



# LUND UNIVERSITY

## External costs of transports imposed on neighbours and fellow road users

Winslott Hiselius, Lena

2005

[Link to publication](#)

*Citation for published version (APA):*

Winslott Hiselius, L. (2005). *External costs of transports imposed on neighbours and fellow road users*. [Doctoral Thesis (compilation), Department of Economics]. Department of Economics, Lund University.

*Total number of authors:*

1

### General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117  
221 00 Lund  
+46 46-222 00 00



# Estimating the relationship between accident frequency and homogeneous and inhomogeneous traffic flows

Lena Winslott Hiselius\*

*Department of Economics or Department of Technology and Society, Lund University, P.O. Box 118, Lund 22100, Sweden*

Received 5 January 2002; received in revised form 19 November 2003; accepted 24 November 2003

---

## Abstract

This paper estimates the relationship between accident frequency and the traffic flow empirically treating the hourly traffic flow in two different ways, as consisting of homogenous vehicles and as consisting of cars and lorries. Rural roads in Sweden are studied using Poisson and Negative Binominal regression models. It is found that important information is lost if no consideration is taken to differences between vehicle types when estimating the marginal effect of the traffic flow. The accident rate decreases when the traffic flow is treated as if homogeneous. However, when cars are studied separately the result suggests that the accident rate is constant or increases. The result with respect to lorries is reversed, indicating a decreasing number of accidents as the number of lorries increases.

© 2004 Elsevier Ltd. All rights reserved.

**Keywords:** Accident prediction models; Road accidents; Hourly traffic flow; Poisson; Negative Binomial

---

## 1. Introduction

Studies that analyse the relationship between accident frequency and traffic flows show a great variation in their results. An early report from Vickery (1968) suggests that the marginal accident rate is 1.5 times the average accident rate. On the other hand, Vitaliano and Held (1991) cannot detect any significant increase in the accident rate as the traffic flow increases and, according to Hauer and Bamfo (1997), and a majority of the results reviewed in Ardekani et al. (1997), the accident rate even decreases with an increasing number of vehicles. Furthermore, only a few empirical studies analyse the effect of the traffic flow separated for different road user groups. One example is the study by Jovanis and Chang (1986), where they found that the accident frequency increased with the number of car and lorry miles travelled. However, it is difficult to distinguish the marginal effect of another car or lorry as the number of miles travelled varies in their study.

There is, thus, an interest in studying the effect of the traffic flow more thoroughly, and the aim of this article is to estimate the relationship between the accident frequency and the traffic flow treated as both homogeneous, i.e. consisting of homogeneous vehicles, and inhomogeneous, i.e.

consisting of cars and lorries. The data is restricted to accidents on sections in rural areas of Sweden.

## 2. Model

The model seeks to describe the relationship between the number of vehicles per hour and the accident frequency. It is developed under the assumption that a limited and homogeneous road system is studied and only the traffic flow affects the number of accidents. In the first case the flow of vehicles per hour,  $q$ , is assumed to be homogeneous, and with all vehicles influencing the occurrence of an accident the same way. In the second case the traffic flow is defined as the number of cars and lorries per hour,  $q_c$  and  $q_l$ . The accident frequency,  $Z$ , is calculated as the expected number of accidents per unit of time and kilometer in order to apply the model to time periods and road systems of different length.

## 3. Data

Data is collected from 83 road sections in rural areas of Sweden, where the number of passing vehicles is continuously counted by the Swedish National Road Administration. The assumption is made that the counted traffic flow at a stationary place is valid along the section. Information is collected on police reported accidents with personal

---

\* Fax: +46-46-12-32-72.

E-mail address: lena.hiselius@tft.lth.se (L.W. Hiselius).



Table 1  
Number of accidents in the data set

Road type	Number of accidents			Total number of daytime accidents
	Total	Single-vehicle	Multi-vehicle	
I	59	29	30	49
II	83	52	31	51
III	179	107	72	111
IV	186	121	65	129

casualties that occurred on the studied road sections from 1989 to the middle of 1995. Given the time and date of the accidents, the hourly traffic flow that prevailed at the time of each accident is obtained. Furthermore, in order to calculate traffic flow frequencies, information on the number of hours that each traffic flow was observed during 1990 is also collected on the assumption that this is a representative year.

