
<th>Road section</th>
<th>Year</th>
<th>AADT</th>
<th>Length (km)</th>
<th>Accident counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>320</td>
<td>3.1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>4100</td>
<td>5.6</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>400</td>
<td>3.1</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>4500</td>
<td>5.6</td>
<td>7</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>450</td>
<td>3.1</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>5000</td>
<td>5.6</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>500</td>
<td>3.1</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>5100</td>
<td>5.6</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1220</td>
<td>4.2</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>7600</td>
<td>3.7</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1200</td>
<td>4.2</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>7400</td>
<td>3.7</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1300</td>
<td>4.2</td>
<td>9</td>
<td>5</td>
<td>3</td>
<td>7200</td>
<td>3.7</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>1250</td>
<td>4.2</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>6800</td>
<td>3.7</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3300</td>
<td>2.7</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>9250</td>
<td>1.9</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3000</td>
<td>2.7</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>9260</td>
<td>1.9</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2700</td>
<td>2.7</td>
<td>0</td>
<td>6</td>
<td>3</td>
<td>8700</td>
<td>1.9</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2700</td>
<td>2.7</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>8900</td>
<td>1.9</td>
<td>3</td>
</tr>
</tbody>
</table>

The model used is \( E(\kappa_{1y}) = d_{1} \alpha_{2} F_{1y}^{\beta} \). We start to find the maximum likelihood value with assigning initial value to the parameters. These parameters are \( \alpha_{1}, \alpha_{2}, \alpha_{3}, \alpha_{4}, \beta, b \). \( \alpha_{1}, \alpha_{2}, \alpha_{3}, \alpha_{4}, \beta \) are from the model, b is parameter of the Gamma distribution. We start with setting \( \beta = 1 \), that is, assume the traffic accident counts would be proportional to traffic flow AADT. Also, use \( \hat{\alpha}_{1} = \hat{\alpha}_{2} = \hat{\alpha}_{3} = \hat{\alpha}_{4} = 0.0002 \). This comes from the observation of road section 4, year 3. It is known that 5.6 kilometer long record accident frequency of 5 with an AADT about 5000 and \( \frac{6}{5.6*5000} \approx 0.0002 \). Also, start with guessing \( \hat{b} = 1 \), this is equivalent to guessing \( E(\kappa_{11}) = \text{Var}(\kappa_{11}) \) since \( b = \frac{E(\kappa_{11})}{\text{Var}(\kappa_{11})} \). With these initial values, we are able to calculate the likelihood value, for example in year 3 of road section 2, \( \hat{E}(\kappa_{2,3}) = d_{2} \hat{\alpha}_{3} F_{2,3}^{\hat{\beta}} = 4.2 * 0.0002 * 1300^{1} = 1.092, \) and \( \hat{C}_{2,3} \) is calculated from \( \hat{E}(\kappa_{2,3}) / \hat{E}(\kappa_{2,1}) = \frac{1.0920}{1.0248} = 1.0656 \). The value of \( \ln(L) \) turns out to be 137.48 in this case.

