A comparison of two methods for estimating the effect of a countermeasure in the presence of regression effects

Stig Danielsson

Reprint from Accident Analysis & Prevention, vol. 18
A COMPARISON OF TWO METHODS FOR ESTIMATING
THE EFFECT OF A COUNTERMEASURE IN THE
PRESENCE OF REGRESSION EFFECTS

STIG DANIELSSON
Swedish Road and Traffic Research Institute (VTI), S-581 01 Linköping, Sweden

(Received 2 March 1983; in revised form 16 May 1985)

Abstract—The report studies the problem of estimating in a non-experimental before-and-after investigation the effect of a countermeasure on the number of traffic accidents at road junctions. The accidents are assumed to occur according to a Poisson process with different intensities at different junctions. The junctions studied in this investigation are assumed to have been selected with the blackspot-technique, i.e. junctions with high numbers of accidents during the before-period have been chosen for the investigation. In the mathematical model this has the consequence that the number of accidents occurring during the before-period at a selected junction has a truncated Poisson distribution. During the after-period the number of accidents has a Poisson distribution (without restrictions), so that the number of accidents on the average decreases between the periods even if the countermeasure has no effect. The magnitude of this regression effect is studied in the report. The observed numbers of accidents during the before and after period are used to estimate the pure effect of the countermeasure both with an intuitive method and with the maximum likelihood method. The characteristics of the two methods of estimation are illustrated with the aid of simulation studies. In general the maximum likelihood method appears preferable, mainly because it produces estimates with higher precision.

1. INTRODUCTION

One problem occurring in many before-and-after studies is the regression effect. This is caused by the selection procedure, which implies that experimental units are drawn with regard to the outcomes of the variable studied during the before-period. Assume, for example, that a study is to be made of the effect on the number of traffic accidents of a certain countermeasure implemented at road junctions. Traditionally these countermeasure-and-effect studies are made by selecting a number of road junctions which have especially high accident levels (i.e. where the number of accidents is considerably above the expected). This means that even if the countermeasure has no effect, the number of accidents at the junctions studied will decrease after the countermeasure has been implemented, since the number of accidents tends to return to the average.

The problem with regression effects in different types of before-and-after studies has been known for a long time. However it is only recently that it has been appreciated by researchers in the traffic safety field how large this problem may be. In empirical studies Brüde and Larsson [1982] observed regression effects of 25—65% for the number of traffic accidents at rural road junctions. Hauer [1980a, 1980b] has described the problem with regression effects in many ways and has given a method for estimating the effects (see Section 4.1). Hauer et al. [1983] have studied the accuracy of the proposed estimator. Hauer’s method is founded on a mathematical model specifying the number of accidents to be truncated Poisson distributed in the before-period. This is also the basis in our paper (see Section 2).

In practice, the regression effect is not of primary interest. Instead, we are concentrating our efforts in estimating the pure effect of the countermeasure. This is done in this paper by utilizing Hauer’s estimator of the regression effect, but also by using the method of maximum likelihood (see Sections 4 and 5). In Section 6 we provide approximate confidence intervals for the countermeasure-effect. The different estimators and confidence intervals are compared in Section 7 by means of simulation studies. In Section 3 we study the expected regression effect by utilizing the truncated Poisson distribution; the purpose is merely to demonstrate the magnitude of the regression effect.
2. FORMULATION OF THE PROBLEM

For the sake of simplicity we shall formulate the problem by means of an example. Assume that we wish to study the effect of a countermeasure implemented at road junctions. Consider junction no. \( i \) and let

\[ X_i = \text{the number of accidents during a period before the countermeasure.} \]

We assume in the model that different \( X_i \) are independent and that they have a Poisson distribution with expected value \( m_i \), i.e. the probability of obtaining \( r \) accidents at the junction is

\[ p_i(m_i) = P(X_i = r) = e^{-m_i} \frac{m_i^r}{r!}, \quad r = 0, 1, 2, \ldots \tag{1} \]

Note that the expected number of accidents \( m_i \) varies from junction to junction, mainly due to different traffic flows.

For simplicity we assume that the period studied after the countermeasure is implemented is of the same length as the before-period, and that the traffic pattern remains the same. For junction no. \( i \) we set

\[ Y_i = \text{the number of accidents after the countermeasure has been implemented.} \]

and can then assume that different \( Y_i \) are mutually independent and also independent of all \( X_i \). In addition all \( Y_i \) are Poisson distributed.