Only accidents occurring on sections without intersections are included. Accidents involving animals are excluded together with accidents that may be considered specific for each road section. Since it is not possible to obtain information on driving speed, the analysis will be made bearing in mind that the estimated effect of the traffic flow may also be an effect of speed adjustment. Moreover, there are several other factors likely to influence the occurrence of accidents, e.g. weather, road conditions, type of vehicle and drivers' characteristics. In order to take some of these factors into account, daylight accidents are studied separately.

Four road types are distinguished. Road type I with speed limit 70 km/h and road width 6–9.7 m, road type II with speed limit 90 km/h and road width 6–7.9 m, road type III with speed limit 90 or 110 km/h and road width 8–13 m, without separated road lanes, and road type IV, motorways, with speed limit 90 or 110 km/h. The numbers of accidents that occurred on the studied sections are presented in Table 1.

For the road types I, II and III the traffic flow is counted as the number of vehicles per hour driving in both directions. Thus, there is no information on the traffic flow per road lane. This is a problem since depending on the distribution of the traffic flow, different types of accidents may occur. The same situation occurs when studying road type IV. Even though the traffic flow is counted for each direction for this road type, there is no information on the traffic in each lane.

#### 4. Accident frequency

In order to analyse the accident frequency,  $Z$ , the distribution model of the number of accidents,  $Y$ , is first discussed. Several authors use a Poisson distribution to approximate the number of accidents occurring during a certain time interval, e.g. Hall (1986) and Nicholson and Wong (1993). The assumption is based on the fact that although the total number of accidents is reasonably stable over a period, each accident is unpredictable. A reasonable assumption is furthermore

that accidents are independent events. However, it may be argued that individual collisions in a multi-collision accident are not mutually independent events. This problem is avoided here, since a traffic accident in this study is defined without consideration being given to the number of vehicles involved. Let  $Y_i$  denote the number of accidents occurring on a specific road site during a given time period. If we assume that the number of accidents follows a Poisson distribution with expected value, and thus variance, equal to  $\lambda_i$ , the probability that the number of accidents during this period will be equal to  $y_i$ , may be written as

$$P(Y_i = y_i) = \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} \quad y_i = 0, 1, 2, 3, \dots$$

There is, however, no reason to believe that  $\lambda_i$  is same at all sites, and there will usually be a variation of the  $\lambda_i$  around a mean,  $\lambda$ . The variation of the  $Y_i$  (the number of accidents occurring at different sites) will thus comprise both the variation of  $y_i$  around  $\lambda_i$  and the variation of  $\lambda_i$  around  $\lambda$ . The variance will therefore tend to be larger than the mean, i.e. we can observe so called over dispersion with respect to the Poisson distribution. The variation in  $\lambda_i$  is often assumed to be Gamma distributed. Assuming  $\lambda_i$  to have a Gamma distributed parameter  $\kappa$ , it can be shown that the number of accidents,  $Y$ , follows a Negative Binomial distribution, e.g. in Kulmala (1995). The probability density function for  $Y$  can now be written as

$$P(Y = y) = \frac{\Gamma(\kappa + y)}{\Gamma(\kappa)y!} \left( \frac{\kappa}{\kappa + \lambda} \right)^\kappa \left( \frac{\lambda}{\kappa + \lambda} \right)^y \quad y = 0, 1, 2, 3, \dots$$

and variance

$$\sigma^2 = \lambda + \frac{\lambda^2}{\kappa}$$

The Negative Binomial distribution is thus an extension of the Poisson distribution, where the gamma parameter,  $\kappa$ , determines the shape of the distribution. As  $\kappa \rightarrow \infty$ , the rate of over dispersion declines and the distribution approaches the Poisson.

