Road traffic congestion and crash severity: econometric analysis using ordered response models

Road Traffic Congestion and Crash Severity: An Econometric Analysis Using Ordered Response Models

Mohammed A Quddus
Chao Wang
Stephen G Ison

Transport Studies Group
Department of Civil and Building Engineering
Loughborough University
Loughborough, Leicestershire
LE11 3TU
UNITED KINGDOM

Abstract:
There is an ongoing debate among transport planners and safety policy makers as to whether there is any association between the level of traffic congestion and road safety. One can expect that the increased level of traffic congestion aids road safety and this is because average traffic speed is relatively low in a congested condition relative to an un-congested condition which may result in less severe crashes. The relationship between congestion and safety may not be so straightforward however as there are a number of other factors such as, traffic flow, driver characteristics, road geometry and vehicle design affecting crash severity. Previous studies have employed count data models (either Poisson or negative binomials and their extensions) while developing a relationship between the frequency of traffic crashes and traffic flow or density (as a proxy for traffic congestion). The use of aggregated crash counts at a road segment level or at an area level with the proxy for congestion may obscure the actual relationship. The objective of this study is to explore the relationship between the severity of road crashes and the level of traffic congestion using disaggregated crash records and a measure of traffic congestion while controlling for other contributory factors. Ordered response models such as *ordered logit models*, *heterogeneous choice models* and *generalised ordered logit* (partially constrained) models suitable for both ordinal dependent variables and disaggregate crash data are used. Data on crashes, traffic characteristics (e.g., congestion, flow, speed) and road geometry (e.g., curvature and gradient) were collected from the M25 London orbital motorway between 2003 and 2006.

Our results suggest that the level of traffic congestion does not affect the severity of road crashes on the M25 motorway. The impact of traffic flow on the severity of crashes however shows an interesting result. All other factors included in the models also provide results consistent with existing studies.

Keywords: Traffic congestion, traffic flow, crash severity, ordered response models, M25 motorway

INTRODUCTION

Two major factors in promoting economic productivity of a healthier economy are *enhanced mobility* and *improved safety*. There is an ongoing debate among transport planners and safety policy makers on the issue as to whether there is any association between mobility and road safety. Previous research suggests that the increased level of traffic congestion (less mobility) improves road safety (Shafer and Rietveld, 1997). This is because average traffic speed is relatively low in a congested condition in contrast to an un-congested condition which may lead to less severe traffic crashes. However, this may increase the occurrence of traffic conflicts often resulting more slight injury crashes. The increased level of traffic congestion reduces mobility which results in an economic loss to society. On the other hand, it is more likely that the level of crash severity would increase if a transport network is not congested. This suggests that the total external costs of crashes may be high in an un-congested condition relative to a congested condition and as such it can be thought that traffic congestion aids road safety but decreases economic productivity.

This poses a potential dilemma for transport policy makers: on the one hand wanting to reduce congestion but this may lead to more severe traffic crashes increasing the total external cost of congestion. In other words, the benefit of reducing congestion might be off-set by more severe crashes (Noland and Quddus, 2005). It is, therefore, important to understand the association between traffic congestion and road safety so that effective policy can be implemented to address both congestion and road safety.

The relationship between traffic congestion and road safety, especially crash severity may not be so straightforward as there are other factors affecting the severity of a crash. This includes other traffic characteristics (e.g., traffic flow and traffic speed), driver characteristics (e.g., seat-belt usage, age, experience, gender and alcohol consumption), vehicle conditions and road geometry (e.g., gradient, curvature, road width). To take all of these factors into account, researchers have employed various statistical models to develop a relationship between crashes and their contributing factors such as traffic characteristics, driver behaviour, vehicle design and road infrastructure. Some of the studies are discussed below.

In order to estimate the external cost caused by road crashes, Peirson et al. (1998) proposed that it is necessary to investigate the relationship between road accidents and traffic flow and found that crash frequency increases (either proportionally or at one and a quarter) if traffic flow increases. Research undertaken by Dickerson et al. (2000) investigated the relationship between the frequency of road crashes and traffic flow with the aim of estimating the change in the external cost of crashes caused by additional traffic flow. Different road types and geographical areas were considered and they found that a strong negative crash externality was associated with high traffic flows.

Ivan et al. (2000) investigated single and multi-vehicle highway crash rates and their relationship with traffic density while controlling for land use, time of day and light conditions. Temporal effects were also considered. For single-vehicle crashes, they found a negative-exponential relationship with the density (volume/capacity ratio), meaning that the crash rate is highest at a low volume/capacity ratio, but this is not fully consistent with the study by Lord et al. (2005) who conducted the analysis on the relationship among crash, density (vehicles per km per lane) and v/c ratio. They found that with v/c ratio increasing, fatal and single-vehicle crashes decreases at some point, and crash rates follows a U-shape relationship. Basically, all these studies suggested a positive relationship between flow and frequency of traffic crashes.

Some studies looked at this issue further by investigating hourly traffic flow and crash rates. For example, Martin (2002) investigated the relationship between crash incidence and traffic flow on French motorways, finding that crash rates are the highest in light traffic

compared to heavy traffic, especially on 3-lane motorways. There is no significant difference between daytime and night-time crashes. However, if crash severity is considered, night-time and light-traffic hourly crashes are much worse. Therefore, the author concluded that light traffic (low traffic flow) is a safety problem both in terms of crash rates and severity. Many things however could affect road safety during night-time such as lighting and as such needs further research. Hiselius (2004), on the other hand, showed the importance of the consideration of traffic flows, i.e., the crash rate would be different depending on whether the traffic flows is homogeneous or not.

It is noticeable that most of the previous studies examined aggregated crash counts (either at a road segment level or at an area level) in developing a relationship among crashes, traffic characteristics and other contributing factors. Moreover, various proxies were used to represent traffic congestion such as traffic flow and density.

Therefore, the primary aim of this study is to investigate the association between the severity (slight, serious and fatal) of individual crashes and the level of traffic congestion measured by total delay. Other contributory factors such as traffic flow, traffic speed, crash characteristics (e.g., a single-vehicle or a multiple-vehicle crash, number of casualties per crash, etc.), weather conditions, light conditions, road surface conditions and road geometry (e.g., gradient and curvature) are also considered while developing the relationship. It should be noted that no attempt is made to estimate the actual probability of a specific accident occurring. Statistical models suitable for both disaggregated crash data and ordered dependent variable (such as slight, serious and fatal) are used.

The paper is organised as follows. The next section provides a discussion of the statistical models used in the study . This is followed by a description of the data used in the analysis. The estimation results along with a discussion on the findings are then presented. The paper ends with conclusions, limitations and future research directions.

ORDERED RESPONSE MODELS (ORM)

The severity of a traffic crash can be expressed by the seriousness of the crash classified as slight, serious and fatal. It is explicitly clear that the dependent variable is categorical and ordered in nature in which a slight injury crash can be coded as 1, a serious injury accident can be coded as 2 and a fatal injury accident can be coded as 3. It should be noted that when a dependent variable is both categorical and ordinal, the distances between categories are unknown. When such an ordinal variable appears on the left-hand side of a statistical model, it is obvious that ordinary least-squares (OLS) estimation suffers from many shortcomings (see Long, 1997 for details). In order to deal with an ordered categorical variable, the use of an ordered logit (an OLOGIT) or ordered probit (an OPROBIT) model is more appropriate (Long 1997; Greene 2000; Gujrati 2003). These models are conditional as it assumed that an accident has already occurred and the factors affecting the accidents are known. However, a recent study byYamamoto et al. (2008) suggested that traditional 'unordered' models may provide unbiased estimate of the parameters, especially in the case for missing data such as under-reporting. Readers are also referred to a number of existing studies that recommend to explore alternative models (Milton et al., 2008; Eluru et al., 2008, Anastasopoulos et al., 2008).