Table 2.12 shows the estimated values of \( E(K_{1,y}) \) and \( C_{1,y} \).
<table>
<thead>
<tr>
<th>Road Section</th>
<th>Year</th>
<th>$\hat{E}(K_{i,y})$</th>
<th>$\bar{c}_{i,y}$</th>
<th>$\sum_{y=1}^{4} \hat{c}_{i,y}$</th>
<th>Road Section</th>
<th>Year</th>
<th>$\hat{E}(K_{i,y})$</th>
<th>$\bar{c}_{i,y}$</th>
<th>$\sum_{y=1}^{4} \hat{c}_{i,y}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.1984</td>
<td>1.0000</td>
<td>0.0000</td>
<td>4</td>
<td>1</td>
<td>4.5920</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0.2480</td>
<td>1.2500</td>
<td>5.0400</td>
<td>2</td>
<td>2</td>
<td>5.4760</td>
<td>0.9737</td>
<td>0.9737</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>0.2790</td>
<td>1.4063</td>
<td>5.6000</td>
<td>3</td>
<td>3</td>
<td>5.3280</td>
<td>0.9474</td>
<td>0.9474</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>0.3100</td>
<td>1.5625</td>
<td>5.7120</td>
<td>4</td>
<td>4</td>
<td>5.6240</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1.0248</td>
<td>1.0000</td>
<td>5.0400</td>
<td>1</td>
<td>1</td>
<td>5.3280</td>
<td>0.9474</td>
<td>0.9474</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1.0080</td>
<td>0.9836</td>
<td>5.4760</td>
<td>2</td>
<td>2</td>
<td>5.3280</td>
<td>0.9474</td>
<td>0.9474</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1.0920</td>
<td>1.0656</td>
<td>5.3280</td>
<td>3</td>
<td>3</td>
<td>5.3280</td>
<td>0.9474</td>
<td>0.9474</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>1.0500</td>
<td>1.0246</td>
<td>5.0320</td>
<td>4</td>
<td>4</td>
<td>5.0320</td>
<td>0.8947</td>
<td>0.8947</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1.7820</td>
<td>1.0000</td>
<td>3.5150</td>
<td>1</td>
<td>1</td>
<td>3.5150</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1.6200</td>
<td>0.9091</td>
<td>3.5188</td>
<td>2</td>
<td>2</td>
<td>3.5188</td>
<td>1.0011</td>
<td>1.0011</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>1.4580</td>
<td>0.8182</td>
<td>3.3060</td>
<td>3</td>
<td>3</td>
<td>3.3060</td>
<td>0.9405</td>
<td>0.9405</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>1.4580</td>
<td>0.8182</td>
<td>3.3820</td>
<td>4</td>
<td>4</td>
<td>3.3820</td>
<td>0.9622</td>
<td>0.9622</td>
</tr>
</tbody>
</table>

Table 2.12. Starting values of $\hat{E}(K_{i,y})$ and $\bar{c}_{i,y}$

Table 2.13 shows the value of the likelihood function. It is divided into four parts:

$$\ln(\mathcal{L}) = \text{Part } 1 + \text{Part } 2 + \text{Part } 3 + \text{Part } 4,$$

$$\text{Part } 1 = \sum_{i=1}^{6} \left( \sum_{y=1}^{4} K_{i,y} \ln(C_{i,y}) \right),$$

$$\text{Part } 2 = b \ln \left( \frac{b}{\hat{E}(\kappa_{i,1})} \right),$$

$$\text{Part } 3 = -\left( \sum_{y=1}^{4} K_{i,y} + b \right) \ln \left( \frac{b}{\hat{E}(\kappa_{i,1})} + \sum_{y=1}^{4} C_{i,y} \right),$$

$$\text{Part } 4 = \ln \left( \frac{\left( \sum_{y=1}^{4} K_{i,y} + b - 1 \right)!}{(b-1)!} \right).$$

The values in Table 2.13 can be calculated by plugging the values from Table 2.11 and Table 2.12 into each part.

<table>
<thead>
<tr>
<th>Road Section</th>
<th>Part 1</th>
<th>Part 2</th>
<th>Part 3</th>
<th>Part 4</th>
<th>Row sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0000</td>
<td>1.6175</td>
<td>-4.6563</td>
<td>0.0000</td>
<td>-3.0389</td>
</tr>
<tr>
<td>2</td>
<td>0.6513</td>
<td>-0.0245</td>
<td>-40.4826</td>
<td>54.7847</td>
<td>14.9289</td>
</tr>
<tr>
<td>3</td>
<td>-1.1739</td>
<td>-0.5777</td>
<td>-16.9512</td>
<td>17.5023</td>
<td>-1.2005</td>
</tr>
<tr>
<td>4</td>
<td>2.2788</td>
<td>-1.5243</td>
<td>-39.1045</td>
<td>54.7847</td>
<td>16.4348</td>
</tr>
<tr>
<td>5</td>
<td>-3.6892</td>
<td>-1.7270</td>
<td>-88.6203</td>
<td>201.0093</td>
<td>106.9727</td>
</tr>
<tr>
<td>6</td>
<td>-0.3360</td>
<td>-1.2570</td>
<td>-22.9166</td>
<td>27.8993</td>
<td>3.3896</td>
</tr>
</tbody>
</table>

