A MULTIFACETED RELATIONSHIP BETWEEN CONGESTION AND ACCIDENTS ON HIGHWAYS: EVIDENCE FROM D-100 HIGHWAY IN ISTANBUL

by

Güner Velioğlu

Submitted to the Social Sciences Institute
in partial fulfillment of the requirements for the degree of
Master of Arts

Sabancı University
Spring 2013-2014

A MULTIFACETED RELATIONSHIP BETWEEN CONGESTION AND ACCIDENTS ON HIGHWAYS: EVIDENCE FROM D-100 HIGHWAY IN ISTANBUL

APPROVED BY:

Eren İnci ....
(Thesis Supervisor)

İnci Gümüş ....

Şerif Aziz Şimşir ....

DATE OF APPROVAL: ....

Acknowledgements

First of all, I am grateful to my thesis advisor Eren İnci for his guidance throughout my M.A. degree. He has helped me in my M.A. thesis with his insightful suggestions, motivated me when I faced challenges in solving problems and executing my ideas, believed in me from the beginning and trained me for my doctoral studies in various respects. I also thank my thesis committee members, İnci Gümüş and Şerif Aziz Şimşir for examining my thesis.

I would like to thank to Traffic Control Center, Istanbul for granting me access to the traffic characteristics and accidents data. I am particularly thankful to Mehmet Mert and Batuhan Altun for their help in supplying process and explanation of the data.

Lastly, I would like to thank my family for their unending support throughout my life.

A MULTIFACETED RELATIONSHIP BETWEEN CONGESTION AND ACCIDENTS ON HIGHWAYS: EVIDENCE FROM D-100 HIGHWAY IN ISTANBUL

Güner Velioğlu

Economics, M.A. Thesis, 2014

Supervisor: Eren İnci

Keywords: Accidents; Property Damage Accidents; Safety; Traffic Congestion

Abstract

Using a comprehensive traffic flow data for D-100 Highway in Istanbul, this thesis empirically analyzes the effects of congestion on highway accident frequency. Prior studies looked at this relationship in the context of injury and fatal accidents and found accident likelihood to be decreasing with congestion. We concentrate on property damage accidents and found that their likelihood increases with congestion. Moreover, in our data, the relationship is nonlinear: congestion increases accident likelihood under hypercongested traffic conditions but it has no significant effect under (normally) congested traffic conditions.

TRAFİK SIKIŞIKLIĞI VE KAZALAR ARASINDAKİ ÇOK TARAFLI İLİŞKİ: İSTANBUL D-100 KARAYOLU'NDAN BULGULAR

Güner Velioğlu

Ekonomi Yüksek Lisans Tezi, 2014

Tez Danışmanı: Eren İnci

Anahtar Kelimeler: Güvenlik; Kazalar; Maddi Hasarlı Kazalar; Trafik Sıkışıklığı

Özet

Bu tez, İstanbul D-100 karayoluna ait kapsamlı trafik karakteristik verisini kullanarak, kaza frekansı ve trafik sıkışıklığı arasındaki ilişkiyi incelemektedir. Geçmiş çalışmalarda bu ilişki sadece ölümlü ve yaralanmalı kazalar bazında incelenmiştir ve ağır kaza ihtimalinin trafik sıkışıklığı ile ters orantılı olduğu bulunmuştur. Bu çalışmada maddi hasarlı kazalar da incelenmiş olup, maddi hasarlı kaza oluşma ihtimalinin trafik sıkışıklığı ile doğru orantılı olduğu bulunmuştur. Bununla birlikte bu ilişkinin doğrusal olmadığı görülmüş olup, çok sıkışık trafikte maddi hasarlı kaza ihtimali trafik sıkışıklığı ile birlikte artar iken, az yoğun trafikte böyle bir ilişki olmadığı saptanmıştır.

Table of Contents

1. Introduction

Every year, countries suffer enormous losses from traffic accidents. These losses are ever increasing with population increase and addition of new drivers to traffic. Miller (2000) finds that the annual cost of traffic accidents is approximately $300 billion in the US. The situation is no different for Turkey. In 2012, 3,750 people were killed, and 268,079 were injured in traffic accidents (TUIK, 2013). Considering the magnitude of losses resulting from traffic accidents, finding methods or policies for reducing their likelihood was the motivation of many studies.

This thesis contributes to the discussions by providing an empirical analysis of the impact of congestion on accident likelihood on segments of D-100 highway in Istanbul. Previous work on this topic concentrates on injury and fatal accidents and finds a negative impact of congestion on accident likelihood. We concentrate on property damage accidents (along with injury and fatal accidents) and find that congestion has a positive impact on those accidents (at least in our data). We also analyze this relationship for the hypercongested and (normally) congested traffic flow periods separately and find a non-linear relationship between accident likelihood and congestion. In particular, likelihood of property damage accidents is increasing with congestion when there is hypercongestion but it there is no significant relationship when the road is normally congested. However, for injury and fatal accidents, the impact is always insignificant in our data.

Before discussing the effect of congestion on road accidents we have to be clear about what we mean by hypercongestion. In the theory of transportation economics, there are two speeds associated with any given level of traffic flow. Since different speeds imply different congestion levels, the high-speed region of traffic flow is termed as congested (or *normally congested* in engineering terminology) while the low speed part is termed as *hypercongested*. We are primarily interested in the *hypercongested* (*congested* in engineering terminology) region, since it is the instance which we encounter frequently, and consider problematic. The significance of such distinction lies in the fact that, in these two congestion types, traffic flow exhibits different characteristics. In the case of normal congestion, road is less occupied by definition and an increase in the demand (*i.e.*, introduction of new vehicles to traffic), does not substantially slow down traffic. Thus flow is increasing. In the case of hypercongestion, the road is occupied more and an addition of new vehicles will slow down the traffic, which will often result in stop-and-go traffic. In congested traffic conditions, the primary risk for drivers is the high speed; in hypercongested traffic conditions, however, the primary concern is the number of surrounding vehicles (*i.e.*, increased interaction).