Therefore, in order to investigate the impact of traffic congestion on the severity of road crashes, the concept of 'ordered' models is retained and the selected model is an OLOGIT model and its various extensions. Although, the OPROBIT model is also suitable

for an ordered categorical variable, the OLOGIT model is selected[1] because both logit and probit formulations provide very similar results.

Assuming that the severity of a road crash is an ordered discrete variable with j categories (slight, serious and fatal), an OLOGIT model (in terms of probability) can be written as (Long, 1997):

$$\Pr(y_i > j \mid X) = g(X_i \boldsymbol{\beta}') = \frac{\exp(X_i \boldsymbol{\beta}' - \phi_j)}{1 + \exp(X_i \boldsymbol{\beta}' - \phi_j)} , \quad j = 1, \dots, m-1 \tag{1}$$

where X_i is a (k×1) vector of observed non-random explanatory variables; $\boldsymbol{\beta}$ is a (k×1) vector of unknown parameters to be estimated; m is the number of categories of the ordinal dependent variable. The parameters of the model ($\boldsymbol{\beta}$) and the cut-points (ϕ_1 and ϕ_2) are estimated by the method of maximum likelihood (Long, 1997). In equation (1), it is assumed that the effects of explanatory variables on the level of severity are assumed to be fixed across observations. However, this may not be true as the effect of a explanatory variable may vary across observations. To overcome this problem, a number of recent studies have suggested to employ random parameters models (e.g., Anastasopoulos and Mannering, 2009; McFadden and Train, 2009; Ben-Akiva et al., 2002).

One of the primary assumptions of an OLOGIT (and an OPROBIT) models is that the error variances are homoskedastic. In the context of ordinary least squares (OLS), a violation of this assumption (i.e., heteroskedasticity) does not bias the estimates, rather it either inflates or underestimates the standard errors. Heteroskedasticity, however, is more problematic in the case for models dealing with categorical dependent variables such as logit or probit and their ordered variants. If variances of the error term are non-constant, not only the standard errors are incorrect, but also the parameters are biased and inconsistent (Keele and Park, 2006). In order to deal with unequal error variances, Williams (2006a) suggests the use of a heterogeneous choice model (HCM) which can be written as:

$$\Pr(y_i > j \mid X) = g\left(\frac{X_i \boldsymbol{\beta}'}{\sigma_i}\right) = \frac{\exp\left(\dfrac{X_i \boldsymbol{\beta}'}{\sigma_i} - \phi_j\right)}{1 + \exp\left(\dfrac{X_i \boldsymbol{\beta}'}{\sigma_i} - \phi_j\right)} , \quad j = 1, \dots, m-1 \tag{2}$$

in which $\ln \sigma_i = Z_i \theta'$ where Z_i is the vector of explanatory variables (either dummy or continuous variables) that effect the error variance (σ_i). Z could either be a subset of X or a set of new variables not included in X.

Another important assumption associated with an OLOGIT (and an OPROBIT) regression is that the relationship between each pair of outcome groups is the same. In the literature, this is known as the *proportional odds assumption* or *the parallel regression assumption* (see Long, 1997 for details).. If the proportional odds assumption is not valid, one needs different models to describe the relationship between each pair of outcome groups.

---

[1] The logit model is preferred to the probit. The logit model assumes that the disturbances are Weibull distributed (Gumbel extreme value type I), whereas the probit model assumes that the disturbances are multivariate normally distributed

Therefore, it is essential to test the proportional odds assumption after estimating an OLOGIT model.

The Brant test (Brant, 1990) could be employ to test the above assumption. A significant test statistic provides evidence that the proportional odds assumption is violated. If this is the case, then the use of the OLOGIT model may lead to incorrect, incomplete or misleading results (Fu, 1998).

A solution is then to employ a *generalised ordered logit (*GOLOGIT*) model* which does not impose the constraints of parallel regressions (Fu, 1998). The unconstrained GOLOGIT model can be rewritten as:

$$\Pr(y_i > j \mid X) = g(X_i \boldsymbol{\beta}_j') = \frac{\exp(X_i \boldsymbol{\beta}_j' - \phi_j)}{1 + \exp(X_i \boldsymbol{\beta}_j' - \phi_j)}, \ j = 1,....,m-1 \tag{3}$$

An issue with this GOLOGIT model is that it estimates far more parameters than is really necessary (Williams, 2006b). For instance, if a dependent variable has 4 categories and there are 10 independent variables, the GOLOGIT model estimates a total of 30 coefficients. This sometimes makes it difficult to interpret the results.

Williams (2006b) then proposes a partially constrained GOLOGOT model known as a *partial proportional odds model* in which only a subset of coefficients are constrained across values of *j* and therefore, is less restrictive than a GOLOGIT. In a recent study, Wang and Abdel-Aty (2008) employed this model to investigate the left-turn crash injury severity at intersections. This model can be rewritten as:

$$\Pr(y_i > j \mid X) = \frac{\exp(X_{1i} \boldsymbol{\beta}_{1j}' + X_{2i} \boldsymbol{\beta}_2' - \phi_j)}{1 + \exp(X_{1i} \boldsymbol{\beta}_{1j}' + X_{2i} \boldsymbol{\beta}_2' - \phi_j)}, \ j = 1,....,m-1 \tag{4}$$

in which the coefficients associated with a subset of independent variables X_2 are the same across values of j and the coefficients related to other independent variables (X_1) differ across values of j.

Eluru et al. (2008) considered a mixed GOLOGIT model for examining pedestrian and bicyclist injury severity level in traffic crashes but could not find any statistically significant unobserved heterogeneity effects on the latent injury risk propensity and the cut-points and therefore, used a GOLOGIT model.

This research examines whether there is any association between the severity of road crashes and the level of traffic congestion employing three ordered response models: (1) an OLOGIT (2) a HCM and (3) a PC – GOLOGIT.

DATA

In order to examine the association between the level of traffic congestion and the crash severity, the UK M25 motorway has been chosen as a case study. The M25 motorway is a 188 km (each direction) London orbital motorway which almost completely encircles London. There are two primary reasons for selecting the M25: (1) this motorway is considered as one of the busiest (about 200,000 vehicles a day in 2003) motorways in Europe and, therefore, there is sufficient spatio-temporal variation in congestion conditions which

allow us to develop statistical models that relate traffic congestion and crash severity. (2) data on traffic characteristics (e.g., traffic congestion, traffic speed, traffic flow) and road geometry (e.g., radius of road curvature, gradient, number of lanes) are available to us for the M25 motorway from 2003 to 2006. However, one of the disadvantages (from a purely research perspective) of using the M25 as a case study is that the number of fatal and serious crashes on the M25 is quite low - there were 23 people killed and 116 seriously injured on the M25 in 2006. In order to tackle this problem, statistical models that can look into individual crash records are used and crash data for multiple years (2003 to 2006) are considered.

STATS19 UK road crash data from 2003 and 2006 were obtained from the UK data archive (UKDA[2]). STATS19 data have three data files: (1) crash data (2) vehicle data and (3) casualty data. A unique crash reference number allows one to integrate these three data files. The vehicle data file contains information regarding driver age and gender. Although driver age and gender affects the severity of a crash, such data cannot be used while analysing multiple-vehicle crashes[3] as there is no information as to whether which driver is at-fault (or not-at-fault) for the crash.

