Highway Accident Data Analysis: Alternative Econometric Methods

PATRICK S. MCCARTHY AND SAMER MADANAT

The past decade has seen significant advances in econometric modeling including the analysis of disaggregate data, the structure of discrete response models, the treatment of simultaneity in linear models, the specification of models based on pooled time series–cross sectional data, and the estimation of models in truncated and censored samples. Furthermore, the data sets available in the field of highway safety include significant amounts of detailed information. However, to date, highway safety analyses using these data sets have not fully exploited state-of-the-art econometric methodologies. The applicability of recently developed econometric methods to the field of highway safety analysis is illustrated. It is anticipated that such applications will improve the accuracy of traffic accident models and lead to more effective policies and investment decisions in the area of highway safety.

Highway safety is an area of research characterized by a disparity between data collection and data analysis. At the state and federal level, significant amounts of detailed information are routinely collected on highway traffic accidents. The amounts and types of data collected are of particular interest to the research community because they enable the researcher to investigate aspects of highway safety using state-of-the-art statistical and econometric methodologies. However, despite significant econometric advances during the past decade that potentially have important implications for improving understanding of factors that affect highway safety, there has been relatively little research identifying and evaluating the potential gain from these new methodologies. This paper constitutes a small step in this direction.

In the following sections, several areas, including policy endogeneity, cross sectional heterogeneity, and small numbers problems, are identified that illustrate problems with existing methodologies and offer alternative econometric techniques to correct the problem. In addition, other econometric issues including sample truncation and ordinality of accident severity data are discussed to illuminate often implicit assumptions associated with existing methodologies.

POLICY ENDOGENEITY AND HIGHWAY SAFETY

Consider the following equation:

$$y_t = \alpha + \beta' x_t + e_t \qquad t = 1, \ldots, T \tag{1}$$

where

y_t = highway safety outcome (e.g., fatality rate),
x_t = vector of k explanatory (exogenous) variables,
e_t = error with mean 0 and constant variance
α = parameter, and
β = parameter vector with k elements.

Assuming away problems of heteroscedasticity and autocorrelation, ordinary least squares (OLS) estimates of the unknown parameters will be best linear unbiased estimates (BLUE). Assume that the kth explanatory variable reflects a policy that was enacted to enhance highway safety (e.g., speed limits laws, mandatory seat belt use laws, minimum drinking age laws, etc.). If the policy was truly exogenous (the original reduction of speed limits in 1975 was a response to the oil crisis instead of to highway safety concerns), then the resulting parameter estimates will be BLUE. Alternatively, however, suppose that the policy was a reaction to concerns about highway safety. Then x_{kt} is itself a function of a set of explanatory variables including v_t. For example, reluctance in the United States to increase speeds after the oil crisis ended was a response to the life-saving effects of the lower speed limit. In this case, Equation 1 is actually a system of two equations that can be expressed as

$$y_t = \alpha + \beta' x_t + e_t \qquad t = 1, \ldots, T \tag{2a}$$

$$x_{kt} = \gamma + \delta' z_t + \Phi y_t + u_t \qquad t = 1, \ldots, T \tag{2b}$$

where

x_{kt} = kth explanatory variable in x_t,
z_t = vector of k' explanatory variables,
u_t = error term with mean 0 and constant variance,
γ = constant term,
δ = vector of k' parameters, and
ϕ = parameter of the endogenous variable y_t.

If the estimation sample is a time series data set, then one could apply Granger causality tests to check for endogeneity between y_t and x_{kt}. Granger (*2*) and Sims (*3*) developed tests to evaluate the direction of causality. To test the hypothesis that "x_{kt} does not cause y_t," regress y_t on lagged values of y_t and lagged values of x_{kt}; run a second regression of y_t on lagged values of y_t only. An F-test based upon the error sum of squares in the unrestricted and restricted regressions, respectively, can be used to test the null hypothesis. A similar set of regressions is run with x_{kt} as the dependent variable and lagged values of x_{kt} and y_t as the explanatory variables. In this case, the null hypothesis is "y_t does not cause x_{kt}." To conclude that "x_{kt} causes y_t," it is necessary that the null hypothesis is rejected in the first set of regressions and accepted in the second set of regressions.

