ROAD ACCIDENTS AND TRAFFIC FLOWS:

AN ECONOMETRIC INVESTIGATION

Andrew Dickerson, John Peirson and Roger Vickerman

April 1998

Abstract

This paper develops an empirical model of the relationship between road traffic accidents and traffic flows. The analysis focuses on the accident externality which is mainly determined by the difference between the marginal and average risks. The model is estimated using a new dataset which combines hourly London traffic count data from automated vehicle recorders together with police records of road accidents. The accident-flow relationship is seen to vary considerably between different road classes and geographical areas. More importantly, even having controlled for these and other differences, the accident externality is shown to vary significantly with traffic flows. In particular, while the accident externality is typically close to zero for low to moderate traffic flows, it increases substantially at high traffic flows.

JEL Classification: C14, C80, D62, R40

Keywords: Road Traffic Accidents, Traffic Flows, Accident Externalities

Acknowledgements: We would like to thank, without implication, the Department of Transport for giving us access to the automatic count data used in this study. Particular thanks are due to Daniel Aromire for his assistance in the provision and explanation of these data. The accident data were supplied by the ESRC Data Archive at the University of Essex. Neither the original collectors of the accident data nor the Archive bear any responsibility for the analyses or interpretations presented here. This research was funded in part by the TRENEN project of the European Commission's Fourth Framework RTD Programme, DG VII (Project ST-96-SC-116). Finally, we are grateful to Alan Carruth for his useful comments on an earlier draft.

Correspondence address: Dr. John Peirson, Department of Economics, Keynes College, University of Kent at Canterbury, Canterbury, KENT CT2 7NP, UK. email: jdp1@ukc.ac.uk; tel: +44 (0)1227 823328; fax: +44 (0)1227 827850.

ROAD ACCIDENTS AND TRAFFIC FLOWS:

AN ECONOMETRIC INVESTIGATION

1. Introduction

The costs of road accidents are generally regarded as being considerable and they are therefore an important element in the analysis of transport projects and the formation of transport policy (Evans, 1994; Department of Transport, 1996). For the United Kingdom, the annual total costs have been estimated at between £5 billion and £26 billion (see Maddison *et al.* 1996; Pearce, 1993; Fowkes *et al.*, 1990; Hopkins and O'Reilly, 1993; Hansson and Marckham, 1992 and Newbery, 1988). New transport projects and polices affect the number of road accidents at the margin. Thus, it is important to investigate the marginal external accident costs of additional road traffic rather than total or average accident costs.

There are three principal stages in the estimation of the external costs of road accidents (Maddison *et al.*, 1996 and Peirson *et al.*, 1998). The first stage is to identify the functional relationship between accidents and vehicular flows. Vickrey (1968, 1969), Jones-Lee (1990), Newbery (1987, 1988), Vitaliano and Held (1991) and, more recently, Jansson (1994) and Schefer and Rietveld (1997) can be regarded as the significant contributors to the formal modelling of this relationship. Second, the different elements of road accident externalities have to be defined. The studies cited above, together with Hansson and Marckham (1992), Jones-Lee *et al.* (1993) and Persson and Ödegaard (1995), have defined and discussed these externalities in a helpful manner. Finally, values have to be placed on these externalities. The literature on this subject is extensive (see, for example, Jones-Lee *et al.*, 1985, Jones-Lee, 1990 and Jones-Lee *et al.*, 1993). This paper is primarily concerned with the first stage, that of determining the nature of the accident-vehicle flow relationship.

We construct a unique and novel dataset for this study by amalgamating two disparate sources of data. Traffic flow data for the period mid-1993 to end-1995 are taken from the Department of Transport's automated recording devices which are distributed throughout London. These vehicle flows are then matched with police data on all road traffic accidents in the corresponding area during the same period.

The statistical analysis of these data reveals two important findings. Firstly, the accident-flow relationship varies significantly between different road classifications and broad geographical areas. This has an important implication for a number of previous studies which have failed to allow for such heterogeneity. Typically, these studies have concluded that there is a near-proportional relation between accidents and flows at all levels of traffic flow and, thus, there is no apparent accident externality. Such a conclusion may be erroneous since it may result from the aggregation of heterogeneous accident-flow relationships which do not exhibit proportionality.

Secondly, under the simple assumption that occupants of additional vehicles internalise the average risk of an accident, the excess of the marginal accident rate over the average rate determines the magnitude of the road accident externality. The data indicate that the ratio of the marginal accident risk to the average accident risk varies significantly with traffic flow, and, thus, this variation is important. In particular, we show that while the externality is typically close to zero for low to moderate traffic flows, it increases substantially at high traffic flows. That is, the relationship between accidents and traffic flows is non-linear.

The remainder of the paper is organised as follows. Previous studies of the relation between accidents and flows are discussed in Section 2. The data are described in Section 3 and the degree of heterogeneity across road class and geographical area is depicted. In Section 4, we estimate the relationship between road accidents and traffic flow using an appropriate econometric specification. The final section draws some conclusions.

2. Previous Studies of Road Accidents and Traffic Flows

Road users impose accident risks on other road users. Such accident risks may give rise to important externalities, but the relationship between road accidents and traffic flows is not well understood and has been the subject of few empirical studies. Maddison *et al.* (1996, p.122) noted that “given the policy relevance of ... the degree of market failure in road transport, it is unfortunate that not more is known regarding traffic flows and accident rates”. Vitaliano and Held (1991, p.373) commented that “there is a significant gap in our understanding of this important facet of highway economics”. The key economic determinant of road accident externalities is the difference between the marginal and average accident rates. Intuitively, an extra vehicle leads to an additional risk of an accident for all vehicles while the extra vehicle faces only the average risk. The difference between the marginal and average accident rates represents the divergence between the social and private costs associated with the extra vehicle (see also Walters, 1961).

In what follows, we construct an illustrative naïve model of road accidents and traffic flows. This model is then extended to allow for variation in road characteristics and driving behaviour, and is related to the extant empirical studies. Assume that all road users are identical, the road network and its use are homogeneous and all vehicles are driven in an identical manner whatever the level of traffic flow F[1]. The potential number of accidents involving two vehicles[2] is proportional to $F(F\text{-}1)/2 \approx ½F^2$ for reasonable values of F. Let the probability of an accident be λ, which will be a function of road conditions, driving behaviour, road network characteristics etc. Then, the expected total number of accidents, N, is given by:

$$N = ½\lambda F^{2} \tag{1}$$

This is a quadratic function[3] and therefore the marginal and average accident rates are not equal. As noted above, the divergence between the marginal ($m = \lambda F$) and average ($a = ½\lambda F$) number of accidents can be interpreted as the extent of the accident externality.

In general, the degree of the externality can be represented by the ratio of the marginal to average accident rates. For example, equation (1) yields $m/a = 2$, which indicates that the full cost of additional vehicles is imposed on other road users. In comparison, a functional specification which generates $m/a = 1$ suggests that there is no externality and the effects of additional vehicles are effectively being completely internalised by existing road users. Thus, if the marginal and average accident rates are equal, then we can conclude that there is no accident externality. Such a result will only emerge if the number of accidents increases exactly proportionally with traffic flow, $N \propto F$.

One effect of higher levels of traffic are to reduce average speeds and make driving behaviour safer, suggesting that λ is a function of F (see Bailey, 1970 and Peirson *et al.*, 1998 for a more

---

[1] This model is in the spirit of Newbery (1987, 1988), Jansson (1994) and Peirson *et al.* (1998).
[2] Approximately two-thirds of all road accidents involve two vehicles (source: own calculation from Road Traffic Accident Statistics 1992-1995).
[3] Shefer and Rietveld (1997) show that the number of two vehicle accidents, N, will increase quadratically with the number of vehicles, F, for a variety of different shapes of road network.

detailed explanation). The form of this function is important since it indicates the degree to which the externality is internalised by road users, and therefore affects the marginal and average accident rates. It therefore determines the degree of the externality. Thus, for example, if $\lambda \propto 1/F$, then the model in equation (1) yields $m/a = 1$, and hence there is no accident externality[4].