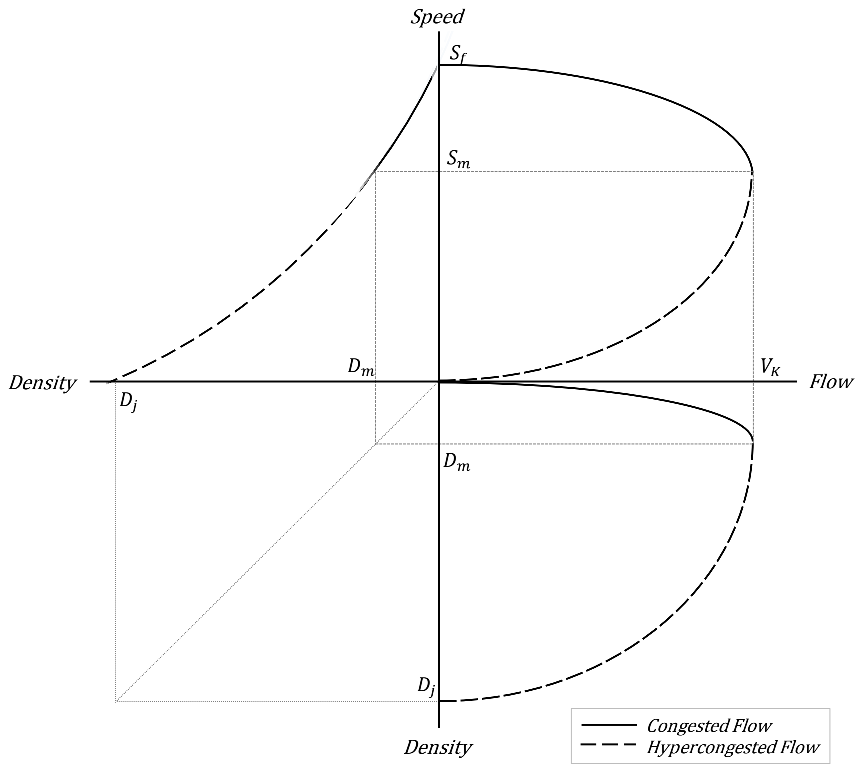

Figure 1. Fundamental Diagram of Traffic Flow[1]

Figure 1 depicts the relationship between traffic variables: Speed, flow and density. It is the result of identity $V \equiv SD$ (For overviews see Small and Verhoef, 2007). For example, first quadrant shows the speed-flow relationship. When the road is free, addition of new vehicles will increase the flow. But after a certain point, road becomes so dense with the addition of vehicles that a new vehicle will considerably slow down the traffic and flow will be decreased.

Shefer and Rietveld (1997) suggest that there is a negative relation between congestion and fatal accidents. This is partially confirmed by the findings of Wang *et al.* (2009) in their analysis of London M25 motorway. While it is intuitive to expect a decrease in the rate of accidents when traffic slows down, this may only hold for the injury accidents. Previous empirical studies do not work on property damage accidents. Even though these

[1] Source: Small and Verhoef (2007). This figure is our elaboration.

accidents are less costly than injury or fatal accidents, their occurrence rate is much higher (about 13 times more common in our data). Thus, their analysis is significantly necessary.

In fact, we find that the congestion-accident occurrence relationship is different in the case of property damage accidents. The intuition is as follows. As congestion increases so does the density of traffic. Average traffic speed would be lower but as the vehicles get closer to one another, the time window for a safe response will be narrow. It may get difficult for a driver to react accurately and on time. Consequently, an increased level of congestion after a certain point might increase the probability of accidents. As Vickrey (1969) noted: “While the incidence of traffic accidents does not arise with traffic density quite as rapidly as do time delays, one does expect, *a priori*, that as vehicle interactions per vehicle-mile increase, accidents per vehicle-mile will also increase.”

We concentrate on a segment of the D-100 Highway in Istanbul. Both the traffic characteristics data and accident data were obtained from the Traffic Control Center, which is a subsidiary of Metropolitan Municipality of Istanbul. By using this data we generated a congestion measure in line with the one introduced by Taylor *et al.* (2000). Considering the accident count variable dependent on congestion and several other control variables, we estimate the frequency of accidents for each region and time interval. Since accident frequency is a count variable, we use count data models which accounts for its distribution.

The non-linear dependence of accident likelihood on congestion implies an optimal level of congestion for traffic safety. It at least means that slower traffic is not necesarily the safer one. Our findings should have important implications on the determination of speed limits and congestion charges.

The rest of this thesis is as follows: Section 2 summarizes the current literature. Section 3 describes the data. Section 4 explains the methods we used in matching accidents and measuring congestion in addition to estimation model we used. Section 5 reports the results and Section 6 concludes.

2. Literature Review

There is a large body of literature analyzing accidents by exploring various contributing risk factors. However, the literature on the relationship between congestion and accidents, especially the empirical work, is relatively recent. This is partly due to the fact that congestion is a complex phenomenon itself and is difficult to measure. Consequently, many studies use "proxies" for congestion.

Ceder and Livneh (1982) investigate the effect of flow on accidents by categorizing them into single and multi vehicle accidents. Ceder (1982) analyses this relationship for different types of traffic flow, and finds a U-shaped relation for free-flow traffic. In particular, he finds that the accident frequency increases faster for congested-flow traffic, which is defined to be the periods during which 95% of total accidents are multi-vehicle accidents and 85% of all accidents resulted from rear-end collisions. Although the criterion presented by Ceder (1982) is reasonable for the determination of different traffic flow modes, it does not enumerate the degree of congestion. This study also demonstrates the importance of investigating the accidents under different flow conditions.

Traffic density is another factor which has been investigated and has a close relation with traffic congestion. Zhou and Sisiopiku (1997) find that there is a U-shaped relationship between volume over capacity ratio (*i.e.*, density). When the road is empty, addition of new vehicles reduces the accident risk; but after a certain point, increasing the flow increases the accident risk. Shefer (1994) also uses the volume over capacity ratio as a proxy for congestion. He considers only fatal accidents and hypothesizes that their frequency decreases after a certain level of traffic congestion, or density. The relationship he suggests is shown in Figure 2. In this figure, the number of road fatalities are plotted against the density of cars on the road. As the density increases, number of road fatalities increase with an increasing rate. Then, a stage is reached where number of fatalities increases with a decreasing rate. However, this relationship is reversed after some critical density level where fatalities drop by density.

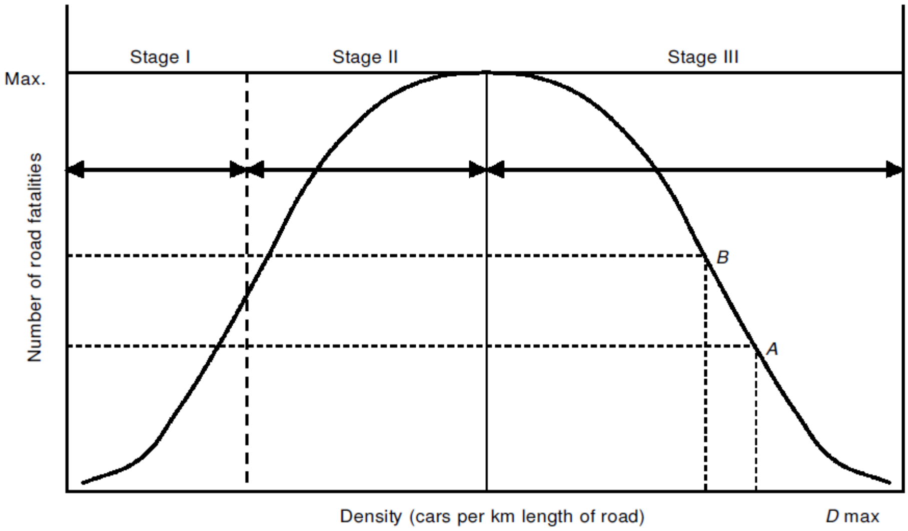

Figure 2. Hypothetical road fatalities rate - density function.