So far the distribution model of the number of accidents has been discussed. However, when analysing accident frequency, there is also need for a regression model that can describe this phenomenon. An exponential function is a common formulation, since this function ensures that the expected number of accidents is a positive number. Furthermore, measuring the explanatory variables on the logarithmic scale,  $G_j = \ln(X_j)$ , the accident frequency may be written as a multiplicative function for which the value of the exponents can be estimated directly

$$E(Y) = \exp \left[ \sum \beta_j \ln(X_j) \right] = \pi X_j^{\beta_j}$$

The result of a regression model is, however, often dependent on the choice of model function, Hauer and Bamfo (1997). Using an exponential regression model, as described



Table 2  
Statistics on traffic flow data

Road type	Number of traffic flow intervals		Average traffic flow			Maximum traffic flow		
	Homogeneous traffic flow	Inhomogeneous traffic flow	Homogeneous traffic flow	Inhomogeneous traffic flow		Homogeneous traffic flow	Inhomogeneous traffic flow	
				Cars	Lorries		Cars	Lorries
I	98	408	153	139	12	2712	2328	312
II	38	159	94	81	8	987	938	132
III	131	595	250	232	24	1687	1645	267
IV	121	458	550	498	55	3088	2937	326

Table 3  
Statistics on total number of accidents (Tot), single-vehicle (Sing) and multi-vehicle accidents (Mult)

Road type	Number of accidents per traffic flow interval, average						Number of accidents per traffic flow interval, variance					
	Homogeneous traffic flow			Inhomogeneous traffic flow			Homogeneous traffic flow			Inhomogeneous traffic flow		
	Tot	Sing	Mult	Tot	Sing	Mult	Tot	Sing	Mult	Tot	Sing	Mult
I	0.6	0.3	0.3	0.1	0.1	0.1	1.4	0.7	0.4	0.4	0.2	0.1
II	2.2	1.4	0.8	0.5	0.3	0.2	4.4	3.8	0.7	1.7	0.9	0.4
III	1.4	0.8	0.5	0.3	0.2	0.1	4.5	2.5	0.9	0.6	0.3	0.2
IV	1.5	1.0	0.5	0.4	0.3	0.1	3.8	2.2	0.7	1.0	0.5	0.2

previously, restricts the model to the description of monotonic functions, i.e. the functions may be either increasing or decreasing, not both. The use of this model function is supported by studying the data visually and by analysing the residuals.

#### 4.1. Variables and regression models

In this study, the relationship between the accident frequency and the traffic flow is estimated empirically. In order to apply the result to time periods and road systems of different length, the accident frequency,  $Z$ , is defined as the expected number of accidents,  $E(Y)$ , per hour and kilometre. A parameter,  $H$ , is also defined, see Appendix A. This parameter describes the exposure, i.e. the number of hours and kilometres that each road type has been studied.

The accident frequency cannot be estimated as a function of the traffic flow directly, since the variable describing the accident frequency is not Poisson or Negative Binomial distributed. The regression model is therefore rewritten as  $E(Y_t) = Z_t \times H_t$  so that the accident frequency may be estimated separately as a function of the traffic flow. The complete regression model for the case of homogeneous traffic flow is estimated as

$$E(Y_t) = \exp[\beta_1 + \beta_2 \ln(q_t) + \beta_3 \ln(H_t)] = [e^{\beta_1} \times (q_t)^{\beta_2}] H_t$$

with the restriction,  $\beta_3 = 1$ . When treating the traffic flow as consisting of cars and lorries, the following regression model is estimated

$$E(Y_t) = \exp[\beta_1 + \beta_2 \ln(q_c)_t + \beta_3 \ln(q_l)_t + \beta_4 \ln(H)_t] \\ = \left[ e^{\beta_1} \times (q_c)_t^{\beta_2} \times (q_l)_t^{\beta_3} \right] H_t$$

with the restriction  $\beta_4 = 1$ .

The regression models are estimated applying both the Poisson and the Negative Binomial distribution. The statistical program LIMDEP version 7 is used. In order to evaluate the goodness-of-fit, four indicators are studied; the scaled deviance statistic, the over dispersion parameter,  $\kappa$ , estimated in the Negative Binomial regression model and two test functions for over dispersion.