It is important to note that simple exponential accident functions of the form

$$N = \lambda F^{\alpha} \tag{2}$$

imply a fixed ratio between the marginal and average accident rates of α for all levels of F. The estimation of accident-flow relationships using equation (2) is therefore very restrictive. Moreover, such relationships are forced to pass through the origin, whereas the (expected) number of accidents may be zero at some positive flow.

As noted in the introduction, there has only been a limited number of studies that have modelled road accidents and the consequent externalities. Vickrey (1968) inferred from evidence on Californian freeway driving in the early 1960s that the marginal accident rate was 1.5 times the average rate. However, Vickrey's work has been criticised for his use of simple arithmetical calculations rather than any formal estimation of an econometric model (Vitaliano and Held, 1991).

Vitaliano and Held (1991) is one of the few studies that econometrically estimates the relationship between road accidents and traffic flows. They investigated annual road accident and traffic flows on urban and rural roads in New York State in 1985. Control variables

---

[4] The marginal and average accident rates are both equal to λ.

included the number of lanes, annual snowfall, speed limit, length of road segment studied, etc, although the slope of the accident-flow relationship was assumed to be constant across all types of roads. They conclude that the relationship between accidents and flows is nearly proportional and thus the external accident effect is close to zero, although when the 19 highest flow observations with more than 50,000 vehicles per day were separately analysed, there is some evidence of a small negative external effect[5]. However, sample separation by vehicular flow rather than road characteristics would appear to be inappropriate, and it is not possible from their results to determine whether the difference is significantly different from the estimated external effect for the remaining 380 observations.

The Department of Transport (1996) COBA manual assumes that, on average, the number of accidents on road links between junctions is proportional to traffic flow. This requires that drivers adjust their driving behaviour to achieve a fixed risk of an accident per vehicle-kilometre travelled at differing traffic flows. For traffic flows at junctions, the accident-flow relationship differs according to the type of junction. Exponential functions of the form of equation (2) were estimated for different types of junction. The estimated α's range between 0.44 and 1.77, but, as noted above, the functional form of equation (2) is very restrictive.

Rather than attempting to estimate the accident-flow relationship, the other notable studies have typically made assumptions about the accident-flow relationship or have drawn on previous research. Citing the above evidence and the US Federal Highway Cost Allocation Study (US Federal Highway Administration, 1982), Newbery (1987) argued that the number of road accidents is proportional to flow. Thus, the average accident rate per vehicle-kilometre

---

[5] Note that this subsample represents only 5% of their 399 sites.

travelled is constant and is equal to the marginal accident rate. By contrast, in a later paper, Newbery (1988) took the ratio of the marginal to average accident rate to be 1.25 - an average of the Department of Transport's (1981) and Vickrey's (1968, 1969) views about the relation between road accidents and traffic flow. Jones-Lee (1990) considered the value of the externality from road accidents where the accident rate per vehicle-kilometre is fixed, but the pedestrian accident rate is proportional to the distance travelled by vehicles. Pearce (1993) argued that a minimum estimate of external costs should include the cost of pedestrian and cyclist fatalities and injuries. Though Pearce discussed other possible external accident costs, no formal model of road accident externalities was developed. Maddison *et al.* (1996) used a similar approach.

There are two important criticisms of the empirical studies discussed above. First, different road types are not distinguished in the analysis. It seems inappropriate to assume that drivers' behaviour, and thus accident rates, do not differ between, say, motorways and small side streets. While some of the differences are undoubtedly random (see Fridstrøm *et al.*, 1995), some are likely to be systematic. In addition, roads have particular design characteristics which determine their ability to carry different levels of traffic flows safely. Thus, it is difficult to believe that the relationship between accidents and flows is the same for all road types. One solution is to specify dummy variables to capture any fixed effects deriving from differences in roads and their impact on behaviour. However, such specifications cannot capture any differences in the *slopes* of the relationship between accidents and flows, and it is these which are of greatest importance given that transport policy is only ever likely to affect traffic flows and accident rates at the margin.

Failure to fully control for the effects of differences in road types can also seriously bias estimates of the underlying accident-flow relationship. Intuitively, observations from low-flow roads will typically be associated with a small number of accidents and thus will be close to the origin, whilst observations from high-flow roads with a greater number of accidents will be some distance away from the origin. Simple aggregate regression techniques which do not distinguish these two road types are therefore likely to produce near-proportional relationships between accidents and traffic flows. This will give near-equality between the marginal and average accident risk rates and will lead to the spurious conclusion that there is no significant accident externality. To avoid this potential bias, empirical models should carefully distinguish the accident-flow relationships for different types of road.

The second point to be noted concerning the previous empirical studies is that the accident-flow relationship should be flexibly specified. In particular, the functional form selected should allow for the possibility that the marginal and average accident rates may vary with traffic flow, rather than imposing the constraint that the marginal to average accident rate ratio is the same at all levels of traffic flow.

3. Data Description and Preliminary Analysis

3.1 Data Description

The data used in this study derive from two sources. First, data on hourly vehicular flows have been provided by the Department of Transport from their automatic count data sites in London for the period July 1993 to December 1995 inclusive. Data are available from a total of 54 sites which are geographically spread fairly evenly throughout the London boroughs. At each site, an automatic data recorder counts each vehicle as it passes and the hourly totals are then transmitted to a central recording device. The data are fairly complete in that there are very few breakdowns in the recording apparatus, but where these do occur, the Department of Transport imputes values from previous flows at the site. Given that we have hourly counts for 30 months, measured in both directions, for over 50 sites, our data on flows comprise in excess of two million hourly observations.

Data on accidents are taken from the Road Accident Data deposited at the ESRC Data Archive by the Department of Transport Road Accident Statistics Branch. These data record every accident and its circumstances occurring on the public highway where at least one road vehicle and one casualty are involved. Details include the date and time (to the nearest hour), the local authority/police force area, road class (although not exact location), the number and type(s) of vehicle(s) involved and some information on the number and severity of casualties[6]. From these data, we select only those accidents that occurred within the London area and during the period covered by the automatic count data. This yields a total of 96,440 accidents.

In order to investigate the relationship between accidents and flows, we need to amalgamate these two sources of data. While there are a number of ways in which this can be accomplished, given the discussion above and the results of previous studies, we choose to use year-month-hour as our basic time dimension on which to merge the two datasets. That is, each observation in the combined dataset comprises the total number of accidents, A_{hmy}, that occurred during a particular hour of the day, h, in a given month, m, and year, y, together with

[6] Certain information, such as the exact location of the accident, is missing, so as to maintain the confidentiality of the victims.

the average vehicular flow, F_{hmy}, in that year-month-hour from the automatic count data[7]. We investigate the sensitivity of our results to this particular choice of unit of observation.

3.2 Preliminary Analysis

The review of the literature on the relationship between accidents and traffic flows in Section 2 suggests that econometric specifications should be capable of allowing for differences between road types and for variations in the ratio of the marginal to average risk rates. Some preliminary evidence on the necessity of the former can be gleaned from an examination of Figure 1. This plots the total number of accidents against the average vehicular flow, where both total accidents, A_{hmy}, and average flows, F_{hmy}, are calculated separately for four different road classifications (A-roads, B-roads, C-roads and Unclassified roads) and are also dichotomised by location into Inner London and Outer London areas[8]. We choose to disaggregate on these two dimensions of the data since accident rates and traffic flows exhibit considerable heterogeneity across both road classification and area as can be seen in Table 1.

<< Table 1 about here >>

<< Figure 1 about here >>

Figure 1 reveals that there are distinct "clusters" of observations. These can clearly be seen to correspond to the eight different road-area classifications as shown in Figure 2, which also

---

[7] Note that we cannot match the two data sources by location since accidents are only located on a road *class* (A-roads, B-roads, C-roads or Unclassified roads) within a borough rather than at an exact location. Moreover, even if the exact geographical location of each accident was known, they could only ever be imperfectly matched by location (presumably using their geographical proximity to one of the limited number of automatic count sites).

