Pergamon

Accid. Anal. and Prev., Vol. 28, No. 6, pp. 739–753, 1996

0001-4575/96 $15.00 + 0.00

PII: S0001-4575(96)00050-4

PREDICTING THE SEVERITY OF MOTOR VEHICLE ACCIDENT INJURIES USING MODELS OF ORDERED MULTIPLE CHOICE

C. J. O'DONNELL* and D. H. CONNOR

Department of Econometrics, University of New England, Armidale, NSW 2351, Australia

(Received 24 June 1996)

Abstract—This paper presents statistical evidence showing how variations in the attributes of road users can lead to variations in the probabilities of sustaining different levels of injury in motor vehicle accidents. Data from New South Wales, Australia, is used to estimate two models of multiple choice which are reasonably commonplace in the econometrics literature: the ordered logit model and the ordered probit model. Our estimated parameters are significantly different from zero at small levels of significance and have signs which are consistent with our prior beliefs. As a benchmark for comparison, we consider the risks faced by a 33-year-old male driver of a 10-year-old motor vehicle who is involved in a head-on collision while travelling at 42 kilometres per hour. We estimate that this benchmark victim will remain uninjured with a probability of almost zero, will require treatment from a medical officer with a probability of approximately 0.7, will be admitted to hospital with a probability of approximately 0.3, and will be killed with a probability of almost zero. We find that increases in the age of the victim and vehicle speed lead to slight increases in the probabilities of serious injury and death. Other factors which have a similar or greater effect on the probabilities of different types of injury include seating position, blood alcohol level, vehicle type, vehicle make and type of collision. Copyright © 1996 Elsevier Science Ltd

Keywords—Injury severity, Road-user attributes, Ordered logit model, Ordered probit model

1. INTRODUCTION

There exists a well-developed literature documenting linkages between road user attributes (e.g. age and sex) and the probability of being involved in a motor vehicle accident. Recent contributions to this literature include the studies by Connolly *et al.* (1989), Cooper (1990), Evans (1990), Levy (1990), Laberge-Nadeau *et al.* (1992), Lassarre (1986), Lloyd (1992), Mannering (1993), Mayhew *et al.* (1986), Stewart (1989) and Vingilis *et al.* (1992). These researchers have analysed traffic accident data using a wide range of statistical techniques and have established that, for example, the probability of being involved in a motor vehicle accident increases with increases in alcohol use and speed.

There also exists a distinct but related literature documenting linkages between road user attributes and the risk of injury given that an accident has occurred. Recent contributions which focus on this aspect of traffic safety include the studies by Evans and Frick (1988, 1992, 1993, 1994), Heulke and Compton (1995), Kim *et al.* (1995) and Shibata and Fukuda (1994). Again, these researchers have analysed traffic accident data using a range of techniques and have established that, for example, the risk of injury in a motor vehicle accident, given that the accident has occured, increases with age and speed (or crash severity).

In this paper we use two econometric models of ordered multiple choice to estimate the linkages between eleven road user attributes and the probabilities of sustaining four different levels of injury, given that an accident has occurred. The contributions of the paper are essentially two-fold. First, our use of econometric models of ordered multiple choice to analyse road accident data is apparently unprecedented: although our models of ordered multiple choice are ideally suited to the analysis of accident data and are well-established in the econometrics literature, we are unaware of any published applications of these models in the area of motor vehicle safety. Second, our simultaneous analysis of eleven road user attributes represents a significant increase in the size and scope of traffic accident studies: most earlier studies have been restricted to an analysis of

*Author for correspondence.

no more than seven road user attributes (e.g. Heulke and Compton 1995; Kim et al. 1995), and many have analysed only one attribute at a time (e.g. Evans and Frick 1988, 1992).

The plan of the paper is as follows. Section 2 outlines our econometric models and the method used to estimate the models' parameters. The specific econometric models we use are the ordered probit model and the ordered logit model, and the parameters are estimated using the well-known method of maximum likelihood. Section 3 describes the data used in our empirical work. The data consists of observations which record the injuries sustained by, and the attributes of, all reported road accident victims on New South Wales roads in 1991. In Section 4 we describe the model selection criterion we used to reduce the number of explanatory variables in each of our multiple choice models, and we also describe the way in which the criterion was implemented. The estimation results are presented in Section 5. The results include estimates of the parameters of the models and estimates of the probabilities of sustaining different levels of injury when the explanatory variables assume different values. In Section 6 we summarize our results and offer some concluding remarks.

2. THE ORDERED LOGIT AND PROBIT MODELS

2.1. *Motivation and structure*

The ordered logit model and the ordered probit model are usually motivated in a latent (ie. unobserved) variables framework. The general specification of each single-equation model is

$$y_i^* = \mathbf{x}_i'\beta + \epsilon_i \tag{1}$$

where y_i^* is a latent variable measuring the risk of injury faced by accident victim i, $\mathbf{x}_i$ is a $(K \times 1)$ vector of observed nonstochastic (i.e. non-random) explanatory variables measuring the attributes of accident victim i, β is a $(K \times 1)$ vector of unknown parameters and ϵ_i is a random error term.

We assume that all error terms have zero mean and that the error terms for different accident victims are uncorrelated. In a small departure from standard models of ordered multiple choice, we also assume that the disturbance term is heteroskedastic (i.e. the variance of the disturbance term can vary from one accident victim to another):

$$Var(\epsilon_i) = Var(y_i^*) = [\exp(\mathbf{z}_i'\gamma)]^2 \tag{2}$$

where $\mathbf{z}_i$ is a $(G \times 1)$ vector of observed nonstochastic explanatory variables, which may or may not include some of the elements of $\mathbf{x}_i$, and γ is a $(G \times 1)$ vector of unknown parameters. Thus, we allow for the possibility that some road user attributes may not only influence the average level of injury risk, but may also influence the variance of injury risk.

We are interested in estimating the unknown parameter vectors β and γ, but cannot apply standard regression techniques to eqn (1) because the dependent variable, y_i^*, is unobserved. Instead our data includes the observed variable y_i, a coded variable measuring the level of injury sustained by accident victim i: $y_i = 1$ if accident victim i requires no medical treatment (a 'non-treated injury'); $y_i = 2$ if he or she requires medical treatment administered by a medical officer such as a doctor, nurse, paramedic or ambulance person (a 'treated injury'); $y_i = 3$ if he or she is admitted to a hospital (an 'admitted injury'); and $y_i = 4$ if he or she dies immediately or within 30 days from injuries attributable to the accident (a 'death').

In any given accident it is reasonable to expect that a high risk of injury, y_i^*, will be translated into a high level of observed injury, y_i, and we formalize this relationship as follows:

$$y_i = \begin{cases} 1 \text{ if } -\infty \le y_i^* \le \mu_1 \text{ (non-treated injury)} \\ 2 \text{ if } \mu_1 \le y_i^* \le \mu_2 \text{ (treated injury)} \\ 3 \text{ if } \mu_2 \le y_i^* \le \mu_3 \text{ (admitted injury)} \\ 4 \text{ if } \mu_3 \le y_i^* \le \infty \text{ (death)} \end{cases} \tag{3}$$

where the threshold values μ_1, μ_2 and μ_3 are unknown parameters to be estimated. Thus, the probability that accident victim i sustains an injury of level j is equal to the probability that the unobserved injury risk, y_i^*, takes a value between two fixed thresholds. Moreover, it is clear that

$$\begin{aligned} P(y_i = 1|\text{accident}) &\equiv P_{1i} = F((\mu_1 - h_i)/g_i) \\ P(y_i = j|\text{accident}) &\equiv P_{ji} = F((\mu_j - h_i)/g_i) \\ &\quad - F((\mu_{j-1} - h_i)/g_i) \quad j = 2, 3 \\ P(y_i = 4|\text{accident}) &\equiv P_{4i} = 1 - F((\mu_3 - h_i)/g_i) \end{aligned} \tag{4}$$

where $h_i \equiv \mathbf{x}_i'\beta$, $g_i \equiv \exp(\mathbf{z}_i'\gamma)$ and $F(x)$ denotes the cumulative distribution function of the random error term ϵ_i evaluated at x. These probabilities will be positive if the threshold parameters satisfy the restriction $\mu_1 < \mu_2 < \mu_3$. Importantly, the inclusion of a constant term in $\mathbf{x}_i$ means we can set any one of the threshold parameters to any arbitrary value, and in the present paper we set $\mu_1 = 0$.

Models such as this, in which the observed dependent variable assumes one of a finite number of discrete values, are collectively known as multiple choice models. A special feature of our model is that the outcomes can be ranked: the ordering $y_i = 1, 2, 3, 4$ represents a ranking of injury severity. Thus, our model is a member of the subclass of multiple choice models known as ordered multiple choice models. Two such models regularly find their way into the empirical literature: the ordered logit model, based on the assumption that the random disturbances, ϵ_i, are independently and identically distributed with the logistic distribution (e.g. Byrne et al. 1991; Meyer and Cooke 1988); and the ordered probit model, based on the assumption that the ϵ_i are multivariate normal (e.g. Spizman and Kane 1992; Hausman et al. 1992). The choice of distributional assumption (logistic or normal) is usually made on empirical grounds, although there is one theoretical rationale for the ordered probit model which has a some appeal: if the disturbance term, ϵ_i, represents the combined influence of many independent factors not formally expressed in the model, central limit theorems can be used to justify the assumption that it is normally distributed. We do not rely on this rationale however, and we use both models in our empirical work.

