A Spatial Generalized Ordered Response Model to Examine Highway Crash Injury Severity

Marisol Castro
The University of Texas at Austin
Dept of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712-1172
Phone: 512-471-4535, Fax: 512-475-8744
Email: m.castro@utexas.edu

Rajesh Paleti
The University of Texas at Austin
Dept of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712-1172
Phone: 512-471-4535, Fax: 512-475-8744
Email: rajeshp@mail.utexas.edu

Chandra R. Bhat*
The University of Texas at Austin
Dept of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712-1172
Phone: 512-471-4535, Fax: 512-475-8744
Email: bhat@mail.utexas.edu

*corresponding author

November 2012

ABSTRACT

This paper proposes a flexible econometric structure for injury severity analysis at the level of individual crashes that recognizes the ordinal nature of injury severity categories, allows unobserved heterogeneity in the effects of contributing factors, as well as accommodates spatial dependencies in the injury severity levels experienced in crashes that occur close to one another in space. The modeling framework is applied to analyze the injury severity sustained in crashes occurring on highway road segments in Austin, Texas. The sample is drawn from the Texas Department of Transportation (TxDOT) crash incident files from 2009 and includes a variety of crash characteristics, highway design attributes, driver and vehicle characteristics, and environmental factors. The results from our analysis underscore the value of our proposed model for data fit purposes as well as to accurately estimate variable effects. The most important determinants of injury severity on highways, according to our results, are (1) whether any vehicle occupant is ejected, (2) whether collision type is head-on, (3) whether any vehicle involved in the crash overturned, (4) whether any vehicle occupant is unrestrained by a seat-belt, and (5) whether a commercial truck is involved.

1. INTRODUCTION

1.1 Background

Roadway crashes are the leading cause of death in the U.S. among individuals 5-24 years of age (Murphy *et al.*, 2012), and impose a tremendous emotional and economic burden on society. This has led to substantial research investments to identify the risk factors associated with the occurrence of, and severity of injuries sustained in, crashes, so that appropriate improvements in vehicle and roadway design may be implemented to reduce the number of crashes as well as the injury severity of those involved in crashes. These efforts, supplemented by reinforcing safety policies and information campaigns, may have contributed (along with economic conditions) to the steady drop since 2005 in the number of roadway crashes and fatalities (see NHTSA, 2012). However, police-reported crashes in 2010 still numbered 5.4 million and resulted in 32,885 fatalities (NHTSA, 2012), underscoring the continued need to better understand the determinants of crash frequency and injury severity.

Transportation and safety researchers have adopted a wide variety of methodological approaches to model crash occurrence and injury severity. In this regard, crash frequency data are in the form of counts, while injury severity is typically reported and collected on an ordinal scale. Also, the factors associated with crash frequency and injury severity suffered in a crash can be quite different. As a result, different modeling mechanisms and different variable specifications are considered for crash frequency and injury severity conditional on crash characteristics. Lord and Mannering (2010) provide a review of methods for crash frequency analysis, while Savolainen *et al.* (2011) present a corresponding review of methods for injury severity analysis. In this paper, the objective is to contribute to the methods for injury severity analysis by proposing an approach to accommodate the dependence in injury severity levels sustained in proximally occurring crashes, and to apply our proposed method to the analysis of highway injury severity data obtained from the crash incident files maintained by the Texas Department of Transportation.

1.2 Injury Severity Analysis: An Overview

There are several methodological issues that need to be considered in injury severity analysis. For example, Ye and Lord (2011) examine the effects of different under-reporting rates of crashes by injury severity level, and Bhat and colleagues (see Eluru and Bhat, 2007 and Paleti *et*

al., 2010) develop methods to recognize the potential endogeneity of "explanatory" variables.[1] In this paper, our emphasis will be on recognizing three other specific econometric issues in injury severity analysis: (1) The nature of the dependent variable (and model flexibility vis-à-vis the nature of the dependent variable), (2) unobserved heterogeneity in the effects of variables, and (3) spatial dependency effects (this last issue also leads to the recognition of heteroscedasticity in the error terms in the underlying latent variable determining injury severity levels).

1.2.1 The nature of the dependent variable

The injury severity level of a traffic crash is in the form of a series of ordinal levels such as no injury, possible injury, non-incapacitating injury, incapacitating injury, and fatal injury. Thus, many injury severity studies in the past have used a standard ordered-response (SOR) model structure (including the ordered logit or the ordered probit) (see, for example, Dissanayake and Ratnayake, 2006, Xie *et al.*, 2009, Christoforou *et al.*, 2010, Haleem and Abdel-Aty, 2010, Quddus *et al.*, 2010, Jung *et al.*, 2010, Paleti *et al.*, 2010, Liu and Donmez, 2011, Zhu and Srinivasan, 2011). However, the SOR model structure is quite susceptible to parameter inconsistency problems caused by varying under-reporting rates (across injury severity levels) in the data (see Ye and Lord, 2011). It is also saddled with a restrictive monotonic form for the effect of variables on injury severity levels. Specifically, as discussed in Eluru *et al.* (2008), the SOR structure holds the threshold values to be fixed across crashes, which will, in general, lead to inconsistent injury risk propensity and threshold values, and inconsistent effects of variables on the probability of different categories of injury severity. Savolainen *et al.* (2011) also point out this limitation of the SOR structure, with the following example. Assume the presence of three ordinal injury severity levels: no injury, some injury, and fatality, and let the deployment (or not) of an airbag be a key attribute influencing the latent injury severity propensity variable in the SOR structure. Then, the SOR structure will require that crashes in which an airbag deployed will entail a higher probability of no injury and a lower probability of fatal injury relative to

---

[1] Endogeneity refers to the situation where explanatory variables may be correlated with the unobserved error term in the dependent variable model. For instance, there is evidence that seat belt non-users tend to be intrinsically unsafe drivers (see Janssen, 1994 and Petridou and Moustaki, 2000). That is, there are personality characteristics of non-seat belt users (such as aggressive driving behavior and risk attitudes) that may not be available in the data being analyzed, and these unobserved factors that affect seat belt non-use also tend to increase injury severity propensity. If this sample selection is ignored, the result is, in general, an artificially inflated estimate of the effectiveness of seat belt use. This has been found by Eluru and Bhat (2007), while Winston *et al.* (2006) demonstrate similar results in the context of air bag effectiveness.

otherwise observationally identical crashes where an airbag did not deploy. On the other hand, it is quite possible that the deployment of an airbag will decrease both the probability of no injury (because airbag deployment by itself can cause minor injuries) as well as fatal injury. This kind of an influence pattern cannot be captured by the SOR structure.

Another structure that has seen substantial use in injury severity analysis is the unordered-response (UR) model structure, including the multinomial logit model or the sequential binary choice model (see Shankar and Mannering, 1996, Ulfarsson and Mannering, 2004, Khorashadi *et al.*, 2005, Rifaat *et al.*, 2011, and Yan *et al.*, 2011), the Markov switching multinomial logit model (Malyshkina and Mannering, 2009), and the nested logit model (Savolainen and Mannering, 2007, Huang *et al.*, 2008, Hu and Donnell, 2010, Patil *et al.*, 2012). The UR model structure is more robust to varying under-reporting rates across injury severity levels, and is also flexible enough to accommodate unrestricted forms for the effects of variables (such as the airbag-related effects discussed earlier). However, it fundamentally does not recognize the ordinal nature of injury severity data, is somewhat more difficult to interpret than the SOR structure, and leads to a proliferation of parameters to be estimated.