In their theoretical work, Shefer and Rietveld (1997) propose a model where accidents are considered to be a function of both speed and density. This makes sense since speeding on an occupied road is much riskier than speeding on a free one. Drivers slow down in an occupied road to reduce accident risk. Consequently, as Verhoef and Rouwendal (2004) argue, the speed-flow curve presented in the first quadrant of Figure 1 may in fact result from individuals' cost minimizing speed choices, instead of simply being a technical relationship. Shefer and Rietveld (1997) provide some empirical evidence based on a simulated dataset that fatality rates are lower at the peak hours. There are two important concerns here. First, density alone does not really measure congestion characteristics accurately. Although it is related to congestion the relationship is by no means monotonic. Second, one may wonder if these results hold with real data.

Other researchers suggest that variance of speed contributes to accident occurrence more than average speed level does. For example, Lave (1985) finds that there is a significant and positive relationship between fatalities and speed variance while the relationship between fatalities and accidents is not significant, which he summarizes as "variance kills, not speed." Traffic engineers also support this finding. Solomon (1964) investigates the relationship between accident rates and variance and obtains the curve shown in Figure 3. This figure shows that accident involvement rate increases as a driver starts to deviate from the median speed. Even though the driver slows down, if he gets away from median speed the accident risk will increase. Hauer (1971) provides a theoretical foundation for this curve by investigating the number of overtakings at different speed levels.

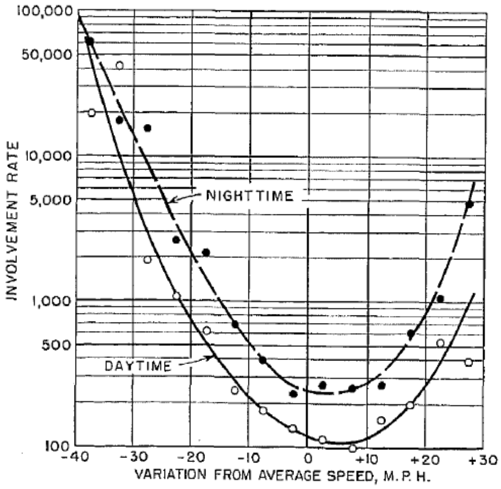

Figure 3. Involvement rate vs. deviation from average speed.

Noland and Quddus (2005) use several types of proxies for congestion in their analysis of various regions of London. They use an indicator variable to separate Inner London and Outer London into two distinct regions which have different congestion levels. Using Negative Binomial model for estimation, they find that there is not any statistically significant relationship between congestion and fatal, serious and slight injury accidents. They also use employment number for different districts of London, but this time they find a significant and positive relationship for all accident types. After controlling employment with district area, however, they again find that there is not any significant relationship with any of the accident types.

Another notable exception in the literature on traffic congestion is the empirical study of London M25 motorway conducted by Wang *et al.* (2009), where the authors employ the congestion index proposed by Taylor *et al.* (2000), which we also use. They develop a method to map the accidents to correct segments of the road, for each of which they have the data for traffic characteristics. Subsequently, they estimate the frequency of accidents using congestion, flow and several other explanatory variables. They also categorize the accidents according to their severities to accurately determine the effect of congestion on different accident types. They find that congestion has no statistically significant effect on accident frequency. They, in fact, find a negative (but insignificant) relationship in the case of serious accidents. This finding, although “insignificant”, is consistent with the hypothesis proposed by Shefer (1994) in that the fatalities decrease while the congestion level increases.

3. Data

Our data is consisted of traffic characteristics and accident occurrences, both collected by the Highway Engineers of the Traffic Control Center in Istanbul, for the years 2012 and 2013. It is a time series of traffic variables – speed, flow and occupancy –, and a binary variable representing the accident occurrence. We obtained the data for a 10 km segment of the D-100 highway in the European side of Istanbul. In this road segment, there are 19 Remote Traffic Microwave Sensors (RTMS) that record the variables at every two minutes. This constructs a panel with approximately 525.000 time periods/observations for each of the 19 sensors.

D-100 highway has six lanes, three on each direction. The microwave sensors have a detection range of eight lanes. Based on test results, this sensor type can provide 90% accuracy in counting traffic flow (Yu and Prevedouros, 2013). The common failures are due to the possibilities where vehicles in different lanes are registered as one because of their proximity, or one of them blocking the view and the other one passes undetected. There is also a slight degradation of performance under congested traffic but there is no degradation under adverse weather conditions. Double counting, which occurs when the detection of a single vehicle is stopped and then resumed thus being recorded as two vehicles, can become an issue under hypercongested traffic conditions. For these reasons, to attain higher levels of accuracy, at every segment two sensors are placed on each side of the road facing each other. Each of them records the same section, whole six lanes, but only the nearest three lanes are kept.

4. Methodology

4.1 Mapping Accidents to Correct Sensors and Time Intervals

1994 accidents occurred on the D-100 highway in 2012 and 2013 (excluding junctions). 140 of them were injury and fatal accidents and 1854 were property damage accidents. In the case of property damage accidents, exact location is not recorded. These accidents are recorded by Traffic Control Center (TCC) in terms of road disturbances. When a minor accident happens, the RTMS assigned to that segment sends a notification, stating that the traffic flow is being disturbed. The TCC client observes the event scene via the cameras placed on each side of the road, based on the spatial relevance of the respective RTMS. The client then records the type of disturbance (Fatal accident, injury accident, chain accident, property damage accident, car breakdown, and road maintenance) stating that such type of event has been observed by that certain camera at that certain time. Consequently, when matching these accidents to correct segments, we assumed that the RTMS that is closest to that camera would provide us the most relevant readings. We matched these accidents assuming that there is a 20-minute recording lag to avoid any reverse-causality (*i.e.*, accident occurrence increasing congestion).

In the case of injury and fatal accidents, the police visits the accident area, and specifies the exact location with Google coordinates. Since these accidents are also recorded by TCC, we used the location data to check whether we are accurately matching accidents to correct sensors. In addition, we again assumed a 20-minute time lag, since it would take some time for the TCC observer to record the events. We tested whether the 20-minute time lag allowance is sufficiently long by examining a sample of 100 accidents on a case by case basis. While allowing for a longer lag certainly reduces any risk of reverse-causality, assigning a time interval much earlier than the actual would weaken the relationship.

Figure 4. D-100 highway and sensors.