Traffic characteristics data such as traffic congestion, traffic speed (km/h) and traffic flow (vehicles/h) were obtained from the UK Highways Agency (UKHA[4]). These data were available from 2003 to 2006 for a total of 72 segments of the M25 (both directions) at 15-minute intervals. Traffic congestion at each of these segments is measured by the total delay (minutes) encountered by all vehicles travelling on that segment. In order to take into account the lengths of the segments, the total delay is averaged over a 10-km stretch of the motorway. Since each segment starts and also terminates at a junction, it is reasonable to assume that delays, traffic speed and traffic flow are the same on different locations of the segment.

Road geometry data such as the radius of road curvature (m) and gradient or vertical grade (%) were also obtained from the UKHA. Since a series of curvatures and gradients were available for a segment, the minimum radius of curvature and the maximum gradient were considered. Data on the number of lanes were also obtained from the UKHA.

Since STATS19 data have the easting and northing coordinates of a crash location, it is possible to identify the motorway segment (out of the 72 segments) on which the crash occurred. Since both crash location data and digital motorway segment data contained errors and the two directions (clockwise and anti-clockwise) of the M25 motorway are treated separately, a matching technique considering the direction(s) of the vehicle(s) just before the crash relative to the direction of the motorway segment (either clockwise or anti-clockwise) and the distance from the crash location to the segment was used to match the crash location onto the correct motorway segment (see Wang et al., 2009 for details).

In order to integrate STATS19 data with the traffic data, a common variable between these datasets was used. There was only one common variable between them which was the *time* epoch as the time of the crash (for the STATS19 data) and the time at which the traffic data were measured (for the UKHA data) were known. Therefore, it was possible to determine the level of congestion, the average traffic speed and traffic flow for each crash record. In order to avoid the impact of the crash itself on the traffic variables, a 30-minute time lag was considered. For instance, if a crash happened at 15:00 then traffic data measured at 14:30 were used when these two datasets were combined.

A total of 3,998 crashes occurred on the M25 motorway between 2003 and 2006. Of which 1.28% were fatal crashes, 8.83% were serious injury crashes and 89.89% were slight injury crashes. Table 1 shows summary statistics of the variables that will be considered in the models.

---

[2] http://www.data-archive.ac.uk/
[3] About 85% of the M25 motorway crashes are multiple-vehicle crashes
[4] http://www.highways.gov.uk/

Table 1 is about here

The combined dataset shows that the average total delay (over all 72 motorway segments) at which fatal crashes occurred is 4 minutes. This increases to 8 minutes for the case of serious injury crashes and 9.6 minutes for the case of slight injury crashes. This suggests that there may be a relationship between total delay and the severity of crashes. This is also true for traffic flow as the mean traffic flow at which fatal crashes happened on the M25 is 2131 veh/h. This increases to 3345 veh/h for the case of serious injury crashes and 3911 veh/h for the case of slight injury crashes suggesting that there is an association between traffic flow and the severity of crashes. The average traffic speed at which fatal crashes occurred is 93km/h. This decreases to 86km/h for the case of serious injuries and 84.5km/h for the case of slight injury crashes.

The M25 motorway has a variable number of lanes with a minimum two-lane and a maximum six-lane (in each direction). According to data from the UKHA, the motorway has three-lane for most of its length (66.8%) and four-lane for 24.9% of its length. There are a few short stretches which are two-lane (2.4%), five-lane (4.2%) and six-lane (1.7%). It might be interesting to see whether road width (number of lanes) has any impact on the level of crash severity.

The details of other explanatory variables can be found in Table 1.

VARIABLES SELECTION AND RESULTS

Before estimating any models using the data, a multicollinearity test among the explanatory variables was carried out suggesting that *fine* and *raining weather conditions* were highly correlated (correlation coefficient: ±0.7) with *dry* and *wet road surface conditions* and as expected, *traffic congestion* was found to be highly and negatively correlated (correlation coefficient: -0.8) with *average traffic speed*. Since our interest is to examine the association between traffic congestion and the severity of road crashes, *average traffic speed* was taken out from the set of explanatory variables along with *weather conditions*. Surprisingly, posted *speed limit* was found to be un-correlated with *average traffic speed*. This may be due to the fact that average speed varies with the level of traffic congestion. When a road is congested, the average speed is much lower than the posted speed limit; whereas motorists would drive faster than the speed limit if a road is not congested. This may also be the results of motorist's perception and psychology, highway hypnosis (see Wertheim, 1978; Cerezuela et al., 2004) and risk compensation (Assum et al., 1999; Dulisse, 1997; Winston et al., 2006). Another interesting observation was that *time of the day* (i.e., peak and off-peak periods) was not correlated with *traffic congestion* as one would expect that traffic congestion is normally high during the peak hours.

The set of un-correlated explanatory variables was used to estimate different ordered response models such as OLOGIT, HCM, GOLOGIT and PC-GOLOGIT. The results are presented in Table 2. The variable – *speed limit* was found to be statistically insignificant in all models and in addition, a log-likelihood ratio (LR) test also confirmed that the inclusion of this variable did not improve the model goodness-of-fit. Therefore, this variable was dropped from all models.

Table 2 shows that the likelihood ratio (LR) Chi-square is higher in the HCM compared with the OLOGIT model. The difference in the LR Chi-squares between these two models was found to be statistically significant (*p*-value<0.01) suggesting that the results of the HCM are much better than those of the OLOGIT model. As noticed, the explanatory variables of the HCM were divided into two classes: (1) the variables affecting the ordinal categorical choice (i.e., fatal, serious and slight injury crashes) and (2) the variables affecting

variances of the error term known as the determinants of variability in the error term across observations. It was expected that all explanatory variables had an impact on the level of crash severity. The set of variables for the error variance equation was identified by a stepwise selection method employing a LR test[5]. The test suggested that *crash category* by vehicle type (either a single-vehicle crash or a multi-vehicle crash) and casualties per crash are the statistically significant variables for the inclusion in the variance equation. This was also confirmed by the HCM estimation results in which both variables were found to be statistically significant in the error variance equation. Therefore, it is reasonable to believe that *crash category* and *casualties per crash* could be a potential source of heteroskedasticity while analysing the severity of crashes.

In addition to the fixed-parameter ordered logit model, the random- parameter ordered logit model was also estimated. It was found that the random effects were statistically insignificant. The mixed multinomial logit model in which certain parameters were assumed to be 'random' was also estimated. Once again, the standard deviations of random parameters were found to be statistically insignificant.

The Brant test (Brant, 1990) was carried out to see whether the proportional odds assumption was violated for the data used in the analysis. A significant test statistic provided evidence that the assumption has been violated. A user-written STATA routine *ologit2* (developed by William (2006)) was used to identify the variables which did not meet the proportional odds assumption. It was found that two explanatory variables (*log of traffic flow* and *number of vehicles involved* in the crash) did not meet the assumption and therefore, their coefficients differed across different thresholds suggesting that the OLOGIT model is a mis-specified model and GOLOGIT or PC-GOLOGOT models should be used.