A third structure that has been used more recently for injury severity modeling, and the one used in the current paper, is the generalized ordered-response (GOR) structure that essentially combines the strengths of the SOR and the UR approaches (see Eluru *et al.*, 2008). Specifically, it strictly recognizes the ordinal nature of injury severity, while also allowing more flexibility than the SOR structure with much fewer parameters than the UR structure. The flexibility is achieved by relaxing the constant threshold assumption (across crashes) of the SOR structure through the parameterization of the thresholds themselves as a function of explanatory variables. One interpretation of the GOR structure is that, given a set of variables that characterize a certain crash context, the underlying latent continuous variable in the ordered-response structure represents the general injury risk propensity from the crash. However, there may be some specific driver and other contextual characteristics that dictate the translation of the general risk propensity into the manifested injury severity level. In the airbag example, the deployment of the airbag may reduce the risk propensity from the primary crash (which gets incorporated through the reduction of the general risk propensity), but there could also be a remnant effect not related to the primary crash that increases the probability of minor injury

relative to no injury (which gets incorporated in the thresholds that map the general risk propensity into the manifested injury levels).[2]

1.2.2 Unobserved heterogeneity in the effects of variables

A majority of injury severity studies to date have assumed that there are no variations in the effects of explanatory variables in the underlying structures for the SOR or UR models. However, it is very likely that there are unobserved crash-specific factors that may moderate the impact of explanatory variables. For example, some angle crashes may lead to injury severities of those involved that may be far more severe than head-on crashes, even if the majority of angle crashes lead to a lesser degree of injury severity. This may be because some angle crashes may lead to a spinning or even an overturning of one or more vehicles involved in the crash, leading to severe injuries. Such possibilities may be reflected by accommodating a random coefficient on the "angle crash" dummy variable (with "rear-end" crashes being the base category) in the underlying risk propensity specification in the SOR and GOR model structures, or multiple random coefficients on the "angle crash" dummy variable in the injury-specific propensity specifications of the UR model structure. While the presence of unobserved heterogeneity effects will be context-dependent, the analyst should consider these effects rather than dismissing them without testing for their presence. In particular, when present, these unobserved heterogeneity effects can have very real implications for the accurate assessment of the effects of variables and for the design of countermeasures to reduce injury severity. This realization has resulted in many more studies in the past five years or so that consider unobserved heterogeneity effects, including in SOR structures (see Srinivasan, 2002, Eluru and Bhat, 2007, Paleti *et al.*, 2010, and Christoforou *et al.*, 2010), in UR structures (see Milton *et al.*, 2008, Anastasopoulos and Mannering, 2011, Chen and Chen, 2011, and Moore *et al.*, 2011), and in GOR structures (see Eluru *et al.*, 2008).

---

[2] The GOR structure discussed here is quite different from other generalizations of the ordered structure used in Quddus *et al.* (2010) and Wang *et al.* (2011). Specifically, the generalization in these other papers, which is based on Fu (1998) and Williams (2006), cannot be cast in the form of a continuous underlying process (risk propensity) that gets mapped to the observed outcomes (injury levels) in the way the SOR model or our GOR structure can be. More generally, the connection between the underlying injury risk propensity scale and the observed injury outcomes becomes ambiguous and unclear in this alternative form of generalization (see Greene and Hensher, 2009; page 198). Further, this alternative generalization has the problem that the probabilities need not be positive for some combinations of explanatory variables, and it is impossible to resolve this problem unless one imposes restrictions on the generalization that brings it back to the simple SOR structure.

1.2.3 Spatial dependency effects

Another methodological challenge in injury severity modeling, as highlighted by Savolainen *et al.* (2011), is the inclusion of spatial dependence. Recognizing spatial dependence is important because crash locations are mapped to a spatial unit of analysis, such as intersections or highway sections. Thus, to analyze injury severity of crashes occurring at urban intersections, the roadway geometrics and signal characteristics of the intersection constitute a part of the explanatory variable vector. Similarly, to analyze injury severity levels for crashes on highways, the explanatory variables will typically include the roadway characteristics of a specific section of the highway around the crash location. In addition to these observed characteristics of the spatial unit to which a crash is mapped, there are likely to be unobserved factors of the spatial unit that can impact injury risk propensity. Now consider two crashes at proximally located intersections or highway sections. It is certainly possible that a certain observed design element at one crash location (say, for example, narrow approach lanes at an intersection or a sharp horizontal curve on a highway section) not only influences injury risk propensity at that location, but also has a "spatial spillover" effect on the injury risk propensity at proximally located crash sites. In addition, there may be common unobserved (to the analyst) location factors (such as the absence of center dividers at adjacent intersection locations or a narrow shoulder on adjacent highway locations) that may lead to a "spatial correlation" effect in the error terms of the injury risk propensity at proximally located crash locations. Ignoring such spatial dependencies will, in general, result in inconsistent and inefficient parameter estimation in non-linear models (see Franzese and Hays, 2008 and LeSage and Pace, 2009 for exhaustive treatments of spatial models in discrete choice contexts). Despite this, there has been no study, to our knowledge, in the injury severity literature that has considered spatial dependency in injury severity modeling.

1.3 The Current Paper

The purpose of this paper is to contribute to the literature on risk factors associated with injury severity. In doing so, we use a GOR model structure that combines the strengths of the SOR and UR approaches, accommodates unobserved heterogeneity in the effect of determinant variables, and explicitly incorporates spatial dependency effects. To our knowledge, this is the first such formulation to be proposed and applied in the injury severity analysis literature. The complex spatial dependences in the resulting multivariate GOR framework lead to an analytically

intractable likelihood function. To overcome this problem, we use a composite marginal likelihood (CML) inference approach for choice models, which is easy to implement and is based on evaluating lower-dimensional marginal probability expressions that do not require simulation (see Bhat *et al.*, 2010, Varin *et al.*, 2011).

The proposed framework and inference approach is applied to study injury severity levels associated with crashes on highways in Austin, Texas. The focus on highway crashes is because highways have high speed limits, which leads generally to more severe injury outcomes than on other roadways (Ma *et al.*, 2009). To put this in perspective, nearly one-half of all fatal crashes in 2009 occurred on roads with posted speed limits of 55 mph or more, which are usually highways (NHTSA, 2011). Further, underreporting is usually lower on highways than on other road classes (see Amoros *et al.*, 2006). The crash data used in the analysis is drawn from the Texas Department of Transportation crash incident files for the year 2009, and includes detailed information on crash, driver, vehicle, roadway segment, and environmental factors.

The rest of the paper is structured as follows. The next section presents the modeling framework. Section 3 describes the data source employed and the sample formation procedures. Section 4 presents the empirical estimation results and the implications of these results for designing countermeasures. The final section summarizes the findings and offers concluding thoughts.

2. MODELING FRAMEWORK

In this section, we first discuss the random coefficients generalized ordered response probit (RC-GORP) model (Section 2.1), and then extend the RC-GORP model to introduce spatial dependence effects (Section 2.2). Finally, we discuss the model estimation approach (Section 2.3).