2.2. Estimation

The parameters of our ordered multiple choice models are estimated by the method of maximum likelihood (ML). In very simple terms, the method of ML is a method for choosing parameter estimates in order to maximize the probability, or likelihood, of observing our given data. A likelihood function is an equation expressing this probability/likelihood as a function of the data and the unknown parameters, and ML estimation involves the systematic evaluation of this function at different points (i.e. sets of parameter values) in order to find the point at which the function is maximized. This set of parameter values then becomes our set of ML estimates. A detailed discussion of ML estimation can be found in any intermediate or advanced econometrics textbook, including the books by Greene (1993) and Judge *et al.* (1985).

For a sample of N accident victims, the log-likelihood function (i.e. the logarithm of the likelihood function) for both the ordered logit and probit models can be written

$$\ln(L)=\sum_{i=1}^{N}\sum_{j=1}^{J} d_{ji} \ln\left[F_{j,i}-F_{j-1,i}\right] \tag{5}$$

where $F_{j,i}\equiv F((\mu_j-h_i)/g_i)$ and d_{ji} is a dummy variable which takes the value one if $y_i=j$ and takes the value zero otherwise[1]. In this paper we maximize this function using the LIMDEP software written by Greene (1992). LIMDEP maximizes $\ln(L)$ using the steepest descent and Davidson–Fletcher–Powell algorithms (see Judge et al. 1985, pp. 958–960), and uses the Berndt *et al.* (1974) (BHHH) estimator to obtain estimates of the standard errors of the ML estimates.

3. DATA

Under NSW law, all road accidents occuring on a public roadway in which a person is injured or where there is more than a certain amount of damage to property must be reported to NSW Police. The accident data collected by the Police is then routinely processed (to ensure confidentiality) by the NSW Roads and Traffic Authority (RTA). For this study, the RTA supplied 28 747 records (or observations), representing a census of reported motor vehicle accident victims on NSW roads in 1991.

Unfortunately, limits to available computer capacity caused us to reduce the size of the RTA data set. Since we were mainly interested in analysing injuries to motor vehicle occupants, we chose to remove observations on motor-cyclists, pedal cyclists, pedestrians and drivers of non-motorized vehicles, leaving a subset of $N=18\,069$ observations for use in our empirical work. Of these 18 069 observations, 2.3% were classified as fatalities ($y_i=4$), 21.4% were classified as admitted injuries ($y_i=3$), 67.8% were classified as treated injuries ($y_i=2$), and 8.5% were classified as uninjured ($y_i=1$). From this breakdown, and from studies by Austin (1995a,b) and Rosman and Knuiman (1994) into the accuracy of United Kingdom and Western Austraan police road injury data, it seems likely that large numbers of uninjured motor vehicle accident victims have not been reported to NSW Police, and are underrepresented in our data set. This underrepresentation does not, however, affect the statistical properties of our maximum likelihood estimators, and should not affect the validity of our results.

Each observation in our data set is a record of the level of injury sustained by, and the attributes of, a motor vehicle accident victim. The attributes of these victims are described by the explanatory variables appearing in Table 1 below. Table 1 presents a definition of each variable together with its mean and standard deviation. These definitions are obtained from a coding manual (Paterson and Baxter 1987) supplied with the data set by the RTA.

It is worth noting from Table 1 that the continuous variables have been scaled (by dividing by multiples of 100) to have means which lie between 0 and 1. The reason for this is that the dummy variables have means between 0 and 1, and ordered multiple choice models are almost never estimable if the variables are of very different magnitudes (Greene 1993, p. 175). The means of the original unscaled

[1]Note that $F_{0,i}\equiv F((-\infty-h_i)/g_i)=F(-\infty)=0$ and that $F_{J,i}\equiv F((\infty-h_i)/g_i)=F(\infty)=1$.

Table 1. Explanatory variables

Variable	Mean	Standard deviation	Description
x_0	1.00	0.000	constant
Continuous variables			
x_1, z_1	0.326	0.188	age of casualty (years) divided by 100
x_2, z_2	0.423	0.325	speed of vehicle (km per hour) divided by 100
x_3, z_3	0.100	0.060	age of vehicle (years) divided by 100
x_4, z_4	0.136	0.056	time of accident (24 hour clock) divided by 10 000
x_5, z_5	0.142	0.152	$=x_1^2=$(age of casualty (years) divided by 100)^2
x_6, z_6	0.022	0.014	$=x_4^2=$ (time of accident (24 hour clock) divided by 10 000)^2
Dummy variables			
x_7	0.035	0.184	=1 if age of casualty unknown; =0 otherwise
x_8	0.006	0.074	=1 if speed of vehicle not stated but excessive; =0 otherwise
x_9	0.112	0.316	=1 if speed of vehicle unknown or not stated; =0 otherwise
x_{10}	0.006	0.075	=1 if seating position is centre-front (CF); =0 otherwise
x_{11}	0.224	0.416	=1 if seating position is left-front (LF); =0 otherwise
x_{12}	0.040	0.197	=1 if seating position is right-rear (RR); =0 otherwise
x_{13}	0.015	0.120	=1 if seating position is centre-rear (CR); =0 otherwise
x_{14}	0.052	0.221	=1 if seating position is left-rear (LR); =0 otherwise
x_{15}	0.006	0.080	=1 if seating position is other than driver, CF, LF, RR, CR or LR; =0 otherwise
x_{16}	0.026	0.159	=1 if non-seating or unknown seating position; =0 otherwise
x_{17}	0.485	0.500	=1 if female; =0 otherwise
x_{18}	0.006	0.080	=1 if $0.05 \leq$ blood alcohol reading ≤ 0.079; =0 otherwise
x_{19}	0.056	0.231	=1 if blood alcohol reading ≥ 0.08; =0 otherwise
x_{20}	0.510	0.500	=1 if blood alcohol reading unknown; =0 otherwise
x_{21}	0.056	0.229	=1 if seatbelt not worn or restraint not fitted; =0 otherwise
x_{22}	0.012	0.111	=1 if child restraint worn; =0 otherwise
x_{23}	0.053	0.225	=1 if vehicle is a light truck; =0 otherwise
x_{24}	0.009	0.092	=1 if vehicle is a heavy rigid truck; =0 otherwise
x_{25}	0.008	0.087	=1 if vehicle is an articulated truck; =0 otherwise
x_{26}	0.010	0.098	=1 if vehicle is a bus; =0 otherwise
x_{27}	0.006	0.080	=1 if vehicle is an emergency vehicle; =0 otherwise
x_{28}	0.005	0.073	=1 if vehicle is other than a car or any vehicle type listed above; =0 otherwise
x_{29}	0.009	0.095	=1 if vehicle is made by Manufacturer *B*; =0 otherwise
x_{30}	0.014	0.117	=1 if vehicle is made by Manufacturer *C*; =0 otherwise
x_{31}	0.215	0.411	=1 if vehicle is made by Manufacturer *D*; =0 otherwise
x_{32}	0.021	0.142	=1 if vehicle is made by Manufacturer *E*; =0 otherwise
x_{33}	0.003	0.054	=1 if vehicle is made by Manufacturer *F*; =0 otherwise
x_{34}	0.050	0.217	=1 if vehicle is made by Manufacturer *G*; =0 otherwise
x_{35}	0.006	0.076	=1 if vehicle is made by Manufacturer *H*; =0 otherwise
x_{36}	0.098	0.298	=1 if vehicle is made by Manufacturer *I*; =0 otherwise
x_{37}	0.101	0.302	=1 if vehicle is made by Manufacturer *J*; =0 otherwise
x_{38}	0.002	0.044	=1 if vehicle is made by Manufacturer *K*; =0 otherwise
x_{39}	0.003	0.056	=1 if vehicle is made by Manufacturer *L*; =0 otherwise
x_{40}	0.016	0.124	=1 if vehicle is made by Manufacturer *M*; =0 otherwise
x_{41}	0.167	0.373	=1 if vehicle is made by Manufacturer *N*; =0 otherwise
x_{42}	0.069	0.254	=1 if vehicle is not made by any of Manufacturers *A* to *N*; =0 otherwise
x_{43}	0.023	0.149	=1 if vehicle weight is over 4.5 tonne (tare); =0 otherwise
x_{44}	0.205	0.404	=1 if first impact is right-angle (RA) vehicle–vehicle (VV) collision; =0 otherwise
x_{45}	0.165	0.371	=1 if first impact is nose–tail (NT) VV collision; =0 otherwise
x_{46}	0.173	0.378	=1 if first impact is VV collision other than RA, NT or head-on; =0 otherwise
x_{47}	0.233	0.423	=1 if first impact is vehicle–object (VO) collision; =0 otherwise
x_{48}	0.110	0.313	=1 if first impact is collision other than VV or VO; =0 otherwise

variables can be obtained easily by multiplying the mean values reported in Table 1 by the appropriate scaling factors (e.g. the mean age of casualties = [mean of x_1] × [scaling factor] = 0.326 × 100 = 32.6 years). The proportions of individuals possessing particular attributes can be obtained easily by reading off the means of the appropriate dummy variables (e.g. the mean of x_7 is 0.035, meaning that the age of casualties was unknown in 3.5% of cases).