If the true state of the world is Equation 2 but the structural relationship between x_{kt} and y_t is ignored, then the OLS estimates of Equation 1 will produce biased and inconsistent parameter estimates (*1*). To avoid the endogeneity bias, the analyst typically

P. McCarthy, Department of Economics; S. Madanat, Department of Civil Engineering, Purdue University, West Lafayette, In 47907.

implements either of two strategies, depending upon the objective. First, the analyst wants to capture the influence of y_t on x_{kt} but the structural relationship between x_{kt} and y_t is not important. In this case, x_{kt} in Equation 2b is simply substituted in Equation 2a and the resulting reduced form equation is estimated. It is important to recognize, however, that in this case the parameter estimates are not capturing the structural impacts of the explanatory variables on highway safety but instead the combined effects of these variables from the system of equations. Second, if one is concerned with the structural effect that a particular (endogenous) policy has on highway safety policy, then the simultaneous structure must be estimated and the structural parameters recovered.

To illustrate, consider the effect that recent relaxation of rural Interstate highway speed limits has on highway safety. Although enacting higher speed limits reflects a state's underlying demand for higher limits, it is also likely that a state's demand for higher speeds depends on the extent to which the affected roads are currently safe. Define speed law to be a variable that reflects the relaxation of rural interstate speed limits. If this variable is entered in an estimating equation for highway safety as a dummy variable that equals 0 in the 55 mph environment and 1 in a 65 mph post-law environment, the resulting estimates could be biased if enactment of relaxed speed limits were, at least in part, a response to changes in highway safety on the 55 mph roads. If so, the relaxed speed limit on rural Interstates is endogenous and the coefficient estimate on a speed limit dummy variable is biased.

To account for endogenous effects of highway safety on speed limit legislation, consider the following system of three equations (*4*):

$$y = \beta_1 x_1 + \alpha(\text{speed law}) + \gamma A + e_1$$

$$A = \beta_2' x_2 + e_2$$

$$A^* = A + \kappa(y^*) \tag{3}$$

where

y = measure of highway safety,
y^* = y in the absence of the relaxed speed limit,
A = latent variable reflecting attitudes towards relaxed speed limits
A^* = latent variable that reflects the demand for speed limit relaxation,
speed limit = dummy variable that equals 1 when the speed limit was relaxed and 0 otherwise,
x_1 = vector of explanatory variables for highway safety,
x_2 = vector of explanatory variables for A, and
e_i ($i = 1,2$) = error term. The relationship between A^* and speed law is given by the following:

$$A^* = A + \kappa(y^*) > 0 => \text{speed law} = 1$$

$$A^* = A + \kappa(y^*) < 0 => \text{speed law} = 0$$

Thus, the hypothesis is that highway safety depends upon a set of explanatory variables, x_1, the speed limit, and the state's preferences for higher speeds, which depend upon a set of explanatory variables, x_2. The state's demand for higher speeds, in turn, reflects its attitudes toward higher speeds and the incidence of accidents, y^*, in the lower speed environment. This produces two estimating equations

$$y = \beta_1' x_1 + \gamma\,(\beta_2' x_2) + \alpha(\text{speed law}) + u_1 \tag{4}$$

and

$$A^* = \beta_2' x_2 + \kappa[\beta_1' x_1 + \gamma(\beta_2' x_2)] + u_2$$
$$= \kappa(\beta_1' x_1) + (1 + \kappa\gamma)\,\beta_2' x_2 + u_2 \tag{5}$$

where $u_1 = e_1 + \alpha e_2$ and $u_2 = e_2 + \kappa e_1$. Estimation is a two-step process. First, Equation 5 is estimated where A^* is replaced with speed law. The predicted value of the dependent variable in Equation 5, speed law, gives the demand for relaxing the speed limit and replaces speed law in the estimating Equation 4. Note that the full set of structural parameters can be recovered. From Equation 4, estimates of β_1' and α are obtained; dividing the coefficient of x_1 in Equation 5 by the coefficient of x_1 in Equation 4 gives κ; dividing the coefficient of x_2 in Equation 4 by that in Equation 5, and knowing κ, enables one to solve for γ; from Equation 4, knowledge of γ produces β_2'.