Table 2 is about here

Table 2 shows that the coefficients of all explanatory variables are different across thresholds for the GOLOGIT model whereas only the coefficients of *log of traffic flow* and *number of vehicles involved* in the crash were different across thresholds for the PC-GOLOGIT model. Although the value of the likelihood ratio (LR) Chi square is higher in the GOLOGIT model relative to the PC-GOLOGIT model, the difference is not statistically significant as the value in the GOLOGIT model is only 10.5 units more for 13 degrees of freedom (*p*-value=0.65). This suggests that the model goodness-of-fit is better in the PC-GOLOGIT model compared to the GOLOGIT model. In terms of both log-likelihood at convergence and LR Chi square, there is no difference between HCM and PC-GOLOGIT models (see Tables 2 and 3). However, HCM handles the effect of heteroskedasticity and PC-GOLOGIT addresses the violation of proportional odds assumption. Since the results from these two models are quite similar in terms of signs and the set of statistically significant variables, the PC-GOLOGIT model will be used to interpret the effects of the explanatory variables on the crash severity.

Table 3 is about here

Using these estimated coefficients and cut points, the probabilities of three different outcomes (slight, serious and fatal) for the given values of explanatory variables were obtained. From these estimated probabilities, factors that are more likely to reduce the probability of a particular level of severity were identified. Table 3 shows the marginal effects

---

[5] By using a user-written STATA routine known as *oglm* by Williams (2006a)

for the probabilities of different outcomes with respect to the statistically significant independent variables. The interpretation of each of the explanatory variables is given below.

Traffic congestion: As discussed previously, traffic congestion at the time of a crash is measured as the total delay encountered by all vehicles travelling on a segment (averaged over a 10km-stretch) where the crash occurred. It was hypothesized that the level of traffic congestion has an impact on the severity of the crash outcome. More specifically, when a crash happens on a road segment with a high level of traffic congestion, one would expect that the severity of the crash would be relatively low resulting in a slight injury crash and vice-versa. However, our data from the M25 motorway do not support this hypothesis and this is the case for all ordered response models estimated in this study.

Figure 1 is about here

Other measurements of traffic congestion such as a congestion index based on actual and free flow travel time proposed by Taylor et al. (1999) was used but also found to be statistically insignificant in all models. One would expect that total delay representing the level of traffic congestion may be correlated with the *time of the day* variable (an indicator variable for peak period) and therefore, this finding may be incorrect. This was not the case however as we found that there is no correlation (correlation coefficient: 0.2) between these two variables for the data we analysed. Figure 1 shows the observed hourly delay (averaged over the four years) just before the crashes happened. It is noticeable that the pattern of delay is different for weekends and weekdays. During the weekdays, the total delay is high at peak periods but this is not the case for the total delay during the weekends. However, all ordered response models were also estimated when the dummy variable for the peak period was being dropped. Nevertheless, the level of traffic congestion was still found to be statistically insignificant. Another possible reason would be the inclusion of traffic flow in the models as one would expect that higher traffic flow (per lane per hour) would be correlated with total delay. Although this was not the case, all models were estimated with the exclusion of total flow variable. The sign of the coefficient was found to be the expected negative sign but again statistically insignificant in all models. In addition, the LR test did not support the exclusion of total flow variable from the set of explanatory variables.

Models were also estimated by other different combinations of explanatory variables but total delay was consistently found to be statistically insignificant.

Traffic flow: This variable was found to be one of the important variables in explaining the severity of a road crash since the exclusion of this variable significantly increased (about 25 units) the log-likelihood at convergence. Not surprisingly, this variable was found to be statistically significant at the 95% confidence level in all models. As discussed, the Brant test suggested that traffic flow influences the cut-off points (thresholds) and, therefore, the impact of this variable on the level of crash severity differed across different thresholds. In general, it was found that if traffic flow increases then the level of crash severity decreases meaning that a road segment with high traffic flow would result in less severe crashes, if all else remain constant. It is noticeable that the value of the coefficient is significantly different (about 70%) between the thresholds. It was found to be more negative in the threshold that divides serious injury and fatal crashes suggesting that it is more likely that higher values on traffic flow increase the likelihood of being a slight injury crash. The impact of traffic flow on the severity of crashes is not uniform across different crash categories and this type of effect could not be found from an OLOGIT model.

Figure 2 is about here

It is also of interest to estimate the probability of a specific crash occurring (i.e., slight, serious and fatal) for a given value of traffic flow. Figure 2 shows how the predicted probabilities of different categories of crashes change with the change in the traffic flow. These probabilities were obtained from the results of the PC-GOLOGIT model. For an average traffic flow (3843 veh/h) on a three-lane stretch of the M25 during weekends, daylight, off-peak periods and at dry weather conditions, the estimated probabilities of different categories of a crash involving a single-vehicle in 2003 are as follows:

$$\Pr(Slight) = 0.8, \ \Pr(Serious) = 0.17 \text{ and } \Pr(Fatal) = 0.03$$

Table 3 shows the marginal effects of the probabilities of a specific crash occurring for changes in the explanatory variables. For instance, the marginal effect for the probability of a slight injury crash occurring with respect to traffic flow (veh/h) is positive and the value is 0.0403, i.e.:

$$\frac{\partial \Pr(y = Slight)}{\partial (TrafficFlow)} = 0.0403$$

The marginal effect for the probability of a serious injury crash occurring with respect to traffic flow (veh/h) is negative and the value is -0.0343, i.e.:

$$\frac{\partial \Pr(y = Serious)}{\partial (TrafficFlow)} = -0.0343$$

The marginal effect for the probability of a fatal crash occurring with respect to traffic flow (veh/h) is negative and the value is -0.006, i.e.:

$$\frac{\partial \Pr(y = Fatal)}{\partial (TrafficFlow)} = -0.006$$

Number of vehicles involved in the crash: Road traffic crash data (2003-2006) from the M25 suggest that single-vehicle crashes were more severe than multi-vehicle crashes. For instance, 3.4% single-vehicle crashes were fatal crashes compared to only 0.9% in the case for multi-vehicle crashes. This is also true for the case of serious injury crashes in which 14% of single-vehicle crashes were serious injury crashes and on the other hand, 8% of multi-vehicle crashes were serious injury crashes. Therefore, it was expected that the models used in this study should be able to pick up such effects.

The dummy variable used to represent a single-vehicle crash was found to be positively associated with the severity of crashes meaning that a single-vehicle crash is likely to result in a more severe crash compared with a multi-vehicle crash, if all else are held constant. As discussed, this variable did not meet the proportional odds assumption and therefore, the coefficient value differs across thresholds. For the first threshold, the value is 0.66 and for the second threshold, this increases to 1.19 suggesting that the effect of a crash involving a single-vehicle on the severity was not uniform. A crash involving a single-vehicle on the M25 is more likely to result in a more severe crash relative to a crash involving multiple-vehicle.

The marginal effects of probabilities of a specific crash occurring for a change from a multi-vehicle crash to a single-vehicle crash are shown in Table 3. When the dummy variable (representing whether a crash is a single-vehicle crash) changes from 0 (a multi-vehicle crash) to 1 (a single-vehicle crash), the predicted probability of outcome: "slight injury" changes by -0.064, "serious injury" changes by 0.05 and "fatality" changes by 0.014, holding all other variables at their means. This finding also confirms that a single-vehicle crash is more likely to result in a higher level of severity on M25. Figure 3 shows observed and predicted probabilities of serious and slight injury categories for both single-vehicle and multi-vehicle crashes. Both actual and predicted probabilities are quite similar indicating that a good measure of fit for the model.