4.2 Congestion Measurement

The measures used in the road accidents literature have been generally inadequte in terms of representing the true nature of congestion. To be sure, these simplified representations have been employed as a result of the limitations in the data (*i.e.*, the data was lacking for realistic/complex models, even if they were to be developed). Therefore, since we would not be able to elaborate on a sophisticated model, we tried to think of a suitable model in line with the available data and that would represent the characteristics of congestion to greatest extent. The main characteristic of congestion is that it decreases the quality of travel by increasing travel time. Therefore, we decided to use on the Congestion Index (CI) introduced by Taylor *et al.* (2000):

$$CI = \frac{T_A - T_0}{T_0}, \tag{1}$$

where T_A is the actual travel time and T_0 is the travel time when the road is completely empty. Since travel time is directly related with travel speed we can rewrite this:

$$CI = \frac{{}^{1}/_{S_A} - {}^{1}/_{S_0}}{{}^{1}/_{S_0}}, \tag{2}$$

where S_A is the actual travel speed and S_0 is the travel speed when the road is completely empty. The choice of free-flow speed S_0 is not an important concern, since it does not play a role when determining the impact of *relative* levels of congestion on accidents. In any given time interval then, all we need to know to calculate the congestion index is the average travel speed of the vehicles. In accordance with this and the availability of speed data, we generated Congestion Index data for every observation in our panel. Simple as it may sound, we believe that the inverse of average travel speed, in small road segments at 2 minutes intervals, would be a *proportionate* proxy for average level of traffic congestion. Certainly, a crucial assumption here is that the road is not free when vehicles are travelling at low speeds.

4.3 Hypercongestion and Its Identification

In our analysis, we splitted the data into two distinct regions of congestion, since we presumed that congestion-accident relationship might change under different congestion modes. As demonstrated in Figure 1, after reaching its capacity flow, the traffic enters into the hypercongested mode with the addition of new vehicles. Therefore, this capacity level is the key in determining the point where traffic starts to change its mode. Daganzo and Geroliminis (2008) show that, by examining the data available from Yokohama, the peak flow point where congestion mode changes does in fact exist. Similarly, we determine whether this also applies to our data and if so which level of speed is the turning point. Figure 5 depicts a representative sample data taken from one of our sensors.[2] We find that 76 kph is the average speed level at which the traffic enters hypercongested mode. Thus, 76 kph is the basis of our data separation into two congestion modes.

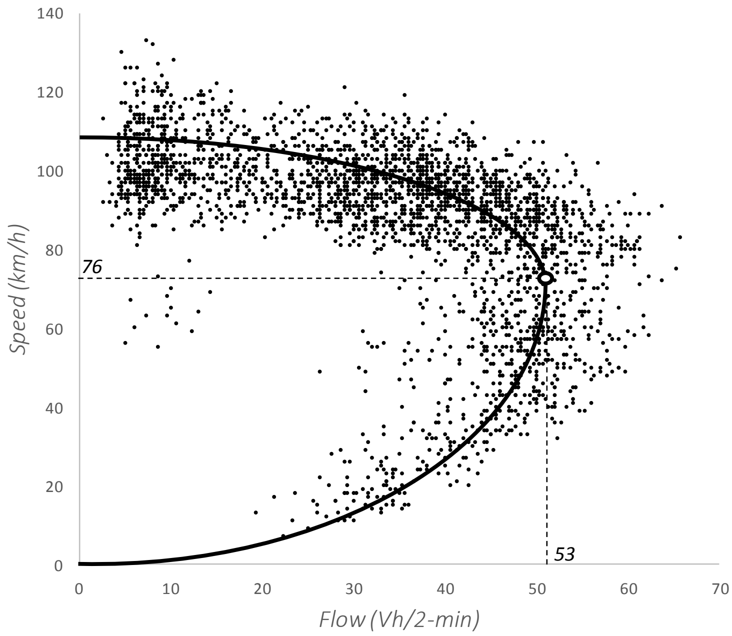

Figure 5. Flow-Speed scatterplot on a sample of our data.

4.4 Estimation Model

When deciding which model would most reliably explain the relationship, we primarily considered the characteristics of our dependent variable, accidents. Frequency of accidents is a discrete count variable and is greater than or equal to zero. Since it is a count variable, it would be expected to follow some variant of Poisson distribution. This is the main reason why many studies employed Poisson-regression models for a long

[2] While the dispersion of data points change between sensors, the curve remains the same. The curve contracts under severe weather conditions, but those constitute a very little fraction of our observations.

time. We can specify a Poisson model for the accident count variable which follows a Poisson distribution as follows:

$$a_i \sim \text{Poisson}(\mu_i),$$

with a probability distribution function *f* such that:

$$f(a_i) = \frac{e^{-\mu_i}\mu_i}{a_i!}, \tag{3}$$

where mean and variance of a_i are equal to Poisson parameter:

$$E[a_i] = Var[a_i] = \mu_i. \tag{4}$$

One crucial constraint of the Poisson model is that, the variance and the mean of the dependent variable is assumed be equal, which is violated for the case of accident counts. Miaou (1994) finds accident count data to be significantly overdispersed (its variance is larger than its mean). To deal with this overdispersion characteristic of accidents and similar count data, the Negative Binomial model was introduced. In order to derive Negative Binomial model, we incorporate a gamma distributed error term to Poisson mean:

$$e^{\varepsilon} \sim Gamma(1, (1/\eta)^2) \tag{5}$$

$$\lambda_i = e^{\beta X_i + \varepsilon_i} \tag{6}$$

$$\ln(\lambda_i) = \beta X_i + \varepsilon_i, \tag{7}$$

where λ_i is the expected number of accidents, β is the coefficients vector that we will estimate, and X_i represents the explanatory variables which have been recorded by sensor *i*. Also known as Poisson-Gamma model, this model allows for overdispersion with parameter ${}^{1}/_{\eta}$ such that:[3]

$$Var[a_i] = E[a_i](1 + \frac{1}{\eta}E[a_i]), \tag{8}$$

where $\eta \geq 0$. While maintaining the skewness characteristics and the mean of Poisson distribution, this model allows for higher variation in data. As $\eta \to 0$, dispersion increases and using Negative Binomial model will be appropriate. As $\eta \to \infty$, this model will be

[3] Derivation of this model is presented in appendix.

equivalent to Poisson model. For this reason, η can be interpreted as an overdispersion parameter.

Table 1 summarizes the accident count variable in our data, which shows that its variance is much larger than its mean (this summary is after the aggregation of data and all accidents are included here). In the percentiles column, we observe that at least one quartile of observations has zero accidents, while the maximum number of accidents observed can be as high as 42. This is in line with the positive skewness (2.69). Considering these facts, our data also confirms that accident count follows a Negative Binomial distribution. This justifies the use of the Negative Binomial model.

Accidents				
	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	456
25%	0	0	Sum of Wgt.	456
50%	2		Mean	**4.25**
		Largest	Std. Dev.	6.51
75%	6	39		
90%	12	40	Variance	**42.32**
95%	17	42	Skewness	2.69
99%	33	42	Kurtosis	12.70

Table 1. Summary of accidents.

For our present analysis, we aggregated the data in the following way: For every hour of the day, we average the Congestion Index and assign the accidents to those congestion levels based on their occurrence time (Also keeping in mind that the real occurrence time v.s. the time that has been actually recorded, in order to avoid causality and biased findings). For each sensor then, we have the count of accidents in a specific time period and the congestion, speed, flow and occupancy levels associated with it.This aggregation method has been employed by the recent studies (for example, Wang *et al.* 2009), so as to prepare the data in accordance with the estimation methods to be used. One reason for this practice is, besides the ease of analysis and interpretation, the count data models are more reliable when the data is not zero-inflated.