## 5. Results

The regression analysis of this study is based on aggregates of hours with similar traffic flows, see Appendix A. Traffic flow statistics per road type are presented in Table 2 together with the number of traffic flow intervals used. In Table 3, statistics on the aggregated number of accidents per traffic flow interval is presented for various accident types<sup>1</sup>. The average number of accidents per traffic flow interval is generally low when inhomogeneous traffic flows or different accidents types are studied.

The results from the regression analysis generally suggest a good fit for both the Poisson and the Negative Binomial models. Distributional assumptions do not seem to affect the results since there are only small differences. The test statistics that are used for over dispersion occasionally contradict the size of the dispersion parameter. Contradictions tend to arise when the scaled deviance is almost the same for the Poisson and the Negative Binomial regression models, suggesting a good fit for both. Since the estimated coefficients are practically the same, only the results of the

<sup>1</sup> Statistics for daytime traffic flows and accidents are left out since these statistics corresponds largely to the presented statistics.



Table 4  
Expected number of accidents occurring all day and in daylight, homogeneous traffic flow<sup>a</sup>

	Expected number of accidents				Expected number of accidents in daylight			
	Road type I	Road type II	Road type III	Road type IV	Road type I	Road type II	Road type III	Road type IV
Constant	−22.21 (−35.60) <sup>b</sup>	−20.70 (−47.33) <sup>b</sup>	−19.90 (−54.71) <sup>b</sup>	−19.58 (−40.30) <sup>b</sup>	−22.07 (−27.99) <sup>b</sup>	−20.45 (−29.08) <sup>b</sup>	−18.99 (−26.99) <sup>b</sup>	−14.83 (−16.88) <sup>b</sup>
ln( <i>q</i> )	0.86 (8.06) <sup>b</sup>	0.49 (4.90) <sup>b</sup>	0.41 (5.99) <sup>b</sup>	0.39 (5.04) <sup>b</sup>	0.86 (6.55) <sup>b</sup>	0.43 (2.80) <sup>b</sup>	0.25 (2.05) <sup>b</sup>	−0.32 (−2.36) <sup>b</sup>
ln( <i>H</i> )	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
d.f.	95	35	128	118	93	34	128	118
Scaled deviation	84.61	22.56	85.75	114.17	86.76	20.05	87.50	122.03
LM-stat	10.62 <sup>c</sup>	3.18	9.02 <sup>c</sup>	4.53	7.17	11.48 <sup>c</sup>	13.88 <sup>c</sup>	5.59
<i>t</i> -Statistics 1	1.14	−1.86	0.12	−0.47	1.38	−1.90	−0.85	0.29
<i>t</i> -Statistics 2	2.26 <sup>c</sup>	−1.46	−1.06	−0.45	2.25 <sup>c</sup>	−2.13 <sup>c</sup>	−1.54	1.72

<sup>a</sup> Values in parenthesis are *t*-statistics.

<sup>b</sup> Significantly different from zero at the 5% level.

<sup>c</sup> The null-hypothesis of over dispersion is significantly rejected at the 5% level.

Table 5  
Expected number of single and multi-vehicle accidents, homogeneous traffic flow<sup>a</sup>