Table 1 also reveals that, while the vector $\mathbf{x}_i=(x_{0i}, x_{1i},\ldots, x_{48i})$ contains a constant term and a large number of both continuous and dummy variables, the vector $\mathbf{z}_i=(z_{1i},\ldots, z_{6i})$ contains continuous variables only. The reasons for this are two-fold: the inclusion of a constant term in $\mathbf{z}_i$ prevents the identification of β; and the inclusion of a full set of dummy variables in $\mathbf{z}_i$ would require the estimation of a prohibitively large number of unknown parameters (almost one hundred).

Several studies, including those by O'Day (1993), Shinar *et al.* (1983) and Rosman and Knuiman (1994), suggest that, although measurements on age and gender are generally error-free, police measurements on alcohol use and several other driver characteristics are often less reliable. Unfortunately, the nature and frequency of errors in the NSW Police data is unknown. Thus, the direction and size of any biases in our results is also unknown. This is a weakness of our study which appears to be shared by all other studies based on police-reported data.

4. MODEL SELECTION

We had no prior beliefs concerning which of the explanatory variables in Table 1 should be included in our ordered multiple choice models, so a subset of variables was chosen according to the Schwarz Bayesian Information Criterion (SBIC) (Schwarz 1978). The SBIC has also been used as an ordered probit model selection criterion by Waterman and Weiss (1992). About one-half of the variables in Table 1 were excluded from the analysis during the course of the SBIC variable selection process.

Under the SBIC, the number of parameters in the model, $K+G$, is chosen to minimize

$$\mathrm{SBIC}(K+G) = -2\ln L(K+G) + (K+G)\ln N \qquad (6)$$

where K is the number of parameters in the vector β, G is the number of parameters in the vector γ, $\ln L(K+G)$ is the maximized value of the log-likelihood function when the model contains $K+G$ parameters, and N is the sample size. It is clear that

$$\mathrm{SBIC}(K+G+1) > \mathrm{SBIC}(K+G) \Rightarrow LR_{K+G+1} \equiv -2\{\ln L(K+G) - \ln L(K+G+1)\} < \ln N \qquad (7)$$

Thus, the smaller model is chosen if LR_{K+G+1} (the likelihood ratio statistic for the exclusion of the $(K+G+1)^{\mathrm{st}}$ variable) is less than $\ln N$.

In practice, the SBIC model selection algorithm begins by including all possible explanatory variables in the model, then reducing the size of the model according to eqn (7). At each step, the likelihood ratio statistic from dropping each variable is calculated, and the variable which yields the smallest likelihood ratio statistic is dropped. The process continues until all such likelihood ratio statistics exceed $\ln N = 9.802$ (this critical value greatly exceeds the 5% critical value which would be used in a standard likelihood ratio test of the null hypothesis that a single parameter equals zero ($\chi^2_{0.05,1} = 3.84$)).

Unfortunately the SBIC requires the estimation of a large number of models and this may be time-consuming and expensive if the numbers of variables and observations are large, as they are in this study. Waterman and Weiss (1992) note that in the general linear model the likelihood ratio statistic is a known function of the ordinary t-statistic and, even though this property does not carry over exactly to ordered multiple choice models, they used asymptotic t-ratios to approximate the likelihood ratio statistics. We used a similar but slightly more accurate procedure: we (only) calculated the exact likelihood ratio statistic for the variable with estimated asymptotic t-ratio closest to zero, and dropped that variable if the likelihood ratio statistic was less than $\ln N$. The time-savings resulting from this short-cut procedure were substantial: each of our models took as much as 7 hours to estimate on a DEC OSF/1 mainframe, and the number of models to be estimated was reduced from a maximum of 2970 to a maximum of 108.

A final comment on the implementation of the SBIC is in order. At each step of the algorithm we considered all the variables in both $\mathbf{x}_i$ and $\mathbf{z}_i$ to be candidates for elimination. The algorithm could, however, have been implemented by first eliminating variables in $\mathbf{x}_i$ then, after determining the composition of $\mathbf{x}_i$, restarting the algorithm to eliminate variables in $\mathbf{z}_i$. Other methods of implementation are also possible, and there is some uncertainty as to whether the use of these different methods will yield different results. We have chosen not to investigate this matter further because the financial costs of using more than one method are quite large, and there is no *a priori* reason for choosing any other method over the one we have used.

5. RESULTS

5.1. *Parameter estimates and injury risk*

Maximum likelihood estimates of the structural parameters β and the error variance parameters γ are presented in Table 2 below. The figures in parentheses are estimated asymptotic t-ratios and are based on estimates of asymptotic standard errors calculated using the BHHH estimator referred to in Section 2.2 above. Note that all the estimates for both the ordered probit and logit models are significantly different from zero at even very small levels of significance.

A pseudo-R^2 (goodness-of-fit) measure is presented in Table 2 even though it is usual practice to ignore such measures in models of ordered multiple choice. Econometricians have ignored these measures for two reasons: there is no generally accepted measure of goodness-of-fit for these types of models, and concerns have been registered in the literature that some potential pseudo-R^2s have empirical and theoretical upper limits which are sometimes substantially less than one (Cohen et al. 1986; Russell and Rives 1979). Recently, however, Veall and Zimmerman (1992) have suggested a pseudo-R^2 which is easy to calculate and, when it takes a value less than about

Table 2. Parameter estimates[a]

	Variable	Ordered probit	Ordered logit
x_0	constant	2.156 (30.10)	2.878 (18.27)
x_2	speed	0.390 (9.35)	0.598 (8.89)
x_3	age of vehicle	0.594 (3.62)	0.960 (3.92)
x_4	time of accident	−2.808 (−4.03)	—
x_5	age of casualty squared	1.031 (14.94)	1.464 (11.87)
$x_6=x_4^2$	time of accident squared	10.464 (3.81)	—
x_8	speed excessive	1.877 (17.22)	2.884 (13.91)
x_9	speed unknown	0.948 (25.76)	1.395 (16.39)
x_{10}	seating position CF	1.121 (10.34)	1.850 (9.57)
x_{11}	seating position LF	1.209 (39.30)	1.955 (18.58)
x_{12}	seating position RR	1.240 (23.08)	1.961 (15.69)
x_{13}	seating position CR	1.285 (13.93)	2.017 (11.72)
x_{14}	seating position LR	1.292 (27.74)	2.067 (16.95)
x_{15}	other seating position	0.458 (3.69)	0.840 (4.17)
x_{16}	non-seating or unknown	1.096 (17.65)	1.722 (13.62)
x_{17}	female	0.068 (3.25)	0.100 (3.29)
x_{19}	blood alcohol ≥0.08	0.191 (4.02)	0.349 (5.28)
x_{20}	blood alcohol unknown	−1.529 (−44.38)	−2.364 (−19.48)
x_{21}	seatbelt not worn	0.493 (11.91)	0.755 (10.90)
x_{23}	light truck	0.150 (3.50)	0.240 (3.83)
x_{39}	Manufacturer L	0.703 (4.14)	1.015 (4.22)
x_{44}	right-angle vehicle–vehicle	−0.557 (−14.88)	−0.804 (−12.24)
x_{45}	nose–tail vehicle–vehicle	−0.646 (−15.57)	−0.911 (−12.31)
x_{46}	other vehicle–vehicle	−0.562 (−14.59)	−0.794 (−11.94)
x_{47}	vehicle–object	−0.244 (−7.28)	−0.313 (−6.24)
x_{48}	not VV or VO	−0.274 (−6.91)	−0.360 (−6.10)
z_1	age of casualty	—	−0.528 (−3.32)
z_2	speed	0.101 (4.96)	0.090 (3.51)
z_4	time of accident	—	−2.358 (−3.67)
z_5	age of casualty squared	0.192 (4.17)	0.878 (4.37)
z_6	time of accident squared	—	9.098 (3.65)
μ_2		2.763 (73.15)	4.002 (20.50)
μ_3		4.456 (72.48)	6.645 (20.53)
Pseudo-R^2		0.360	0.356

[a] Figures in parentheses are estimated asymptotic t-ratios. Note that μ_1 has been normalised to zero.

0.6, seems able to mimic the OLS-R^2 which would be obtained if the underlying latent variable, y_i^*, was in fact observed. If we bear in mind that our data set contains observations on a cross-section of road accident victims, we can conclude from the Veall and Zimmerman pseudo-R^2 values reported in Table 2 that the degree of variation in injury severity explained by variations in the attributes of victims is reasonably high.

The remainder of this section is devoted to a discussion of the signs and magnitudes of our estimated coefficients.

We held prior beliefs concerning the signs of a small subset of coefficients, and the estimates reported in Table 2 are consistent with those beliefs. Note from eqn (1) that estimates which are greater than zero imply that increases in the associated (nonnegative) variables tend to lead to increases in injury risk, y_i^*. Thus, the estimates reported in Table 2 confirm our prior belief that the average risk of injury increases with the values of the continuous variables speed (x_2), age of vehicle (x_3) and age of casualty squared (x_5). The estimated dummy variable coefficients confirm our prior belief that the average risk of injury increases when the accident victim does not wear a seatbelt ($x_{21}=1$) or is travelling in a light truck ($x_{23}=1$), and falls when the accident victim is involved in any type of collision other than a vehicle–vehicle head-on ($x_{44}=1,...., x_{47}=1$ or $x_{48}=1$). These findings are consistent with the findings of several earlier studies, including Shibata and Fukuda (1994) (speed), Evans (1987) (seatbelts) and Heulke and Compton (1995) (seatbelts).