Figure 3 is about here

Road surface conditions: Two types of *road surface conditions* (at the instant of a crash) are considered: (1) dry and (2) wet. An indicator variable was used to characterise them in the models where 0 means dry road surface conditions and 1 means wet road surface conditions. This indicator variable was highly and positively correlated (correlation coefficient = 0.7) with weather conditions (specifically with "raining") suggesting that wet surface conditions are mostly due to the result of *raining weather conditions*. Our results suggest that wet road surface conditions reduce the level of M25 crash severity compared with dry surface conditions. This finding is consistent with other studies (e.g., Quddus et al., 2002; Duncan et al., 1998; Shankar and Mannering, 1996). Quddus et al. (2002) argued that this is likely to be an effect of reduced speed levels.

All models were also estimated with weather conditions categorised as fine, raining, snowing and others (while the indicator variable for road surface conditions was being dropped). The results (not shown here for brevity) also suggest that "raining" weather conditions reduce the level of M25 crash severity compared with "fine" weather conditions whilst "snowing" and "others" weather conditions were not statistically significant.

Number of lanes: As discussed, the number of lanes within the M25 varies from two- to six-lane in each direction. STATS19 UK road crash data (2003 to 2006) suggest that 71% of all serious injury crashes and 67% of all fatal crashes happened on the three-lane stretches of the motorway. In order to see whether the variability in lanes has an impact on the level of crash severity, a categorical variable with three categories was used in the models. The categories were: (1) three-lane (or less) (2) four-lane (3) five- lane (or higher) and the second category was taken as a reference case. The results suggest that the level of crash severity on the stretches with three-lane (or less) was statistically and significantly different than that of on the stretches with four-lane. There was no difference in the crash severity between stretches with four-lane and stretches with five-lane (or higher) given that all other variables included in the models were held constant.

Crashes on the stretches of M25 with three-lane appear to increase crash severity levels. This is also reflected in the signs of the marginal effects of this categorical variable (three-lane, five-lane where four-lane is taken as a reference) on the probabilities of different injury crashes (see Table 3). The sign is negative for the marginal effect of the probability of a slight injury crash and positive for marginal effects of the probabilities of a serious injury crash or a fatal crash.

Time trend: Both a time trend variable representing the month in which the crash occurred and a categorical variable representing the crashes that occurred in different years

were tested and the results were found to be very similar. Therefore, models only with a categorical variable are shown for brevity. This categorical variable has four categories such as year 2003, year 2004, year 2005 and year 2006. The first category was used as a reference case. It can be seen that the coefficients for 2004, 2005 and 2006 are negative suggesting that there is a downward trend in injury severity[6]. Some factors (not included in the models) that vary over time are leading to this trend. The signs of the marginal effects of probabilities for either a serious or a fatal crash occurring are also negative for 2004, 2005 and 2006 (see Table 3). This finding is consistent with other studies that used STATS19 data (e.g., Noland and Quddus, 2004).

Number of casualties per crash: About 65% of the crashes occurred on the M25 between 2003 and 2006 had a single casualty, 22% of the crashes had two casualties, 7% of the crashes had three casualties and 3% of the crashes had four casualties. The average number of casualties per crash is 1.55. A slight injury crash with a single casualty and a slight injury crash with multiple casualties were taken as an identical dependent variable in our models. This might be a problem given that a slight injury crash with multiple casualties is considered to be more severe than a slight injury crash with a single casualty. In order to control for such effects, a continuous variable representing the number of casualties per crash was used in the models. The results suggest that the level of severity increases with the increase in the number of casualties per crash. Figure 4 shows how the predicted probabilities of different categories of crashes change with the change in the number of casualties per crash.

Figure 4 is about here

Other factors: Other factors that had an effect on the severity of road crashes were found to be the radius of *road curvature*, *day of the week* and *light conditions*. However, all of these variables were statistically significant only at the 90% confidence level. If the radius of road curvature of a road segment increases then the severity of a crash that occurred in that segment also increases suggesting that crashes on a straighter road segment (relatively high radius of curvature) were more severe than a curved road segment (relatively low radius of curvature). Although one would think that this is a surprising result, existing studies (Haynes et al., 2007 and Wang et al., 2008) while examining the frequency of killed and seriously injured (KSI) crashes at the area-wide level also found that curved roads are safer than straighter roads. Milton and Mannering (1998) also found that sharp horizontal curves tend to decrease accident frequency. While comparing with a disaggregated crash data, this result is inconsistent with the finding of Quddus et al (2002) that reported that bends in the road appear to result in more severe injuries while investigating the severity of road crashes in an urban area. Since the characteristics of road configuration between an urban area and a motorway (M25 in our case) is quite different, a dissimilar result can be expected.

Looking at *day of the week effects* (weekdays = 1 and weekend = 0), more severe crashes are predicted during weekdays. However, this finding is not consistent with the finding of Gray et al. (2008) who found that crashes in Great Britain are more severe on Fridays, Saturdays and Sundays (relative to Mondays). In order to investigate this, a categorical variable of seven categories representing seven days of the week was also examined (the results are not shown for brevity). However, less severe crashes are predicted on Sundays, Tuesdays, Fridays and Saturdays (relative to Mondays). This again suggests that

---

[6] The coefficient for 2006 is statistically significant at only 80% confidence level

day of the week effects on the crash severity between motorways and other types of roadways may be different.

The variable *light conditions* (darkness=1, daylight=0) allows us to investigate the effect of the level of light on the injury severity. Less severe injury crashes are predicted during darkness. This finding is also not consistent with that of Gray et al. (2008) while analysing the crash severity in London.

CONCLUSIONS

Disaggregated crash data from the M25 motorway have been used to investigate the association between the severity of road crash and the level of traffic congestion. This has been done while controlling for other contributory factors such as traffic characteristics (e.g., traffic flow), road geometry (e.g., curvature and gradient) and crash characteristics (e.g., single-vehicle or multi-vehicle). Statistical models such as ordered logit models, heterogeneous choice models and partially constrained generalised ordered logit (PC-GOLOGIT) models suitable for an ordered response variable have been employed. Our results suggest that ordered logit models are not appropriate for the data we analysed. Both heterogeneous choice models and partially constrained generalised ordered logit models have fitted the data equally. The results are also consistent between these two models. Our results suggest that that the level of traffic congestion (measured by total delay or congestion index) does not affect the severity of road crashes on the M25 motorway. However, the impact of traffic flow on the severity of crashes shows an interesting result. While previous studies show a positive association between the frequency of traffic crashes and traffic flow, our disaggregated analysis suggests that increased traffic flow reduces the severity of crashes. The PC - GOLOGIT model has also been used to estimate the change in the relative probability of three different levels of severity of a crash for given values of explanatory variables.

The factors that result in less severe crashes have been found to be traffic flow, radius of road curvature, darkness light conditions, wet road surface conditions and time trend. The factors resulting in high severe crashes have been found to be three-lane stretches of the motorway, single-vehicle crash and weekdays. The gradient of road segment and time of the day have found to be insignificant.

One of the limitations of this study is that traffic flow and speed values were assigned to crashes based on segment measurements. However, segments would not necessarily have uniform conditions over 10km length if queues are present.

There are a number of ways to extend the analysis used in this study. It would be interesting to analyse single-vehicle crashes separately as there is clear evidence that crashes involving a single-vehicle are more severe than those involving multiple-vehicle. In such an analysis, the effects of driver age and gender on the severity of a crash can also be estimated. Another possible extension would be to consider crashes that occurred on different types of roads such as motorways, A roads, B roads and minor roads.

REFERENCES

Brant, R., 1990, "Assessing proportionality in the proportional odds models for ordinal logistics regression", *Biometrics*, 46, 1171-1178.

Clogg, Clifford C., and Edward S. Shihadeh. 1994. *Statistical Models for Ordinal Variables*. SAGE Series on Advanced Quantitative Techniques, 4.