We assume that the effect of the countermeasure is proportionally of the same magnitude at all junctions, i.e. we assume that \( Y_i \) has a Poisson distribution with parameter \( \alpha \cdot m_i \). The appropriateness of this assumption may be questionable. Probably the effect of the countermeasure varies with different \( m_i \), according to, among other factors, the varying number of other countermeasures which have presumably been implemented at the junctions. Having no knowledge of these factors there is no justification in working with any other model than this simple multiplicative model. The model implies that the probability of obtaining \( r \) accidents at the junction after the countermeasure is

\[ p_i(\alpha m_i) = P(Y_i = r) = e^{-\alpha m_i} \frac{(\alpha m_i)^r}{r!}, \quad r = 0, 1, 2, \ldots \tag{2} \]

The countermeasure will have had no effect if \( \alpha = 1 \), a positive effect if \( \alpha < 1 \) and a negative effect if \( \alpha > 1 \).

Using a sample of road junctions we wish to estimate \( \alpha \) and assess the uncertainty of the estimate. The estimation problem is simple if the sample is made without regard to the observed numbers of accidents during the before-period. However this is not the case in practice. Instead junctions with especially high numbers of accidents, "blackspots," are chosen.

We make the assumption that junction no. \( i \) will be included in the selection only if the number of accidents \( X_i \) is at least \( k_i \). This means that for junction no. \( i \) in the sample the number of accidents \( X_i \) before the countermeasure is at least \( k_i \) and the probability of \( r \) accidents is

\[ p'_i(m_i) = P(X'_i = r) = \frac{1}{\sum_{j=k_i}^{k_{i+1}}} e^{-m_i} \frac{m_i^r}{r!}, \quad r = k_i, k_{i+1}, \ldots \tag{3} \]

\( X'_i \) is said to have a truncated Poisson distribution.

The number of accidents \( Y_i \) after the countermeasure is not influenced by the selection procedure and so the distribution of \( Y_i \) will still be as stated in (2).

Before we begin estimating \( \alpha \) we shall study in more detail the significance of the truncated Poisson distribution and the regression effect accompanying it.
3. THE TRUNCATED POISSON DISTRIBUTION AND EXPECTED REGRESSION EFFECT

We shall now limit the study to a given junction. Let $X'$ have a truncated Poisson distribution so that

$$p_r'(m) = P(X' = r) = \frac{p_r(m)}{q_r(m)}, \quad r = k, k+1, \ldots$$

where

$$p_r(m) = \frac{e^{-m} m^r}{r!} \quad \text{and} \quad q_r(m) = \sum_{r=k}^{\infty} p_r(m); \quad k \geq 1. \quad (4)$$

Also introduce $\mu_k(m)$ as the expected value of $X'$. The bias $B$ can then be written

$$B = \frac{\mu_k(m) - m}{m} \cdot 100. \quad (5)$$

i.e. $B$ is the expected difference between the numbers of accidents before and after an ineffective countermeasure. The expected regression effect $R$ is defined according to

$$R = \frac{B}{\mu_k(m)} \cdot 100. \quad (6)$$

If we can determine $\mu_k(m)$ it will be easy to study $R$. Different expressions of the expected value of a truncated Poisson distribution is given in many places in the literature, see for example Hauer [1980b], Hauer et al. [1982] and Selvin [1974].

$$\mu_k(m) = \sum_{r=k}^{m} r p_r'(m) = \frac{m}{q_k(m)} \cdot q_{k-1}(m) = m + k \frac{p_k(m)}{q_k(m)}. \quad (7)$$

Thus $B$ has the simple expression (cf. Hauer et al. [1983])

$$B = k \frac{p_k(m)}{q_k(m)} \quad (8)$$

while the regression effect will be

$$R = \left(1 - \frac{q_k(m)}{q_{k-1}(m)}\right) \cdot 100 = \frac{p_{k-1}(m)}{q_{k-1}(m)} \cdot 100. \quad (9)$$

It is also interesting to determine the dispersion obtained in the truncated Poisson distribution, which, of course, is less than the dispersion in an unrestricted Poisson distribution. Making use of the results of Hauer et al. [1983] it can be shown that the variance of $X'$ is

$$\text{Var}[X'] = m \left(1 - k \frac{p_k(m)}{q_k(m)}\right) + k^2 \frac{p_k(m)}{q_k(m)} \left(1 - \frac{p_k(m)}{q_k(m)}\right). \quad (10)$$

For different $m$ and $k$ it is then easy to compute the parameters derived above, and the results are given in Table 1.