	Expected number of single-vehicle accidents				Expected number of multi-vehicle accidents			
	Road type I	Road type II	Road type III	Road type IV	Road type I	Road type II	Road type III	Road type IV
Constant	−21.04 (−29.69) <sup>b</sup>	−20.31 (−39.54) <sup>b</sup>	−19.15 (−47.61) <sup>b</sup>	−19.04 (−35.17) <sup>b</sup>	−25.84 (−21.11) <sup>b</sup>	−23.41 (−27.27) <sup>b</sup>	−23.67 (−29.67) <sup>b</sup>	−23.14 (−22.04) <sup>b</sup>
ln( <i>q</i> )	0.52 (3.85) <sup>b</sup>	0.28 (2.25) <sup>b</sup>	0.16 (2.00) <sup>b</sup>	0.23 (2.63) <sup>b</sup>	1.35 (7.02) <sup>b</sup>	0.88 (4.88) <sup>b</sup>	0.93 (6.73) <sup>b</sup>	0.78 (4.88) <sup>b</sup>
ln( <i>H</i> )	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
d.f.	95	35	128	118	95	35	128	118
Scaled deviation	49.73	19.96	69.92	83.78	67.75	21.55	88.08	111.83
LM-stat	3.48	5.44	5.71	4.33	11.97 <sup>c</sup>	5.54	3.32	8.03 <sup>c</sup>
<i>t</i> -Statistics 1	1.10	−1.86	−1.30	−0.07	0.02	−1.58	0.89	−1.18
<i>t</i> -Statistics 2	1.95	−1.45	−1.88	−0.30	0.46	−2.09 <sup>c</sup>	−0.22	−0.53

<sup>a</sup> Values in parenthesis are *t*-statistics.

<sup>b</sup> Significantly different from zero at the 5% level.

<sup>c</sup> The null-hypothesis of over dispersion is significantly rejected at the 5% level.



Table 6

Expected number of accidents occurring all day and in daylight, inhomogeneous traffic flow<sup>a</sup>

	Expected number of accidents				Expected number of accidents in daylight			
	Road type I	Road type II	Road type III	Road type IV	Road type I	Road type II	Road type III	Road type IV
Constant	−31.54 (−16.42) <sup>b</sup>	−23.48 (−33.03) <sup>b</sup>	−21.16 (−50.56) <sup>b</sup>	−22.84 (−35.38) <sup>b</sup>	−31.14 (−15.65) <sup>b</sup>	−24.18 (−23.84) <sup>b</sup>	−21.42 (−29.97) <sup>b</sup>	−19.52 (−21.60) <sup>b</sup>
$\ln(q_c)$	3.62 (7.10) <sup>b</sup>	1.43 (7.28) <sup>b</sup>	1.04 (9.96) <sup>b</sup>	1.90 (13.04) <sup>b</sup>	3.60 (6.95) <sup>b</sup>	1.68 (6.18) <sup>b</sup>	1.16 (7.36) <sup>b</sup>	1.30 (7.02) <sup>b</sup>
$\ln(q_l)$	−2.66 (−5.40) <sup>b</sup>	−0.94 (−5.69) <sup>b</sup>	−0.77 (−9.07) <sup>b</sup>	−1.84 (−14.44) <sup>b</sup>	−2.66 (−5.41) <sup>b</sup>	−1.26 (−5.40) <sup>b</sup>	−0.92 (−7.85) <sup>b</sup>	−1.61 (−11.39) <sup>b</sup>
$\ln(H)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
d.f.	404	155	591	454	396	154	571	454
Scaled deviation	56.31	35.44	160.23	153.97	55.49	26.54	113.88	178.19
LM-stat	61.97 <sup>c</sup>	17.52 <sup>c</sup>	12.55 <sup>c</sup>	14.40 <sup>c</sup>	54.67 <sup>c</sup>	7.46	5.54	19.70 <sup>c</sup>
<i>t</i> -Statistics 1	0.60	−1.21	1.46	−0.39	0.74	−2.11 <sup>c</sup>	−2.11 <sup>c</sup>	1.25
<i>t</i> -Statistics 2	3.69 <sup>c</sup>	0.03	0.94	0.63	4.28 <sup>c</sup>	−3.26 <sup>c</sup>	−2.43 <sup>c</sup>	1.79

<sup>a</sup> Values in parenthesis are *t*-statistics.<sup>b</sup> Significantly different from zero at the 5% level.<sup>c</sup> The null-hypothesis of over dispersion is significantly rejected at the 5% level.