[8] Thus there are 2.5 (years) x 12 (months) x 24 (hours) x 4 (roads) x 2 (areas) = 5760 accident-flow observations in Figure 1. The definitions used for Inner London and Outer London are in the Appendix.

reveals significant differences in the marginal accident-flow relationship according to different road types and area[9]. Estimating an aggregate accident-flow relationships for the data in Figure 1 can obviously yield misleading conclusions given the heterogeneity evident in Figure 2. Any aggregate relationship looks to be near-proportional, whereas the disaggregated relationships would appear to be: (i) different from one another; (ii) non-proportional; and (iii) (perhaps) non-linear. To illustrate the problem, the fitted values from both linear and cubic polynomial regressions are super-imposed on Figure 1[10]. Both yield functional relationships which indicate that the average and marginal accident rates are very similar and fairly constant over the range of data. Of course, this exactly mirrors the results of previous studies. Table 2 reports the results of polynomial regressions of this kind. Evaluated at the mean accident and flow rates, the ratio of marginal to average accident rates are 1.042, 0.834 and 0.861 for the linear, quadratic and cubic regressions respectively, from which one might easily conclude that the accident externality is negligible (or even slightly positive)[11]. These results therefore encompass previous studies which have used aggregated data.

<< Figure 2 about here >>

<< Table 2 about here >>

It is useful to demonstrate formally that the accident-flow relationship differs significantly between the various road and area classifications depicted in Figure 2. This can be

---

[9] Compare, for example, Inner London A-roads (top left) and Outer London Unclassified roads (bottom right).

[10] The econometric methodology should really take account of the fact that the number of accidents is bounded below by zero (effectively we are estimating the realisation of the accident propensity and should therefore use a Tobit model). However, despite the high level of disaggregation used here, only 7% our observations have zero accidents and hence we restrict ourselves to classical least squares estimation.

[11] Note that only the cubic polynomial passes the RESET misspecification test.

accomplished in a number of ways. First, note that simple polynomial regressions as in Table 2 with intercept dummy variables capturing the road-area combination are always rejected against a more general specification in which the slope as well as the intercept of the accident-flow relationship is allowed to differ between the different roads and areas[12]. Table 3 reports the results of these regressions for the quadratic polynomial. The dummy variable specification in column 1 in which we allow only the intercept to differ between the $4 \times 2 = 8$ road-area combinations fails the RESET test, while the fully interactive specification (which is equivalent to estimating the quadratic relationship separately for all 8 road-area combinations) satisfies this criterion. Moreover, the interactive effects are individually and jointly significant; a test for the equality of the linear F terms across all 8 road-area combinations yields a F-statistic of $F(7,5736) = 19.98$ [p=0.00], while the equality of the F^2 terms is also comprehensively rejected ($F(7,5736) = 32.15$ [p=0.00]). Additionally, these interactive terms also differ from each other - that is, not only are the accident-flow relationships jointly different from that for A-roads in Inner London (the base category in Table 3), but pairwise tests reveal that B-roads in Inner London differ from B-roads in Outer London etc, and that Inner London roads are collectively different from Outer London roads. These differences are presumably because of the rather differ composition and densities of traffic flows and the design and structure of the different types of roads in Inner and Outer London. These results thus demonstrate that the accident-flow relationship differs substantively and statistically significantly between the 8 different road-area classifications.

<< Table 3 about here >>

[12] Separately distinguishing roads and areas is not adequate (statistically) since the data reveal that it is the *combination* of these two characteristics which is important.

4. Specification of the Accident-Flow Relationship

We could simply use the results in the second column of Table 3 to calculate the marginal and average accident rates at, for example, the average traffic flow in each road-area combination. However, such a specification has two undesirable properties. First, it is well known that polynomial specifications are very sensitive to observations at the ends of the data range (see, for example, Royston and Altman, 1994a). Even fractional polynomial[13] estimation is not immune to such end-effect and/or outliers (Royston and Altman, 1994b). Secondly, the specification of simple polynomials does not allow the relationship between marginal and average accident rates to vary in an unrestricted manner with traffic flow as discussed in Section 2 above. In particular, polynomials force the marginal accident rate to be either increasing or decreasing between turning points. For example, they do not allow a constant marginal rate over some range of traffic flow.

Therefore, in order to facilitate the identification of an appropriate functional specification, we first employ a non-parametric (kernel) estimator of the relationship between accidents and traffic flows (Härdle, 1990). Effectively, this simply fits a “smooth” curve through the observations, where the degree of smoothness depends on the particular kernel method chosen. While the choice of smoothing algorithm is essentially arbitrary, we choose a method that is robust to outliers but has “locality” - that is, it tends to follow the data closely. Such a method is provided by Cleveland (1979)[14]. The kernel estimates of the relationship between

---

[13] That is, non-integer exponents in the polynomial function.

[14] This particular smoother is based on the fitted values from a locally weighted regression. A centred subset of the data is used in a (polynomial) regression, with each observation in this subset weighted according to the inverse of its distance from the central point. The estimated regression is then used to predict the smoothed value at the central point only and the procedure is repeated for each observation in the dataset (see Härdle, 1990, for further details of the algorithm).

the number of accidents and traffic flows for each of the road-area classifications are shown in Figure 3. The bold line is the kernel estimate while the light lines are 95% confidence bounds on this curve. As can be seen, in almost all cases, the relationship looks near-proportional for low to moderate traffic flows (although of course the marginal rates differ substantially given the different scales on the axes). However, at high traffic flows, there is evidence to suggest that the accident-flow relationship changes[15]. Any parametric specification therefore needs to be able to accommodate this observation. In particular, smooth (i.e. continuously differentiable) polynomial functions may not be sufficiently flexible to capture this characteristic of the data.

<< Figure 3 about here >>

We therefore need to select a flexible specification in which the relation between the marginal and average accident rates is allowed to differ across traffic flows. A simple, semi-parametric, specification is to use piecewise linear splines. These maintain the continuity of the accident-flow functional relationship and also provide the necessary flexibility. Moreover, they permit the average and marginal rates to be directly calculated. The visual inspection of the data provided by the kernel estimates, together with some pre-testing, indicates that a maximum of three spline segments (i.e. two nodes) is sufficient for all road-area combinations.

The results from estimating spline functions are illustrated in Figure 4. All 8 estimated regressions pass the RESET test for misspecification, and thus can be regarded as econometrically satisfactory. We scale the graphs for ease of reference, but it should be noted

[15] The use of a narrower bandwidth increases the slope of the functions at high flows. Thus, the computed accident-flow relationship deviates even further from proportionality.

that the slopes of the spline functions are generally significantly different between the road-area classes. As can be seen, the third segment in each road-area classification typically has a slope greater than that of the first two segments. Thus at high vehicle flows, the marginal accident rate is greater than the average accident rate. The exception is C-roads in Inner London where the third segment is downward sloping. This is purely a site-specific problem in that we only have one representative site for this particular road-area combination. Hence the results for this road-area classification should be treated with some caution[16].

<< Figure 4 about here >>

In order to better evaluate the differences between the marginal and average accident rates at different flows, Table 4 tabulates the marginal accident rate, *m*, (the slope of each spline segment), the average accident rate, *a*, (evaluated at the mid-point of each spline segment), and the ratio of the marginal to average accident rate, *m*/*a*, for each road-area combination. That is, we present estimates of *m*/*a* at the first, third and fifth sextiles of the data.