The table shows that the regression effect may be very important. When $m$ is larger than $2k$ the regression effect is often quite small. When $m$ has the magnitude $k$ a regression effect of 20–40% is obtained. For increased truncations the effect becomes greater and in extreme but nevertheless realistic cases effects of 50–70% may occur.
Table 1. Expected value, standard deviation and expected regression effect in a truncated Poisson distribution

<table>
<thead>
<tr>
<th>m</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.45</td>
<td>0.2</td>
<td>1.10, 82</td>
<td>2.07, 90</td>
<td>3.18, 94</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.71</td>
<td>0.5</td>
<td>1.27, 61</td>
<td>2.18, 77</td>
<td>3.13, 84</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1.00</td>
<td>1.0</td>
<td>1.58, 37</td>
<td>2.39, 58</td>
<td>3.29, 70</td>
<td>4.23, 76</td>
<td>5.19, 81</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1.22</td>
<td>1.5</td>
<td>1.93, 22</td>
<td>2.64, 43</td>
<td>3.47, 57</td>
<td>4.37, 66</td>
<td>5.30, 72</td>
<td>6.25, 76</td>
<td>-</td>
</tr>
<tr>
<td>1.41</td>
<td>2.0</td>
<td>2.31, 14</td>
<td>2.91, 31</td>
<td>3.66, 46</td>
<td>4.53, 56</td>
<td>5.43, 63</td>
<td>6.36, 69</td>
<td>7.31, 73</td>
</tr>
<tr>
<td>1.73</td>
<td>3.0</td>
<td>3.16, 5</td>
<td>3.56, 16</td>
<td>4.17, 28</td>
<td>4.91, 39</td>
<td>5.73, 48</td>
<td>6.60, 55</td>
<td>7.51, 60</td>
</tr>
<tr>
<td>2.00</td>
<td>4.0</td>
<td>4.07, 2</td>
<td>4.32, 7</td>
<td>4.77, 16</td>
<td>5.38, 26</td>
<td>6.10, 34</td>
<td>6.91, 42</td>
<td>7.76, 48</td>
</tr>
<tr>
<td>2.24</td>
<td>5.0</td>
<td>5.03, 1</td>
<td>5.18, 3</td>
<td>5.48, 9</td>
<td>5.95, 16</td>
<td>6.57, 24</td>
<td>7.28, 31</td>
<td>8.07, 38</td>
</tr>
<tr>
<td>2.83</td>
<td>8.0</td>
<td>8.00, 0</td>
<td>8.02, 0</td>
<td>8.09, 1</td>
<td>8.24, 3</td>
<td>8.51, 6</td>
<td>8.91, 10</td>
<td>9.42, 15</td>
</tr>
<tr>
<td>3.16</td>
<td>10.0</td>
<td>10.00, 0</td>
<td>10.00, 0</td>
<td>10.02, 0</td>
<td>10.08, 0</td>
<td>10.19, 2</td>
<td>10.41, 4</td>
<td>10.72, 7</td>
</tr>
<tr>
<td>3.87</td>
<td>15.0</td>
<td>15.00, 0</td>
<td>15.00, 0</td>
<td>15.00, 0</td>
<td>15.00, 0</td>
<td>15.00, 0</td>
<td>15.00, 0</td>
<td>15.00, 0</td>
</tr>
<tr>
<td>4.47</td>
<td>20.0</td>
<td>20.00, 0</td>
<td>20.00, 0</td>
<td>20.00, 0</td>
<td>20.00, 0</td>
<td>20.00, 0</td>
<td>20.00, 0</td>
<td>20.00, 0</td>
</tr>
</tbody>
</table>

\[ \hat{\alpha} = \frac{Y}{\hat{m}} \]  
\[ \hat{R} = \frac{X' - \hat{m}}{X'} \cdot 100 \]

where \( \hat{m} \) is a suitable estimator of \( m \). The estimator of the regression effect \( R \) in (6) is

We should note that there is a regression effect not only on the expected value but also on the standard deviation. The reduction in the standard deviation is fairly important; it is not unrealistic that \( \sqrt{\hat{m}} \) may be halved. This may have serious consequences for statistical methods using predicted number of accidents. Unless adjustment is made for the regression effects, incorrectly small prediction limits will be obtained in addition to an incorrectly high predicted value.