Table 7

Expected number of single and multi-vehicle accidents, inhomogeneous traffic flow<sup>a</sup>

	Expected number of single-vehicle accidents				Expected number of multi-vehicle accidents			
	Road type I	Road type II	Road type III	Road type IV	Road type I	Road type II	Road type III	Road type IV
Constant	−29.00 (−12.57) <sup>b</sup>	−22.85 (−27.23) <sup>b</sup>	−20.36 (−42.99) <sup>b</sup>	−22.08 (−30.04) <sup>b</sup>	−37.41 (−10.80) <sup>b</sup>	−26.58 (−20.18) <sup>b</sup>	−24.99 (−29.02) <sup>b</sup>	−26.89 (−20.47) <sup>b</sup>
$\ln(q_c)$	2.90 (4.65) <sup>b</sup>	1.14 (4.82) <sup>b</sup>	0.76 (6.10) <sup>b</sup>	1.70 (9.92) <sup>b</sup>	4.69 (5.23) <sup>b</sup>	1.96 (5.57) <sup>b</sup>	1.61 (8.38) <sup>b</sup>	2.40 (8.64) <sup>b</sup>
$\ln(q_l)$	−2.29 (−3.83) <sup>b</sup>	−0.85 (−4.34) <sup>b</sup>	−0.73 (−7.09) <sup>b</sup>	−1.82 (−11.91) <sup>b</sup>	−3.15 (−3.67) <sup>b</sup>	−1.11 (−3.68) <sup>b</sup>	−0.83 (−5.69) <sup>b</sup>	−1.90 (−8.23) <sup>b</sup>
$\ln(H)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
d.f.	404	155	591	454	404	155	591	454
Scaled deviation	59.34	47.30	228.55	174.27	78.34	49.39	192.30	190.31
LM-stat	13.54 <sup>c</sup>	13.95 <sup>c</sup>	8.93	11.15 <sup>c</sup>	9.94 <sup>c</sup>	7.19	13.67 <sup>c</sup>	56.60 <sup>c</sup>
<i>t</i> -Statistics 1	0.02	−0.03	0.03	−0.04	−0.01	−0.05	0.04	0.02
<i>t</i> -Statistics 2	0.30 <sup>c</sup>	0.02	0.20 <sup>c</sup>	−0.02	0.12	−0.22 <sup>c</sup>	−0.09	0.08

<sup>a</sup> Values in parenthesis are *t*-statistics.<sup>b</sup> Significantly different from zero at the 5% level.<sup>c</sup> The null-hypothesis of over dispersion is significantly rejected at the 5% level.







for single-vehicle accidents, the estimated exponents of the traffic flow are significantly less than one for all road types.

### 5.2. Inhomogeneous traffic flow

Table 6 shows the results for all accidents and accidents occurring in daylight only when separating the traffic flow for cars and lorries. The estimated coefficients are significantly different from zero for all road types. For road type I and IV the exponents are significantly different from 1, indicating that the expected number of accidents increases more than proportionally with the number of cars per hour, i.e. that the accident rate increases. However, for the road types II and III the null-hypothesis cannot be rejected. The estimated exponent for the flow of lorries is generally negative and different from zero. This suggests that an increasing number of lorries per hour will lower the expected number of accidents independently of the flow of cars. When studying accidents in daylight, the results are not significantly different.

The estimates when studying single and multi-vehicle accidents are presented in Table 7. The value of the exponent for the number of cars per hour is generally different from 1 when both single and multi-vehicle accidents are studied. Accordingly, the expected number of accidents increases more than proportionally for both single and multi-vehicle accidents. However, the exponent for cars is significantly lower for single-vehicle accidents than for multi-vehicle accidents. The exponent for the number of lorries per hour is also negative here. The value of the exponent is generally, however not significantly, lower when multi-vehicle accidents are studied compared to when all accidents are studied together.

The difference between the studied road types is also tested. The results for the road types I and IV are significantly higher and lower than the results for the road type II and III, respectively. This corresponds to the results when all accidents are studied together.