We had no prior beliefs concerning the signs of the remaining coefficients. Thus it was interesting to find that the average risk of injury increases when the accident victim is seated anywhere other than in the drivers position ($x_{10}=1,...., x_{15}=1$ or $x_{16}=1$), is a female ($x_{17}=1$), has a blood alchohol reading of greater than 0.08 ($x_{19}=1$) or is travelling in a vehicle made by Manufacturer L ($x_{39}=1$). Our finding that the safest seating position is the driving position is slightly inconsistent with the findings of Evans and Frick (1988) who find that the drivers position is the safest position only when vehicles are impacted from the rear or passenger side. Our different finding may result from the use of a different methodology, may be due to measurement errors in our data, or may reflect differences in United States and Australian vehicle designs. Our finding that females tend to face a higher risk of injury than males is consistent with the findings of Evans (1991) and Kim et al. (1995), and seems to imply that females are generally less able to tolerate certain types and levels of physical trauma. The apparent tendency for injury risk to increase when blood alchohol levels exceed 0.08 is consistent with the findings of Waller *et al.* (1986), Evans and Frick (1993) and Kim *et al.* (1995), and seems to imply an inability on the part of the body to overcome physical trauma when suffering the effects of alcohol. Finally, our finding that the average risk of injury increases when the accident victim is travelling in a vehicle made by Manufacturer L is a result for which there appears to be no compelling intuitive explanation, and it might be interesting to determine whether this result is peculiar to our particular variable selection process.

The relative magnitudes of the estimated coefficients are also of interest. Because injury risk, y_i^*, is specified as a linear function of the explanatory variables, the relative magnitudes of the estimated dummy variable coefficients are, in most cases, a measure of the relative impacts of these variables on the average risk of injury. For example, because the probit estimate of the coefficient of the dummy variable recording a centre-front seating position ($\hat{\beta}_{10}=1.121$) is about 2.3 times larger than the estimated coefficient of the dummy variable recording seatbelt use ($\hat{\beta}_{21}=0.493$) we conclude on the basis of our probit estimates that the increase in injury risk faced by an individual in the centre-front seating position is about 2.3 times higher than the increase in injury risk faced by an accident victim who does not wear a seatbelt, all other things being equal. Most of the estimated dummy variable coefficients can be compared in this way and the influences of different dummy variables on average injury risk can be ranked. The only exceptions are comparisons involving the estimated coefficients of the dummy variables measuring excessive or unknown speed. These dummy variables, x_8 and x_9, only take the value one when the continuous variable speed, x_2, takes the value zero, so the effects of these variables on injury risk would need to be calculated using $\hat{\beta}_8-\hat{\beta}_2x_2$ and $\hat{\beta}_9-\hat{\beta}_2x_2$, rather than simply using $\hat{\beta}_8$ and $\hat{\beta}_9$. Ignoring the special cases of excessive or unknown speed, our comparisons of the remaining dummy variable coefficients lead us to conclude that seating positions appear to have the greatest influence on injury risk, and gender appears to have the smallest influence.

It is readily apparent that the conclusions we are able to draw from the logit and probit estimates presented in Table 2 are qualitatively identical in all respects except for the effects of time. Estimates from the probit model reveal that the mean level of injury risk is a function of time but the variance of injury risk is not. Estimates from the logit model reveal the converse: the variance of injury risk is a function of time but the mean is not. Interestingly, both the probit estimate of the mean level of injury risk and the logit estimate of the variance of injury risk are minimized during the middle of the day: the continuous variable recording the time of the accident enters the functions $\mathbf{x}_i'\beta$ (probit) and $\mathbf{z}_i'\gamma$ (logit) as a quadratic, and both these quadratic functions reach their minimum values when the time variable takes a value of approximately 0.13 (one o'clock in the afternoon). These results may reflect variable delays in the availability of medical treatment at different times of the day (through variations in emergency vehicle response times and variations in the demand and supply of hospital services). These delays in medical treatment may compound the severity of accident injuries and lead to an increase in both the mean and variance of injury risk.

Our final comment on the signs and magnitudes of the estimated coefficients is that the probit estimates of the structural parameters β are always smaller than the corresponding nonzero logit estimates. Together with the fact that the probit quadratic in time is nonpositive over the permissible range of the time variable (the interval from 0.00 to 0.24), this implies that probit predictions of injury risk will always be less than the corresponding logit predictions. Of course the probit estimates of the threshold parameters μ_2 and μ_3 are also smaller, so lower predictions of injury risk, y_i^*, will not necessarily translate into lower predictions of injury severity, y_i. Our logit and probit estimates of the probabilities of sustaining various levels of road accident injury, P_{ji}, will be discussed immediately below.

5.2. *Injury probabilities*

A useful starting point for a discussion of injury probabilities is to consider the characteristics of the accident victim when all dummy variables in the models take the value zero. Such an accident victim

(a) is a male,
(b) is the driver of any type of motor vehicle other than a light truck,
(c) wears a seatbelt,
(d) has a blood alcohol reading less than 0.08,
(e) is involved in a vehicle–vehicle head-on collision, and
(f) drives a vehicle made by any manufacturer other than Manufacturer L.

Hereinafter, the term 'benchmark victim' will be used to refer to an accident victim who has these characteristics and, in addition,

(g) is of average age (33 years),
(h) is travelling at average speed (42 kilometres per hour),
(i) is travelling in a vehicle of average age (10 years), and

(j) is involved in an accident at the average time of approximately 1:30 p.m.

In the early part of this section we present estimates of the probabilities that a benchmark victim will sustain different types of accident injury. We then present estimates of these injury probabilities when certain dummy variables take the value one. The final part of this section presents estimates of changes in injury probabilities resulting from changes in the values of the continuous variables.

Estimates of the probabilities that our benchmark victim will sustain various levels of injury in his motor vehicle accident are reported in the first row of Table 3 below: the estimated probability that he will remain uninjured is almost zero, the estimated probability that he will sustain an injury which requires treatment is approximately 0.7, the estimated probability that he will sustain an injury which requires admission to a hospital is approximately 0.3, and we estimate that he will be killed with a probability of almost zero. Notice from Table 3 that all these probabilities are significantly different from zero at usual levels of significance. Also notice that, even though we found in Section 5.1 that our probit predictions of injury risk will always be lower than our logit predictions, the estimated probabilities of sustaining various levels of road accident injury, $\hat{P}_{ji}$, are similar for both models. Thus, as we suspected, the probit estimates of the threshold parameters μ_2 and μ_3 are small enough relative to the logit estimates for both sets of predicted probabilities to be roughly equivalent.

The remaining rows of Table 3 report estimates of injury probabilities when certain dummy variables take the value one. Again, estimated probabilities from both the probit and logit models are similar and are significantly different from zero at usual levels of significance. The interpretation of these probability estimates is straightforward: the second row of estimates, for example, reveals that when an

Table 3. Injury probabilities[a]