2.1 The Random Coefficients Generalized Ordered Response Probit (RC-GORP) Model

Let q ($q = 1,2,...,Q$) be an index to represent crashes and let k ($k = 1,2,...,K$) be an index to represent injury severity level (of the most severely injured person in the crash). For instance, in the empirical context of this paper, the index k takes the following values: no injury ($k = 1$), possible injury ($k = 2$), non-incapacitating injury ($k = 3$) and incapacitating or fatal injury ($k = 4$). The equation system for the RC-GORP model then takes the following form:

$$y_q^* = \boldsymbol{\beta}_q' \boldsymbol{x}_q + \varepsilon_q, \ y_q = k \ \text{if} \ \psi_{q,k-1} < y_q^* < \psi_{q,k} \qquad (1)$$

where y_q^* is the latent underlying injury risk propensity for crash q, $\boldsymbol{x}_q$ is an ($L \times 1$)-column vector of exogenous attributes (excluding a constant), and $\boldsymbol{\beta}_q$ is a corresponding ($L \times 1$)-column vector of crash-specific variable effects. ε_q is a standard normal random error term that impacts the latent propensity and that is assumed to be identically and independently distributed across crashes q. The latent propensity y_q^* is next mapped to the observed injury level y_q through the thresholds $\psi_{q,k}$ ($\psi_{q,0} = -\infty$ and $\psi_{q,K} = \infty$; $-\infty < \psi_{q,1} < \psi_{q,2} < ... < \psi_{q,K-1} < \infty \ \forall q$) in the usual ordered response fashion. Unlike the SOR model, the GORP model allows these thresholds to vary across individuals, adding flexibility to the SOR model. Following Eluru *et al.* (2008), we parameterize the threshold associated with the kth injury severity level as a function of a relevant exogenous variable vector $\boldsymbol{z}_{qk}$ (excluding a constant):

$$\psi_{q,k} = \psi_{q,k-1} + \exp(\alpha_k + \boldsymbol{\gamma}_k \boldsymbol{z}_{qk}). \qquad (2)$$

In the above equation, α_k is a scalar, and $\boldsymbol{\gamma}_k$ is a vector of coefficients, associated with injury severity level $k = 1, 2, ..., K-1$. The above parameterization immediately guarantees the ordering condition on the thresholds for each and every crash, while also enabling the identification of parameters on variables that are common to the $\boldsymbol{x}_q$ and $\boldsymbol{z}_{qk}$ vectors. For identification reasons, we adopt the normalization that $\psi_{q,1} = \exp(\alpha_1)$ for all q (equivalently; all elements of the vector $\boldsymbol{\gamma}_1$ are normalized to zero, which is innocuous as long as the vector $\boldsymbol{x}_q$ is included in the risk propensity equation).

Finally, to allow for heterogeneity among observations, the parameter $\boldsymbol{\beta}_q$ is defined as a realization from a multivariate normal distribution with mean vector $\boldsymbol{b}$ and covariance matrix $\boldsymbol{\Omega}$.[3] Then, we can write $\boldsymbol{\beta}_q = \boldsymbol{b} + \widetilde{\boldsymbol{\beta}}_q$, where $\widetilde{\boldsymbol{\beta}}_q \sim MVN_L(0, \boldsymbol{\Omega})$ (MVN_L represents the multivariate normal distribution of dimension L). If this multivariate distribution become degenerate, then $\boldsymbol{\beta}_q = \boldsymbol{b} \ \forall q$, and the RC-GORP model collapses to the GORP model. Further, in the GORP

---

[3] For ease of presentation, we will treat all elements of $\boldsymbol{\beta}_q$ as random, but this is not necessary; the researcher can fix some elements of $\boldsymbol{\beta}_q$ and let the remaining elements be random.

model, if all elements of $\boldsymbol{\gamma}_k$ are zero for all k, the result is the standard ordered-response probit (SORP) model.

2.2 The Spatial RC-GORP (or SRC-GORP) Model Formulations

In the current paper, we introduce spatial dependency through the latent risk injury propensity variable, using the RC-GORP framework presented in Section 2.1. The reasons for such spatial dependency effects through the latent risk propensity have already been discussed in Section 1.2.3. However, we assume no spatial dependence caused through the thresholds that determine the translation of the latent risk propensity to the observed injury levels. Our expectation is that the variables embedded solely in the $\boldsymbol{z}_{qk}$ vectors (that is, those variables that only affect the thresholds and do not affect the latent risk propensity) are specific crash-related factors that should not have location dependency effects (for instance, if a particular crash results in the deployment of airbags, it is likely to impact the thresholds that translate the risk propensity to observed injury severity levels for that specific crash, but should not impact the risk propensity at another proximate crash location).

In the spatial econometrics literature (see Anselin, 1988), it is common to consider two types of spatial dependency formulations. The first formulation, referred to as the *spatial lag structure*, assumes that the injury risk propensity at one crash location is influenced by the injury risk propensity at other proximal crash locations. Such a structure allows both observed variables as well as unobserved variables at proximal crash locations to impact injury severity at a specific crash location. As explained in Section 1.2.3, this can be the case for injury severity modeling. But if there is little variation in the observed explanatory variables through space (for example, roadway geometry may change little between proximally spaced crash roadway segments of a highway in the observed crash data), it can be difficult to estimate a spatial lag model. The second formulation commonly used in spatial econometrics is the *spatial error formulation*, in which the dependency in injury risk propensity across proximally located crash locations originates purely from unobserved (to the analyst) factors. However, this spatial error formulation assumes only generic unobserved correlation in injury risk propensity across proximally located crash sites, and does not consider that there could be correlation effects in the unobserved factors that moderate the effects of observed variables. That is, there may be space-associated peculiarities that may exacerbate or temper the effects of roadway geometry and other

crash-specific characteristics on injury risk propensity across all proximally located crash sites. To allow for this possibility, we also propose, apparently for the first time in the spatial econometrics literature, a *spatial intermediate model* that expands the spatial error model to include spatial correlation in the effects of observed variables.

In the following presentation, we first discuss the spatial lag formulation and then proceed to discuss the minor changes to obtain the spatial error and the spatial intermediate formulations.

2.2.1 The spatial lag formulation

The spatial lag formulation superimposes spatial dependency on the RC-GORP section of Section 2.1 as follows:

$$y_q^* = \delta \sum_{q'=1}^{Q} w_{qq'} y_{q'}^* + \boldsymbol{\beta}_q' \boldsymbol{x}_q + \varepsilon_q, \ y_q = k \ \text{if} \ \psi_{q,k-1} < y_q^* < \psi_{q,k}, \tag{3}$$

where the $w_{qq'}$ terms are the elements of an exogenously defined distance-based space weight matrix $\mathbf{W}$ corresponding to the locations of crashes q and q' (with $w_{qq} = 0$ and $\sum_{q'} w_{qq'} = 1$), and δ $(0 < \delta < 1)$ is the spatial autoregressive parameter. The latent injury risk propensity representation of Equation (3) can be written equivalently in vector notation as:

$$\boldsymbol{y}^* = \delta \mathbf{W} \boldsymbol{y}^* + \mathbf{x} \boldsymbol{b} + \tilde{\mathbf{x}} \tilde{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}, \tag{4}$$

where $\boldsymbol{y}^* = (y_1^*, y_2^*, ..., y_Q^*)'$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_Q)'$ are $(Q \times 1)$ vectors, $\mathbf{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_Q)'$ is a $(Q \times L)$ matrix of exogenous variables for all Q units, $\tilde{\mathbf{x}}$ is a $(Q \times QL)$ block-diagonal matrix with each block-diagonal of size $(1 \times L)$ being occupied by the vector $\boldsymbol{x}_q'$ $(q = 1, 2, ..., Q)$, and $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}}_1', \tilde{\boldsymbol{\beta}}_2', ..., \tilde{\boldsymbol{\beta}}_Q')'$ is a $(QL \times 1)$ vector. Through simple matrix algebra manipulation, Equation (4) may be re-written as:

$$\boldsymbol{y}^* = \mathbf{S}\left(\mathbf{x} \boldsymbol{b} + \tilde{\mathbf{x}} \tilde{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}\right), \tag{5}$$

where $\mathbf{S} = [\mathbf{I}_Q - \delta \mathbf{W}]^{-1}$ is a $(Q \times Q)$ matrix and $\mathbf{I}_Q$ is an identity matrix of size Q. The vector $\boldsymbol{y}^*$ is multivariate normally distributed as $\boldsymbol{y}^* \sim MVN_Q(\boldsymbol{B}_{LAG}, \boldsymbol{\Sigma}_{LAG})$, where

$$B_{LAG} = \mathbf{S}\mathbf{x}\,b \text{ and } \Sigma_{LAG} = \mathbf{S}\left[\widetilde{\mathbf{x}}\left(I_Q \otimes \Omega\right)\widetilde{\mathbf{x}}' + I_Q\right]\mathbf{S}'. \tag{6}$$