5. Results

Out of 19 sensors, 1 is eliminated due to deficiencies in its readings (our purpose was to construct a balanced panel). During the analysis we also dropped some other sensors, because of all zero outcomes (accidents). The panel data we analyzed, then, has hourly aggregations of traffic variables from 16 to 17 different sensors for different accident types, respectively. Table 2 presents the summary of our explanatory variables. All of these variables are continuous, except Daylight, which is a binary variable. Congestion Index measures the time delay compared to free flow speed. Its between-sensor variance is lower than within-sensor variance, because the traffic flow of one sensor directly affects another's. This is the case for other variables as well. In a way, a low between-sensor variance indicates that the road segments they are watching have similar characteristics.

Table 2. Summary of variables (All sensors are included).

Variable		Mean	Std. Dev.	Min	Max	Observations
Congestion Index	overall	0.8597	0.6479	0.1898	3.3366	N = 456
	between		0.3365	0.3925	1.6757	n = 19
	within		0.5588	-0.1933	2.9841	T = 24
Flow	overall	32.2205	15.4532	3.4144	73.5422	N = 456
	between		7.6849	14.7279	54.3604	n = 19
	within		13.5177	-6.0454	51.4024	T = 24
Speed Variance	overall	4.0831	1.3782	0.4854	8.1998	N = 456
	between		0.8859	2.3713	5.9580	n = 19
	within		1.0744	1.5992	7.4642	T = 24
Daylight	overall	0.5000	0.5006	0.0000	1.0000	N = 456
	between		0.0000	0.5000	0.5000	n = 19
	within		0.5006	0.0000	1.0000	T = 24
Weather (L. Risk)	overall	0.0601	0.0295	0.0230	0.1429	N = 456
	between		0.0086	0.0505	0.0901	n = 19
	within		0.0282	0.0157	0.1176	T = 24
Weather (M. Risk)	overall	0.0753	0.0148	0.0502	0.1503	N = 456
	between		0.0092	0.0638	0.1083	n = 19
	within		0.0118	0.0419	0.1173	T = 24
Weather (H. Risk)	overall	0.0088	0.0028	0.0024	0.0194	N = 456
	between		0.0011	0.0063	0.0121	n = 19
	within		0.0026	0.0038	0.0160	T = 24

Weather variable represents 35 categories of weather conditions, which we categorized according to their severity (Low, medium and high risk). For example, when a snowstorm happens, a high risk weather event, the Weather (H) variable is assigned one. Even though these different risk weather dummies are binary variables themselves, when aggregated, they give us the probability of observing a certain type of weather event in a specific period of time.

Following is the correlation table which signals a collinearity problem:

Variable	Congestion	Flow	Speed Var.	Daylight	W (L)	W (M)	W (H)	VIF
Congestion	1.00							1.87
Flow	0.54	1.00						3.81
Speed Variance	0.52	0.62	1.00					1.86
Daylight	0.39	0.63	0.43	1.00				2.10
W (L)	-0.49	-0.79	-0.50	-0.70	1.00			4.14
W (M)	0.09	-0.26	-0.11	-0.31	0.33	1.00		1.45
W (H)	-0.27	-0.52	-0.29	-0.50	0.60	0.48	1.00	1.89

Table 3. Correlation matrix and VIFs

It is debatable whether the magnitude of multicollinearity here should enforce us to study the effect of each of these variables separately (It was not deemed necessary in previous studies). The findings from one variable are likely to reinforce the others, since traffic variables are very interrelated. But this might go other way around when they are estimated simultaneously. We present the VIF values on the right column. A high VIF value indicates that some variables which should have been significant may turn out to be insignificant. There is not any definite threshold, but a VIF less than 5 is not considered critical, and in some cases a VIF of 10 is also acceptable (Haan, 2002). Thus we decided that simultaneous analysis of these variables would not create a big problem. In any case, we studied them both together and separately, and the significance of our estimates did not change. Therefore, we only present our results when they are analysed together.

Furthermore, most of the variables that will be presented here are logged versions of the real ones. This does not change the sign or significance of their estimates, but gives us an idea about the relative magnitude of their contribution to accident occurrence.

5.1 Property Damage Accidents

A main feature of our analysis is the investigation of property damage, injury/fatal accidents separately. As previously mentioned, we split our data into two main parts with respect to their congestion levels (congested vs. hypercongested). We first present our results for property damage accidents. We then present the results for fatal/injury accidents and discuss their relevance. Finally, we provide some possible policy implications.

Table 4 – Estimates for property damage accidents.

Variables	Negative Binomal (FE) *All*		Negative Binomal (FE) *Hypercongested Region (<76 kph)*		Negative Binomal (FE) *Congested Region (>76 kph)*	
	Coefficient	S. D.	Coefficient	S. D.	Coefficient	S. D.
Congestion Index	0.4612**	0.2140	1.0554***	0.1802	-8.2880*	4.4884
Flow	1.7850***	0.2470	1.7661***	0.2607	0.5153	0.3981
Speed Variance	0.5787*	0.3063	0.5858*	0.3222	0.7136	0.6878
Daylight	0.1700*	0.1005	0.1714*	0.1044	0.1471	0.2296
Weather (L)	-0.0260	0.0321	-0.0202	0.0309	-0.0154	0.0325
Weather (M)	-0.0132	0.0266	-0.0607**	0.0245	-0.0227	0.0264
Weather (H)	-0.2354	0.1605	-0.2601**	0.1221	0.1867*	0.1033
Intercept	-4.9514***	0.9811	-5.1885***	0.9579	2.9456	2.6526
Log-likelihood	-617		-581		-259	

***Significantly different from zero (At the 99% interval)

**Significantly different from zero (At the 95% interval)

*Significantly different from zero (At the 90% interval)

Table 4 presents the results for property damage accidents. Findings for the hypercongested part are more representative of the relation between congestion and property damage accidents. This is because majority of property damage accidents occured when the average speed level of traffic was below 76 kph.[4] The results of the regression of Accidents dependent on Congestion Index and Traffic Flow suggests a positive relation for both variables. More importantly, their relation to accident counts turned out to be significant. This means, in the case of property damage accidents, the

[4] Results for different thresholds are presented in Appendix B.

number of accidents increases as congestion increases. This result was not anticipated by previous studies since property damage accidents were, simply, not analyzed. In addition, we also observe that splitting the data into two parts is worthwhile since it strenghtened the results. Although there is no change in the number of significant variables, there is a considerable increase in their significance level (This is the case for congestion index and speed variance variables, two key explanatory variables in our analysis).

This finding also indicates that more congestion does not necessarily mean higher safety. While increasing the level of congestion may prevent serious accidents as a result of the overall slowing down of the traffic (Shefer, 1994), it gives rise to less serious but more frequent property damage accidents. To support this claim, we generated another variable called Tailgating Hazard which represents the risk level of average following distance of vehicles:

$$Tailgating\ Hazard = \frac{Speed^2}{Following\ Distance} \equiv Constant * Speed^2 * \frac{Flow}{Speed} \quad (9)$$

where speed, flow and following distance variables are averages of the vehicles which passed a certain sensor/segment over 2 minutes. A common phenomenon in highly congested traffic is that, vehicles follow each other very closely. For instance, when driving at 100 kph, recommended following distance is approximately 100 meters. For 50 kph, it is 45 meters.[5] While drivers are usually more risk averse at higher speeds, they do not keep up with the safe following distance at low speed levels. Thus, particularly during peak hours, it is common to observe vehicles travelling faster than 50 kph but following each other at distances lower than 15 meters.