	Ordered probit				Ordered logit			
	Non-treated injury	Treated Injury	Admitted Injury	Death	Non-treated Injury	Treated Injury	Admitted Injury	Death
Benchmark individual	0.015	0.652	0.311	0.022	0.017	0.662	0.302	0.019
	(10.12)	(63.22)	(30.52)	(10.72)	(11.33)	(57.40)	(26.42)	(11.32)
Speed excessive ($x_2=z_2=0$; $x_8=1$)	0.000	0.110	0.558	0.332	0.000	0.074	0.614	0.312
	(2.10)	(5.46)	(27.41)	(8.57)	(4.17)	(5.32)	(20.27)	(7.33)
Seating position centre-front ($x_{10}=1$)	0.000	0.267	0.566	0.166	0.002	0.183	0.662	0.154
	(2.63)	(7.73)	(51.28)	(6.13)	(4.23)	(5.68)	(61.62)	(5.27)
Seating position left-front ($x_{11}=1$)	.000	0.241	0.571	0.187	0.002	0.165	0.663	0.171
	(5.62)	(18.63)	(73.40)	(14.56)	(7.94)	(14.79)	(64.26)	(12.30)
Seating position right-rear ($x_{12}=1$)	0.000	0.232	0.572	0.195	0.002	0.164	0.663	0.172
	(4.32)	(13.23)	(72.14)	(11.45)	(6.61)	(10.41)	(64.03)	(9.50)
Seating position centre-rear ($x_{13}=1$)	0.000	0.219	0.573	0.206	0.001	0.155	0.662	0.182
	(5.80)	(19.86)	(72.32)	(16.44)	(4.83)	(6.55)	(60.17)	(6.49)
Seating position left-rear ($x_{14}=1$)	0.000	0.217	0.573	0.209	0.001	0.147	0.661	0.191
	(6.75)	(23.38)	(72.12)	(18.43)	(6.95)	(11.23)	(59.09)	(10.54)
Other seating position ($x_{15}=1$)	0.005	0.497	0.443	0.056	0.006	0.428	0.515	0.051
	(2.77)	(10.57)	(12.60)	(3.96)	(9.13)	(21.95)	(31.66)	(9.42)
Female ($x_{17}=1$)	0.013	0.631	0.331	0.025	0.015	0.637	0.327	0.022
	(9.68)	(57.49)	(31.46)	(10.79)	(10.90)	(51.73)	(26.90)	(11.47)
Blood alcohol reading ≥ 0.08 ($x_{19}=1$)	0.010	0.591	0.367	0.033	0.011	0.570	0.390	0.029
	(6.63)	(31.35)	(22.01)	(7.91)	(8.99)	(27.49)	(20.11)	(9.27)
Seatbelt not worn ($x_{21}=1$)	0.004	0.484	0.452	0.059	0.007	0.452	0.495	0.046
	(6.29)	(25.57)	(30.96)	(9.59)	(8.51)	(21.09)	(26.05)	(10.00)
Light truck ($x_{23}=1$)	0.011	0.605	0.355	0.030	0.012	0.600	0.362	0.025
	(7.10)	(35.48)	(23.04)	(8.25)	(8.89)	(30.44)	(19.26)	(9.24)
Manufacturer L ($x_{39}=1$)	0.002	0.408	0.503	0.086	0.005	0.378	0.555	0.062
	(1.96)	(6.63)	(13.35)	(3.67)	(3.33)	(5.66)	(10.72)	(3.62)
Right-angle vehicle–vehicle ($x_{44}=1$)	0.051	0.780	0.164	0.005	0.043	0.805	0.145	0.007
	(13.20)	(153.40)	(23.13)	(8.62)	(13.18)	(126.67)	(21.53)	(10.06)
Nose-tail vehicle–vehicle ($x_{45}=1$)	0.060	0.791	0.145	0.004	0.048	0.816	0.130	0.006
	(13.36)	(164.05)	(20.06)	(8.11)	(13.09)	(129.13)	(18.77)	(9.79)
Other vehicle–vehicle ($x_{46}=1$)	0.051	0.781	0.163	0.005	0.042	0.804	0.146	0.007
	(12.87)	(149.02)	(22.04)	(8.30)	(12.83)	(122.65)	(20.69)	(9.92)
Vehicle–object ($x_{47}=1$)	0.027	0.720	0.242	0.012	0.024	0.731	0.232	0.013
	(11.71)	(93.62)	(28.37)	(9.78)	(12.46)	(85.75)	(26.93)	(10.89)

[a] Numbers in parentheses are estimated asymptotic t-ratios and are based on estimated standard errors calculated by linear approximation.

accident victim has all the characteristics of a benchmark victim but is otherwise travelling at an excessive speed ($x_2=z_2=0$, $x_8=1$), the probability that he will require admission to a hospital is almost double that of a benchmark victim at 0.6, and the probability that he will be killed increases from practically zero to more than 0.3. By way of further example, the next several rows in Table 3 reveal that the effects of different seating positions on the probabilities of serious injury and death are similar but perhaps not quite as dramatic as the effects of excessive speed. In particular, it appears that the left-rear seating position ($x_{14}=1$) is the most dangerous seating position: an accident victim who has all the attributes of a benchmark victim but is otherwise travelling in the left-rear seating position will sustain an admitted injury with a probability of between 0.6 (probit) and 0.7 (logit) (up from 0.3), and will be killed with a probability of approximately 0.2 (up from practically zero). Finally, the nose–tail vehicle–vehicle row of Table 3 reveals that when an accident victim has all the attributes of a benchmark victim but has a nose–tail rather than a head-on vehicle–vehicle collision, the probability that he does not require treatment increases from almost zero to 0.1, the probability that he sustains a treated injury increases from less than 0.7 to approximately 0.8, the probability that he sustains an admitted injury falls from 0.3 to approximately 0.1, and the probability he is killed falls marginally to something very close to zero. Thus, for an accident victim to possess all the characteristics of a benchmark victim but otherwise be involved in a nose–tail vehicle–vehicle collision is, in a sense, desirable.

By now it should be apparent that a relationship exists between the signs of the dummy variable coefficients reported in Table 2 and differences between at least some of the probabilities reported in Table 3. Indeed, we can formally establish such a relationship through a closer examination of eqn (4). Noting that the value of $g_i \equiv \exp(z_i'\gamma)$ does not depend on the values of any dummy variables, it is clear that for all dummy variables other than those recording excessive or unknown speed (see our discussion in Section 5.1):

$$\mathrm{sign}(P_{1i}|x_{ji}=1-P_{1i}|x_{ji}=0)\neq\mathrm{sign}(\beta_j)$$
$$=\mathrm{sign}(P_{4i}|x_{ji}=1-P_{4i}|x_{ji}=0). \quad (8)$$

In words, if a dummy variable coefficient is positive (negative), the probability that an accident victim will remain uninjured, P_{1i}, will decrease (increase) and the probability that he or she will be killed, P_{4i}, will increase (decrease) when the dummy variable takes the value one. No such direct relationship exists between the signs of the dummy variable coefficients and the changes in the remaining probabilities, P_{2i} and P_{3i}.

A useful way of summarizing the relationships between our dummy variable coefficient estimates and our estimates of different injury probabilities is displayed in Fig. 1 below. Because our probit and logit estimates are so similar and we wish to avoid unnecessary repetition, Fig. 1 summarizes the results from the probit model only. The estimated probabilities that our benchmark victim will sustain different levels of injury are found from Fig. 1 by reading off the probabilities corresponding to a coefficient value of zero: as we already know, these probabilities are estimated to be $\hat{P}_{1i}=0.0$, $\hat{P}_{2i}=0.7$, $\hat{P}_{3i}=0.3$ and $\hat{P}_{4i}=0.0$ (rounded to one decimal point). The estimated injury probabilities for other victims are found

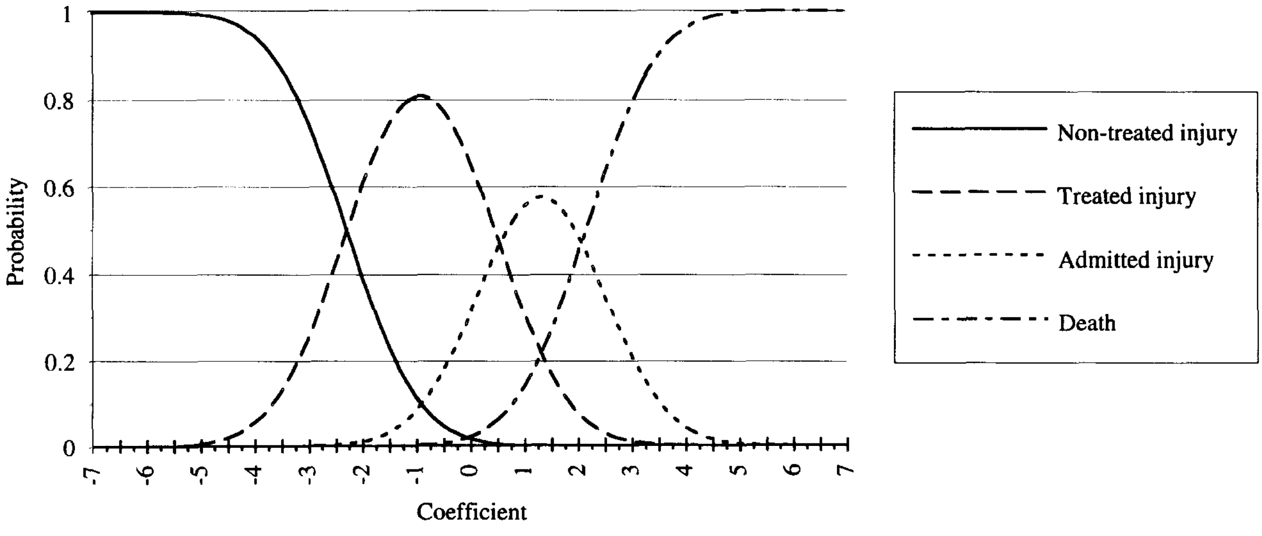

Fig. 1. Injury probabilities (benchmark).

by reading off the probabilities corresponding to the appropriate estimated dummy variable coefficients. For example, when an accident victim has all the attributes of a benchmark victim but otherwise is not wearing a seatbelt ($x_{21}=1$), the estimated injury probabilities are found by reading off the probabilities corresponding to the estimated coefficient value of $\hat{\beta}_{21}=0.493$: again, as we already know, these probabilities are estimated to be $\hat{P}_{1i}=0.0$, $\hat{P}_{2i}=0.5$, $\hat{P}_{3i}=0.5$ and $\hat{P}_{4i}=0.1$ (again, rounded to one decimal point). Interestingly, when we begin to use Fig. 1 in this way we find that it bears out the relationships established by eqn (8) above: positive (negative) dummy variable coefficients mean a decrease (increase) in P_{1i} and an increase (decrease) in P_{4i}. Moreover, the larger the dummy variable coefficient, the larger the effect. Of course this is not the case for the intermediate probabilities, P_{2i} and P_{3i}: these probabilities may increase and then decrease (or vice versa) as the absolute value of the coefficient increases. That is, the relationships between the signs of the dummy variable coefficients and the changes in the intermediate probabilities, P_{2i} and P_{3i}, are ambiguous. Finally, and in view of our discussion in Section 5.1, it should be noted that it is not appropriate to use Fig. 1 to read off the injury probabilities for a road user travelling at an unknown or excessive speed ($x_8=1$ or $x_9=1$).