4. Estimation of the effect of a countermeasure and the regression effect for a single selected junction

We define \( X' \) and \( Y \) in accordance with Sections 2 and 3 and note that the expectation of \( Y \) is \( \alpha \cdot m \). A very natural estimator of \( \alpha \) is

\[ \hat{\alpha} = \frac{Y}{\hat{m}} \]  
\[ \hat{R} = \frac{X' - \hat{m}}{X'} \cdot 100 \]

and we see that the main problem is to estimate \( m \).

Below we describe three methods of estimating \( m \) and \( \alpha \). The first two methods are motivated on intuitive grounds, while the last method is the method of maximum likelihood.

4.1 Hauer’s method

Hauer [1980a] and [1980b] proposes an estimator for the sum of the accidents at \( n \) junctions (all truncated at the same \( k \)). This result can also be found in an article by Robbins [1977].
The proposal can be interpreted as a desire to estimate

\[
\frac{p_k(m)}{q_k(m)}
\]

with a suitable relative frequency. However if only one measurement has been made at the junction the estimator will be

\[
\text{Estimated } \left( \frac{p_k(m)}{q_k(m)} \right) = \begin{cases} 
1 & \text{if } X' = k \\
0 & \text{if } X' > k
\end{cases}
\]

i.e. the estimator of \( m \) will be (when \( E[X'] \) is estimated with \( X' \))

\[
\hat{m} = \begin{cases} 
0 & \text{if } X' = k \\
X' & \text{if } X' > k
\end{cases}
\]

(cf. Hauer et al. [1983]).

As is proven for example by Hauer et al. [1983] the estimator is unbiased with variance

\[
\text{Var}[\hat{m}] = m(1 + B).
\]

This large variance is discussed in detail by Hauer et al. [1983].

4.2 The moment method

Instead of estimating the conditional probability

\[
\frac{p_k(m)}{q_k(m)}
\]

and then implicitly solving \( m \) as above, it would be better to try to estimate \( m \) directly.

An estimator of \( m \) with the moment method means that an attempt is made to solve the equation

\[
X' = E[X'] = m + B
\]

with regard to \( m \).

The discussion of the solution of eqn (15) is carried out in Section 4.3.

4.3 The method of maximum likelihood

Most often better estimators are obtained with the maximum likelihood method (ML method) than with the moment method. The ML method implies that the values which maximizes the likelihood (probability of the observed values) are used as the estimators of the actual parameters. In this case it is therefore possible to estimate both \( m \) and \( \alpha \). The likelihood for the observed values \( x' \) and \( y \), corresponding to \( X' \) and \( Y \) is [see (3) and (1)].

\[
L(m, \alpha) = p_{y(m)} \cdot q_{x(m)}
\]

Maximizing \( L \) is equivalent to maximizing the log-likelihood

\[
\ell = \ln L = (x' + y) \ln (m) - \ln \left( \sum_{r=k}^{\infty} \frac{m^r}{r!} \right) - \alpha m + y \ln \alpha - \ln (x!) - \ln (y!).
\]

Differentiating (17) with respect to \( m \) and \( \alpha \) and equating to zero gives:

\[
\alpha = \frac{y}{m}
\]

\[
x' = m \cdot \frac{q_{y-1}(m)}{q_{x}(m)}.
\]
The second equation in (18) can be found in Cohen [1972] and Selvin [1974] as the ML equation for estimating \( m \), when observations are made solely on a truncated Poisson distribution.

A straightforward calculation shows that the solutions \( a^* \) and \( m^* \) of (18) maximizes the log-likelihood, and therefore are the ML estimators. It is interesting to note, that \( a^* \) agrees with the intuitive \( \hat{a} \) in (11) and that \( m^* \) is identical to the moment estimator (15).

With a little calculation it can be seen that the function

\[
f(m) = \frac{c_k - 1(m)}{c_k(m)}
\]

is steadily increasing and has the minimal value \( k \) at \( m = 0 \). Therefore, the ML estimator \( m^* \) is unique. Furthermore, the function \( f(m) \) lies above the line \( y = m \) and has an asymptote at \( y = x' \). Since the estimator \( m^* \) is zero for \( x' = k \), it follows that \( m^* \leq \hat{m} \), where \( \hat{m} \) is Hauer’s estimator in (13). But \( \hat{m} \) is an unbiased estimator of \( m \), and so the ML estimator \( m^* \) will on average underestimate \( m \).