The result using an inhomogeneous traffic flow differs greatly to that of homogeneous traffic flow. Since the estimates raise certain questions the data is also studied visually. Figs. 1–4 present the number of accidents per hour and kilometre in relation to the flow of lorries per hour. Figs. 5–8 present the number of accidents per hour and kilometre in relation to the flow of cars per hour. The figures are drawn for the average number of cars (Figs. 1–4) and lorries (Figs. 5–8) per hour for each road type  $\pm 20\%$ .<sup>3</sup> Generally, there are few accidents in the data set and when restricting the analysis to certain traffic flows, the number of accidents is reduced even further. In spite of a limited number of observations, the plots can be argued to indicate the type of relation between the accident frequency and the flow of cars and lorries.

Data presented in Figs. 1–4 indicates that the number of accidents is constant or decreasing with an increasing

number of lorries on the road, whereas in Figs. 5–8 the number of accidents is increasing with an increasing number of cars. Consequently, the visual analysis is not in conflict with the estimates received in the regression analysis.

## 6. Discussion

The estimated relationship between the expected number of accidents per hour and kilometre, and the traffic flow differs considerably depending on whether different types of traffic modes are considered or not, i.e. if a homogeneous or a inhomogeneous traffic flow analysis is carried out. Differences appear both in relation to accidents occurring throughout the day, accidents occurring in daylight, single- and multi-vehicle accidents. In the homogeneous traffic flow analysis regarding accidents that have occurred throughout the day, we reject the null-hypothesis that the expected number of accidents increases proportionally with the traffic flow for the road types II, III and IV. The value of the exponent is less than 1. This result is in line with a number of studies reviewed in Satterthwaite (1981), Ardekani et al. (1997) and Hauer and Bamfo (1997).

As cars constitute the main part of the traffic flow, one may expect the outcome, with respect to the number of cars per hour, to be similar to that of the homogeneous traffic flow analysis. However, here the suggestion is that the expected number of accidents increases in proportion or more, to the number of cars per hour. Furthermore, studying the effect of an increasing number of lorries one might expect that the presence of more lorries on the road would increase the number of accidents since the incidence of possibly dangerous overtaking manoeuvres increases. The result of this study indicates, however, that at a given number of cars per hour, the expected number of accidents will decrease with increasing number of lorries per hour. This safety effect of lorries may be regarded as counter intuitive at first glance. It is, however, possible that the number of accidents decreases as an effect of a speed reduction as the speed limit for lorries is lower than for cars. An increasing number of lorries per hour will accordingly slow down the average speed. The result with respect to the flow of lorries may also be a result of people's unease when sharing the road-space with a lorry causing the attention to increase. There is, furthermore, a possibility that the number of lorries per hour is correlated with an unknown factor that lowers the number of accidents. Hours with many lorries may coincide with hours with good road conditions and hours when experienced drivers are driving.

It is obvious that there are shortcomings in analysis indicating that some of the results in this paper can be questioned. For instance, the model for inhomogeneous traffic flow analysis may be considered as a better model in relation to differences between traffic modes. As this model can be assumed to have more explanatory power, no over dispersion is expected in the Poisson regression model. However, in the inhomogeneous traffic flow model, over dispersion becomes

<sup>3</sup> The same pattern is given irrespective of the flow of cars and lorries used.



evident, while in the homogeneous traffic flow model, when the traffic flow is treated without consideration to different types of traffic modes, there is generally no over dispersion.

In addition, the accident sample is rather small even though data on accidents for 6.5 years is used. According to Table 3, the average numbers of accidents per traffic flow interval is generally less than 1, indicating low statistical power in especially the analysis of inhomogeneous traffic flows. Random fluctuations may then be a source of error. The maximum hourly flow of lorries is furthermore a fraction of that of cars which may lead to unstable results.<sup>4</sup>

Consequently, the marginal effect of the flow of different vehicle types may have been miscalculated and exaggerated in the regression analysis due to shortcomings in the data set and in the regression models. However, visual analysis of the inhomogeneous traffic flow data, presented in Figs. 1–8, indicates that the flow of lorries affects the accident frequency differently than the flow of cars. The result implies that important information is lost if no consideration is taken to vehicle type. Literature of today shows limited interest in studying traffic safety models distinguishing between different vehicle types, though. This is a shortcoming since the characteristics of our vehicles and their driving pattern differ in a number of respects, undoubtedly affecting the accident frequency in various ways. Other model structures ought to be applied in future research including, for instance, speed as an explanatory variable.