So far in our discussion of Fig. 1 we have considered value changes in only one dummy variable at a time. However, we can also use Fig. 1 to find sets of injury probabilites when several dummy variables take the value one simultaneously: because the effects of our estimated dummy variable coefficients are additive, we simply read off the probabilities corresponding to the sum of the relevant estimated dummy variable coefficients. For example, when an accident victim has all the attributes of a benchmark victim but has a blood alchohol reading greater than 0.08 ($x_{19}=1$), does not wear a seatbelt ($x_{21}=1$) and is travelling in a car made by Manufacturer L ($x_{39}=1$), the estimated injury probabilities are found by reading off the probabilities corresponding to $\hat{\beta}_{19}+\hat{\beta}_{21}+\hat{\beta}_{39}=1.387$: these probabilities are estimated to be $\hat{P}_{1i}=0.0$, $\hat{P}_{2i}=0.2$, $\hat{P}_{3i}=0.6$ and $\hat{P}_{4i}=0.2$. Thus, Fig. 1 allows us to illustrate the cumulative effects of different dummy variables (injury risk factors) on injury probabilities. Again, it is not appropriate to use Fig. 1 to read off the injury probabilities for a road user travelling at an unknown or excessive speed.

The remainder of this section is concerned with changes in injury probabilities arising from changes in the values of the continuous variables. We first consider the effects of differential changes in the values of these variables. We then consider the effects of certain large changes in the continuous variables, either in isolation or in combination with changes in the values of certain dummy variables.

The elasticities which measure the effects of a one-percentage change in the continuous variable x_{ji} on the probabilities of different injuries are given by:

$$\frac{\delta P_{1i}}{\delta x_{ji}}\frac{x_{ji}}{P_{li}}=\left[-f(-h_i/g_i)\left[\frac{h_{ji}g_i-h_ig_{ji}}{g_i^2}\right]\right]\times\frac{x_{ji}}{P_{li}}$$

$$\frac{\partial P_{ji}}{\partial x_{ji}}\frac{x_{ji}}{P_{ji}}=\left[f((\mu_{j-1}-h_i)/g_i)\left[\frac{h_{ji}g_i-(h_i-\mu_{j-1})g_{ji}}{g_i^2}\right]\right.$$
$$\left.-f((\mu_j-h_i)/g_i)\left[\frac{h_{ji}g_i-(h_i-\mu_j)g_{ji}}{g_i^2}\right]\right]$$
$$\times\frac{x_{ji}}{P_{ji}}\quad j=2,3$$

$$\frac{\partial P_{4i}}{\partial x_{ji}}\frac{x_{ji}}{P_{4i}}=\left[f(\mu_3-h_i)/g_i\left[\frac{h_{ji}g_i-(h_i-\mu_3)g_{ji}}{g_i^2}\right]\right]\times\frac{x_{ji}}{P_{4i}}$$

where $h_{ji}\equiv\delta h_i/\delta x_{ji}$, $g_{ji}\equiv\delta g_i/\delta x_{ji}$, $\mu_1=0$ and $f(x)$ denotes the probability density function of ϵ_i evaluated at x. These expressions have been evaluated using our probit estimates of the unknown parameters and using benchmark values of the continuous and dummy variables (i.e. continuous variables have been set to their average values and the dummy variables set to zero). The results are reported in Table 4. All the elasticities reported in Table 4 are plausible and most are significantly different from zero at usual levels of significance. Furthermore, all estimates are significantly less than one in absolute value, which means that a one-percentage change in the value of any continuous variable will lead to a smaller percentage change (i.e. less than 1%) in the injury probabilities (i.e. in econometrics jargon, the responses of injury probabilities to changes in the continuous variables are inelastic).

The sign patterns evident in Table 4 are of some interest and in most cases can be traced back to the signs of the coefficient estimates reported in the top half of Table 2: if the estimated coefficient of the continuous variable x_{ji} is positive then an increase in x_{ji} will lead to an increase in injury risk and, provided the effect of x_{ji} on the variance of injury risk is not too large, this will lead to a decrease in the probability of no injury and an increase in the probability of death (again, the changes in the intermediate probabilities are ambiguous). Thus, because the estimated coefficients of variables measuring speed (x_2), age of vehicle (x_3) and age of casualty (x_5) are all positive, we observe that marginal increases in any of these variables will give rise to

Table 4. Elasticities [a]

	Ordered probit, % change in probability of				Ordered logit, % change in probability of			
	Non-treated injury	Treated injury	Admitted injury	Death	Non-treated injury	Treated injury	Admitted injury	Death
1% increase in speed	−0.161	−0.093	0.162	0.564	−0.142	−0.107	0.214	0.442
	(−2.13)	(−11.57)	(9.53)	(9.69)	(−2.42)	(−11.27)	(10.39)	(8.14)
1% increase in age of vehicle	−0.144	−0.028	0.056	0.131	−0.110	−0.033	0.071	0.113
	(−3.45)	(−3.60)	(3.61)	(3.44)	(−3.79)	(−3.69)	(3.69)	(3.78)
1% increase in time of accident	−0.009	−0.002	0.004	0.008	−0.057	0.005	−0.004	−0.056
	(−0.16)	(−0.16)	(0.16)	(0.16)	(−0.65)	(0.64)	(−0.64)	(−0.64)
1% increase in age of casualty	−0.304	−0.118	0.214	0.677	−0.298	−0.120	0.254	0.423
	(−3.86)	(−16.33)	(15.06)	(10.76)	(−4.04)	(−12.76)	(14.19)	(5.72)

[a] Numbers in parentheses are estimated asymptotic t-ratios and are based on estimated standard errors calculated by linear approximation.

marginal increases in the probability of death. The only other continuous variable, the variable which measures the time of the accident (x_4), enters the probit model as a quadratic, and this quadratic reaches a minimum at $x_4=0.13$. Because the probit elasticities reported in Table 4 are calculated using a value of x_4 greater than 0.13, we observe that an increase in the time of accident will also give rise to a marginal increase in the probability of death. If the probit elasticities reported in Table 4 had been calculated using a value of x_4 less than 0.13, we would have observed that an increase in the time of accident will give rise to a marginal decrease in the probability of death.

The relative magnitudes of the estimated elasticities reported in Table 4 are also of interest insofar as they measure the relative importance of different continuous variables on injury probabilities. Thus, it is estimated that a one-percentage increase in the age of the casualty will give rise to a larger increase in the probability of death than a one-percentage increase in vehicle speed. The estimated effects of a one-percentage increase in either the age of the vehicle or the time of the accident appear to be negligible by comparison.

We conclude this section by considering the effects on injury probabilities of certain large changes in the values of selected continuous variables. Specifically, we consider the effects of an increase in speed from 42 to 100 kilometres per hour, a decrease in the age of the vehicle from ten to five years, and an increase in the age of the casualty from 33 to 50 years. All calculations are based on estimates obtained from the ordered probit model.

The effects of an increase in vehicle speed from 42 kilometres per hour (the benchmark value) to 100 kilometres per hour can be illustrated using Fig. 2, which should be interpreted in the same way as Fig. 1. Thus, when an accident victim has all the characteristics of a benchmark victim but is otherwise travelling at 100 kilometres per hour, the estimated probabilities of sustaining different levels of injury are found by reading off the probabilities corresponding to a coefficient value of zero: $\hat{P}_{1i}=0.0$, $\hat{P}_{2i}=0.6$, $\hat{P}_{3i}=0.4$ and $\hat{P}_{4i}=0.0$. Notice that these probabilities are only slightly different from the benchmark probabilities of $\hat{P}_{1i}=0.0$, $\hat{P}_{2i}=0.7$, $\hat{P}_{3i}=0.3$ and $\hat{P}_{4i}=0.0$ (see Fig. 1). By way of further example, when an accident victim has all the characteristics of a benchmark victim but is otherwise travelling at 100 kilometres per hour, has a blood alcohol reading greater than 0.08 ($x_{19}=1$), does not wear a seatbelt ($x_{21}=1$) and is travelling in a car made by Manufacturer L ($x_{39}=1$), the estimated injury probabilities are found by reading off the probabilities corresponding to $\hat{\beta}_{19}+\hat{\beta}_{21}+\hat{\beta}_{39}=1.387$: we obtain $\hat{P}_{1i}=0.0$, $\hat{P}_{2i}=0.2$, $\hat{P}_{3i}=0.5$ and $\hat{P}_{4i}=0.3$. Again, these probabilities are only slightly different from the estimated injury probabilities faced by a casualty who has similar attributes but is otherwise travelling at 42 kilometres per hour: recall that these estimated probabilities are $\hat{P}_{1i}=0.0$, $\hat{P}_{2i}=0.2$, $\hat{P}_{3i}=0.6$ and $\hat{P}_{4i}=0.2$ (see Fig. 1). Thus, the estimated effects on injury probabilities of this large increase in speed are quite small. Indeed, it is only after close scrutiny of Figs 1 and 2 that we detect any difference in the two sets of injury probability curves: a 58 kilometre per hour increase in speed causes all curves to shift marginally to the left, implying only a small decrease in the probability of no injury and only a small increase in the probability of death.