We have seen above that the ML estimator and the moment estimator of \( m \) are identical. Comparing Hauer’s estimator and the ML estimator, the former at first sight seems better because it is unbiased.

However, our emphasis is to estimate the effect of the countermeasure, and both from a theoretical and practical point of view, \( \alpha \) is not estimable when \( x' = k \). But if \( x' > k \) the estimate of \( m \) with Hauer’s method is simply \( x' \) and so the estimate of \( \alpha \) is not corrected at all for the regression effect (this effect is estimated to zero). On the contrary, the ML estimate of \( m \) never equals \( x' \) in this case and the estimate of the regression effect is strictly positive. In this regard the ML estimator of \( \alpha \) is better than Hauer’s estimator.

Now, the situation discussed in this section is not very practical. In most cases the countermeasure is implemented at several junctions. This is the situation actually treated by Hauer [1980a, 1980b], and obviously an estimate of \( \alpha \) with Hauer’s method would be desirable.

5. MAXIMUM LIKELIHOOD ESTIMATION OF THE EFFECT OF A COUNTERMEASURE IMPLEMENTED AT SEVERAL SELECTED JUNCTIONS

Consider the selected junction no. \( i, i = 1, 2, \ldots, n \), and define \( X'_i \) and \( Y_i \) in accordance with Sections 2 and 3. Observe that we assume a common effect of the countermeasure at each junction.

For the observations

\[
(x'_1, y_1), (x'_2, y_2), \ldots, (x'_n, y_n)
\]

calculations for the likelihood as in Section 4.3 give

\[
L(\alpha, m_1, \ldots, m_n) = \prod_{i=1}^{n} p_{x'_i}(m_i)p_{y_i}(\alpha m_i)
\]

\[
l = \ln L = -\alpha \sum_{i=1}^{n} m_i + \sum_{i=1}^{n} (x'_i + y_i) \ln m_i
\]

\[
- \sum_{i=1}^{n} \ln \left( \sum_{r=k}^{m_i} \frac{m_i!}{r!} \right) + \sum_{i=1}^{n} y_i \ln \alpha - \ln \left( \prod_{i=1}^{n} x'_i! y_i! \right)
\]

\[
\begin{align*}
\frac{\partial l}{\partial \alpha} &= -\sum m_i + \sum \frac{y_i}{\alpha} = 0 \\
\frac{\partial l}{\partial m_i} &= -\alpha + \frac{x'_i + y_i}{m_i} - \frac{q_{k-1}(m_i)}{q_k(m_i)} = 0; \quad i = 1, 2, \ldots, n.
\end{align*}
\]
Solving (19) gives

$$\alpha^* = \frac{\sum y_i}{\sum m_i^*}$$

(20)

where $m_1^*, \ldots, m_n^*$ are obtained from the equations

$$x_i + y_i = m_i \cdot \left( \frac{q_{k-1}(m_i)}{q_k(m_i)} + \frac{\sum y_i}{\sum m_i} \right); \quad i = 1, 2, \ldots, n.$$  

(21)

Obviously we do not now obtain the same type of equations as in (18). The equation system above requires in principle a simultaneous solution with regard to all $m_i$ and will therefore be numerically more troublesome than the corresponding problem in Section 4.3. We can rewrite the equations as

$$m_i \cdot \frac{q_{k-1}(m_i)}{q_k(m_i)} = (x_i' + y_i) - \alpha^* m_i$$

(22)

from which we can see that the same solution as in (18) is obtained only if all $y_i = 0$. In the case where $y_i = 0$ but at least one other $y_j > 0$ we see that the ML estimate $m_i^*$ is now smaller. Generically we can see that it is the relation between $y_i$ and $\alpha^* m_i$ that determines whether the particular estimate becomes larger or smaller than that in (18). In the extreme case where $x_i' = k_i, m_i = 0$ will be the ML solution if $y_i = 0$ but not, however, if $y_i > 0$. We see also that the ML estimators $m_1^*, \ldots, m_n^*$ are not now independent stochastic variables since $m_i^*$ will depend on $\Sigma y_i$ and $\Sigma m_i^*$.