## Acknowledgements

Financial support from the Swedish National Road Administration is gratefully acknowledged. I am grateful for insightful suggestions from Risto Kulmala, and Karin Brundell-Freij. Any errors are my sole responsibility.

## Appendix A

Traffic flow data from the stationary counters are applied to calculate traffic flow frequencies for each road section, i.e. set of hours of different traffic volumes. An interval level of 25 vehicles per hour is used when treating the traffic flow as homogeneous. When treating the traffic flow as

inhomogeneous, an interval level of 25 cars per hour is used together with an interval level of five lorries per hour. The traffic flow frequency for interval  $t$  is denoted by  $\tau_t$ . For each road type the number of kilometres,  $\eta$ , is calculated for each section, i.e.  $\eta_r$  kilometres for road type  $r = I, II, III$  and  $IV$ . The number of kilometres is then multiplied by the traffic flow frequencies for the studied road section. The analysis is based on four road types and each road type is represented on several road sections. If  $s$  is the number of sections of road type  $r$ , the exposure,  $H$ , is calculated as the total number of hours and kilometres studied.

$$H_{r,t} = \sum_{s=1}^n \eta_{r,s} \times \tau_{t,s}$$

for  $t = 1, \dots, p$  and  $r = I, \dots, IV$

The traffic flow variables are calculated as the average number of vehicles,  $q$ , and the average number of cars and lorries per hour,  $q_c$  and  $q_l$ , for each traffic flow interval.

The number of accidents throughout the day and in daylight is calculated by aggregation using the same traffic flow interval as when calculating the traffic flow frequencies. These variables are used as dependent variables. When calculating the exposure for daytime traffic, frequencies of traffic flows observed between 6 a.m. and 6 p.m. are used.

## References

- Ardekani, S., Hauer, E., Jamei, B., 1997. In: Gartner, N. et al. (Eds.), *Traffic Impact Models in Traffic Flow Theory—A State-of-the-Art Report*, Oak Ridge National Laboratory.
- Hall, R.D., 1986. *Accidents at Four-arm Single Carriageway Urban Traffic Signals*. Transport and Road Research Laboratory Contractor Report, 65.
- Hauer, E., Bamfo, J., 1997. In: *Proceedings of the ICTCT conference on Two Tools for Finding what Function Links the Dependent Variable to the Explanatory Variables*, Lund.
- Jovanis, P.P., Chang, H.L., 1986. *Modelling the relationship of accidents to miles travelled*. Transportation Res. Rec. 1068, 42–51.
- Kulmala, R., 1995. *Safety at Rural Three- and Four-arm Junctions*, VTT-Publications, 233.
- Nicholson, A., Wong, Y.D., 1993. *Are accidents Poisson distributed? a statistical test*. *Accid. Anal. Prev.* 1, 91–97.
- Satterthwaite, P.S., 1981. *A Survey of Research into Relationship Between Traffic Accidents and Traffic Volumes*, Transport and Road Research Laboratory Contractor Report, 692.
- Vickery, W.S., 1968. *Automobile Accidents, Tort Law, Externalities and Insurance*. *Journal of Law and Contemporary problems* 33, 464–484.
- Vitaliano, D.F., Held, J., 1991. *Road Accident External Effects: an Empirical Assessment*. *Applied Economics* 23, 373–378.
- Winslott, L., 1998. *The external costs of traffic accidents: an empirical analysis of the traffic flow*. Department of Economics Working Paper Series, 1.

<sup>4</sup> In this type of data one could also expect a correlation between the flow of cars and lorries, leading to spurious effects in the regression analysis. No correlation can be detected, however. Even though high flows of lorries are more frequently found at high flows of cars there are also hours in the data set with high flows of lorries and low flows of cars.