In the estimations, the components of x_1 would include standard determinants of highway safety (e.g., young drivers per capita, alcohol consumption, per capita income, and time trend). x_2 reflects a state's attitudes towards raising the speed limit and could be measured by two variables: excess speed, defined as the extent to which observed average speeds on 55-mph roads exceed the 55 mph limit, and speed variance, the variance of speed on 55-mph roads. Since the primary benefit of an increased speed limit is travel time savings, observed average speeds that are above the mandated 55-mph limit are tantamount to the driving population revealing its demand for higher speeds. This suggests that excess speed is positively correlated with drivers' sentiments toward raising the speed limit. On the other hand, authorities have a responsibility for providing safe driving environments and will not be inclined to raise the speed limit if it is believed to compromise highway safety. Thus, the net effect of excess speed on the demand for raising the speed limit is ambiguous and depends upon the magnitudes of these two effects.

Lave (*5*) has shown that increases in speed variance, all else constant, reduce highway safety. Consistent with this, the demand for raising the speed limit would be expected to be negatively related to speed variance.

Using this methodology, Saffer and Grossman (*6*) estimate a model in which highway safety and a state's drinking age policy are endogenous. McCarthy and Ziliak (*7*) use a similar framework to analyze the simultaneity between highway safety and the formation of Mothers Against Drunk Driving chapters.

TRUNCATION

Most analyses of the policy effects on highway safety base these results upon an OLS model (simple or reduced form) in which the dependent variable is some measure of highway safety. Because highway safety policy strives to reduce the incidence of the most serious accidents, namely, those involving a fatality, a frequently used measure of highway safety is some function of highway fatalities (fatal accidents, fatality rate, fatalities per capita.) By limiting the analysis to accidents involving a fatality, the sample is truncated from below because it excludes observations on all individuals who have experienced nonfatal accidents in the sample period. Thus, the estimates of the effect of a policy on highway safety are likely to be biased. Figure 1 illustrates this graphically. By excluding those accidents below severity level SC, the effect of increasing speed on highway safety is seen in the figure to bias the slope parameter downward and the intercept upward.

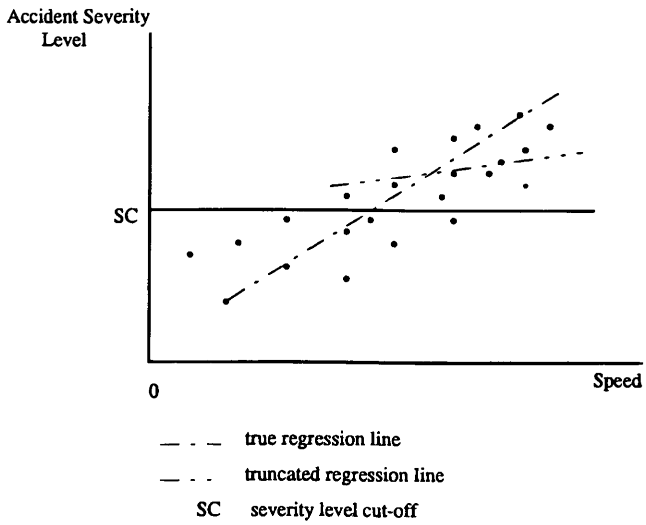

FIGURE 1 Potential bias from truncation.

To quantify the potential bias, consider a problem in which the dependent variable y is highway safety, x is a vector of K independent variables and β is a vector of parameters to be estimated. The underlying sample is a cross section of states or counties. Oftentimes, an OLS model of the following form is estimated

$$y_i = \beta' x_i + u_i \qquad i = 1, \ldots, \mathrm{N}$$

where u_i is a normally distributed error term with 0 mean and constant variance. Let y_i be measured as the number of fatalities or the fatality rate for Cross Section i. Define SC to be the level of truncation (e.g., AIS severity level) such that all accidents for which $y_i \leq \mathrm{SC}$ (e.g., AIS $\leq$ 4) are eliminated. The density function for the truncated variable y_i is

$$g(y_i) = \frac{(1/\sigma)\phi[(y_i - \beta' x_i)/\sigma]}{1 - \Phi\,[(\mathrm{SC} - \beta' x_i/\sigma)]} \qquad y_i > \mathrm{SC}$$
$$= 0 \qquad \text{otherwise} \tag{6}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the density function and the distribution function of the standard normal, respectively (*4*). The log-likelihood for this function is given as

$$\log \mathrm{L} = -\mathrm{N} \ln[(2\pi)^{\frac{1}{2}}\sigma] - (\tfrac{1}{2}) \sum_n \left(\frac{y_i - \beta' x_i}{\sigma} \right)^2 - \sum_n \ln\left[1 - \Phi\left(\frac{\mathrm{SC} - \beta' x_i}{\sigma}\right)\right]$$