<< Table 4 about here >>

---

[16] The distribution of the 54 sites between road types and area is as follows:

	road classification			
area	A-roads	B-roads	C-roads	U-roads
Inner	13	2	1	7
Outer	21	2	2	6

Hence, the results for A-roads and Unclassified-roads are based on traffic flows averaged across rather more sites than those for B-roads and C-roads and thus can perhaps be regarded with greater confidence, although the Department of Transport chooses all the sites for its automatic recorders to be fairly representative of traffic flows throughout London. An additional problem is that there is some evidence that there are inconsistencies in the recording of accidents on C-roads and U-roads (particularly in urban areas as in this study), perhaps because the exact classification of any road is largely unknown to the police when they complete the Road Accident Report Form (STATS 19) (for example, road classifications are not recorded on OS maps; see Lupton *et al.*, 1997).

The results confirm what is apparent in Figure 4. For low to moderate traffic flows on each road-area classification, the marginal and average accident rates are approximately equal, and hence the ratio of the margin al to average rate is close to unity. Thus, at low to moderate flows, there would appear to be no significant accident externality, although the average and marginal rates can be seen to differ substantially between different road classes and London area. However, at high flows, the *m*/*a* ratio is considerably greater than unity for almost all road-area classifications, indicating that there are substantial negative accident externalities at high traffic flow rates. It is only by allowing the *m*/*a* ratio to differ by traffic flow that this relationship becomes apparent. In comparison, a similar 3-segment piecewise linear spline function applied to the aggregate data depicted in Figure 1 actually indicates a *falling m*/*a* ratio at high flows (mirroring the shape of the cubic polynomial illustrated in Figure 1). This confirms the danger in using aggregated accident-flow relationship to estimate the degree of accident externality.

Finally, we assess the robustness of our results to the aggregation methodology that we have selected. Clearly there are differences in traffic flows by time-of-day, and one might suspect there to be differences in accident rates according to the season of the year (due to driving conditions and daylight hours). If these differences systematically impinge upon accidents and traffic flows, then they represent omitted factors from our analysis. However, if they are completely stochastic influences, they will only affect the error variances. In order to assess the importance of time-of-day and seasonal effects, we re-estimate the spline functions using a fixed-effects panel estimator[17]. Our fixed-effects are all 20 combinations of 5 time-of-day

[17] In fact, the (more efficient) random effects specification is not rejected for many of the road-area classifications, although the coefficients on the variables of interest are little changed from the fixed-effects estimates.

periods (am-peak, inter-peak, pm-peak, evening and night) and 4 seasons (spring, summer, autumn and winter). The results are qualitatively similar to those in Table 4. In particular, the marginal to average risk ratio is increasing in traffic flows, and at high flows, there is a clear negative accident externality with the *m*/*a* ratio in excess of unity[18].

5. Conclusions

Knowledge of the relationship between road accidents and traffic flows is crucial to the estimation of the magnitude of road accident externalities. In spite of its importance, there has been little econometric investigation of this relationship. Most previous research has either found a near proportional relationship between accidents and flows or assumed such a relationship. This implies a zero or very low road accident externality.

Utilising a unique dataset, this paper finds that there are important differences in the accident-flow relationships between different types of roads and geographical areas. Our results indicate that while there is a near-proportional relationship at low to moderate traffic flows, the marginal accident rate rises substantially above the average rate at higher traffic flows. This implies that there is a substantial negative road accident externality at high traffic flows. This result is robust to specifying time-of-day and seasonal effects, and kernel estimators with

[18] The key difference in the panel estimates is that the estimated *m*/*a* ratios are typically substantially less than unity at low traffic flow rates. This would appear to be due to the fact that low flows are almost exclusively associated with the night period and, thus, the lower accident rates at low flows are more properly due to a fixed night effect rather than being attributable to low traffic flows *per se*. Hence the slope of the estimated accident-flow relationship which ignores this fixed effect (as presented in Figure 4 and Table 4) is biased upwards at low flows. These panel results thus suggest that the difference in the magnitude of the externality between low and high flows is even greater than that reported in Table 4.

narrower bandwidths and spline functions with more segments and nodes. Moreover, the estimated differences in externalities at high and low flows are larger in such specifications.

These results have important implications for transport policy. Such policy should take account of the fact that the magnitude of road accident externalities varies between road classifications and geographical areas and changes with traffic flows. In particular, policies which attempt to internalise accident externalities, for example through road pricing, need to be carefully devised (see CEC, 1996). Our results parallel the findings on congestion externalities in that only at high traffic flows is there a large external accident cost (see Maddison *et al.*, 1996; Newbery, 1988).

It should be noted that our analysis does not distinguish between the severity of road accidents. As traffic flows increase, the distribution of accidents types between slight, severe and fatal accidents may shift towards less serious accidents. Thus, though the marginal accident rate may be increasing at high flows, the marginal accident *cost* may be constant or even decreasing. We intend to investigate this possibility in future work.

REFERENCES

Bailey, M. (1970), Comment on Vickrey's paper, in Margolis, J. (ed.) *Analysis of Public Output*, NBER, New York, pp.336-338.

CEC (1996), *Towards fair and efficient pricing in transport: policy options for internalising the external costs of transport in the European Union*, COM (95) 691 Final Report, Brussels.

Cleveland, W.S. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots", *Journal of the American Statistical Association*, 74, pp.829-836.

Department of Transport (1981 and revisions), *COBA 9*, Department of Transport, London.

Department of Transport (1996), *COBA 10*, Department of Transport, London.

Department of Transport Road Accident Statistics Branch *Road Accident Data* 1992-1995, [computer file] Colchester, Data Archive.

Evans, A. (1994), Editorial on transport safety, *Journal of Transport Economics and Policy*, XXIIX, pp.3-6.

Fowkes, A.S., Nash, C.A. and Tweddle, G. (1990), "The Track and External Costs of Road Transport", ITS Working Paper 312, University of Leeds.

Fridstrøm, L., Ifver, J., Ingebrigtsen, S., Kulmala, R. and Thomsen, L.K. (1995), "Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts", *Accident Analysis and Prevention*, 27, pp.1-20.

Hansson, L. and Marckham, J. (1992), *Internalisation of External Effects in Transportation*, UIC, Paris.

Härdle, W. (1990), *Applied Nonparametric Regression*, CUP, Cambridge.

Hopkins, M. and O'Reilly, D. (1993), "Revaluation of the Cost of Road Accident Casualties: 1992 revision", TRL RR378, Transport Research Laboratory, Crowthorne.

Jansson, J.O., (1994), "Accident Externality Charges", *Journal of Transport Economics and Policy*, XXIIX, pp.31-43.

Jones-Lee, M.W. (1990), "The Value of Transport Safety", *Oxford Review of Economic Policy*, 6, pp.39-60.

Jones-Lee, M.W., Hammerton, M. and Philips, P.R. (1985), "The Value of Safety: Results of a National Sample Survey", *Economic Journal*, 95, pp.49-72.

Jones-Lee, M.W., Loomes, G., O'Reilly, D. and Philips, P. (1993), "The Value of Preventing Non-Fatal Road Injuries: Findings of a Willingness-To-Pay National Sample Survey", TRL Working Paper WP/SRC/2.

Lupton, K., Jarrett, D.F. and Wright, C.C. (1997), “The Consistency of Road Accident Variables in Great Britain: 1995”, Technical Report No: 1997/6, Transport Management Research Centre, Middlesex University.

Maddison, D., Pearce, D., Johansson, O., Calthrop, E., Litman, T., and Verhoef, E., (1996), *Blueprint 5: The True Costs of Road Transport*, Earthscan, London.

Newbery, D. (1987), “Road User Charges in Britain”, Centre for Economic Policy Research, Discussion Paper 174, April 1987.

Newbery, D. (1988), “Road User Charges in Britain”, *Economic Journal*, 98, pp.161-176.

Pearce, D.W. (1993), *Blueprint 3: Measuring Sustainable Development*, Earthscan, London.

Peirson, J., Skinner, I. and Vickerman, R. (1998), “The microeconomic analysis of the external costs of road accidents”, *Economica*, forthcoming.