The effects of a decrease in vehicle age from ten years (the benchmark value) to five years are depicted in Fig. 3, and the effects of an increase in the age of the casualty from 33 years (the benchmark value) to 50 years are depicted in Fig. 4. Close scrutiny of Figs 1 and 3 reveals an extremely small rightward shift in the injury probability curves, so

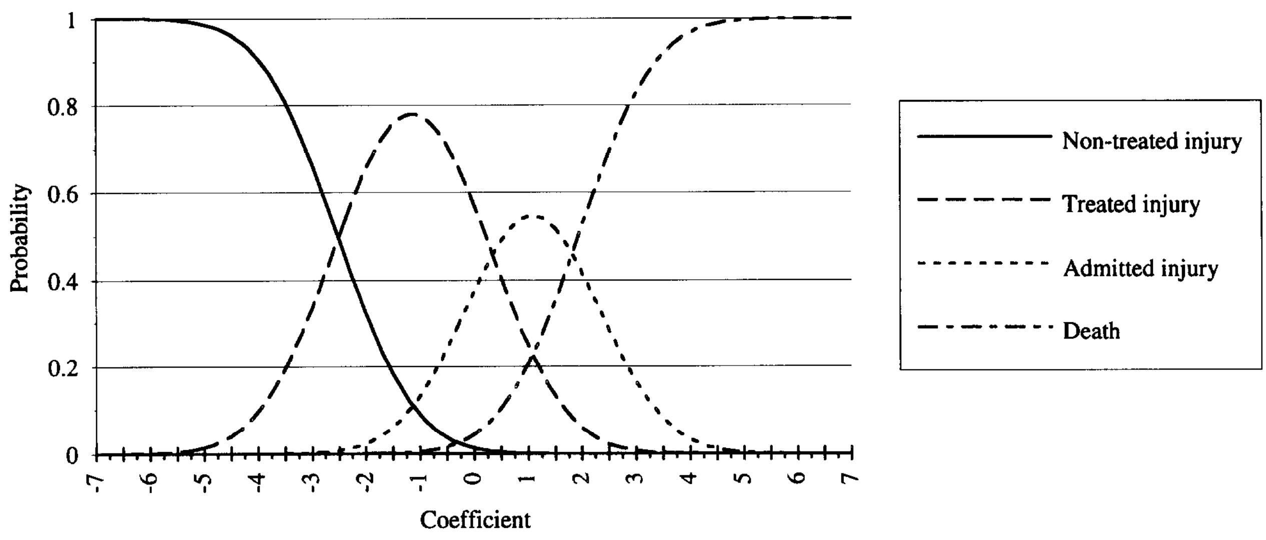

Fig. 2. Injury probabilities (speed = 100 kilometres per hour).

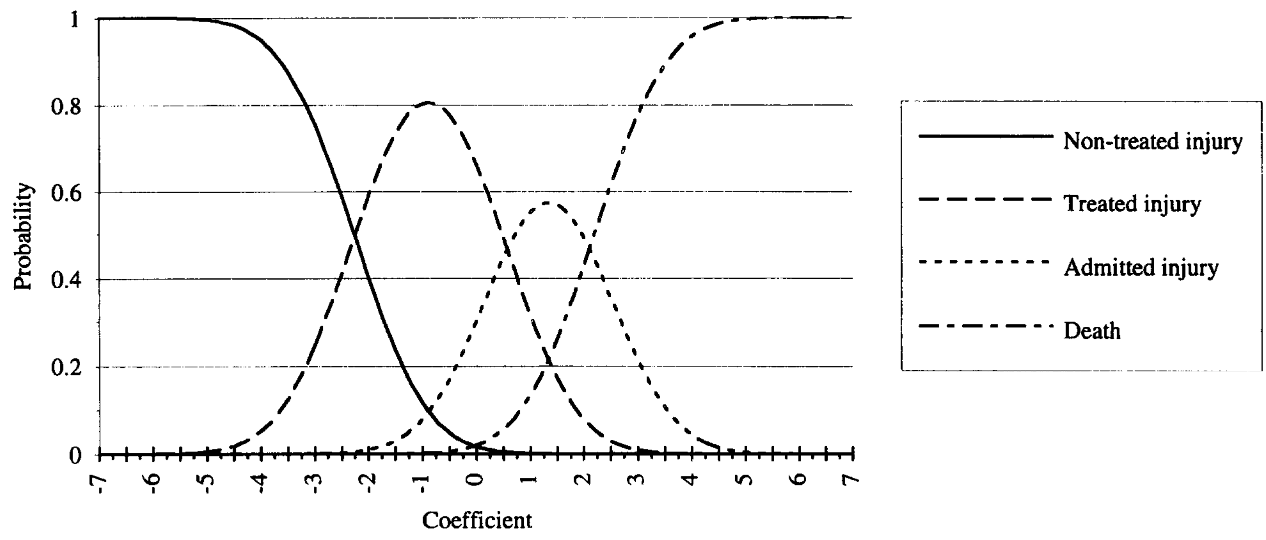

Fig. 3. Injury probabilities (vehicle age = 5 years).

small that the injury probabilities corresponding to a coefficient value of zero are the same as the benchmark probabilities when rounded to one decimal point: $\hat{P}_{1i}=0.0$, $\hat{P}_{2i}=0.7$, $\hat{P}_{3i}=0.3$ and $\hat{P}_{4i}=0.0$. Close scrutiny of Figs 1 and 4 reveals that an increase in the age of the casualty causes a leftward shift in the injury probability curves, and the estimated injury probabilities corresponding to a coefficient value of zero become $\hat{P}_{1i}=0.0$, $\hat{P}_{2i}=0.5$, $\hat{P}_{3i}=0.4$ and $\hat{P}_{4i}=0.1$. Thus, the effects on injury probabilities of large increases in age are significant, and even larger than the effects of a 58 kilometre per hour increase in speed. Also note that, as large as the effects of age might be, they are still quite small relative to the effects of changes in several dummy variables: recall, for example, that when an accident victim has all the attributes of a benchmark victim but is otherwise travelling in the left-rear seating position, the injury probabilities are estimated to be $\hat{P}_{1i}=0.0$, $\hat{P}_{2i}=0.2$, $\hat{P}_{3i}=0.6$ and $\hat{P}_{4i}=0.2$.

6. CONCLUSION

In this study we have estimated the parameters of two ordered multiple choice models using a large and highly disaggregated cross-section data set. The focus of our attention has been on the conditional probabilities of sustaining different levels of injury in a motor vehicle accident, and how these probabilities vary with variations in the attributes of road users. For illustrative purposes we defined the attributes of a 'benchmark' victim: a 33-year-old male driver of a

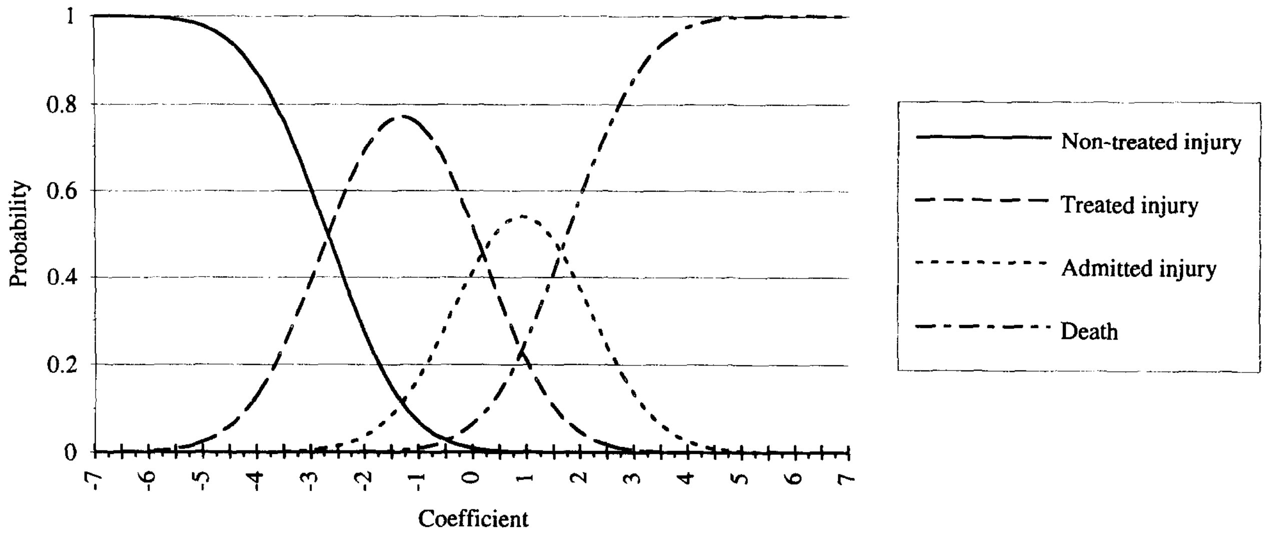

Fig. 4. Injury probabilities (age of casualty = 50 years).