2.2.2 The spatial error formulation

The spatial error formulation considers spatial dependency only in the ε_q terms:

$$y_q^* = \delta\sum_{q'=1}^{Q} w_{qq'}\varepsilon_{q'} + \boldsymbol{\beta}_q'\mathbf{x}_q + \varepsilon_q, \ y_q = k \ \text{ if } \psi_{q,k-1} < y_q^* < \psi_{q,k}, \tag{7}$$

which can be written more compactly in vector notation as:

$$\boldsymbol{y}^* = \mathbf{x}\,b + \widetilde{\mathbf{x}}\,\widetilde{\boldsymbol{\beta}} + \mathbf{S}\boldsymbol{\varepsilon}. \tag{8}$$

Then, the latent variable is multivariate normally distributed as $\boldsymbol{y}^* \sim MVN_Q(\boldsymbol{B}_{ERROR}, \boldsymbol{\Sigma}_{ERROR})$, where

$$B_{ERROR} = \mathbf{x}\,b \text{ and } \Sigma_{ERROR} = \widetilde{\mathbf{x}}\left(I_Q \otimes \Omega\right)\widetilde{\mathbf{x}}' + \mathbf{S}\mathbf{S}'. \tag{9}$$

2.2.3 The spatial intermediate formulation

The spatial intermediate formulation considers spatial dependency not only through the ε_q terms, but also in the unobserved factors moderating the effects of variables:

$$y_q^* = \delta\sum_{q'=1}^{Q} w_{qq'}(\widetilde{\boldsymbol{\beta}}_{q'}'\mathbf{x}_{q'} + \varepsilon_{q'}) + \boldsymbol{\beta}_q'\mathbf{x}_q + \varepsilon_q, \ y_q = k \ \text{ if } \psi_{q,k-1} < y_q^* < \psi_{q,k}, \tag{10}$$

which can be written more compactly in vector notation as:

$$\boldsymbol{y}^* = \mathbf{x}\,b + \mathbf{S}\left[\widetilde{\mathbf{x}}\,\widetilde{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}\right]. \tag{11}$$

Then, the latent variable is multivariate normally distributed as $\boldsymbol{y}^* \sim MVN_Q(\boldsymbol{B}_{INT}, \boldsymbol{\Sigma}_{INT})$, with

$$B_{INT} = \mathbf{x}\,b \text{ and } \Sigma_{INT} = \mathbf{S}\left[\widetilde{\mathbf{x}}\left(I_Q \otimes \Omega\right)\widetilde{\mathbf{x}}' + I_Q\right]\mathbf{S}'. \tag{12}$$

10

2.3 Model Estimation

The parameter vector to be estimated in the model is denoted $\theta = (b', \overline{\Omega}', \delta, \gamma', \alpha')'$, where $\overline{\Omega}$ is a column vector obtained by vertically stacking the upper triangle elements of the matrix Ω, $\gamma = (\gamma_2', \gamma_3', ..., \gamma_{K-1}')'$, and $\alpha = (\alpha_1, \alpha_2, ..., \alpha_{K-1})'$. Several restrictive models are obtained from the spatial model formulations developed here. If $\delta = 0$, all the three spatial formulations in Section 2.2 collapse to the RC-GORP model of Section 2.1. If, in addition, the elements of $\overline{\Omega}$ corresponding to the non-diagonal elements of Ω are zero, but not the diagonal elements, it represents the RC-GORP without correlation across the unobserved heterogeneity effects of variables. Further, if all elements of $\overline{\Omega}$ are zero, the result is the GORP model. Finally, if all elements of γ are also zero, the result is the SORP model.

The likelihood function $L(\theta)$ for the SRC-GORP model takes the following form:

$$L(\theta) = P(y = m) = \int_{D_{y^*}} \phi_Q(y^* \mid B, \Sigma) dy^*, \tag{13}$$

where $y = (y_1, y_2, ..., y_Q)'$, $m = (m_1, m_2, ..., m_Q)'$ is the corresponding $(Q \times 1)$ vector of the actual observed injury severity levels, D_{y^*} is the integration domain defined as $D_{y^*} = \{y^* : \psi_{q, m_q - 1} < y_q^* < \psi_{q, m_q}, \forall q = 1, 2, ..., Q\}$, $\phi_Q(.)$ is the multivariate normal density function of dimension Q, and B and Σ refer to $[B_{LAG}, \Sigma_{LAG}]$ for the spatial lag formulation, $[B_{ERROR}, \Sigma_{ERROR}]$ for the spatial error formulation, and $[B_{INT}, \Sigma_{INT}]$ for the spatial intermediate formulation.

The rectangular integral in the likelihood function is of dimension Q, which can become problematic from a computational standpoint. Further, the use of traditional numerical simulation techniques can lead to convergence problems during estimation even for moderately sized Q (Müller and Czado, 2005, Bhat *et al.*, 2010). The alternative is to use the composite marginal likelihood (CML) approach (see Bhat *et al.*, 2010 and Varin *et al.*, 2011). The CML is a simple estimation approach that can be used when the full likelihood function is near impossible or plain infeasible to evaluate due to underlying complex dependencies, as is the case of Equation (13). The CML approach, which belongs to the more general class of composite likelihood function approaches (see Lindsay, 1988), is based on maximizing a surrogate likelihood function that

compounds much easier-to-compute, lower-dimensional, marginal likelihoods. In the current study we use the pairwise CML method based on the product of the injury severity likelihood contributions from pairs of crashes:

$$L_{CML}(\boldsymbol{\theta}) = \prod_{q=1}^{Q-1} \prod_{q'=q+1}^{Q} P\big([\boldsymbol{y}]_q = [\boldsymbol{m}]_q, [\boldsymbol{y}]_{q'} = [\boldsymbol{m}]_{q'}\big)$$

$$= \prod_{q=1}^{Q-1} \prod_{q'=q+1}^{Q} \big[\Phi_2(\varphi_q, \varphi_{q'}, v_{qq'}) - \Phi_2(\varphi_q, \mu_{q'}, v_{qq'}) - \Phi_2(\mu_q, \varphi_{q'}, v_{qq'}) + \Phi_2(\mu_q, \mu_{q'}, v_{qq'})\big],$$

(14)

where $\varphi_q = \dfrac{\psi_{q,m_q} - [\boldsymbol{B}]_q}{\sqrt{[\boldsymbol{\Sigma}]_{qq}}}$, $\mu_q = \dfrac{\psi_{q,m_q-1} - [\boldsymbol{B}]_q}{\sqrt{[\boldsymbol{\Sigma}]_{qq}}}$, $v_{qq'} = \dfrac{[\boldsymbol{\Sigma}]_{qq'}}{\sqrt{[\boldsymbol{\Sigma}]_{qq'}}\sqrt{[\boldsymbol{\Sigma}]_{qq'}}}$.