The equation system (22) which gives the ML estimates of $m_1, \ldots, m_n$ does not in practice require to be solved simultaneously with regard to all $m_i$. Instead (for $i = 1, 2, \ldots, n$) each one-dimensional equation can be solved by fixing $\alpha^*$ [the equation will be of the same type as (18)]. Different $\alpha^*$ are then tested and the solution $(m_1^*, m_2^*, \ldots, m_n^*)$ is accepted when the deviation between the fixed $\alpha^*$ and $\sum y_i / \sum m_i^*$ becomes sufficiently small.

An iterative method for solving (22) can be found by noting that summing (22) gives:

$$\sum m_i \frac{q_{k-1}(m_i)}{q_k(m_i)} = \sum x_i'.$$

(23)

Equation (23) is also obtained by summing (18) and therefore the solutions $m_i$ of (18) can be used as initial values in an iterative method for solving (22) using eqn (20) for successive calculations of $\alpha$.

Having the point estimator of $\alpha$ we turn to the problem of assessing the uncertainty of $\alpha^*$. Asymptotic theory for ML estimators cannot be used because we do not have repeated observations on the same distribution. Instead, we try to use general large sample theory.

6. ASSESSING THE UNCERTAINTY OF THE ESTIMATED COUNTERMEASURE EFFECT. CONFIDENCE LIMITS

The ML estimator of $\alpha$ is given by formula (20). Without a formal proof we conjecture that

$$\left( \sum Y_i, \sum m_i^* \right)$$

is asymptotically bivariate normal

with mean vector

$$\begin{pmatrix} \alpha \sum m_i \\ \sum m_i \end{pmatrix}$$

and covariance matrix

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$
where

$$\sigma_{11} = \alpha \sum m_i.$$ 

(24)

With this assumption

$$\alpha^* = \frac{\sum Y_i}{\sum m_i^*}$$

is asymptotically normal, \(N(\mu, \sigma)\) (see Rao, 1973, p. 387). It is easy to derive

$$\mu = \alpha$$

$$\sigma^2 = \frac{\sigma_{11}}{\left(\sum m_i\right)^2} + \frac{\sigma_{22} \cdot \alpha^2}{\left(\sum m_i\right)^2} - 2 \frac{\sigma_{12} \cdot \alpha}{\left(\sum m_i\right)^2}.$$ 

(25)

Furthermore, we conjecture that

$$\frac{\alpha^* - \alpha}{\sigma^*}$$ 

(26)

is asymptotically \(N(0, 1)\), where \(\sigma^*\) is identical to \(\sigma\) in (25) but \(\alpha\) and \(\sum m_i\) have been replaced by \(\alpha^*\) and \(\sum m_i^*\). Using (26) a confidence interval for \(\alpha\) will be

$$\alpha = \alpha^* \pm \text{const} \cdot \sigma^*.$$ 

(27)

Formula (26) is valid if \(\sigma^2\) is a continuous function of \(\alpha\) and \(\sum m_i\). Unfortunately it is not easy to derive expressions for \(\sigma_{12}\) and \(\sigma_{22}\). Presumably \(\sigma_{12}\) is slightly positive and can be cancelled in formula (25), resulting in a confidence interval (27) slightly too broad. We propose the following estimator of \(\sigma_{22}\).

$$\sigma_{22}^* = \frac{n}{n - 1} \sum (m_i^* - \bar{m})^2$$

where

$$m^* = \frac{1}{n} \sum m_i^*.$$ 

(28)

It is likely that \(\sigma_{22}^*\) overestimates \(\sigma_{22}\) (an estimator with independent terms \(m_i^*\) always gives overestimates). Therefore, with

$$(\sigma^*)^2 = \frac{\sum Y_i}{\sum m_i^*} \left(1 + \frac{\sum Y_i}{\sum m_i^*} \cdot \frac{n}{n - 1} \sum (m_i^* - \bar{m})^2\right)$$ 

(29)

it is likely that the confidence interval (27) is too broad.