We found that this variable has a positive and significant effect on accident occurrence. For every 1% increase in tailgating hazard, accident frequency increases by 1.22%. This indicates that as vehicles start to violate their safe following distance for their respective speeds, the likelihood of accidents increases. Moreover, we find a positive and significant relation between occupancy and accident frequency. The accident number increases as the density of the traffic increases. This confirms the reasoning of Vickrey: If the interaction number increases, the likelihood of accidents will also increase.

---

[5] Theoretically, the safe following distance for 50 kph should be one fourth of the 100 kph speed ($E_k = \frac{1}{2}mv^2$). But other factors such as driver's reaction time remain relatively fixed over different levels of speed.

Speed variance has been, again, proved to be another significant contributor to property damage accidents. This variable in a way covaries with congestion since it will be harder for drivers to coordinate as the vehicle number in traffic increases. This finding tells us that speed laws should be carefully designed and should aim to increase coordination between drivers. If one sets the speed limit too low, this might result in a denser traffic which would make it harder to coordinate with other vehicles. In fact, in highways where drivers choose to travel with the highest speed (*i.e.*, speed limit), implementing a minimum speed law would be reasonable, since slow drivers contribute to accident occurrence just as the fast ones do.

Overall, we found that the slower traffic does not necessarily have to be the safer one. And ignoring this fact may have costly implications. For example, setting the speed limit low would be counterintuitive for increased traffic safety (In addition to time costs).

5.2 Fatal/Injury Accidents

Table 5 – Estimates for fatal and injury accidents.

Variables	Negative Binomal (FE) *All*		Negative Binomal (FE) *Hypercongested Region (<76 kph)*		Negative Binomal (FE) *Congested Region (>76 kph)*	
	Coefficient	S. D.	Coefficient	S. D.	Coefficient	S. D.
Congestion Index	-1.0220	0.6592	-0.9534	0.6865	5.9041	6.8336
Flow	0.9605*	0.5783	0.9124*	0.5105	0.4101	0.5389
Speed Variance	0.2163	0.8996	-0.7272	1.1213	1.7746*	0.9525
Daylight	0.0391	0.2763	-0.2562	0.2767	-0.2274	0.3134
Weather (L)	-0.1015	0.0984	-0.1339	0.0856	0.0354	0.0486
Weather (M)	0.1759**	0.0805	0.0422	0.0658	-0.0033	0.0557
Weather (H)	-0.1282	0.4870	0.0133	0.2117	-0.1242	0.2805
Intercept	0.3147	9.7897	2.3394	5.7537	6.3274	1386.2
Log-likelihood	-198		-184		-135	

***Significantly different from zero (At the 99% interval)

**Significantly different from zero (At the 95% interval)

*Significantly different from zero (At the 90% interval)

In the case of injury/fatal accidents, we find the impact of congestion to be negative; but insignificant. One does expect, *a priori*, that slower traffic would warrant safety against serious accidents. But the cause of this relation's insignificance is not obvious. There may be a number of reasons for this. One of them might be the aggregation of data, which in general reduces the predictive power of the analysis. In addition, since serious accidents are more idiosyncratic, *randomness* and other unmeasured factors may play a greater part in their occurrence (Fridstrom and Ingebrigsten, 1991). Randomness itself is a phenomenon we use when we can not readily observe the effect of contributing factors of an outcome. The fact that accident occurrence depends on many factors, especially in the case of fatal accidents, makes it difficult to ascertain its likelihood by using a small number of explanatory variables. Certainly, there might be other unobservable (unrecordable) factors which play a greater role in their occurrence.

Another reason could be the way traffic sensors record traffic variables. For example, if a driver crashes for overspeeding, say while driving with 120 kph in a traffic which flows with 80 kph, we associate that accident with an *average* speed of 80 kph, not the original speed of culprit. Because the sensor averages the speed levels of every vehicle which passes in the last two minutes, the speed level which resulted in accident gets averaged out.

Finally, we regress fatal/injury accidents on tailgating hazard measure. We, again, find that there is a positive and significant relation. For every 1% increase in average tailgating, serious accident frequency increases by 0.23%. Considering the property damage accidents as well, this implies that increasing tailgating hazard increases the risk of accident, whichever its type. The intuition is straightforward: For a given level of speed, getting closer to the vehicle in the front or increasing the speed while keeping the same distance with that vehicle, increases the accident risk.

6. Conclusion

In this study, we tried to answer two questions: How do different levels of congestion affect the likelihood of accidents? Does this relation change for different types of accidents?

We find that congestion and accident likelihood relationship is not straightforward and that it differs for different congestion types. Overall, controlling for other contributing factors, we found a non-linear relationship between congestion level and accident frequency. Slower is not always safer, and it is worthwhile to think of the policy implications, such as speed limits. It suggests that setting the speed limit lower for traffic safety is counter-intuitive.

We also find that the impact of congestion on road accidents substantially changes with accident severity. While it is suggested by previous studies that increased congestion reduces the likelihood of fatal/injury accidents, this relation is reversed in the case of property damage accidents. In terms of fatal/injury accidents, our results are consistent with the previous work, which documents a negative relationship (although "insignificant"). For property damage accidents, to our knowledge, this study is the first to investigate the relation. Certainly, our findings are yet to be affirmed/refuted by further research.

One limitation of our study, and other similar empirical studies, is that only one particular road is examined while several factors -such as speed limits, temporal shape of travel demand, road infrastructure- vary greatly between countries, and over time. Thus, further analyses of roads with distinct characteristics are required to reach a universal consensus.

7. Appendix A: Negative Binomial Model

Since Negative Binomial model is derived from Poisson model, many of their properties are similar. It is also known as Poisson-Gamma model since these two distributions are used in its derivation. The purpose of this combination was to develop a Poisson model which is able to account for overdispersion in count data, its main feature which makes it popular for accident frequency analysis. There are two parametrizations that have been proposed and they are referred to as NB1 and NB2 (Cameron and Trivedi, 1998). The derivations we conducted in this section are based on Lord and Park (2010). In our analysis we employed the NB2 model, so we shall describe it here.