Persson, U. and Ödegaard, K. (1995), “External cost estimates of road traffic accidents: an international comparison”, *Journal of Transport Economics and Policy*, XXIX, pp.291-304.

Ramsey, J.B. (1969), “Tests for specification errors in classic least squares regression analysis”, *Journal of the Royal Statistical Society Series B*, 31, pp.350-371.

Royston, P. and Altman, D.G. (1994a), “Regression using fractional polynomials of continuous: parsimonious parametric modelling”, *Applied Statistics*, 43, pp.429-467.

Royston, P. and Altman, D.G. (1994b), “Using fractional polynomials to model curved regression relationships”, *Stata Technical Bulletin*, 21, pp.11-23.

Shefer, D. and Rietveld, P. (1997), “Congestion and safety on highways: towards an analytical model”, *Urban Studies*, 34, pp.679-692.

US Federal Highway Administration (1982), *Final Report on the Federal Highway Cost Allocation Study*, USGPO, Washington.

Vickrey, W. (1968), “Automobile Accidents, Tort Law, Externalities and Insurance: An Economist’s Critique”, *Journal of Contemporary Law and Problems*, pp.464-484.

Vickrey, W. (1969), “Congestion Theory and Transport Investment”, *American Economic Review*, 59, pp.251-260.

Vitaliano, D.F. and Held, J. (1991), “Road Accident External Effects: An Empirical Assessment”, *Applied Economics*, 23, pp.373-378.

Walters, A.A. (1961), “The Theory and Measurement of Private and Social Cost of Highway Congestion”, *Econometrica*, 29, pp.676-699.

TABLE 1

Accident Rates and Traffic Flows by Area and Road Classification

Area	Road Classification							
	A-roads		B-roads		C-roads		U-roads	
Inner London	*34.11*	466.80	*4.56*	181.55	*4.46*	69.07	*6.33*	79.83
Outer London	*50.12*	721.65	*7.61*	253.71	*10.36*	214.42	*16.07*	78.62

Notes:
1. The *average number of accidents is in italics* and average traffic flows are underlined.

TABLE 2

Estimates of the Aggregate Accident-Flow Relationship

	Dependent Variable: A_{hmy}					
	linear		**quadratic**		**cubic**	
constant	-0.7017	(0.1484)	2.4642	(0.1660)	4.3328	(0.1720)
F	0.0674	(0.0006)	0.0388	(0.0015)	0.0081	(0.0028)
F^2 x10^{-3}	-	-	0.0293	(0.0016)	0.1127	(0.0085)
F^3 x10^{-6}	-	-	-	-	-0.0529	(0.0058)
	Diagnostics					
R^2	0.7956		0.8127		0.8173	
RESET	22.92		-3.00		0.33	
obs	5760		5760		5760	
m/a	1.042		0.834		0.861	

Notes:
1. Huber-White heteroskedastic-consistent standard errors in parentheses.
2. RESET is a variant of Ramsey's (1969) general misspecification test. It is distributed as N(0,1) under the null hypothesis of no misspecification; 5% critical values ±1.96.
3. m/a is the ratio of the marginal to average accident rate evaluated at the mean traffic flow.

TABLE 3

Distinguishing Road and Area Classifications in the Accident-Flow Relationship

	Dependent Variable: A_{hmy}			
	dummy variable specification		**fully interactive specification**	
constant	6.3766	(0.4384)	6.6620	(0.8400)
roadB×inner	-11.4972	(0.3928)	-6.3050	(0.8647)
roadC×inner	-5.4231	(0.4103)	-6.5262	(0.8470)
roadU×inner	-4.1010	(0.4186)	-6.4253	(0.8621)
roadA×outer	-4.3962	(0.6110)	-2.3965	(1.1280)
roadB×outer	-12.8405	(0.4202)	-6.3127	(0.8575)
roadC×outer	-7.6775	(0.4043)	-5.7169	(0.8677)
roadU×outer	5.6920	(0.5587)	-5.2069	(0.8813)
F	0.0488	(0.0015)	-0.0118	(0.0066)
roadB×inner×F			0.0285	(0.0079)
roadC×inner×F			0.0958	(0.0092)
roadU×inner×F			0.0565	(0.0128)
roadA×outer×F			0.0563	(0.0084)
roadB×outer×F			0.0421	(0.0075)
roadC×outer×F			0.0384	(0.0084)
roadU×outer×F			0.0994	(0.0182)
$F^2 \times 10^{-3}$	0.0190	(0.0017)	0.1269	(0.0089)
roadB×inner×$F^2 \times 10^{-3}$			-0.0996	(0.0167)
roadC×inner×$F^2 \times 10^{-3}$			-0.3373	(0.0497)
roadU×inner×$F^2 \times 10^{-3}$			0.1699	(0.0824)
roadA×outer×$F^2 \times 10^{-3}$			-0.1066	(0.0099)
roadB×outer×$F^2 \times 10^{-3}$			-0.1317	(0.0120)
roadC×outer×$F^2 \times 10^{-3}$			-0.0676	(0.0175)
roadD×outer×$F^2 \times 10^{-3}$			0.7608	(0.1279)
	Diagnostics			
R^2	0.8812		0.9214	
RESET	5.079		0.956	
obs	5760		5760	

Notes:
1. See Table 2.
2. roadJ, J=A,B,C,U are dummy variables denoting the road classification, while inner and outer are dummy variables denoting the London area.
3. The omitted (base) category is A-class roads in Inner London.

TABLE 4

Estimates of the Marginal and Average Accident Risk

			Spline Segment		
road	**area**	**risk**	**(1) low flow**	**(2) moderate**	**(3) high flow**
A	inner	*m*	0.0540	0.0648	0.1674
		a	0.0549	0.0572	0.0773
		m/a	0.9845	1.1325	2.1654
A	outer	*m*	0.0686	0.0521	0.1134
		a	0.0721	0.0650	0.0717
		m/a	0.9521	0.8023	1.5820
B	inner	*m*	0.0253	0.0198	0.0400
		a	0.0252	0.0237	0.0260
		m/a	1.0051	0.8374	1.5405
B	outer	*m*	0.0315	0.0257	0.0280
		a	0.0343	0.0307	0.0292
		m/a	0.9178	0.8378	0.9585
C	inner	*m*	0.0595	0.0572	-0.0741
		a	0.0717	0.0629	0.0347
		m/a	0.8309	0.9089	-2.1378
C	outer	*m*	0.0417	0.0463	0.0823
		a	0.0470	0.0451	0.0525
		m/a	0.8868	1.0268	1.5670
U	inner	*m*	0.0663	0.0896	0.1267
		a	0.0652	0.0730	0.0863
		m/a	1.0169	1.2284	1.4686
U	outer	*m*	0.1445	0.2462	0.3555
		a	0.1731	0.1873	0.2316
		m/a	0.8346	1.3144	1.5353

Notes:

1. *m* denotes marginal risk; *a* denotes average risk; *m*/*a* is the ratio of the marginal to average risk.
2. Low, moderate and high flows correspond, respectively, to the first, third and fifth sextiles of the data.