which can be shown to be globally concave. Standard Newton-Raphson techniques can be used to obtain the maximum likelihood parameter estimates for β and σ. Once estimated, the parameter estimates are used to obtain the conditional mean and variance of y_i. In particular,

$$E(y_i | y_i > \mathrm{SC}) = \beta' x_i + \sigma\lambda(t_i)$$
$$V(y_i | y_i > \mathrm{SC}) = \sigma^2[1 - \delta(t_i)] \tag{7}$$

where $\lambda(t_i) = \phi(t_i)/1 - \Phi(t_i)$ and $\delta(t_i) = \lambda(t_i)(\lambda(t_i) - t_i)$ [which lies in the open unit interval (*1*)] and $t_i = \mathrm{SC} - \beta' x_i/\sigma$. Thus, the bias is $\sigma\lambda\,(t_i)$, which can be shown to be increasing in SC. By excluding accidents with severity levels below SC, the effect of explanatory variables, including policy variables, on highway safety will be biased. To determine the marginal effect of an increase in x_{ki} ($i = 1, \ldots, I$; $k = 1, \ldots, K$), differentiate the conditional mean with respect to x_{ki}. This gives

$$\text{marginal effect of } x_{ki} = \beta_k[1 - \delta(t_i)] \tag{8}$$

which is less than β_k since $\delta(t_i)$ lies between 0 and 1. In the subpopulation, the marginal effect of each of the k explanatory variables on y_i is less than the coefficient β_k.

To illustrate the potential importance of this bias, return to the graph showing the effect of increasing speeds on highway safety. If the analyst is concerned only with the subpopulation of accidents above severity level SC, then the marginal effect identified in Equation 8 is the relevant effect; alternatively, if the objective is to identify the effect for the entire population, then β_k is the relevant effect. Thus, an inference on the highway safety effects of higher speeds drawn from an analysis based upon a subpopulation of fatal accidents will understate the effect if applied to the entire population.

In the literature, criteria for truncation include accident severity (*8–10*), age of driver (*11–14*), alcohol involvement (*15–17*), number of vehicles involved (*18*), and vehicle size (*19–22*).

ENDOGENOUS STRATIFICATION

A related problem is endogenous stratification. As indicated, most models identify fatalities (or fatality rate, fatal accident rate) as a measure of highway safety to the general exclusion of other accidents (serious injury, minor injury, and property damage accidents). In this case, the sample is stratified according to accident severity. Lower severity accidents are often undersampled or completely absent. A generalization of the truncated sample is to analyze a model that identifies various severity strata and their sampled proportions. This may be an important stratagem for reducing costs because of the size of accident data files (statewide as well as nationally). In California, for example, there are more than 500,000 accidents annually.

To illustrate, suppose that a researcher uses the regression model $y_i = \beta' x_i + u_i$ ($i = 1, \ldots, N$) to identify the determinants of statewide highway safety. u_i is normally distributed with mean 0 and variance σ^2. Instead of obtaining the complete set of accident records, the analyst takes a p_1 and p_2 percent sample of fatal ($y_i > \mathrm{SC}$) and nonfatal accidents ($y_i \leq \mathrm{SC}$) respectively. Note that for the truncated model in the previous section, p_1 percent = 1 and p_2 percent = 0.

The density function for y_i is now given as (*4*)

$$g(y_i) = \frac{p_1 f(y_i)}{p_1 Pr(y_i > S) + p_2 Pr(y_i \leq S)} \qquad y_i > \mathrm{SC}$$
$$= \frac{p_2 f(y_i)}{p_1 Pr(y_i > S) + p_2 Pr(y_i \leq S)} \qquad y_i \leq \mathrm{SC} \tag{9}$$

where $f(y_i)$ is the density of y_i in the population. Substituting for $f(y_i)$ and t_i (defined as in the previous section) gives

$$g(y_i) = \frac{p}{p + (1-p)\Phi(t_i)} \frac{1}{\sigma} \phi\left(\frac{y_i - \beta' x_i}{\sigma}\right) \qquad y_i > SC$$
$$= \frac{1}{p + (1-p)\Phi(t_i)} \frac{1}{\sigma} \phi\left(\frac{y_i - \beta' x_i}{\sigma}\right) \qquad y_i \leq SC \qquad (10)$$