Dickerson, A., Peirson, J. and Vickerman, R. 2000, "Road accidents and traffic flows: an econometric investigation", *Economica*, 67(265), 101-121.

Duncan, C. S., Khattak, A. J., and Council, F. M., 1998, "Applying the ordered probit model to injury severity in truck– passenger car rear-end collisions." *Transportation Research Record*, 1635, 63–71.

Eluru, N., Bhat, C.R., Hensher, D.A., 2008, "A mixed generalised ordered response model for examining pedestrian and bicyclists injury severity level in traffic crashes", *Accident Analysis and Prevention*, 40(3), 1033-1054.

Fu, V., 1998. "Estimating Generalized Ordered Logit Models". *Stata Technical Bulletin* 44: 27-30. In *Stata Technical Bulletin Reprints*, vol 8, 160-164. College Station, TX: Stata Press.

Greene, W. H. 2000. "Econometric analysis." (4th ed). New Jersey: Prentice-Hall.

Gray, R., Quddus, M.A., Evans, A., 2008, "Injury severity analysis of accidents involving young male drivers in Great Britain", *Journal of Safety Research*, 39(5), 483-495.

Gujarati, D. 2003. "Basic econometrics." (4th ed), MaGraw Hill.

Haynes, R., Jones, A., Kennedy, V., Harvey, I. and Jewell, T. 2007, "District variations in road curvature in England and Wales and their association with road-traffic crashes", *Environment and planning A,* vol. 39, no. 5, pp. 1222-1237.

Hiselius, L.W. 2004, "Estimating the relationship between accident frequency and homogeneous and inhomogeneous traffic flows", *Accident Analysis & Prevention,* vol. 36, no. 6, pp. 985-992.

Ivan, J.N., Wang, C. and Bernardo, N.R. 2000, "Explaining two-lane highway crash rates using land use and hourly exposure", *Accident Analysis & Prevention,* vol. 32, no. 6, pp. 787-795.

Keele, L., Park, D., 2006, "Difficult choices: an evaluation of heterogeneous choice models", Working paper, Department of Political Science, Ohio State University. Available at http://www.nd.edu/~rwilliam/oglm/ljk-021706.pdf, Accessed on 18th July 2008.

Long, J.S., 1997. "Regression models for categorical and limited dependent variables." Thousand Oaks, CA, Sage Publications.

Lord, D., Manar, A. and Vizioli, A. 2005, "Modeling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments", *Accident Analysis & Prevention,* vol. 37, no. 1, pp. 185-199

McCullagh, P. and J.A. Nelder. 1989, "*Generalized Linear Models",* Chapman and Hall: London

Martin, J. 2002, "Relationship between crash rate and hourly traffic flow on interurban motorways", *Accident Analysis & Prevention,* vol. 34, no. 5, pp. 619-629.

Milton, J., Mannering, F., 1998, "The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies", *Transportation,* vol. 25, no. 4, pp. 395-413.

Milton, J.C., Shankar, V.N., Mannering, F.L., 2008, "Highway accident severities and the mixed logit model: an exploratory empirical analysis", *Accident Analysis and Prevention*, 40(1), 260 – 266.

Noland, R.B., Quddus, M.A., 2004, "Improvements in medical care and technology and reductions in traffic-related fatalities in Great Britain". *Accident Analysis Prevention*, 36 (1), 103–113.

Noland, R.B. and Quddus, M.A. 2005, "Congestion and safety: a spatial analysis of London", *Transportation Research Part A: Policy and Practice,* vol. 39, no. 7-9, pp. 737-754.

Peirson, J., Skinner, I. and Vickerman, R. 1998, "The microeconomic analysis of the external costs of road accidents", *Economica,* vol. 65, no. 259, pp. 429-440.

Quddus, M.A., Noland, R.B., Chin, H.C., 2002, "An analysis of motorcycle injury and vehicle damage severity using ordered probit models", *Journal of Safety Research*, 33(4), 445-462

Shankar, V., & Mannering, F., 1996, "An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity". *Journal of Safety Research*, 27(3), 183– 194.

Shefer, D. and Rietveld, P. 1997, "Congestion and safety on highways: towards an analytical model", *Urban Studies,* vol 34, no. 4, pp. 679-692.

Taylor, M.A.P., Woolley, J.E. and Zito, R. 2000, "Integration of the global positioning system and geographical information systems for traffic congestion studies", Transportation Research Part C: Emerging Technologies, 8(1-6), 257-285.

Wang, X., Abdel-Aty, M., 2008, "Analysis of left-turn severity by conflicting pattern using partial proportional odds models", *Accident Analysis and Prevention*, 40(5), 1674-1682.

Wang, C., Quddus, M.A. and Ison, S.G., "The effects of area-wide road speed and curvature on traffic casualties in England", *Journal of Transport Geography*(in press)

Williams, R., 2006a, "OGLM: Stata module to estimate Ordinal Generalized Linear Models." Available at http://econpapers.repec.org/software/bocbocode/s453402.htm, accessed on 18th July 2008.

Williams, R., 2006b, "Generalized ordered logit/ partial proportional odds models for ordinal dependent variables." *The Stata Journal* 6(1):58-82.

Yamamoto, T., Hashiji, J., Shankar, V.N., 2008, "Underreporting in traffic accident data, bias in parameters and the structure of injury severity models", *Accident Analysis and Prevention*, 40(4), 1320-1329.

Table 1: Summary statistics of variables included in the models

Variables	Mean	Std. Dev.	Min	Max
The level of crash severity*	1.113807	0.3555267	1	3
Level of traffic congestion (minutes)	9.4295	19.7400	0	448.33
Traffic flow in vehicles/h	3838.84	1730.82	0	8251
Average traffic speed (km/h)	84.7135	22.1328	5.1	136.08
Radius of road curvature in m	755.22	237.72	24.96	1815.26
Gradient (%)	3.3171	0.9986	0.7	6.1
Number of casualties per crash	1.5515	1.0055	1	17
Speed limit (km/h)	110.3	10.1100	32.2 (20mph)	112.6 (70mph)
Categorical or dummy variables	**Description**			
Number of lanes				
Three-lane (or less)	1=Three-lane or less (count= 2420), 0=otherwise (count=1578)			
Four-lane	1=Four-lane (count=1218), 0=otherwise (count=2780)			
Five-lane (or more)	1=Five-lane or more (count=360), 0=otherwise (count=3638)			
Time of the day	1=peak (count=1574), 0=off-peak (count=2424)			
Day of the week	1=weekdays (count=3137), 0=weekends (count=861)			
Light conditions	1=darkness (count=1328), 0=daylight (count=2670)			
Weather conditions				
Fine	1=Fine (count=3313), 0=otherwise (count=685)			
Raining	1=Raining (count=547), 0=otherwise (count=3451)			
Snowing	1=Snowing (count=13), 0=otherwise (count=3985)			
Others (Fog/mist)	1=others (count=125), 0=otherwise (count=3873)			
Road surface conditions	1=wet road surface (count=1072, dry road surface=2926)			
Number of vehicles involved	1=single-vehicle (count=528), 0=multi-vehicle(3470)			
Year of the crash				
Year 2003	1=Year 2003 (count=956), 0=otherwise (count=3042)			
Year 2004	1=Year 2004 (count=1067), 0=otherwise (count=2931)			
Year 2005	1=Year 2005 (count=1039), 0=otherwise (count=2959)			
Year 2006	1=Year 2004 (count=936), 0=otherwise (count=3062)			