Total Sum | 137.4866

Table 2.13. Values for the four components of the log likelihood function
The value of the likelihood varies along with the changing of the parameter values \( \alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta, b \). The question is when the likelihood function reaches the maximum. The solver function in excel is a tool to find the maximum. The way to apply is shown here:

1. Choose the value that needs to be maximized. Select the Solver function in the data column. A window will pump out by pressing the “solver” bottom. In this window, one can set the target cell. In this spreadsheet, the target cell is P17, the corresponding value is the maximum likelihood function value.

2. Set the values that will be changed to achieve the maximum or (minimum). In this example, the values are \( \bar{\alpha}_1, \bar{\alpha}_2, \bar{\alpha}_3, \bar{\alpha}_4, \bar{\beta}, \bar{b} \) which correspond to the cells L5, M5, N5, O5, P5, and Q5 respectively. The “Solver Parameters” window also provides space for you to add the constraints for the parameters.

3. After setting the objective values and the dependent values in the window, press the “Solve” bottom and Excel will iterate until reach the maximum. It will shows on the spreadsheet all the corresponding values of the maximum.

Spread sheet1. The road section data and estimate of \( E(K_{i,y}) \) and \( C_{i,y} \).
Spread sheet 2. Output of the Maximum Likelihood Estimate

The result is \( \ln(L) = 142.3781 \), when

\[
\hat{\alpha}_1 = 0.0080, \hat{\alpha}_2 = 0.0069, \hat{\alpha}_3 = 0.0074, \hat{\alpha}_4 = 0.0086, \hat{\beta}_1 = 0.6480, \hat{b}_1 = 3
\]

It is necessary to emphasize the role of the initial value. In this example, the likelihood function is a nonstandard type of function. Local maximum values might be reached with different set of initial values. It is better to have the initial values as close to the true values as possible.

With the estimated parameters, the next thing is to estimate \( \kappa_{t,1} \kappa_{t,2} \ldots \kappa_{t,L} \) in the before period and predict \( \kappa_{t,L+1} \) in the after period for the treatment site. From the last step, we now have obtained \( \hat{\alpha}_1, \ldots, \hat{\alpha}_L, \hat{\alpha}_{L+1}, \hat{\beta}, \hat{b} \). The model \( E(\kappa_{t,y}) = d_t \alpha_y F_{t,y} \beta \) gives the estimate of \( E(\kappa_{t,y}) \) of year 1...L+1. By equation (2.24), \( \frac{E(\kappa_{t,y})}{E(\kappa_{t,1})} = C_{t,y} \), the estimate of \( C_{t,1}, \ldots, C_{t,L+1} \) can be calculated. Section 2.2.2 provides two types of estimation methods: Maximum Likelihood estimate and Empirical Bayes estimate. It is suggested to use the Empirical Bayes estimation method since it provides the variance of the estimation. The estimator is \( \hat{\kappa}_{t,1} = \frac{\hat{b} + \sum_{y=1}^{L+1} \kappa_{t,y}}{\hat{\alpha}_t + \sum_{y=1}^{L+1} C_{t,y}} \), where \( \hat{\alpha}_t = \frac{\hat{b}}{E(\kappa_{t,1})} \). This estimator estimates the expected accident number of the treatment site \( t \) of year 1. It’s variance is given by equation (2.34), \( \text{Var}(\kappa_{t,1}) = \frac{\hat{b} + \sum_{y=1}^{L+1} \kappa_{t,y}}{(\hat{\alpha}_t + \sum_{y=1}^{L+1} C_{t,y})^2} \). For the
coming year L+1, by the assumption

\[
\frac{k_{l,y}}{k_{l,1}} = c_{l,y}
\]

which appears in equation (2.24), then

\[
\hat{\kappa}_{t,L+1} = \hat{\kappa}_{t,1} \hat{c}_{t,L+1}.
\]

\[
\text{Var}(\kappa_{t,L+1}) = \hat{c}_{t,L+1}^2 \text{Var}(\kappa_{t,1}).
\]