where $p = (p_1/p_2)$. These expressions can be used to form the log-likelihood function, which is then maximized with respect to β and σ (if p is not known, the log-likelihood function can be maximized with respect to p as well). The conditional means for this model are

$$E(y_i|x_i, y_i > SC) = \beta' x_i + E(u_i|y_i > SC)$$
$$= \beta' x_i + \sigma \frac{\phi(t_i)}{1 - \Phi(t_i)} \qquad (11a)$$

$$E(y_i|x_i, y_i \leq S) = \beta' x_i + E(u_i|y_i \leq S)$$
$$= \beta' x_i - \sigma \frac{\phi(t_i)}{\Phi(t_i)} \qquad (11b)$$

and the unconditional mean is a weighted average of the conditional means

$$E(y_i|x_i) = \beta' x_i + \sigma \frac{p_1\phi(t_i) - p_2\phi(t_i)}{p_1[1 - \Phi(t_i)] + p_2\Phi(t_i)}$$
$$= \beta' x_i + \sigma\,\gamma(t_i) \qquad (12)$$

Similar to the comments made in the previous section, depending on whether the analyst is concerned about the marginal effect of x_{ki} on the estimation subpopulation or its effect on the entire population, the appropriate marginal effect is obtained by differentiating Equations 11 and 12, respectively, with respect to x_{ki}. It can be shown that the marginal effect of the unconditional mean with respect to x_{ki} is $\beta_k\,[1 - \delta'(t_i)]$ where $\delta'(t_i)$ equals $\gamma(t_i)\,[\gamma(t_i) - t_i]$, which is similar to the expression for $\delta(t_i)$ given below Equation 7. Also note that if $p_1 = 1$ and $p_2 = 0$, then the marginal effect obtained from Equation 12, $\beta_k\,[1 - \delta'(t_i)]$ is identical to the conditional marginal effect (Equation 8) for the truncated model in the previous section. That is, $\delta'(t_i) = \delta(t_i)$.

Although this would appear to be a useful procedure for obtaining meaningful highway safety results while reducing the effort and computational burden associated with analyzing statewide or national accident records, the authors are not aware of any studies in the highway safety literature that use this methodology.

CROSS SECTION–TIME SERIES

Continuous Dependent Variable

State and national highway agencies routinely collect highway safety data that are organized into monthly and annual reports. In that these reports often discriminate by state, by county within state, by type of road, by various socioeconomic characteristics, and along numerous other dimensions, the information represents a panel of data—a time series of data across a set of cross section units.

When a time series of cross sections is available, ordinary least is generally not appropriate because it ignores the heterogeneity in the cross-sectional units. There are generally two methodologies for estimating panel data. First, a fixed-effects approach includes dummy variables for each of the cross-sectional units. This model assumes that differences between cross-sectional units can be captured by a parametric shift in the regression line. If all cross-sectional units are represented in the sample (e.g., all states in the nation, all vehicle types, all times of day), then a fixed-effects approach may be appropriate because it embodies all the differences among the cross-sectional units. However, if the cross sections represent a sample from a larger population (e.g., a subset of states), then it may be more appropriate to assume that the cross-sectional heterogeneity is randomly distributed across cross-sectional units. This latter approach represents a random effects specification.

In general, a cross section–time series model can be expressed as

$$y_{it} = \alpha + \sum_{j=1}^{k} \beta_j x_{it,j} + \epsilon_{it} + \eta_i \qquad (13)$$

where

- y_{it} $(i = 1, \ldots, N; t = 1, \ldots, T)$ = highway safety outcome for cross section i and time period t,
- $x_{it,j}$ $(i = 1, \ldots, N; t = 1, \ldots, T; j = 1, \ldots, k)$ = jth explanatory variable for cross section i and time period t,
- α = constant term and $\beta_j (j = 1, \ldots, k)$ is a parameter that reflects the marginal effect of the jth explanatory variable on the highway safety outcome,
- ϵ_{it} = error term associated with cross section i and time period t with mean 0 and constant variance, and
- η_i = term specific to cross section i.

In the absence of any cross-sectional heterogeneity, η_i is equal to 0, and OLS is used to estimate the model. For a fixed-effects specification, η_i is a parameter that is estimated along with β_j, where η_i represents a parallel shift in the regression line for cross-section unit i. In a random effects model, η_i is assumed to be a random term with mean 0 and constant variance that is specific to cross section unit i. Notice that cross-sectional heterogeneity is confined to the error term in the random effects model, whereas in the fixed effects model it is explicitly represented as a parametric shift in the regression line.