*1=Slight (count=3594), 2=Serious (count=353), 3=Fatal (count=51)

Table 2: Model estimation results for OLOGIT, HCM, GOLOGIT and PC- GOLOGIT

Severity of road crashes		OLOGIT	HCM	GOLOGIT		PC-GOLOGIT		
		Coef	**Coef**	**Threshold between slight and serious injury crashes (y>1)**	**Threshold between serious and fatal crashes (y>2)**	**Coefficients not varying by threshold**	**Threshold between slight and serious injury crashes (y>1)**	**Threshold between serious and fatal crashes (y>2)**
Factors affecting the ordinal categorical choice (X)	Level of traffic congestion (minutes)	-0.0005	-0.0000	-0.0005	-0.0110	-0.0006	-	-
	ln(Traffic flow in veh/hr)	-0.5342***	-0.7124***	-0.510***	-0.771***	-	-0.509***	-0.869***
	ln(Radius of curvature in m)	0.2317*	0.2644*	0.228*	-0.0981*	0.225*	-	-
	Gradient (%)	0.0520	0.0664	0.0519	0.0495	0.0514	-	-
	Number of lanes: three lanes or less	0.45933***	0.5269***	0.469***	-0.1180	0.461***	-	-
	five lanes or higher	0.3302	0.3999	0.3300	0.1950	0.3250	-	-
	Time of the day: peak	-0.0668	-0.0225	-0.0672	-0.3610	-0.0714	-	-
	Day of the week:weekdays	0.2071*	0.3228*	0.202*	-0.0226*	0.2*	-	-
	Lighting conditions: darkness	-0.2074*	-0.2907*	-0.211*	0.336	-0.202*	-	-
	Road surface conditions: wet	-0.3798***	-0.4593***	-0.375***	-0.4420	-0.377***	-	-
	Number of vehicles involved: single	0.6712***	-0.2835	0.661***	0.975***	-	0.658***	1.199***
	Number of casualties	0.3141***	0.2819***	0.308***	0.388***	0.312***	-	-
	Year 2004	-0.3959***	-0.4407***	-0.392**	-0.7710	-0.395***	-	-
	Year 2005	-0.3207***	-0.3414*	-0.323**	-0.3610	-0.323**	-	-
	Year 2006	-0.1848	-0.2547	-0.1970	0.1910	-0.1880	-	-
	Constant	-	-	-0.4670	1.5710	-0.4530	-	-
Factors affecting the error variance	Number of vehicles involved: single	-	0.4630***	-	-	-	-	-
	Number of casualties	-	0.0646***	-	-	-	-	-
Statistics	cut-point1	0.3064**	-0.5852**	-	-	-	-	-
	cut-point2	2.5968***	2.3400***	-	-	-	-	-
	Observations	3837	3837	3837		3837		
	Log-likelihood at convergence	-1306.55	-1299.21	-1294.29		-1300.03		
	Likelihood Ratio (LR) Chi-Square	155.7	170.4	180.20		169.68		
	Degrees of freedom (dof)	15	17	30		17		
	McFadden pseudo Rho-square	0.096	0.099	0.09		0.10		

*** $p<0.01$, ** $p<0.05$, * $p<=0.1$

Table 3: Marginal effects

| Marginal Effects	Explanatory variables	OLOGIT	HCM	PC-GOLOGIT
Slight injury crashes	ln(Traffic flow in veh/hr)	0.0422**	0.0484**	0.0403**
	ln(Radius of curvature in m)	-0.0183*	-0.0180*	-0.0178*
	Number of lanes: three lanes or less	-0.0352**	-0.0347**	-0.0353**
	Day of the week:weekdays	-0.0156*	-0.0206**	-0.0151*
	Lighting conditions: darkness	0.0159*	0.0191*	0.0156*
	Road surface conditions:wet	0.0280**	0.0291**	0.0278**
	Number of vehicles involved:single	-0.0652**	-0.0745**	-0.0638**
	Number of casualties	-0.0248**	-0.0313**	-0.0247**
	Year 2004	0.0291**	0.0280**	0.0291**
	Year 2005	0.0238**	0.0219**	0.0239**
	Year 2006	0.0140	0.0165*	0.0143
Serious injury crashes	ln(Traffic flow in veh/hr)	-0.0372**	-0.0437**	-0.0343**
	ln(Radius of curvature in m)	0.0161*	0.0162*	0.0161
	Number of lanes: three lanes or less	0.0310**	0.0313**	0.0321**
	Day of the week:weekdays	0.0138*	0.0186**	0.0137*
	Lighting conditions: darkness	-0.0141*	-0.0173*	-0.0142*
	Road surface conditions: wet	-0.0247**	-0.0263**	-0.0253**
	Number of vehicles involved: single	0.0570**	0.0474**	0.0496**
	Number of casualties	0.0219**	0.0269**	0.0192
	Year 2004	-0.0257**	-0.0253**	-0.0265**
	Year 2005	-0.0210**	-0.0198**	-0.0218**
	Year 2006	-0.0124	-0.0149*	-0.0130
Fatal crashes	ln(Traffic flow in veh/hr)	-0.00502**	-0.00476**	-0.006**
	ln(Radius of curvature in m)	0.00218*	0.00177*	0.0016
	Number of lanes: three lanes or less	0.00417**	0.00340**	0.00322**
	Day of the week:weekdays	0.0018*	0.002*	0.0014*
	Lighting conditions: darkness	-0.00189*	-0.00187*	-0.00142*
	Road surface conditions: wet	-0.00330**	-0.00283**	-0.00252**
	Number of vehicles involved: single	0.00817**	0.0271**	0.0141**
	Number of casualties	0.0029**	0.0043**	0.0021
	Year 2004	-0.00343**	-0.00272**	-0.00264**
	Year 2005	-0.00280**	-0.00213*	-0.00217**
	Year 2006	-0.0017	-0.0016	-0.0013
	Observations	3837	3837	3837