FIGURE 1

Accidents *vs* Flows: Aggregated

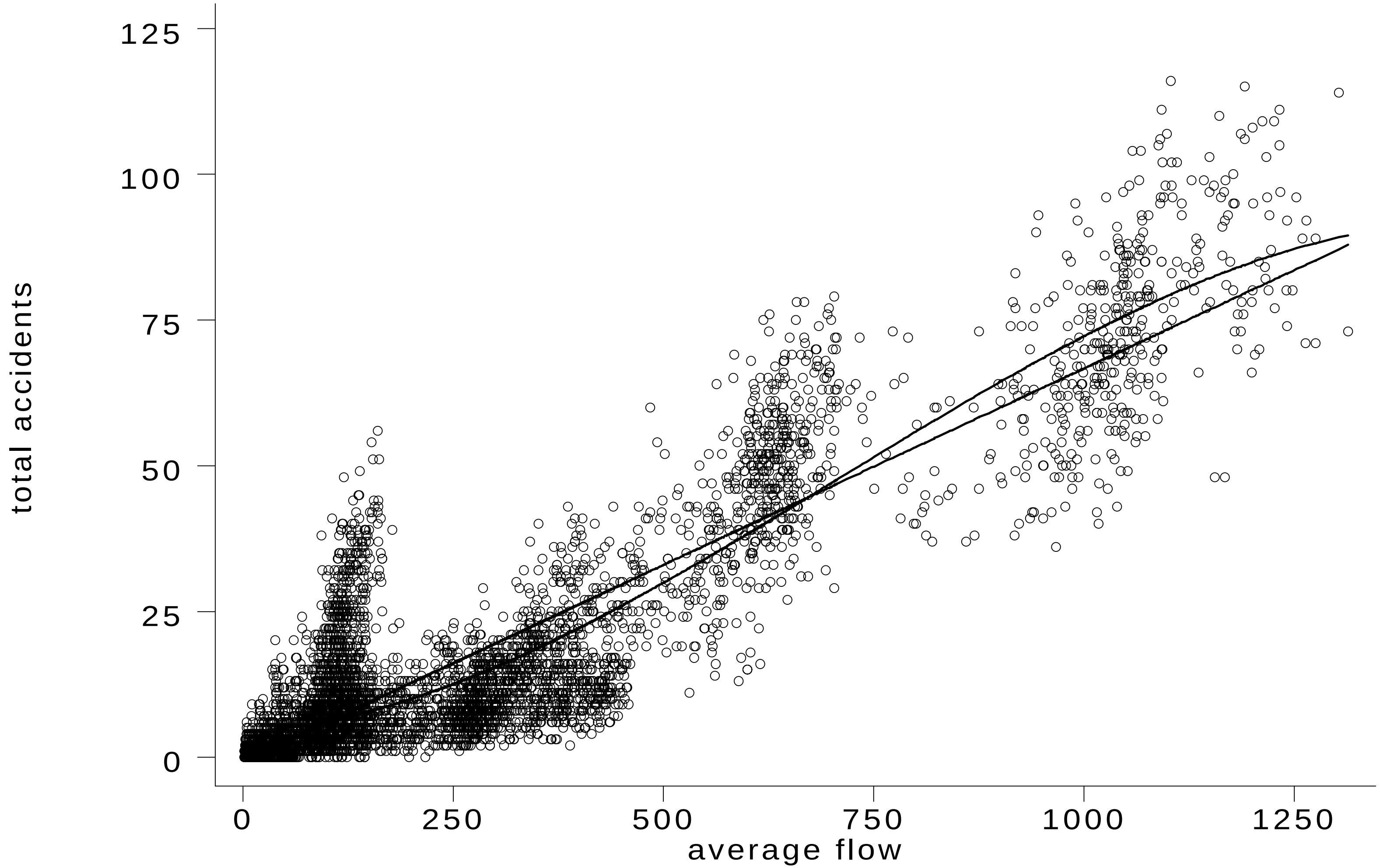


FIGURE 2

Accidents *vs* Flows: Disaggregated

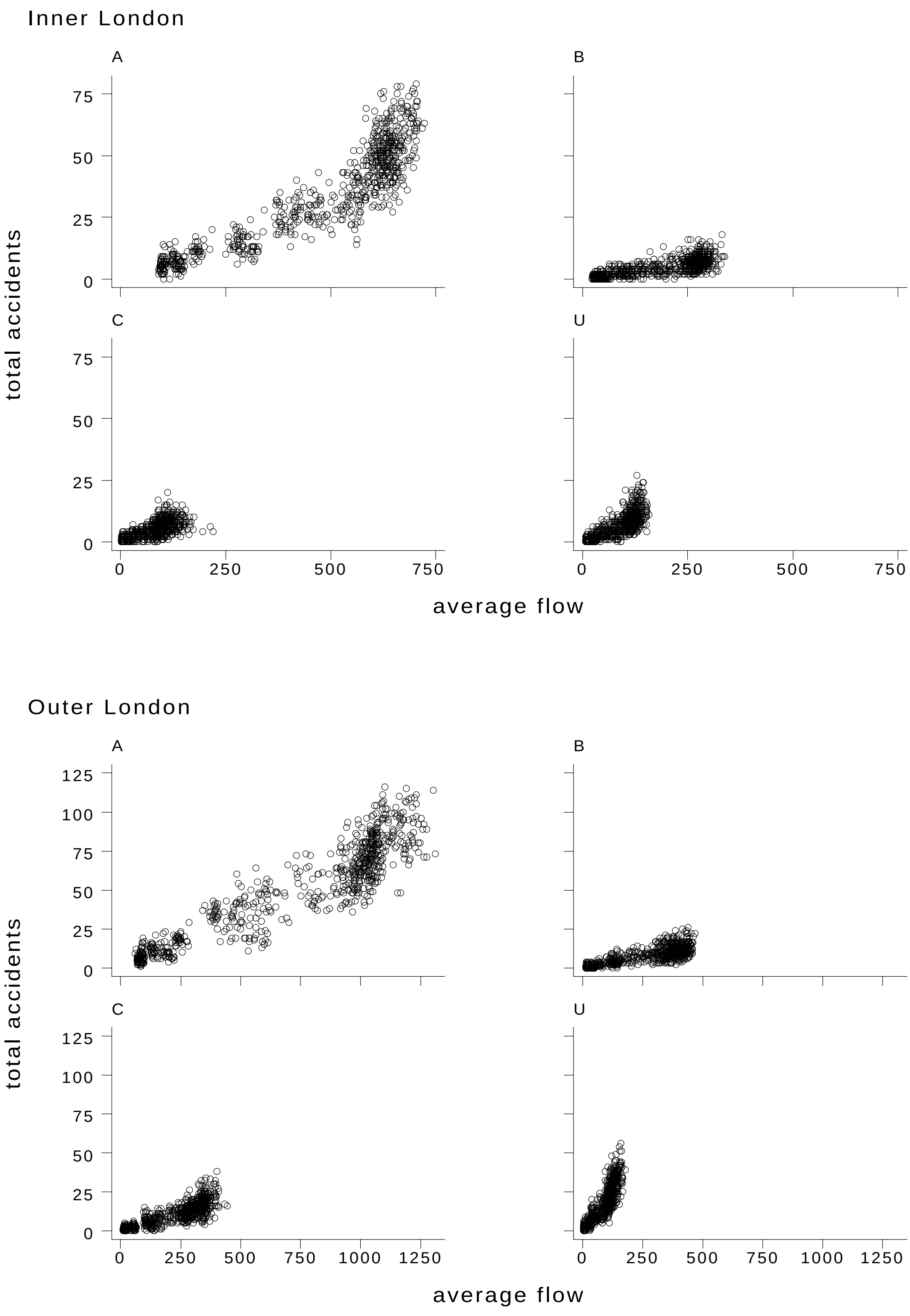


FIGURE 3

Accidents *vs* Flows: Non-parametric Kernel Estimates

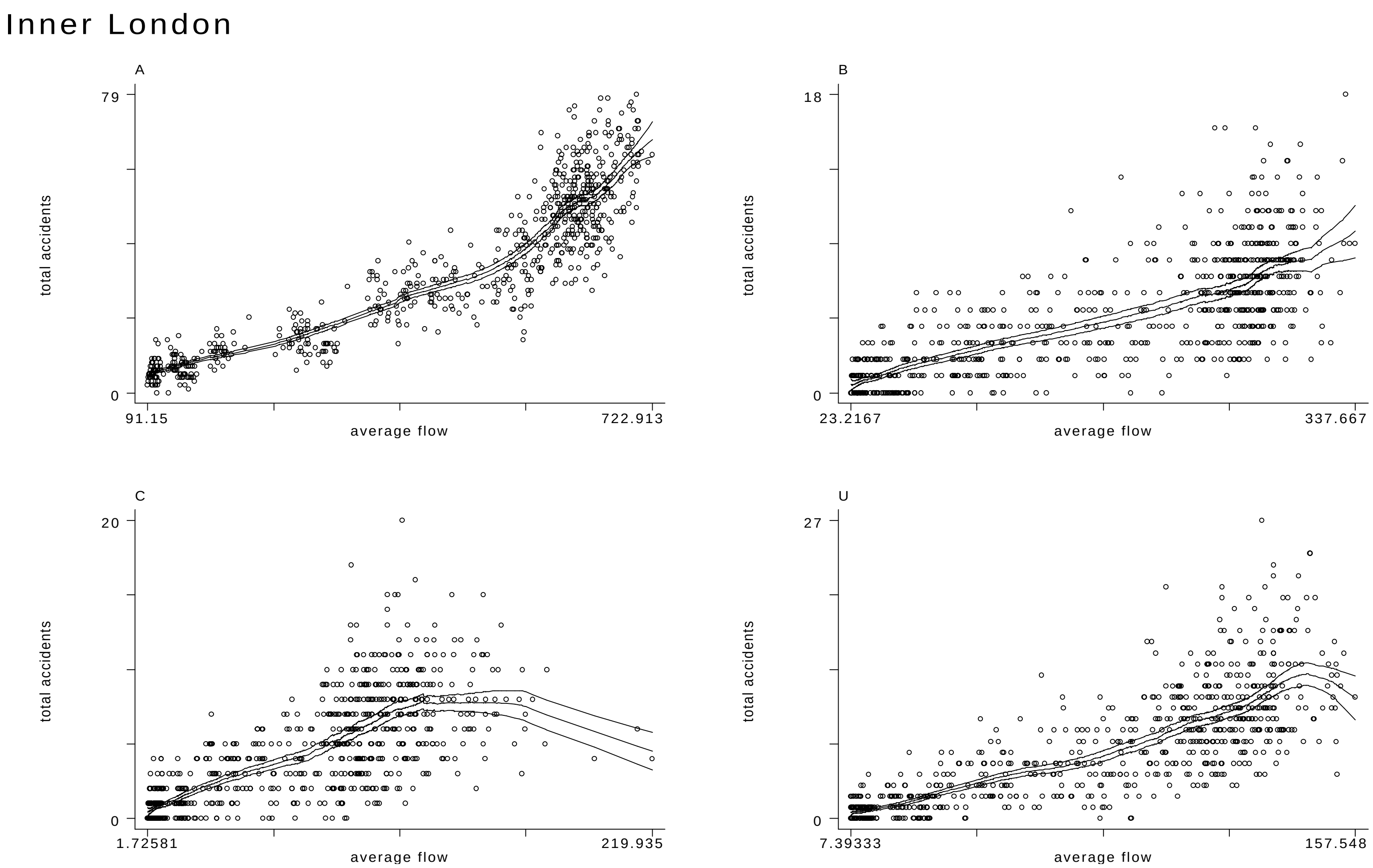


FIGURE 3 (cont.)