In general, there are advantages and disadvantages to either approach. A fixed-effects specification entails a potentially large decrease in degrees of freedom if there are a high number of cross sections in the sample. In addition, fixed effects models cannot be estimated if any of the explanatory variables is constant throughout the sample period. On the other hand, if the fixed effects parameters are correlated with the included variables but omitted from the model, then a random effects specification leads to biased parameter estimates (*23*). Hausman (*24*) developed a specification test, based on a chi-squared statistic, to test the null hypothesis that the cross section–specific parameters in a fixed effects model are independent of the included explanatory variables. Accepting the null hypothesis would be consistent with a random effects specification, whereas rejecting the null hypothesis would argue for a fixed effects specification.

There have been a number of recent examples in the literature (*14,16,17*) of panel data analyses using accident data.

Discrete Dependent Variable

An interesting variation of that problem occurs when the dependent variable takes on very small integer values. For example, suppose a researcher is interested in modeling the incidence of countywide alcohol-related fatal accidents or countywide fatal accidents among teenagers. If this study were undertaken by state transportation departments, it is likely that in many states there would be a large number of counties in which very few or no fatal accidents occurred. As an example, consider Indiana, which has 92 counties. In 1989, there were 99 alcohol-related fatal accidents and 2923 alcohol-related injury accidents statewide, which represents an average of just over 1 and 31 alcohol-related fatal accidents and injury accidents, respectively.

One methodology for modeling these accidents is to estimate a logit model that defines the dependent variable y_{it} to be one if an accident in cross section i and time period t involved a fatality and 0 otherwise. In particular, the probability of a fatal accident is given by

$$P(\text{fatal accident}) = P(y_{it} = 1) = \frac{e^{\beta' x_{it}}}{1 + e^{\beta' x_{it}}}$$

$$i = 1, \ldots, N; t = 1, \ldots, T \quad (14)$$

Similar to the continuous case discussed, if the cross-section units are heterogeneous and the heterogeneity is ignored, then estimating Equation 14 will lead to inconsistent parameter estimates.

Consider an alternative model that incorporates cross-sectional parameters, α_i ($i = 1, \ldots, N$) to reflect the underlying heterogeneity. Then Equation 14 becomes

$$P(\text{fatal accident}) = P(y_{it} = 1) = \frac{e^{\alpha_i + \beta' x_{it}}}{1 + e^{\alpha_i + \beta' x_{it}}}$$

$$i = 1, \ldots, N; t = 1, \ldots, T \quad (15)$$

For large N and small T (≤ 5) Chamberlain (25) devised a method for estimating this model that is based on conditional maximum likelihood functions that do not depend on the heterogeneity parameters. Moreover, on the basis of a Hausman test of the null hypothesis that the cross-section units are homogeneous, it is possible to test a standard logit specification in Equation 14 against the alternative specification given by Equation 15.

To date, the authors are aware of no studies in highway safety using a panel logit methodology.

MODELS WITH ORDINAL DEPENDENT VARIABLES

Accident data generally obtained from police records are disaggregate data. However, when such data are used to analyze the effect of various factors on accident severity, they are usually aggregated and analyzed by using classical statistical methods such as multivariate regression. These methods are limited to the analysis of continuous variables, such as the total number of accidents, or the total number of fatalities and hence require that the data on individual accidents be aggregated before analysis. This is especially the case when accident data are recorded on an ordinal instead of a continuous scale. For example, the National Safety Council (26) devised a scheme for injury classification—no injury, possible injury, non incapacitating injury, incapacitating injury, and fatal injury.

With such a scale, an order is established between different categories of injury, but the distance between any two numbers on the scale is of unknown size. As such, these data cannot be analyzed by using traditional statistical methods, except by aggregation and subsequent loss of information.

In a recent paper, Nassar et al. (27) proposed using a sequential logit approach for modeling accident severity using disaggregate accident data. Such a model structure implies that an accident moves up the scale of severity, starting from the least severe. After each move, the accident either moves up one more notch or stays at its current level of severity. By assuming independence across error terms of the different logit models, the authors end with a model that is a product of binary logits. Each logit model is of the form

$$P(S_m | S_{@m-1}) = \frac{\exp(\sum_j \beta_{jm} X_{jm})}{1 + \exp(\sum_j \beta_{jm} X_{jm})} \quad (16)$$

$P(S_m | S_{@m-1})$ = probability of experiencing injury severity level m given that the impact is sufficient to produce at least an injury of severity level $m - 1$.
X_{jm} = impact of factor j on severity level m.
β_{jm} = coefficient associated with factor j on severity level m.