Finally, an estimator of \(\alpha\) with Hauer's method is by analogy with (20)

$$\hat{\alpha} = \frac{\sum Y_i}{\sum \hat{m}_i}.$$
where

\[ \hat{m}_i = \begin{cases} 0 & \text{if } x'_i = k_i \\ x'_i & \text{if } x'_i > k_i. \end{cases} \]  

(30)

By using arguments similar to those above we can derive the following confidence interval for \( \alpha \).

\[ \alpha = \hat{\alpha} \pm \text{const.} \sqrt{\frac{\sum Y_i}{\left( \sum \hat{m}_i \right)^2} \left( 1 + \frac{\sum Y_i}{n \left( \sum \hat{m}_i \right)^2} \right) \frac{n}{n-1} \sum (\hat{m}_i - \hat{m})^2} \]

where

\[ \hat{m} = \frac{1}{n} \sum \hat{m}_i. \]  

(31)

In order to gain a clear picture of the properties of the estimators of \( \alpha \) we must carry out a simulation study since the results in this section are asymptotic and the calculations leading to the confidence intervals are rather crude. Nevertheless, we can state (using Jensen’s inequality) that \( \hat{\alpha} \) always overestimates \( \alpha \) and this is probably often the case of \( \alpha^* \) too. It is however harder to state anything about the confidence intervals given in (27) and (31).

7. RESULTS OF THE SIMULATION STUDIES

A large number of simulation studies have been made to assess the properties of the estimates \( \alpha^* \) and \( \hat{\alpha} \). Poisson-distributed data \( y_1, y_2, \ldots, y_n \) with expected values \( \alpha m_1, \alpha m_2, \ldots, \alpha m_n \) have been simulated, in addition to corresponding truncated Poisson-distributed data \( x_1, x_2, \ldots, x_n \) [See (2) and (3)]. The truncation points \( k_i, i = 1, 2, \ldots, n \) have been selected according to

\[ k_i = \begin{cases} 3 \text{ with probability 0.2} \\ \max (3, [m_i + 2\sqrt{m_i}] + 1) \text{ with probability 0.8.} \end{cases} \]

This means very hard restrictions on selected junctions and thus large regression effects can be expected.

For each data series it is easy to calculate \( \hat{\alpha} \) in (31) while the ML estimate \( \alpha^* \) in (20) requires a large number of numerical calculations (see Section 5). The approximate confidence limits in (27) and (31) are then easy to calculate. We have also calculated the estimate of \( \alpha \) which does not take into account the regression effect, i.e.

\[ \alpha' = \frac{\sum y_i}{\sum x_i}. \]  

(32)

Table 2a. Simulation results for a wide range of expected accident numbers and the countermeasure effect 20%

<table>
<thead>
<tr>
<th>Expected value</th>
<th>Exact confidence limit (95 %)</th>
<th>Approximate confidence limit (95 %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\alpha} )</td>
<td>0.84 ± 0.03</td>
<td>0.28</td>
</tr>
<tr>
<td>( \alpha^* )</td>
<td>0.79 ± 0.02</td>
<td>0.15</td>
</tr>
<tr>
<td>( \alpha' )</td>
<td>0.47 ± 0.01</td>
<td></td>
</tr>
</tbody>
</table>

\( \alpha = 0.8, m_i = 0.4, 0.5, \ldots, 2.9, 3.0, 3.5, \ldots, 20.0, n = 61. \)
Table 2b. Simulation results for small accident numbers and the countermeasure effect 20%

<table>
<thead>
<tr>
<th></th>
<th>Expected value</th>
<th>Exact confidence limit</th>
<th>Approximate confidence limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.94 ± 0.05</td>
<td>1.06</td>
<td>0.85 ± 1.74</td>
</tr>
<tr>
<td>$\alpha^*$</td>
<td>0.84 ± 0.02</td>
<td>0.38</td>
<td>0.45 ± 0.20</td>
</tr>
<tr>
<td>$\alpha'$</td>
<td>0.13 ± 0.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\alpha = 0.8, m_i = (0.05, 0.10, 0.15, \ldots, 1.00) \times 3, n = 60 (\times 3$ means that each $m_i$ is repeated 3 times and independent observations are made for each (repeated) $m_i$).

The whole procedure is then repeated $N$ times. For the results below we have generally selected $N = 100$ in order to achieve reasonable calculation times. The mean of the $\alpha$-estimates is then an approximation of the expected value of the estimates. The confidence limits (95%) have also been determined as $\pm 1.96 \text{SD}/\sqrt{N}$, where $\text{SD}$ is the standard deviation of the $\alpha$-estimates. The "exact" confidence limit for $\alpha$ is calculated as $1.96 \cdot \text{SD}$. Furthermore the averages of the approximate confidence limits and their prediction limits ($\pm 1.96$ standard deviation of the confidence limits) are also calculated.