10-year-old motor vehicle who is restrained by a seatbelt, has a blood alcohol reading less than 0.08, and is involved in a head-on collision while travelling at 42 kilometres per hour.

We estimated that a road accident victim with all the attributes of the benchmark victim will remain uninjured with a probability of almost zero, will require treatment from a medical officer with a probability of approximately 0.7, will be admitted to hospital with a probability of approximately 0.3, and will be killed with a probability of almost zero. We found that these probabilities change with changes in seating positions (the most dangerous seating position is the left-rear seating position), gender (females are more likely to sustain serious injuries than males), blood alcohol levels (victims with blood alcohol levels greater than 0.08 are more likely to sustain serious injuries than victims with blood alchohol levels of less than 0.08), seatbelt usage (failure to wear a seatbelt increases the probability of death), vehicle type (light trucks are more dangerous than any other type of motor vehicle), vehicle make (vehicles made by Manufacturer *L* are more dangerous than any other make of motor vehicle), and type of collision (vehicle–vehicle head-on collisions are more dangerous than any other type of collision). We also found that the probabilities of serious injury and death increase with increases in speed, the age of the vehicle, the age of the casualty and the time of the accident. Indeed, we found that the effects of increasing the age of the casualty from 33 to 50 years are greater than effects of an increase in vehicle speed from 42 to 100 kilometres per hour, and that these latter effects are surprisingly small. Finally, we found that the effects of changes in the age of the vehicle, the age of the casualty and speed are smaller than the effects of changes in many other road user attributes including seating positions and seatbelt use.

Although most of our findings are consistent with the findings of earlier studies, including the recent studies of Heulke and Compton (1995) and Kim *et al.* (1995), our findings in relation to seating position are slightly inconsistent with the findings of Evans and Frick (1988). We attribute this to differences in methodology, possible measurement errors in our data, and possible differences in United States and Australian vehicle designs. Further refinement of our methodology, and the replication of our study using alternative data sets, may help validate our results.

Two distinct but related methodological issues arose in our empirical work. First, our model selection criterion is statistically sound but is nevertheless *ad hoc*, and it is not clear how our results have been influenced by our choice of criterion and/or our choice of implementation method. Second, there is no generally accepted measure of goodness-of-fit for models of ordered multiple choice, so it has been difficult to judge just how well our models explain variations in injury severity. Research into these issues is straightforward but beyond the scope of the present paper.

Researchers contemplating these methodological issues, or seeking to refine our methodology in other ways, might also give some consideration to the functional form of $h_i = h(\mathbf{x}_i, \beta)$. In particular, consideration might be given to the use of a more general functional form which explicitly allows for interaction effects: failure to explicitly allow for interaction between the variables measuring seating position and type of collision may be one explanation for the differences between our own findings on seating position and the findings of Evans and Frick (1988). Researchers might also consider the use of a Fourier series to model the time-of-day effect, instead of the relatively inflexible quadratic functional form.

Acknowledgements—The authors wish to thank the referees of this journal for their helpful comments and suggestions.

REFERENCES

Austin K. The identification of mistakes in road accident records: part 1, locational variables. Accident Analysis and Prevention 27 (2):261–276; 1995a.

Austin K. The identification of mistakes in road accident records: part 2, casualty variables. Accident Analysis and Prevention 27 (2):277–282; 1995b.

Berndt E.; Hall B.; Hall R.; Hausman J. Estimation and inference in nonlinear structural models. Annals of Economic and Social Measurement 3/4:653–666; 1974.

Byrne P. J.; Gempesaw II C. M.; Toensmeyer U. C. An evaluation of consumer pesticide residue concerns and risk information sources. Southern Journal of Agricultural Economics 23 (2):167–174; 1991.

Cohen M. A.; Tell E. J.; Wallack S. S. Client-related risk factors of nursing home entry among elderly adults. Journal of Gerontology 41:785–792; 1986.

Connolly M. A.; Kimball A. W.; Moulton L. H. Alcohol and traffic safety: a sensitivity analysis of data from composite sources. Accident Analysis and Prevention 12:1–31; 1989.

Cooper P. J. Differences in accident characteristics among elderly drivers and between elderly and middle-aged drivers. Accident Analysis and Prevention 22:499–508; 1990.

Evans L. Fatality risk reduction from safety belt use. Journal of Trauma 27 (7):746–749; 1987.

Evans L. The fraction of traffic fatalities attributable to alcohol. Accident Analysis and Prevention 22:587–602; 1990.

Evans L. Traffic safety and the driver. New York: Van Nostrand Reinhold; 1991.

Evans L.; Frick M. C. Seating position in cars and fatality risk. American Journal of Public Health 78 (11):1456–1458; 1988.

Evans L.; Frick M. C. Car size or car mass: which has greater influence on fatality risk. American Journal of Public Health 82 (8):1105–1112; 1992.

Evans L.; Frick M. C. Alcohol's effect on fatality risk from a physical insult. Journal of Studies on Alcohol 54:441–449; 1993.

Evans L.; Frick M. C. Car mass and fatality risk: has the relationship changed. American Journal of Public Health 84 (1):33–36; 1994.

Greene W. H. Econometric Analysis (2nd ed.). New York: Macmillan; 1993.

Greene W. H. LIMDEP Version 6.0. New York: Econometric Software Inc; 1992.

Hausman J. A.; Lo A. W.; MacKinlay A. C. An ordered probit analysis of transaction stock prices. Journal of Financial Economics 31 (3):319–379; 1992.

Heulke D. F.; Compton C. P. The effects of seat belts on injury severity of front and rear seat occupants in the same frontal crash. Accident Analysis and Prevention 27:835–838; 1995.

Judge G. G.; Griffiths W. E.; Hill R.C.; Lütkepohl H.; Lee T. C. The Theory and Practice of Econometrics (2nd ed.). New York: John Wiley and Sons; 1985.

Kim K.; Nitz L.; Richardson J.; Li L. Personal and behavioural predictors of automobile crash and injury severity. Accident Analysis and Prevention 27:469–481; 1995.

Laberge-Nadeau C.; Maag U.; Borbeau R. The effects of age and experience on accidents with injuries: should the licensing age be raised? Accident Analysis and Prevention 24:107–116; 1992.

Lassarre S. The introduction of the variables 'traffic volume', 'speed' and 'belt-wearing' into a predictive model of the severity of accidents. Accident Analysis and Prevention 18:129–134; 1986.

Levy D. T. Youth and traffic safety: the effects of driving age, experience, and education. Accident Analysis and Prevention 22:327–334; 1990.

Lloyd C. J. Alcohol and fatal road accidents: estimates of risk in Australia 1983. Accident Analysis and Prevention 24:339–348; 1992.

Mannering F. L. Male/female driver characteristics and accident risk: some new evidence. Accident Analysis and Prevention 25:77–84; 1993.

Mayhew D. R.; Donelson A. C.; Beirness D. J.; Simpson H. M. Youth, alcohol and relative risk of crash involvement. Accident Analysis and Prevention 18:273–287; 1986.

Meyer D.; Cooke W. Economic and political factors in formal grievance resolution. Industrial Relations 27 (3):318–335; 1988.

O'Day J. Accident data quality. National Cooperative Highway Research Program: Synthesis of Highway Practice 192. Transportation Research Board. Washington DC: National Academy Press; 1993.

Paterson R., Baxter D. Coding manual for traffic accident information (users edition). Sydney: Traffic Authority of New South Wales; 1987.

Rosman D. L.; Knuiman M. W. A comparison of hospital and police road injury data. Accident Analysis and Prevention 26 (2):215–222; 1994.

Russell L. M.; Rives N. W. Household migration plans: a multivariate probit model. Sociological Methods and Research 8:95–109; 1979.

Schwarz G. Estimating the dimensions of a model. The Annals of Statistics 6 (2):461–464; 1978.

Shibata A.; Fukuda K. Risk factors of fatality in motor vehicle traffic accidents. Accident Analysis and Prevention 26 (3):391–397; 1994.

Shinar D.; Treat J. R.; McDonald S. T. The validity of police reported data. Accident Analysis and Prevention 15:175–191; 1983.

Spizman L. M.; Kane J. Loss of future income in the case of personal injury of a child: parental influence on a child's future earnings. Journal of Forensic Economics 5:159–168; 1992.

Stewart R. J. Estimating the effects over time of alcohol on injury severity. Accident Analysis and Prevention 21:575–579; 1989.

Veall M. R.; Zimmerman K. F. Pseudo-R^2s in the ordinal probit model. Journal of Mathematical Sociology 16 (4):333–342; 1992.

Vingilis E.; Liban C. B.; Blegfen H.; Colbourne D.;

Reynolds D. Introducing beer sales at a Canadian ball park: the effect on motor vehicle accidents. Accident Analysis and Prevention 24:521–526; 1992.
Waller P. F.; Stewart J. R.; Hansen A. R.; Stutts J. C.; Popkin C. L.; Rodgman E. A. The potentiating effects of alcohol on driver injury. Journal of the American Medical Association 256:1461–1466; 1986.
Waterman D.; Weiss A. A. The effects of vertical integration between pay cable networks and cable television systems. Mimeo 1992.