In the above expression, $[\boldsymbol{B}]_q$ represents the qth element of the column vector $\boldsymbol{B}$, while $[\boldsymbol{\Sigma}]_{qq'}$ represents the qq'^{th} element of the matrix $\boldsymbol{\Sigma}$. The pairwise marginal likelihood function of Equation (14) comprises $Q(Q-1)/2$ pairs of bivariate probability computations, which can itself become quite time consuming. However, previous studies (Varin and Vidoni, 2009, Bhat *et al.*, 2010, Varin and Czado, 2010) have shown that spatial dependency drops quickly with inter-observation distance. Therefore, there is no need to retain all observation pairs because the pairs formed from the closest crashes provide much more information than pairs from crashes that occur far from one another. We examine this issue by creating different distance bands and, for each specific distance band, considering only those pairings in the CML function that are within the distance band. To do so, construct the matrix $\boldsymbol{R}$ of dimension $Q \times Q$ with its qth column filled with a $Q \times 1$ vector of zeros and ones as follows: if the observational unit q' is not within the specified threshold distance of unit q, the q'^{th} row has a value of zero; otherwise, the q'^{th} row has a value of one. Then, the CML function gets modified as follows:

$$L_{CML}(\boldsymbol{\theta}) = \prod_{q=1}^{Q-1} \prod_{q'=q+1}^{Q} \big[P\big([\boldsymbol{y}]_q = [\boldsymbol{m}]_q, [\boldsymbol{y}]_{q'} = [\boldsymbol{m}]_{q'}\big)\big]^{[\mathbf{R}]_{qq'}}$$

$$= \prod_{q=1}^{Q-1} \prod_{q'=q+1}^{Q} \big[\Phi_2(\varphi_q, \varphi_{q'}, v_{qq'}) - \Phi_2(\varphi_q, \mu_{q'}, v_{qq'}) - \Phi_2(\mu_q, \varphi_{q'}, v_{qq'}) + \Phi_2(\mu_q, \mu_{q'}, v_{qq'})\big]^{[\mathbf{R}]_{qq'}}.$$

(15)

Under usual regularity assumptions (Molenberghs and Verbeke, 2005, Xu and Reid, 2011), the CML estimator of $\boldsymbol{\theta}$ is consistent and asymptotically normal distributed with

asymptotic mean $\boldsymbol{\theta}$ and covariance matrix given by the inverse of Godambe's (1960) sandwich information matrix $\mathbf{G}(\boldsymbol{\theta})$ (see Zhao and Joe, 2005):

$$\mathbf{V}_{CML}(\hat{\boldsymbol{\theta}}) = [\mathbf{G}(\boldsymbol{\theta})]^{-1} = [\mathbf{H}(\boldsymbol{\theta})]^{-1}\mathbf{J}(\boldsymbol{\theta})[\mathbf{H}(\boldsymbol{\theta})]^{-1},\tag{16}$$

where $\mathbf{H}(\boldsymbol{\theta}) = E\left[-\dfrac{\partial^2 \mathrm{log}L_{CML}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]$ and $\mathbf{J}(\boldsymbol{\theta}) = E\left[\left(\dfrac{\partial \mathrm{log}L_{CML}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\left(\dfrac{\partial \mathrm{log}L_{CML}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right)\right]$. The variance

matrix $V_{CML}(\hat{\boldsymbol{\theta}})$ is computed for each distance band.

The matrix $\mathbf{H}(\boldsymbol{\theta})$ can be estimated in a straightforward manner using the Hessian of the negative of $\mathrm{log}L_{CML}(\boldsymbol{\theta})$, evaluated at the CML estimate $\hat{\boldsymbol{\theta}}$. This is because the information identity remains valid for each pairwise term forming the composite marginal likelihood. Thus, $\mathbf{H}(\boldsymbol{\theta})$ can be estimated as:

$$\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}}) = -\left[\sum_{q=1}^{Q-1}\sum_{q'=q+1}^{Q}\dfrac{\partial^2 \mathrm{log}L_{CML,qq'}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]_{\hat{\boldsymbol{\theta}}},\tag{17}$$

where $L_{CML,qq'}(\boldsymbol{\theta}) = \left[P\big([\boldsymbol{y}]_q = [\boldsymbol{m}]_q, [\boldsymbol{y}]_{q'} = [\boldsymbol{m}]_{q'}\big)\right]^{[\mathbf{R}]_{qq'}} | \boldsymbol{\theta}$.

The estimation of the matrix $\mathbf{J}(\boldsymbol{\theta})$ is, however, not straightforward because of the underlying spatial dependence among counts. We use the method proposed by Bhat (2011), which entails the construction of suitable overlapping subgroups of the count data that may be viewed as independent replicated observations.[4]

One final important issue that we have not discussed thus far is the positive definiteness of the matrix $\boldsymbol{\Omega}$. Once this is ensured, the positive definiteness of $\boldsymbol{\Sigma}$ is ensured as long as $0 < \delta < 1$. In our estimation, the positive definiteness of $\boldsymbol{\Omega}$ is guaranteed by writing the logarithm of the pairwise-likelihood in terms of the Cholesky-decomposed elements of $\boldsymbol{\Omega}$ and maximizing with respect to these elements of the Cholesky factor. Essentially, this procedure entails passing the Cholesky elements as parameters to the optimization routine, constructing the $\boldsymbol{\Omega}$ matrix internal to the optimization routine, then computing $\boldsymbol{\Sigma}$, and finally selecting the appropriate elements of the matrix for the pairwise likelihood components. To ensure the

---

[4] An additional piece of information that readers may find useful. The maximum likelihood (ML) function for the GORP model is the same as the CML function in Equation (15) after imposing the constraint $\delta = 0$ and raising the CML function to the power of $1/(Q-1)$. Thus, for the GORP model, the CML and the ML estimates are identical.

constraints on the autoregressive term δ, we parameterize it as $\delta = e^{\tilde{\delta}}/(1+e^{\tilde{\delta}})$. Once estimated, the $\tilde{\delta}$ estimate can be translated back to estimate δ.

3. DATA

3.1 Sample Formation

The crash data used in the analysis is drawn from the Texas Department of Transportation (TxDOT) Crash Records Information System (CRIS) for the year 2009, which was the latest year of CRIS crash data available at the time of this study. The Department of Public Safety in Texas officially maintains the records of the crashes reported by police and drivers that involve property damage of more than $1,000 and/or the injury (including fatal injury) of one or more individuals.[5] The CRIS compiles police and driver reports into multiple text files, including complete crash, person (drivers, passengers and non-occupants), vehicle, and weather-related details. The injury level sustained by each injured individual in each crash is also recorded. TxDOT overlays the crash locations from the crash files to a Geographic Information System (GIS)-based street network, identifies crash locations on the street network, and subsequently extracts information on road design, road geometry, and traffic conditions for each crash. The CRIS contains the characteristics of crashes occurring all over Texas (that is, more than 520,000 crashes in 2009), along with supplementary information as just identified.