Let y_i be a random count variable which follows Poisson distribution,

$$y_i \sim Poisson(\mu_i)$$

$$p(y_i) = \frac{e^{-\mu_i}\mu_i}{y_i!}, \tag{A1}$$

where μ_i is the mean. To derive Negative Binomial model we introduce error to this Poisson mean:

$$ln\lambda_i = \beta x_i + \varepsilon_i \tag{A2}$$

$$\lambda_i = e^{(\beta X_i)}e^{\varepsilon_i} \tag{A4}$$

$$\lambda_i = \mu_i \epsilon_i, \tag{A5}$$

where ϵ_i ~ *Gamma(1,(1/η))*. The marginal distribution of y_i can be obtained by integrating the error term, ϵ_i,

$$f(y_i) = \int_0^\infty p(y_i) r(\epsilon_i) d\epsilon_i. \tag{A6}$$

In equation (A6), $p(y_i)$ and $r(\epsilon_i)$ are Poisson and Gamma distributions, respectively. Combination of these constitutes the Negative Binomial distribution. Here $p(y_i)$ is already given in terms of λ_i. In order to find *f(y_i)* we have to write the Gamma distribution as a function of λ_i as well. If we assume that the variable ϵ_i follows a Gamma distribution with two parameters η and δ:

$$g(\epsilon_i) = \frac{\delta^\eta}{\Gamma(\eta)} \epsilon_i^{\eta-1} e^{-\epsilon_i \delta}, \qquad \eta, \delta, \epsilon_i > 0, \tag{A7}$$

where $E[\epsilon_i] = {}^{\eta}/_{\delta}$ and $\lambda_i = \mu_i \epsilon_i$.

Then we can rewrite the Gamma distribution as a function of λ_i, which gives us the following PDF:

$$g(\lambda_i) = \frac{({}^{\eta}/_{\mu_i})^\eta}{\Gamma(\eta)} \lambda_i^{\eta-1} e^{-\frac{\lambda_i}{\mu_i}\delta}. \tag{A8}$$

Combining equations (A1) and (A8) into equation (A6) gives us the distribution of y_i:

$$f(y_i) = \int_0^\infty \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \frac{({}^{\eta}/_{\mu_i})^\eta}{\Gamma(\eta)} \lambda_i^{\eta-1} e^{-\frac{\lambda_i}{\mu_i}\delta} d\lambda_i. \tag{A9}$$

It can be shown that equation (A9) can be rewritten as (since Gamma function is a version of factorial function):

$$f(y_i) = \frac{({}^{\eta}/_{\mu_i})^{\eta}}{\Gamma(\eta)\Gamma(y_i+1)} \int_0^{\infty} e^{-\lambda_i(1+\frac{\eta}{\mu_i})} \lambda_i^{y+\eta-1} d\lambda_i \tag{A10}$$

$$f(y_i) = \frac{(\frac{\eta}{\mu_i})^{\eta}(1+\frac{\eta}{\mu_i})^{-(\eta+y_i)}\Gamma(y_i+\eta)}{\Gamma(\eta)\Gamma(y_i+1)}. \tag{A11}$$

Finally, we can express the probability distribution function of NB2 model as follows:

$$f(y_i) = \frac{\Gamma(y_i+\eta)}{\Gamma(y_i+1)\Gamma(\eta)} \left(\frac{\eta}{\eta+\mu_i}\right)^{\eta} \left(\frac{\mu_i}{\eta+\mu_i}\right)^{y_i}, \tag{A12}$$

with variance:

$$Var[y_i] = \mu_i + \frac{\mu_i^2}{\eta} = \mu_i \left(1 + \frac{1}{\eta}\right). \tag{A13}$$

so that the conditional variance of y_i given x_i is a quadratic function in the conditional mean. Variance rapidly increases as the mean does. Here $1/\eta$ is usually called as dispersion parameter. The larger $1/\eta$ is the greater the overdispersion. As this parameter goes to zero, NB model gets closer to Poisson model. So if $1/\eta$ is different from zero, using the Negative Binomial model would produce more reliable results. The model is estimated with maximum likelihood estimation. The likelihood function can be written as:

$$L(\eta, \beta) = \prod_{i=1}^{N} \frac{\Gamma(\eta+y_i)}{\Gamma(\eta)y_i!} \left(\frac{\eta}{\eta+\mu_i}\right)^{\eta} \left(\frac{\mu_i}{\eta+\mu_i}\right)^{y_i} \tag{A14}$$

where *N* is the number of observations in our analysis.

8. Appendix B: Results for Different Congestion Mode Thresholds

We use 76 kph as the threshold level of speed which traffic changes its flow mode. We found this level according to our data (*i.e.*, the point where the flow curve is bent). However, this point might be different for different roads and results may change. In this section we present results if we were to take this point differently. We estimate for the speed levels of 60 kph and 90 kph, the neighbourhood which traffic may change its mode. Our purpose is to show that our main results are robust to reasonable changes in this threshold speed level. As long as two parts of the splitted data cover the majority of different flow modes, results do not change. We firstly present the case for property damage accidents.

8.1 Property Damage Accidents

Table 6 – Estimates for 60 kph threshold.

Variables	Negative Binomal (FE) *All*		Negative Binomal (FE) *Hypercongested Region (<60 kph)*		Negative Binomal (FE) *Congested Region (>60 kph)*	
	Coefficient	S. D.	Coefficient	S. D.	Coefficient	S. D.
Congestion Index	0.4612**	0.2140	1.3745***	0.2126	-1.1069	1.7332
Flow	1.7850***	0.2470	2.0678***	0.4274	1.5134***	0.2857
Speed Variance	0.5787*	0.3063	0.7421*	0.4561	0.2003	0.3467
Daylight	0.1700*	0.1005	0.2477*	0.1325	0.3452**	0.1504
Weather (L)	-0.0260	0.0321	-0.0140	0.0310	0.0518	0.0394
Weather (M)	-0.0132	0.0266	-0.0400**	0.0196	0.0029	0.0329
Weather (H)	-0.2354	0.1605	-0.2203***	0.0708	0.0590	0.1928
Intercept	-4.9514***	0.9811	-8.4218***	1.5986	-2.9953**	1.4818
Log-likelihood	-617		-453		-487	

***Significantly different from zero (At the 99% interval)

**Significantly different from zero (At the 95% interval)

*Significantly different from zero (At the 90% interval)

Table 7 – Estimates for 90 kph threshold.

Variables	Negative Binomal (FE) *All*		Negative Binomal (FE) *Hypercongested Region (<90 kph)*		Negative Binomal (FE) *Congested Region (>90 kph)*	
	Coefficient	S. D.	Coefficient	S. D.	Coefficient	S. D.
Congestion Index	0.4612**	0.2140	1.0276***	0.1822	0.1975	1.7111
Flow	1.7850***	0.2470	1.2376***	0.2325	-0.0176	0.5720
Speed Variance	0.5787*	0.3063	0.7688**	0.2997	0.9277	1.4358
Daylight	0.1700*	0.1005	0.2101**	0.1044	-0.2233	0.4053
Weather (L)	-0.0260	0.0321	-0.0510	0.0334	-0.0724	0.0764
Weather (M)	-0.0132	0.0266	-0.0554**	0.0276	-0.1286	0.0935
Weather (H)	-0.2354	0.1605	-0.1433	0.1578	0.0858	0.0793
Intercept	-4.9514***	0.9811	-3.6995***	0.9176	14.790	1094.1
Log-likelihood	-617		-648		-72	

***Significantly different from zero (At the 99% interval)

**Significantly different from zero (At the 95% interval)

*Significantly different from zero (At the 90% interval)

Both Table 6 and 7 show that under hypercongested flow, the impact of congestion on property damage accidents is more significant.