Outer London

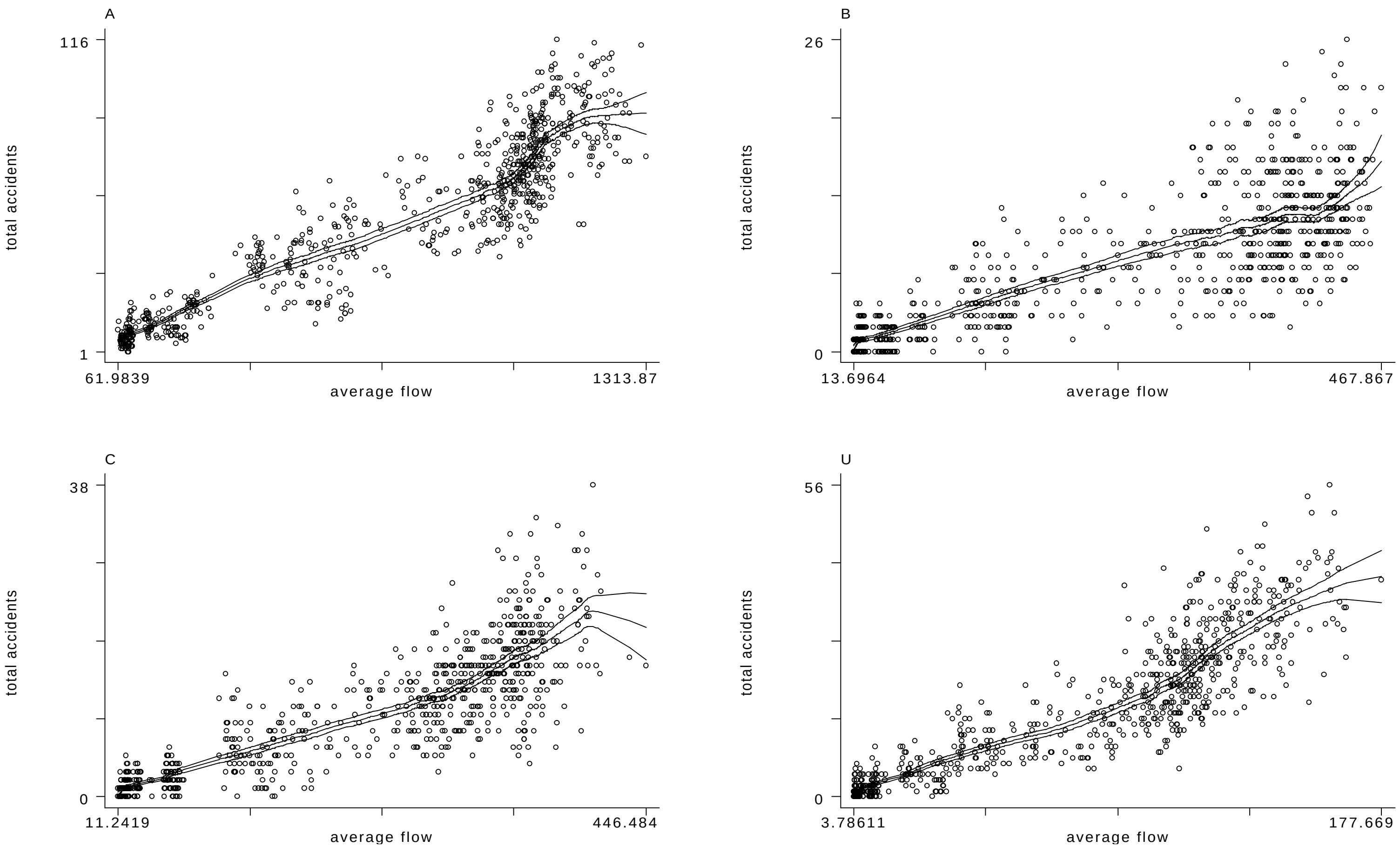


FIGURE 4

Accidents *vs* Flows: Piecewise Linear Splines

Inner London

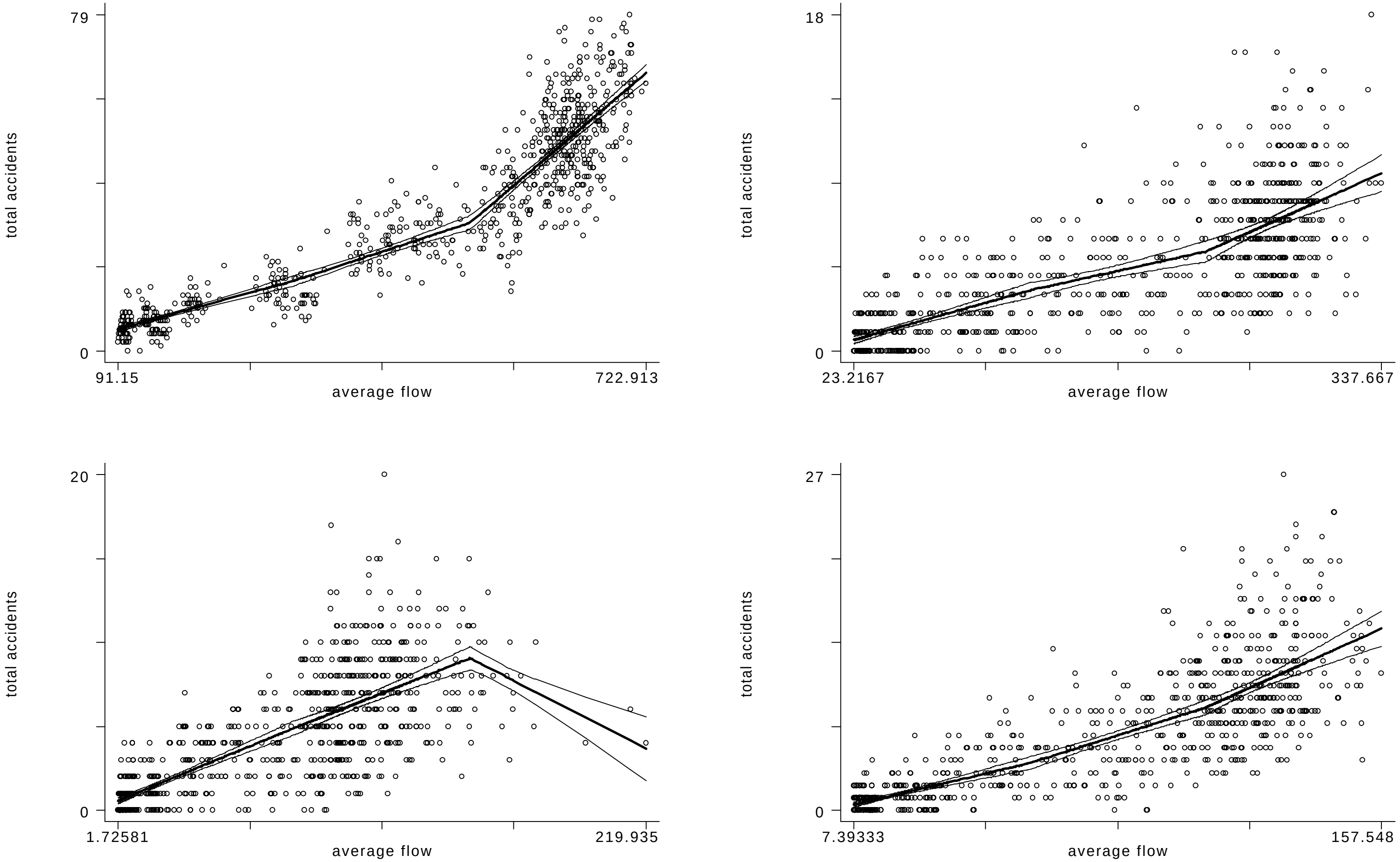
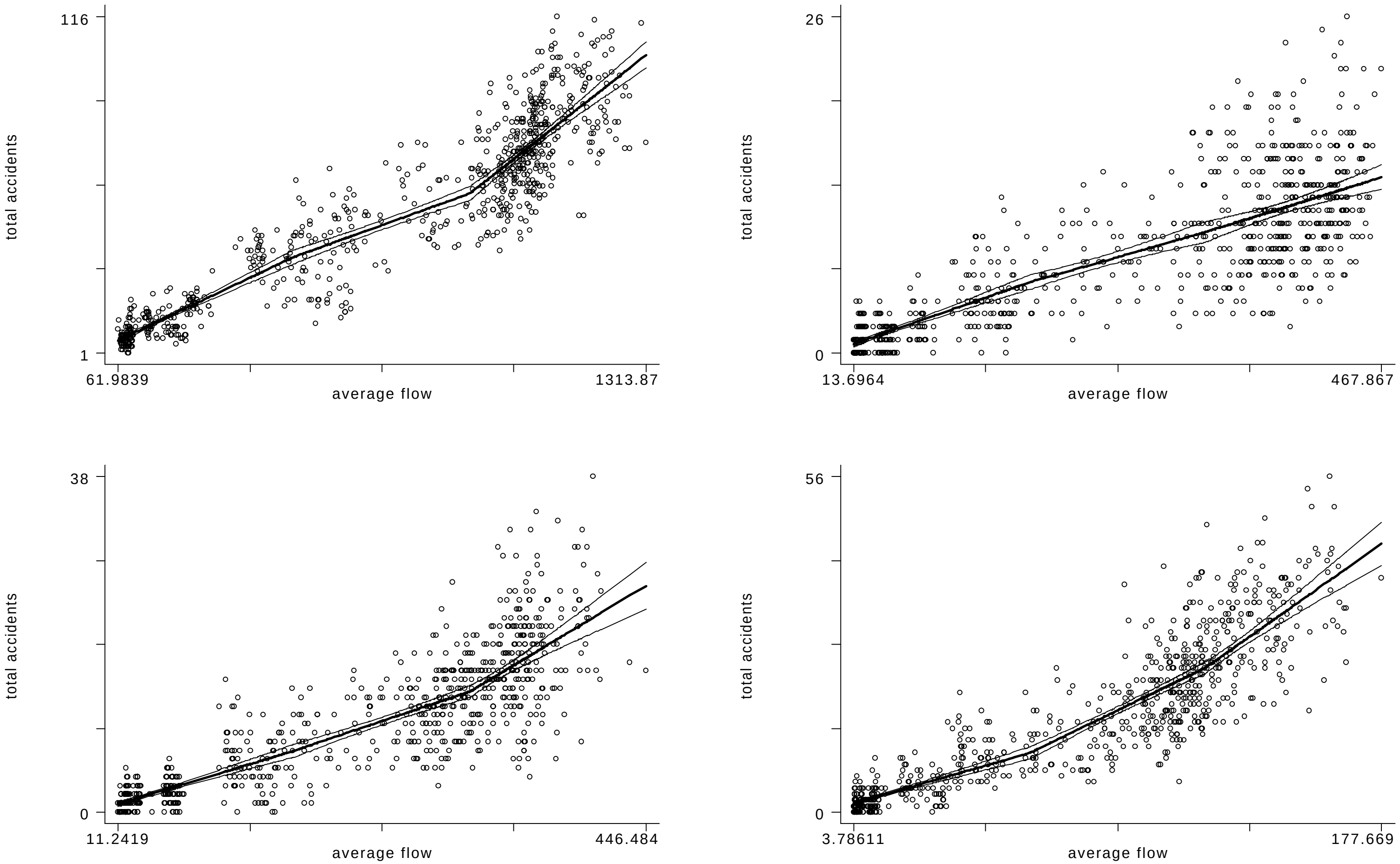


FIGURE 4 (cont.)

Outer London



APPENDIX

Definitions of Inner London and Outer London

Inner London Boroughs	No. of Sites	Outer London Boroughs	No. of Sites
Camden	3	Barking and Dagenham	3
Hackney	0	Barnet	3
Hammersmith and Fulham	0	Bexley	1
Haringey	0	Brent	2
Islington	3	Bromley	1
Kensington and Chelsea	3	Croydon	2
Lambeth	2	Ealing	0
Southwark	3	Enfield	3
Tower Hamlets	1	Greenwich	0
Wandsworth	5	Harrow	1
Westminster	3	Havering	0
		Hillingdon	2
		Hounslow	3
		Kingston	1
		Lewisham	1
		Merton	0
		Newham	3
		Redbridge	1
		Richmond	2
		Sutton	1
		Waltham Forest	1

Note: The definition of "inner" and "outer" London used here differs slightly from traditional local government administration agglomerations since Lewisham and Newham are defined as being in outer rather than inner London. However, this allocation accords with the Department of Transport's analysis of traffic flows and traffic speeds in London, and hence is the one that we use here.