For the current study, crashes occurring in Austin, Texas were extracted from the CRIS data base with a focus on non-bridge related and non-intersection related crashes (TxDOT defines a crash as being intersection-related if it occurs within the curbline limits of intersections or on one of the approaches/exits to the intersection within 200 feet from the intersection center point). Also, crash records involving pedestrians, bicyclists, motorcyclists and trains were removed because the nature and characteristics of the injury severity sustained by road-users in such crashes are likely to be quite different from the injury severity sustained by motor vehicle occupants in crashes involving only motor vehicles (Bagdadi, 2012). Finally, records with

---

[5] Clearly, by construction, there is an under-reporting of the "no injury" category in the CRIS database. A study of the effect of such under-reporting in the generalized ordered-response framework of this paper has not been undertaken in the literature, and would be a good avenue for future research. But, given the flexibility of the generalized ordered-response framework, it should be more robust to under-reporting than the standard ordered-response framework. Also, from a policy standpoint, there is more incentive to focus on countermeasures to reduce severe injuries in crashes, and so our injury severity analysis could be viewed as focused on the population of crashes that are biased toward higher injury severity.

incomplete or inconsistent information on crash, highway design, driver and vehicle, and environmental factors (weather/lightning conditions and traffic volume) were removed from the sample. After the data cleaning process, the final sample for analysis included 2,087 crashes. As just discussed, the unit of analysis used in our empirical analysis corresponds to a traffic crash. Specifically, following previous studies (Lee and Mannering, 2002, Dissanayake and Ratnayake, 2006, Milton *et al.*, 2008, Mujalli and de Oña, 2011, Anastasopoulos and Mannering, 2011), the dependent variable is the injury severity level of the most severely injured individual (could be a driver or a passenger) in each crash.[6] Injury severity is recorded in five ordinal categories: (1) no injury (52.7% of the sample), (2) possible injury (27.7%), (3) non-incapacitating injury (17.2%), (4) incapacitating injury (1.7%), and (5) fatal injury (0.7%). In the current analysis, we converted this five-level ordinal categorization into a four-level scheme by combining the incapacitating and fatal categories into a single level (because of the very low share of crashes with a fatal injury).[7] Figure 1 shows the study area, which corresponds to the city limits of Austin (the shaded portion in the figure). The highways in the study area have also been identified. To avoid clutter and keep the presentation simple, we are showing only the incapacitating and fatal crashes in the figure. One can notice that there is a clustering of the crashes of high injury severity, which may be in part due to similar observed characteristics of highway segments in close proximity and/or similar unobserved characteristics of the highway segments in close proximity (leading to spatial dependency effects).[8]

The next section discusses additional sample details on relevant exogenous variables in our analysis.

---

[6] Admittedly, the use of such a dependent variable leads to a loss in information on the distribution of injury severities across individuals involved in a crash, and does not explicitly recognize that multiple occupants may suffer the same level of injury severity. It also precludes the use of each occupant's characteristics as determinants of that occupant's injury severity level. Extension of the proposed approach to include the injury severity of each occupant of each vehicle involved in a crash is left for future work. While there has been some recent research focusing on all drivers involved in a crash (for example, Helai *et al.*, 2008), or all occupants of all vehicles involved in a crash (for example, Eluru *et al.*, 2010), these efforts have ignored spatial dependency effects.

[7] A similar approach has been adopted in many earlier studies to circumvent the statistical issues that arise from having a very low number of crashes in one or more categories; see, for example, Savolainen and Mannering (2007), Eluru *et al.* (2008), Milton *et al.* (2008) and Chen and Chen (2011).

[8] Technically speaking, one needs to see crashes of all injury severities to make the determination that there is spatial clustering by injury severity level. But showing crashes of all injury severity levels makes things cluttered. Suffice it to say that we observed the same kind of clustering for crashes of all injury severity levels, which translates to potential spatial dependence effects for the underlying injury severity propensity.

15

3.2 Sample Characteristics

Several types of exogenous variables were considered in the empirical analysis, including crash characteristics, highway design attributes, characteristics of drivers involved in the crash, characteristics of vehicles involved in the crash, and environmental factors. Table 1 presents the sample characteristics of selected exogenous variables within each of these categories of variables.[9]

Crash characteristics include (a) the crash location (where crash occurred in relation to the roadway), (b) the first point of impact in the crash, (c) the collision type, (d) the number of vehicles involved in the crash, and (e) the following variables related to counts across all vehicles involved in the crash: the number of passengers, the number of unrestrained occupants, and the number of occupants ejected. Among these, the first variable, crash location, is represented in one of three discrete states as on-roadway (first point of impact happened on the portion of the highway normally used for vehicular travel), off-roadway (first point of impact happened off the portion of the highway normally used for vehicular travel), or on the shoulder or median. The first point of impact in the crash is characterized in one of four discrete states – (with) moving vehicles, (with) fixed objects (such as protection barriers, signal posts, and trees), vehicle overturning with no specific point of impact, and other points of impact (such as with animals and parked vehicles). Collision type is identified in one of six possible states: rear-end (the front of one moving vehicle crashes into the back of another moving vehicle travelling in the same direction), one stopped-vehicle (the moving vehicle collides with a stopped vehicle), single vehicle (only one moving vehicle is involved in the crash), sideswipe (the sides of two vehicles travelling in the same direction collide, one vehicle impacting the side of the other vehicle), head-on (the front of two vehicles travelling in opposite directions collide), and others (such as turn crashes – at least one vehicle attempts to make a turn in front of an opposing vehicle – and angle crashes – two vehicles approach from non-opposing angular directions).[10] Table 1 shows

---

[9] Many explanatory variables were not statistically significant in the final model specification; the sample characteristics of these variables are not presented in Table 1 to conserve on space. Among these variables were: number of occupants who experienced an airbag deployment, age of vehicles involved in the crash, roadway surface type (flex base, concrete and roadbed soil, or asphalt base), horizontal curvature length, horizontal curvature degree, horizontal curvature type (right curve, left curve and no curve), and road surface conditions (dry, wet, icy and others).

[10] To some extent, the categories defined for collision type overlap with the categories defined for the first point of impact in the crash. For example, the collision types 'rear-end', 'sideswipe' and 'head-on' fall within the 'moving vehicle' designation as the first point of impact. Similarly, the collision type 'single vehicle' overlaps with the

that more than 81% of the crashes occurred on the roadway, while the remaining crashes occurred off the roadway (about 16%) and on medians or shoulders (about 3%). The first point of impact for the vast majority of crashes was a moving vehicle, with a sizeable share of crashes with a fixed object. With respect to collision type, one third of crashes were rear-end crashes, and about a fifth each were "one stopped-vehicle", "single vehicle", and sideswipe crashes. The descriptive statistics of other count variables related to crash characteristics are provided in the bottom sub-table of Table 1, and indicate a preponderance of two-vehicle crashes in the sample and a mean of 2.85 individuals (an average of two drivers and 0.85 passengers) involved in a crash. The statistics also show, as expected, a very low proportion of crashes in which occupants were either unrestrained or ejected.

Highway design attributes include indicator variables for specific highways in Austin to account for possible design and operational factors unique to the highways not captured in the CRIS database, median type (in the two states of barrier and no barrier), inside and outside shoulder widths (the inside shoulder is to the left of the direction of movement, while the outside shoulder is to the right of the direction of movement), roadway width (for both travel directions, not including shoulders or median width), median width, and number of lanes (total number of lanes in both travel directions). Table 1 shows that a large proportion of the crashes occurred on Interstate 35 (42.8%), and more than 15% of crashes occurred on two other major highways (US Highway 183 and Texas State Highway Loop 1). Interestingly, about 60% of crashes occurred at locations with a median barrier. The descriptive statistics in the bottom sub-table show that outside shoulders are, on average, wider than inside shoulders. The average roadway width is 78.67 feet, with a minimum of 20 feet and a maximum width of 156 feet. The average median width is 35.39 feet, with a large variation from 0 feet (no median) to 378 feet. The number of lanes varies from 2 to 11, with a mean value of 5.73 lanes.