We can also use the same algorithm to get the estimate of \( \hat{\kappa}_{i,L+1} \) for each reference site where \( i=1,\ldots,n \). But so far at this step, what we care about is the estimate of \( \kappa_{t,L+1} \).

At this moment, \( \hat{\kappa}_{t,L+1} \) — the forecast value of expected accident count in the after period with no treatments is available to use.

**Step 3: Forecast the expected accident count of treatment site after treatment**

Different from the previous step, the accident counts of the treatment site after treatment can neither be observed nor be extrapolated. The critical point is that we do not have any information about what will happen in the post-treatment site. However, similar information can be gained from the sites with the same treatment. Therefore, it is necessary to introduce another reference group (G2). G2 is a group of sites that have similar traits with treatment site and have been treated at least from the beginning of the study period with certain improvement(s). G1 different from G2 in the sense that treatment has been applied to G2 before study begins while G1 are not been treated throughout the study period. Suppose there are n reference sites in G2, their accident counts are listed below:

\[
K_{1,1}^*, K_{1,2}^*, \ldots, K_{1,L}^*
\]

\[
K_{n,1}^*, K_{n,2}^*, \ldots, K_{n,L}^*
\]

Where ‘*’ is a special sign to differentiate second step from first step. At this point, according to different situation, we have three recommended methods.

1. The ideal case. Suppose there exists a site that is similar enough to the treatment site in many aspects such as road type, road section length, traffic flow and other sundry factors.
The only difference is that it has been treated before the study period. In this case it is reasonable to assume that the observed accident frequency between this site and the treatment site are close enough. By introducing this site as a reference site, and then extrapolate to predict accident count for next year. This predicted value would be the estimate of accident count for the treatment site of the next year.

\[ K_{1,1}^*, K_{1,2}^*, ..., K_{1,L}^* \xrightarrow{\text{extrapolation}} K_{1,L+1}^* \]

\[ \hat{K}_{t,L+1}^* = K_{1,L+1}^* \] \hspace{1cm} (2.39)

2. However, in most of the cases, it is hard to find such a site that similar enough to the target site. In this case, a large sample size of the second reference group would be a better choice. “Large” means sufficient enough to get rid of the random effect and other sundry effects. And also the mean of the road section length and the mean of the traffic flow of those reference sites are close to the treatment site \( (E(d_i) = d_t, E(F_{i,y}) = F_{t,y}) \). To estimate the expected accident count of treatment site after treatment, first extrapolate each reference site to get the predicted accident count. And then take the average of those predicted accident counts. This average value will serve as the estimated expected accident count of the treatment group in the after period.

\[ K_{1,1}^*, K_{1,2}^*, ..., K_{1,L}^* \xrightarrow{\text{extrapolation}} K_{1,L+1}^* \]

\[ \vdots \]

\[ K_{n,1}^*, K_{n,2}^*, ..., K_{n,L}^* \xrightarrow{\text{extrapolation}} K_{n,L+1}^* \]

\[ \hat{K}_{t,L+1}^* = \frac{\sum_{i=1}^{B} K_{i,L}^*}{n} \] \hspace{1cm} \text{Var}(\hat{K}_{t,L+1}^*) = \frac{\sum_{i=1}^{B} (K_{i,L}^* - \bar{K}_L)^2}{n-1} \] \hspace{1cm} (2.40)