With such a modeling approach, the richness of information available at the disaggregate level is exploited. Sequential choice models, however, are restrictive in the sense that they assume independence of the error terms across moves, for each accident, which may be an unrealistic assumption.

To relax the assumption of independence that the sequential logit approach imposes on the error terms, models with ordinal dependent variables such as the ordered logit should be used. Such an approach was specifically developed for models in which the dependent variable is ordinal, such as the accident severity ratings described (28). These models do not assume that the observed rating is the result of a sequence of move-ups; instead, the assumption is that the ratings represent a discretization of an underlying latent severity scale that is continuous. By using such an approach, it is possible to estimate jointly the parameters of the different severity factors and the thresholds that separate the successive severity ratings on the underlying latent scale. Mathematically, let the continuous underlying accident severity be denoted by y^*. Then, we have that

$$y^* = \sum_j \beta_j X_j + \epsilon$$

where

X_j = impact factor j,
β_j = coefficient associated with factor j, and
ϵ = random disturbance.

The process giving rise to the observed severity levels S_m ($m = 1, \ldots, M$) may be viewed in terms of y^* crossing some of the $M - 1$ threshold values. Specifically, we have that

$$m = 1 \text{ if } -\infty < \sum_j \beta_j X_j + \epsilon < t_1^*$$

$$m = 2 \text{ if } t_1^* < \sum_j \beta_j X_j + \epsilon < t_2^*$$

•

•

•

$$m = M \text{ if } t_{M-1}^* < \sum_j B_j X_j + \epsilon < +\infty$$

Thus, the probability of observing an injury severity S_m is

$$P(S_m) = P(t_{m-1}^* < \sum_j \beta_j X_j + \epsilon < t_m^*)$$
$$= P(\epsilon < t_m^* - \sum_j \beta_j X_j) - P(\epsilon < t_{m-1}^* - \sum_j \beta_j X_j)$$

If the error terms ϵ's are independently and identically distributed logistically, then the probabilities of various severity are given by an ordered logit model (*28*).

CONCLUSION

An overview has been presented of the potential application of some recent developments in econometric methodology to the field of highway safety analysis. Although no empirical work was presented, the data required to perform the analyses discussed are readily available to highway safety researchers.

In addition to the presented methodologies, there have been other modeling techniques, including empirical Bayesian analysis and Poisson methodologies (and variants thereof), which have been successfully although not frequently used to study the effect of traffic improvements at highway intersections (*29–32*). However, because these techniques are more familiar to traffic safety analysts than those identified here, they have been omitted from the overview.

Because major policy and investment decisions are often made by state and federal agencies on the basis of the results of highway safety analyses, the importance of accuracy in such analyses can hardly be overemphasized. By using state-of-the-art econometric methods such as those described herein, researchers can improve the level of accuracy in highway safety analysis.