The *characteristics of all drivers involved in the crash* are considered in the model specification. When a crash involves only one vehicle, the characteristics of the vehicle's driver are used to develop the demographic (age and gender) and alcohol usage variables defined in Table 1. When crashes involve two or more vehicles, several combination variables characterizing the attributes of the drivers were originally created (this step is needed to reconcile

---

designations of 'fixed object' and 'vehicle overturning'. During the estimation process, we were careful to avoid the simultaneous use of variables with substantial collinearity.

the multiple driver presence with the single injury outcome that constitutes the dependent variable in our analysis). The combination variables were tested for statistical significance and then appropriately simplified based on the test results to generate the final driver-related variables presented in Table 1. The age-related variables associated with drivers in the final specification are the dummy variables "teenage driver" (at least one driver involved in the crash is aged 19 years or less) and "driver aged more than 55 years" (at least one driver involved in the crash is aged more than 55 years). The gender of drivers involved in the crash is classified in three categories: only male drivers, both female and male drivers, and only female drivers. Driving under the influence (DUI) of alcohol is in the two states of "at least one driver DUI of alcohol" and "no driver DUI of alcohol."[11] Table 1 shows that about 15% of the crashes involved at least one teenage driver, while almost 20% of crashes involved at least one driver aged more than 55 years. In terms of the gender of drivers, about 39% of crashes involved only male drivers, about 38% involved both male and female drivers, and only about 23% of the crashes involved only female drivers. Crashes in which one or more drivers were DUI of alcohol were uncommon, accounting for 6.6% of the sample.[12]

The *characteristics of* the *vehicles involved in the crash* include vehicle body type (sedans, pickups, sport utility vehicles, station wagons, and commercial trucks) and vehicle age (in 5-year intervals). As with the characteristics of drivers, we developed many combination variables to characterize the attributes of the vehicles involved in the crash and tested these for statistical significance. However, the only vehicle body type in the final specification is a dummy variable for "commercial truck", which takes the value of one when at least one vehicle involved in the crash is a commercial truck, and zero otherwise.[13] According to Table 1, truck-involved

---

[11] Texas law defines DUI of alcohol based on the following blood alcohol contents (BAC): for drivers aged under 21 year, $BAC \geq 0.02$; for drivers aged 21 years or more, $BAC \geq 0.08$; for commercial drivers of any age, $BAC \geq 0.04$.

[12] An important point to note here is that we use driver characteristics as determinants of the injury level of the most severely injured person in the crash. We do so for two reasons. First, the driver is the one in control of the vehicle, and so her or his driving habits/reflexes are likely to impact the injury severity level of all occupants (and that of the most severely injured occupant). Second, in 94% of all crashes involving at least one non-driver vehicle occupant, it was a driver who incurred the most severe injury. But, of course, one still has to be somewhat circumspect in interpreting the effects of the driver demographic variables because it is possible that the one suffering the most severe injury is not a driver. The results discussed in Section 4.2.3 regarding driver effects should be viewed in this cautionary light.

[13] A commercial truck, as used in this study, refers to trucks used to transport cargo and excludes pickup trucks, fire trucks and police trucks.

crashes comprised about 4% of all crashes. Vehicle age attributes do not appear in Table 1, because these were not statistically significant in the final model specification.

Environmental factors at the time of the crash include weather, lighting and traffic conditions. Weather conditions are categorized into normal conditions or rainy/foggy conditions. Lighting conditions (type and level of light that existed at the time of the crash) are classified into daylight, night–lighted, night–not lighted, and other lighting conditions, such as dawn, dusk and dark with unknown lighting. The traffic conditions on the highway on which the crash occurred is proxied by the average daily traffic volume on the highway. Table 1 reveals, not surprisingly, that a vast majority of crashes occurred under normal weather conditions and in daylight times. The descriptive statistics for average daily traffic volume in Table 1 show an average of 109,576 vehicles, with a variation of almost 50% of the mean. The daily average percentage of trucks, single-unit or combo-unit, is relatively low, not surpassing 10% of the total traffic volume.

Finally, given that the unit of analysis used in this study is a crash, the distance between crash locations needed to construct the spatial weight matrix was obtained from the latitude and longitude coordinates (in degrees) of the crash location provided by TxDOT in the CRIS database. These coordinates were first translated into *x-y* coordinates and then the Euclidean distance was computed for each pair of crash locations. The average distance between crash locations is 5.0 miles with a minimum distance of 0.1 miles (that is, some crashes occurred at about the same location) and a maximum distance of 20.7 miles. The distance between crash locations was used as a measure of spatial proximity, and formed the basis to develop the spatial weight matrix.

4. ESTIMATION RESULTS

4.1 Variable Specification

The selection of variables included in the final specification was based on previous research, intuitiveness, and parsimony considerations. For categorical exogenous variables, if a certain level of the variable did not have sufficient observations, it was combined with another appropriate level; and if two levels had similar effects, they were combined into one level. For continuous variables, we tested alternative linear and non-linear functional forms, including dummy variables for different ranges. To obtain the final variable specification, the exogenous

variables were considered both in the latent variable (with random coefficients) and in the threshold specification. In total, we considered about 130 different variables for inclusion in each threshold function and in the latent propensity function. The final specification was obtained after extensive specification testing and functional form testing, and after ensuring that variables were not substantially correlated with each other (for example, as noted in footnote 8). The final specification for the SRC-GORP model has 44 highly significant variables (including three constants). Interestingly, even though we tested extensively for unobserved heterogeneity effects in the latent injury risk propensity, no statistically significant random coefficients turned up.[14] Thus the SRC-GORP model collapses to the S-GORP model. However, this result is specific to the current empirical context. In general, it is important to consider unobserved heterogeneity effects in variable impacts, as undertaken here, and then drop these effects if not significant in a specific context rather than *a priori* ruling out such effects. Indeed, some recent injury severity studies in other empirical contexts (see, for example, Anastasopoulos and Mannering, 2011) have identified substantial unobserved heterogeneity in variable effects.

A number of functional forms were considered in the construction of the spatial weight matrix based on a continuous (and decaying) representation of distance, including inverse distance (and its higher orders) and inverse of exponential distance. Among these different weight matrix specifications, the inverse distance cubed-based specification gave the best fit, suggesting a rather rapid fading of spatial dependence. We also explored alternative distance bands to select the pairs of observations (*i.e.*, crash locations) for inclusion in the composite marginal likelihood (CML) estimation (see Bhat, 2011 for details). The optimal distance band was selected by minimizing the trace of the variance-covariance matrix. In the current effort, we explored four distance bands (2, 5, 10, 20.7 miles), the last of which corresponded to considering all pairs of highway crash locations. Among the many distance bands, the best estimator efficiency was obtained with a distance band of 5 miles. Finally, we compared the three non-nested spatial formulations (spatial lag, spatial error, and spatial intermediate formulations) using the composite likelihood information criterion (*CLIC*). The spatial error structure turned out to be the best formulation to represent spatial effects in the current empirical context, indicating that only unobserved factors are responsible for generating spatial dependency in crash injury

---

[14] We also tested for unobserved heterogeneity without accommodating for spatial effects. Even in this case, we did not find statistically significant unobserved heterogeneity on any exogenous variable.

severity levels at proximally located crash locations. This is not entirely surprising given that there is little variation in the observed explanatory variables between proximally spaced crash roadway segments of a highway.