Table 2a gives the results for $\alpha = 0.8$, i.e. the effect of the countermeasure $1 - \alpha = 20\%$. We have chosen

$$m_i = 0.4, 0.5, \ldots, 2.9, 3.0, 3.5, 4.0, \ldots, 19.5, 20.0.$$  

The sample size $n$ will then be 61.

We see that in this case with widely varying $m_i$ the traditional estimate $\alpha'$ has a regression effect of just over 30%. The ML estimator $\alpha^*$ generally has correct expected value, while Hauer's estimator $\alpha$ somewhat underestimates the effect of the countermeasure ($1 - \alpha$). The ML estimator $\alpha^*$ is more precise (has a smaller dispersion) than $\alpha$ as is shown by the exact confidence limits. The approximate confidence limits are somewhat too large. We see also that with Hauer's method no positive effect of the countermeasure will be ascertained in this case, although this may be possible with the ML method.

The large $m_i$ dominate the above results. We have also made a simulation with extremely small $m_i$ values (where we carried out $N = 400$ simulations in order to obtain reasonably safe results). In Table 2b we see that $\alpha^*$ is incomparably the best estimator in every way. However

Table 2c. Simulation results for different sample sizes and the countermeasure effect 20%

<table>
<thead>
<tr>
<th></th>
<th>Expected value</th>
<th>Exact confidence limit</th>
<th>Approximate confidence limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.89 ± 0.07</td>
<td>0.73</td>
<td>0.69 ± 1.00</td>
</tr>
<tr>
<td>$\alpha^*$</td>
<td>0.83 ± 0.03</td>
<td>0.34</td>
<td>0.37 ± 0.16</td>
</tr>
<tr>
<td>$\alpha'$</td>
<td>0.29 ± 0.01</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$n=27$

$\alpha$ | 0.84 ± 0.05  | 0.49                  | 0.43 ± 0.45                |
| $\alpha^*$ | 0.82 ± 0.03  | 0.25                  | 0.26 ± 0.09                |
| $\alpha'$ | 0.29 ± 0.01  |                       |                            |

$n=54$

$\alpha$ | 0.84 ± 0.03  | 0.33                  | 0.34 ± 0.21                |
| $\alpha^*$ | 0.84 ± 0.03  | 0.25                  | 0.22 ± 0.06                |
| $\alpha'$ | 0.29 ± 0.01  |                       |                            |

$n=81$

$\alpha$ | 0.83 ± 0.03  | 0.25                  | 0.26 ± 0.11                |
| $\alpha^*$ | 0.85 ± 0.02  | 0.20                  | 0.17 ± 0.04                |
| $\alpha'$ | 0.29 ± 0.01  |                       |                            |

$n=135$

$\alpha = 0.8, m_i = (0.4, 0.5, \ldots, 3.0) \times k, k = 1, 2, 3, 5.$
we must mention that in this situation it is quite impossible to ascertain a positive effect of the countermeasure, while the traditional estimator points to an effect of almost 90%.

In order to illuminate the extent to which the sample size influences the comparison between the properties of the two estimators the simulations in Table 2c were carried out. We see that both methods consistently have a certain tendency of overestimating \( \alpha \) (i.e. to underestimate the effect of the countermeasure). For small samples the ML estimator \( \alpha^* \) will be closer to \( \alpha \) than Hauer's estimator \( \hat{\alpha} \). For the largest sample the situation will be the reverse. From the exact confidence limits we can state that \( \alpha^* \) is throughout a considerably more precise estimator than \( \hat{\alpha} \). For small samples \( \alpha^* \) has twice as high precision. The approximate confidence limits agree well with the exact limits.

Finally it is interesting to see how the magnitude of the effect of the countermeasure influences the comparison between the two estimation methods. The results are given in Table 2d. For large \( \alpha \), \( \alpha^* \) will be closer to the true value than \( \hat{\alpha} \), while the situation will be the reverse for small values of \( \alpha \). Throughout, \( \alpha^* \) is more precise than \( \hat{\alpha} \); furthermore the dispersion of \( \alpha^* \) does not appear to be influenced by the magnitude of \( \alpha \). The approximate confidence limits agree well with the exact limits.

Acknowledgements—I am grateful to the referee for helpful comments and for proposing the iterative method described in Section 5, eqn (23).

REFERENCES