3. Sometimes the restrictions are tight thus the number of the qualified reference sites is limited. In this case the size of the reference group is not large enough to remove the random effect. However, if the reference sites all have similar road section length and traffic flows and close to the treatment group, we could first extrapolate to forecast the occurrence of crash and then apply Hauer’s method to calculate the expected accident count for each reference group in order to adjust the regression to the mean effect.
averages of the estimated expected accident count of the reference sites would be the wanted value. The brief algorithm is listed below:

\[ K_{1L}^*, K_{2L}^*, \ldots, K_{IL}^* \rightarrow K_{iL+1}^* \quad \text{where } i=1, \ldots, n \]

\[ \hat{\alpha}_1, \ldots, \hat{\alpha}_L, \hat{\alpha}_{L+1}, \hat{\beta}, \hat{b} \text{ and } \hat{K}_{iL+1}^* \text{ for } i=1, \ldots, n \]

The detailed algorithm to calculate \( \hat{\alpha}_1, \ldots, \hat{\alpha}_L, \hat{\alpha}_{L+1}, \hat{\beta}, \hat{b} \) and \( \hat{K}_{iL+1}^* \) is the similar to the one showed in step 2.

\[ \hat{K}_{iL+1}^* = \frac{\sum_{i=1}^{n} \hat{K}_{iL+1}}{n}. \quad (2.41) \]

If the road section length and the traffic flow are different among the reference sites, then similar to the previous situation, after the extrapolation, we need to use Hauer’s method to adjust the regression to the mean. However, the difference in road section length and traffic flow is still a disturbing factor. The best way to solve this problem is to use the road section length and the traffic flow of the treatment site to calculate its own expected accident count in the coming year. The algorithm is listed below:

\[ K_{1L+1}^*, K_{2L+1}^*, \ldots, K_{IL+1}^* \rightarrow K_{iL+1}^* \quad \text{(where } i=1, \ldots, n) \]

\[ \hat{\alpha}_1, \ldots, \hat{\alpha}_L, \hat{\alpha}_{L+1}, \hat{\beta}, \hat{b} \]

\[ \hat{K}_{iL+1}^* = \hat{E}( \kappa_{iL+1} ) = d_t \hat{\alpha}_{L+1} F_{iL+1} \hat{\beta}. \quad (2.42) \]

The detailed algorithm to calculate \( \hat{\alpha}_1, \ldots, \hat{\alpha}_L, \hat{\alpha}_{L+1}, \hat{\beta}, \hat{b} \) is the similar to the one showed in step 2.

In the previous section 2.2.2 we have discussed the relationship between \( \hat{U}_{i,y} \) and \( \hat{E}( \kappa_{i,y} ) \), normally they are not equal to each other. However, since we could not get enough information of the post-treatment site, there is no way to distinct \( \hat{U}_{i,y} \) from
Therefore, we will assume that they are equal. That is

\[ \hat{\mathbb{E}}(\kappa_{i,y}) = \hat{\mathbb{E}}(\kappa_{i,y}). \]  

(2.43)

4. Conclusion

We have discussed the reactive and proactive procedures to improve transportation safety. The proactive procedure is good in the sense it makes predictive analysis before hand and chooses to implement the most effective strategies to prevent accident from happening. The contribution of our study is to provide an accurate input for the optimization model proposed in section 2. We found Hauer’s coherent method combined with extrapolation will do the job. We also corrected the incomplete proofs and examples for the reference book-- Observational before-after study’s in road safety by Hauer,E. Future work in this direction may include the following: First, though the RTM problem is properly settled by the EB method, the potential existence of crash migration problems still need to be discussed and solved. Second, the Hauer’s coherence method imposes some assumptions about the probability distribution of crash occurrences. The estimate tends to be inaccurate when the assumption is not met. Moreover, section 3 proposed only the methodology to forecast accident of the coming year for road sections. When deal with intersections or other types of traffic designs, more specific models need to be proposed. In addition, a more refined extrapolation method needed to be discussed to better predict or forecast the accident count for the coming year.
REFERENCES