** p<0.05, * p<0.1
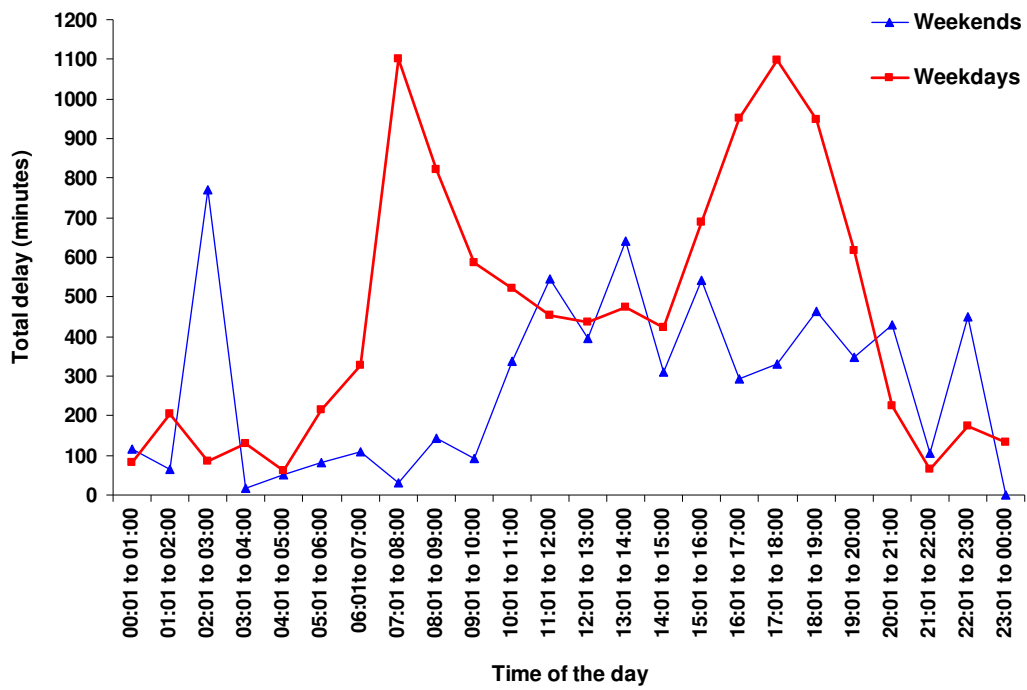
**Figure 1: Average hourly delay (in minutes) just before the crashes happened on M25
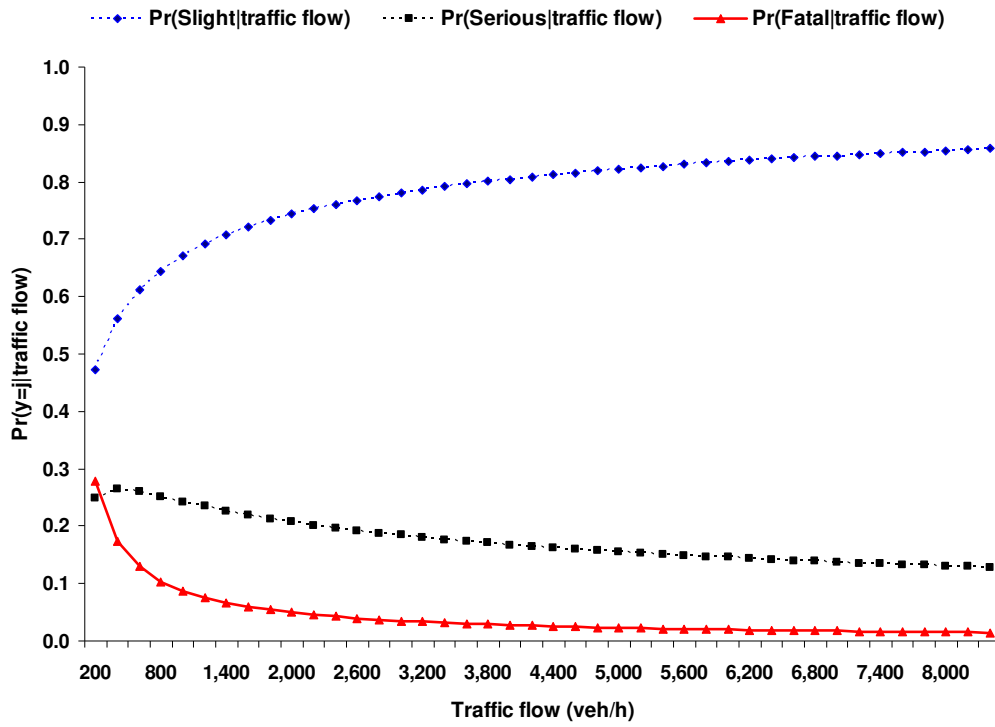(2003 – 2006)**

Figure 2: The predicted probabilities of different categories of road crashes for different values of traffic flow on a 3-lane road (using the PC- GOLOGIT model)
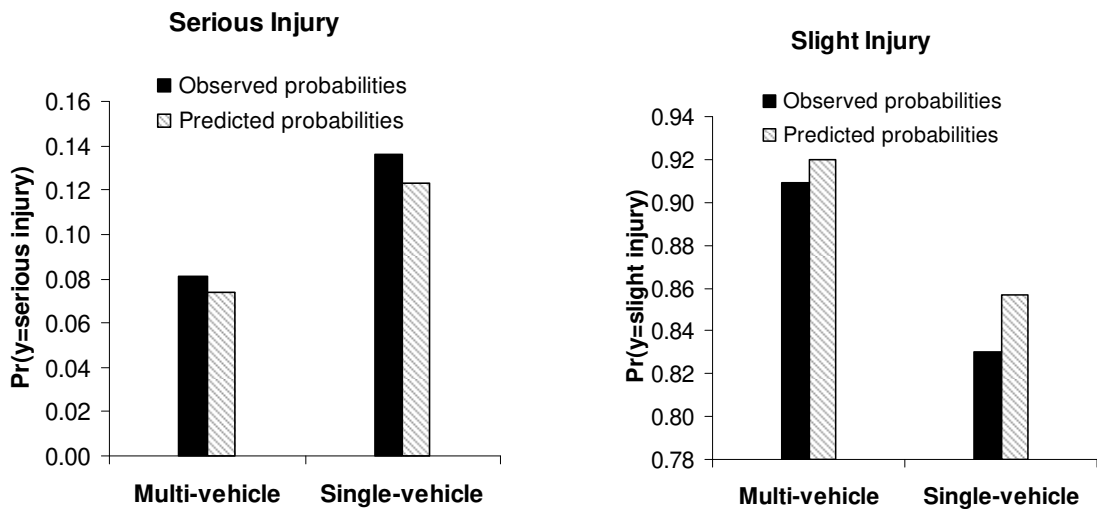
Figure 3: Actual vs predicted probabilities for the dummy variable representing whether a crash is a single-vehicle crash (all other vehicles at their means)
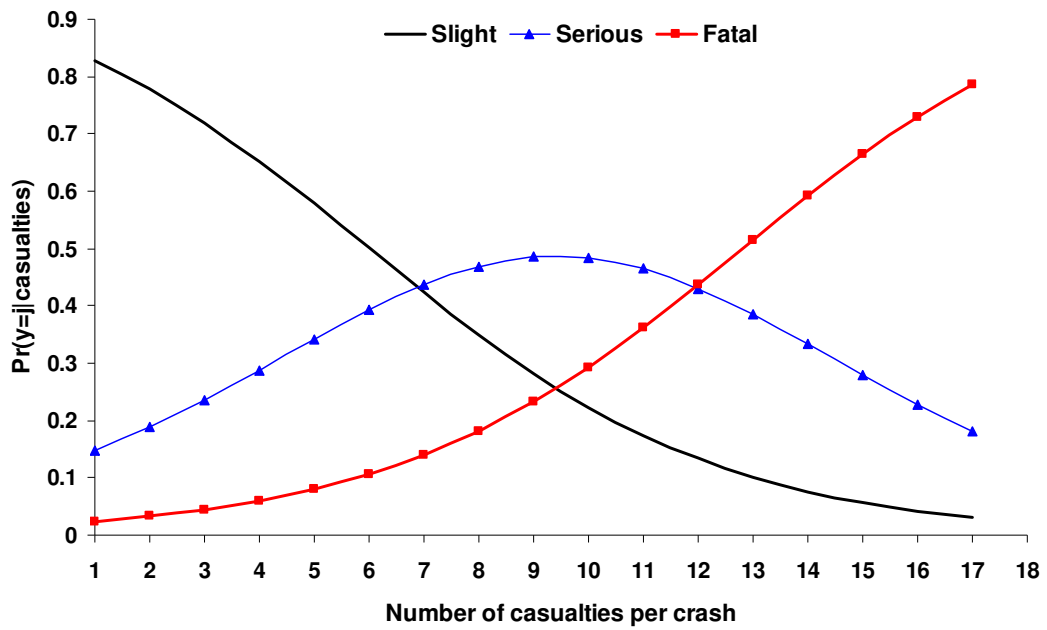
Figure 4: Predicted probabilities for number of casualties per crash (on a three-lane stretch of M25)