REFERENCES

1. Greene, W. H. *Econometric Analysis.* Macmillan Publishing Company, New York, 1990.
2. Granger, C. W. J. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica*, Vol. 37, 1969, pp. 424–438.
3. Sims, C. A. Money, Income, and Causality. *American Economic Review*, Vol. 62, 1972, pp. 540–552.
4. Maddala, G. *Limited Dependent and Qualitative Variables in Econometrics.* Cambridge University Press, Cambridge, England, 1983.
5. Lave, C. Speeding, Coordination, and the 55 MPH Speed Limit. *American Economic Review* Vol. 75, 1985, pp. 1159–1164.
6. Saffer, H., and M. Grossman. Drinking Age Laws and Highway Mortality Rates: Cause and Effect. *Economic Inquiry*, Vol. 25, 1987, pp. 403–417.
7. McCarthy, P., and J. Ziliak. The Effect of MADD on Drinking-Driving Activities—An Empirical Study. *Applied Economics*, Vol. 9, 1990, pp. 1215–1227.
8. Crandall, R., et al. *Regulating the Automobile.* The Brookings Institution, Washington D.C., 1986.
9. Graham, J., and S. Garber. Evaluating the Effects of Automobile Safety Regulation. *Journal of Policy Analysis and Management*, Vol. 3, 1984, pp. 206–221.
10. Zlatoper, T. Regression Analysis of Time Series Data on Motor Vehicle Deaths in the United States. *Journal of Transport Economics and Policy*, Sept. 1984, pp. 263–274.
11. Williams, A., et al. The Effect of Raising the Legal Minimum Drinking Age on Youthful Auto Fatalities. *Journal of Legal Studies*, Vol. 12, 1983, pp. 169–179.
12. Cook, P., and G. Tauchen. The Effect of Minimum Age Drinking Age Legislation on Youthful Auto Fatalities, 1970–77. *Journal of Legal Studies*, Vol. 13, 1984, pp. 169–190.
13. Males, M. The Minimum Purchase Age for Alcohol and Young-Driver Fatal Crashes: A Long-Term View. *Journal of Legal Studies*, Vol. 15, 1986, pp. 181–217.
14. Saffer, H., and M. Grossman. Beer Taxes, the Legal Drinking Age, and Youth Motor Vehicle Fatalities. *Journal of Legal Studies*, Vol. 16, 1987, pp. 351–374.
15. Wilkinson, J. Reducing Drunken Driving: Which Policies Are Most Effective? *Southern Economic Journal*, Vol. 54, 1987, pp. 322–334.
16. Chaloupka, F., et al. *Alcohol Control Policies and Motor Vehicle Fatalities.* Working Paper 3831, National Bureau of Economic Research, Inc., 1991.
17. McCarthy, P. The Effect of Higher Rural Interstate Speed Limits on Alcohol-Related Accidents. *Journal of Health Economics*, Vol. 11, 1993, pp. 281–299.
18. Colon, I. The Alcohol Beverage Purchase Age and Single-Vehicle Highway Fatalities. *Journal of Safety Research*, Vol. 15, 1984, pp. 159–162.
19. Campbell, K. L., et al. *Analysis of Accident Rates of Heavy-Duty Vehicles.* Final Report. University of Michigan Transportation Research Institute, Ann Arbor, 1988.
20. Stein, H. S., and I. S. Jones. Crash Involvement of Large Trucks by Configuration: Case Control Study. *American Journal of Public Health*, Vol. 78, 1988, pp. 491–498.
21. Jovanis, P., H. Chang, and I. Zabaneh. A Comparison of Accident Rates for Two Truck Configurations. In *Transportation Research Record 1249*, TRB, National Research Council, Washington D.C., 1989, pp. 18–29.
22. Mingo, R., et al. *Safety of Multi-Unit Combination Vehicles: Report on Findings.* Final Report, Association of American Railroads, Sept. 1990.
23. Judge, G. G., et al. *Introduction to the Theory and Practice of Econometrics.* 2nd Edition. John Wiley and Sons, Inc., New York, 1988.
24. Hausman, J. Specification Tests in Econometrics. *Econometrica*, Vol. 46, 1978, pp. 69–85.
25. Chamberlain, G. Analysis of Covariance with Qualitative Data. *Review of Economic Studies*, Vol. 47, 1980, pp. 227–238.
26. National Safety Council. *Manual of Classification of Motor Vehicle Traffic Accidents.* 5th edition, 1989.
27. Nassar, S., F. Saccomanno, and J. Shortreed. Nested Logit Model of Road Accident Severity. Presented at 72nd Annual Meeting of Transportation Research Board, Washington, D.C., 1993.
28. Barnard, P. How Some Advances in Econometrics Might Be Used in Road Crash Analysis. Australian Road Research Board Report, ARR 171, 1989.
29. Hauer, E. An Application of the Likelihood/Bayes Approach to the Estimation of Safety Countermeasures Effectiveness. *Accident Analysis and Prevention*, Vol. 15, 1983, pp. 287–298.
30. Hauer, E. On the Estimation of the Expected Number of Accidents. *Accident Analysis and Prevention*, Vol. 18, 1986, pp. 1–12.
31. Ben-Akiva, M., A. Ceder, L. Cheng, and M. Liss. Estimation of Intersection Traffic Safety Improvements. Final Report, Massachusetts Department of Public Works, 1988.
32. Hauer, E., and B. Persaud. How to Estimate the Safety of Rail-Highway Grade Crossings and the Safety Effect of Warning Devices, In *Transportation Research Record 1114*, TRB, National Research Council, Washington, D.C., 1990, pp. 131–140.

Publication of this paper sponsored by Committee on Traffic Records and Accident Analysis.