8.2 Fatal/Injury Accidents

Table 8 – Estimates for 60 kph threshold.

Variables	Negative Binomal (FE) *All*		Negative Binomal (FE) *Hypercongested Region (<60 kph)*		Negative Binomal (FE) *Congested Region (>60 kph)*	
	Coefficient	S. D.	Coefficient	S. D.	Coefficient	S. D.
Congestion Index	-1.0220	0.6592	0.8817	0.8511	-2.2263	2.9604
Flow	0.9605*	0.5783	1.8798**	0.7934	0.9510*	0.4904
Speed Variance	0.2163	0.8996	1.1824	1.7269	-0.3626	0.7194
Daylight	0.0391	0.2763	-0.1195	0.3878	-0.0116	0.2808
Weather (L)	-0.1015	0.0984	0.0082	0.0733	0.0009	0.0787
Weather (M)	0.1759**	0.0805	0.0342	0.0481	0.0791	0.0669
Weather (H)	-0.1282	0.4870	0.1511	0.1129	-0.3218	0.3926
Intercept	0.3147	9.7897	7.7915	788.45	0.1354	2.5079
Log-likelihood	-198		-83		-226	

***Significantly different from zero (At the 99% interval)

**Significantly different from zero (At the 95% interval)

*Significantly different from zero (At the 90% interval)

Table 9 – Estimates for 90 kph threshold.

Variables	Negative Binomal (FE) *All*		Negative Binomal (FE) *Hypercongested Region (<90 kph)*		Negative Binomal (FE) *Congested Region (>90 kph)*	
	Coefficient	S. D.	Coefficient	S. D.	Coefficient	S. D.
Congestion Index	-1.0220	0.6592	-0.2381	0.5277	-3.3725	2.2631
Flow	0.9605*	0.5783	0.7120	0.4588	1.4103*	0.7456
Speed Variance	0.2163	0.8996	-0.6905	0.8213	1.7607	1.6284
Daylight	0.0391	0.2763	-0.2071	0.2602	-0.1023	0.5514
Weather (L)	-0.1015	0.0984	-0.0804	0.0864	-0.0192	0.0285
Weather (M)	0.1759**	0.0805	0.0785	0.0821	-0.0191	0.0893
Weather (H)	-0.1282	0.4870	-0.0551	0.4048	-0.0169	0.1322
Intercept	0.3147	9.7897	1.0181	2.5979	-4.3238	3.1313
Log-likelihood	-198		-239		-52	

***Significantly different from zero (At the 99% interval)

**Significantly different from zero (At the 95% interval)

*Significantly different from zero (At the 90% interval)

In fatal and injury accidents as well, we do not observe a major change (in terms of significant variables and their sign). We then conclude that minor changes in threshold level would not affect the results.

9. References

1. Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. *Accident Analysis and Prevention* 32: 633-642.
2. Cameron, C.A., Trivedi, P.K., 1986. Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests. *Journal of Applied Econometrics* 1: 29-53.
3. Ceder, A., Livneh, M., 1982. Relationships between road accidents and hourly traffic flow – I. Analysis and interpretation. *Accident Analysis and Prevention* 14: 19-34.
4. Ceder, A., 1982. Relationships between road accidents and hourly traffic flow – II. Probabilistic approach. *Accident Analysis and Prevention* 14: 35-44.
5. Daganzo, C.F., Geroliminis, N., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. Transportation Research Part B, 42: 771-781.
6. Fridstrom, L., Ingebrigtsen, S., 1991. An aggregate accident model based on pooled, regional time-series data. *Accident Analysis and Prevention* 23: 363-378.
7. Haan, C.T., 2002. *Statistical Methods in Hydrology*. Iowa State University Press, Ames, Iowa.
8. Hauer, E., 1971. Accidents, overtaking, and speed control. *Accident Analysis and Prevention* 3: 1-12.
9. Lave, C.A., 1985. Speeding coordination and the 55-mph limit. *American Economic Review* 75: 1159-1164.
10. Lord, D., Mannering F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice* 44: 291-305.
11. Lord, D., Park, B.J. Appendix D-Negative Binomial Regression Models and Estimation Methods. *In Crime Stat* Version 3.3 Update Notes: Part 2: Regression Modeling, The National Institute of Justice, Washington, D.C. 2010.
12. Lord, D., Washington, S.P., Ivan, J.N., 2005. Poisson, Poisson-Gamma and Zero Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory. *Accident Analysis and Prevention* 37: 281-296.

13. Miaou, S., 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accident analysis and prevention* 26: 471-482.
14. Miller, T. R., 2000. Variations between countries in values of statistical life. *Journal of Transport Economics and Policy* 34: 169–188.
15. Noland, R.B., Quddus, M.A., 2005. Congestion and safety: A spatial analysis of London. *Transportation Research Part A: Policy and Practice* 39: 737-754.
16. Rouwendal, J., Verhoef, E.T., 2004. A behavioral model of traffic congestion Endogenizing speed choice, traffic safety and time losses. *Journal of Urban Economics* 56: 408-434.
17. Solomon, D., 1964. *Accidents on main rural highways related to speed, driver and vehicle.* Federal Highway Administration, U.S. Department of Transportation.
18. Shefer, D., Rietveld, P., 1997. Congestion and safety on highways: towards an analytical model. *Urban Studies* 34: 679-692.
19. Shefer, D., 1994. Congestion, air pollution and road fatalities in urban areas. *Accident Analysis and Prevention* 26: 501-509.
20. Small, K., Chu X., 2003. Hypercongestion. *Journal of Transport Economics and Policy* 37: 319-352.
21. Small, K., Verhoef, E., 2007. *The Economics of Urban Transportation*. Routledge, Abingdon, UK.
22. Taylor, M.A.P., Wooley, J.E., Zito, R., 2000. Integration of the global positioning system and geographical information systems for traffic congestion studies. *Transportation Research Part C: Emerging Technologies* 8: 257-285.
23. TUIK, 2013. *Yearly Accident Casualty Report.* Turkish Statistical Institute. Retrieved from: http://www.tuik.gov.tr/PreTablo.do?alt_id=1051
24. Wang, C., Quddus, M. A., Ison, S. G., 2009. Impact of traffic congestion on road accidents: A spatial analysis of the M25 motorway in England. *Accident Analysis and Prevention* 41: 798-808.
25. Vickrey, W. S., 1969. Congestion theory and transport investment. *The American Economic Review* 59: 251-260.
26. Yu, X., Prevedouros, P.D., 2013. Performance and challenges in utilizing non-intrusive sensors for traffic data collection. *Advances in Remote Sensing* 2: 45-50.

27. Zhou, M., Sisiopiku, V., 1997. Relationship between volume-to-capacity ratios and accident rates. *Transportation Research Record: Journal of Transportation Research Board.* TRB, National Research Council, Washington, D.C., 1581, 47-52.