4.2 Estimation Results Analysis

The final estimation results correspond to the S-GORP model specification, and are presented in Table 2. The table is organized in three major numeric columns. The first major column provides the parameter estimates characterizing the latent injury risk propensity. The second major column provides the α_k scalar and the γ_k parameters in the second threshold ($\psi_{q,2}$) demarcating the possible injury and non-incapacitating injury levels. The final major column presents the corresponding estimates for the third threshold ($\psi_{q,3}$) demarcating the non-incapacitating and incapacitating/fatal injury levels (for ease in presentation, we will refer to the non-incapacitating level as the NI level and the incapacitating/fatal level as the IF level in the rest of this paper). The first row of the table reports the constant for the first threshold ($\psi_{q,1}$) between no injury and possible injury levels (this threshold does not include a γ_k ($k=1$) vector because they are fixed to zero for reasons of identification). In the table, for categorical variables, the base category is presented in parenthesis. For example, for crash location, the base category is "on-roadway". The effects of variables from each variable category are discussed in turn in the next few sections.

4.2.1 Crash characteristics

In the category of crash location, the results indicate that off-roadway crashes lead to higher injury risk propensity than on-roadway crashes (a similar result was found by Dissanayake and Ratnayake, 2006 and Yamamoto *et al.*, 2008). In addition, this variable also appears in the threshold between the possible and NI injury levels (see the second major column). The negative sign of the coefficient on this threshold has the effect of moving this threshold to the left. Further, given the way the thresholds are parameterized in Equation (2), the negative parameter on off-roadway crashes also moves the threshold demarcating the NI and IF levels to the left by the same amount as the second threshold. The net result of all these effects is that off-roadway crashes reduce the probability of the no injury level and increase the probability of the IF injury level, relative to on-roadway crashes. The higher probability of serious injury for crashes off-

roadway may be attributable to the rough pavement off-roadway and/or vehicles losing control and running off-the-road at high speeds. There is no effect of "shoulder or median" crashes on the long term risk propensity, but there is a negative effect on the threshold between the possible and NI injury levels. Following the same logic as earlier, this implies that crashes occurring on the shoulder or median of highways have a lower probability (relative to those on-roadway) of leading to possible injury and a higher probability of resulting in IF injuries (with no change in the probability of the "no injury" level). The differential effects of the crash location variable on the probability of sustaining injury of different severity levels highlights the GORP model flexibility (relative to the case of fixed thresholds).

The results related to the first point of impact in the crash reflect a lower probability of no injury and a higher probability of the IF injury level if the crash involves the overturning of one or more vehicles. This is not surprising, since overturning increases the chances that the occupants - drivers and passengers - have contact with the roof of the vehicle (Hu *et al.*, 2005), which can result in traumatic head, spine, and brain injuries injury. Collision type also affects the injury severity of traffic accidents. In particular, (a) a collision of a moving vehicle with a stopped vehicle (*i.e.*, one stopped-vehicle) is less likely (relative to a rear-end collision) to result in IF injuries, and more likely to result in NI injuries, with no impact on the probability of the two lower injury levels, (b) a sideswipe crash results, in general, in a lower probability of possible injury and a higher probability of the IF injury level relative to rear-end collision, possibly because of vehicle spinning and getting out of control in sideswipe crashes, and (c) a head-on crash, as expected, results in a high injury risk propensity, with a substantially higher probability of IF crashes relative to a rear-end collision (see also Dissanayake and Ratnayake, 2006, Wang and Abdel-Aty, 2008, Rana *et al.*, 2010, Mujalli and de Oña, 2011, and Zhu and Srinivasan, 2011 for a similar result).

A crash involving multiple vehicles increases the probability of the most severe IF injury category, presumably because of multiple injury exposure instances (see Hu and Donnell, 2010 and Chen and Chen, 2011). The effect of the number of passengers is best captured using both continuous and dummy variable representations. The results show that the injury risk propensity is lowest when there are no passengers or one passenger (across all vehicles involved in the crash), and highest when there are two passengers (assuming that a vehicle cannot carry more than five passengers). To be specific, the overall effect of the number of passengers on the latent

risk propensity is -0.6397 for crashes with zero passengers, -0.3786 for crashes with one passenger, -0.4184 for crashes with two passengers, -0.2226 for crashes with three passengers, -0.2968 for crashes with four passengers, and -0.3710 for crashes with five passengers. This non-linear effect of the number of passengers has not been considered in earlier studies, and is perhaps the reason for the seeming inconsistency in the direction of effect of this variable on injury risk propensity (Schneider IV *et al.*, 2009 and Dupont *et al.*, 2010 suggest that number of passengers increases injury risk, while Yamamoto *et al.*, 2008 and Malyshkina and Mannering, 2009 find an opposite effect). Finally, in the category of crash characteristics, a crash involving at least one (seat-belt) unrestrained occupant and/or ejected occupant results in an increase in severe injury, as also observed in earlier studies (Wang and Abdel-Aty, 2008 and Anastasopoulos and Mannering, 2011). This result emphasizes the importance of seat-belt usage to reduce injury severity levels.

4.2.2 Highway design attributes

A number of highway design attributes were significant in the final model specification. Highway-specific indicator variables show a higher injury severity propensity for crashes occurring on Interstate 35, U.S. Highway 183, U.S. Highway 290 and Texas State Highway Loop 1 compared to other highways in Austin. Additionally, the highway-specific indicator variables affect the threshold parameterization. These indicator variables are capturing the mean effect of all unobserved factors not considered in our analysis (such as traffic congestion effects) and do not have substantive interpretations. Inside and outside shoulder width have opposite effects on the latent injury risk propensity: crashes on highways with wide inside shoulders have a reduced injury risk propensity (relative to crashes on highways with relatively narrow inside shoulders), while crashes on highways with wide outside shoulders have a higher injury risk propensity (relative to crashes on highways with a relatively narrow outside shoulder). Broad inside shoulders perhaps serve as safety cushions not only for providing separation between the two directions of travel, but also additional room for drivers to take injury-minimizing evasive actions to lessen the impact of a crash as it starts to happen. Broad outside shoulders can also provide the option to lessen the impact of a crash, but may also encourage non-emergency stopping that when combined with exit/entrance ramp traffic traveling at relatively high speeds lead to higher injury risk propensity. Future research should study the effects of inside and

outside shoulder widths in greater detail, and understand the underlying causes for these effects, especially because there has been little systematic investigation into the effect of the inside shoulder width and inconsistent conclusions regarding the effect of outside shoulder width (ranging from no effect in Milton *et al.*, 2008 and Mujalli and de Oña, 2011 to high severity risk in Wang *et al.*, 2009 and Zhang *et al.*, 2011 to low severity risk in Lee and Mannering, 2002). Wider roadways and medians at the crash location result in a lower injury risk propensity, possibly because they provide drivers more physical safety margins (see also Anastopoulos and Mannering, 2011 for a similar result). Finally, crashes on highways with more lanes, in general, result in a reduced probability of NI injuries, and a higher probability of IF injuries, suggesting that roads with more lanes may lead to high-speed, high injury severity crashes due to conflict points arising when lane-changing or maneuvering to exit/enter a highway (Kermanshah *et al.*, 2011).

4.2.3 Characteristics of drivers involved in the crash

The age of drivers involved in the crash has a statistically significant effect on crash injury severity levels. The involvement of one or more teenager drivers in a crash increases the probability of the "no injury" level, and reduces the probability of the "possible" injury level substantially (owing to the negative effect on the threshold between the possible and NI injury categories). It also has a general tendency to reduce the probability of the IF injury category (due to the higher negative magnitude on the risk propensity than on the threshold between the possible and NI injury categories), but less so for this IF category than for the "possible" injury level. In effect, these results are suggesting that teenage drivers and vehicle occupants, perhaps because of their physical flexibility, come out of crashes less severely injured than other vehicle occupants, but also that the benefits of their physical flexibility may be tempered by the severity of the crash impact (because teenage drivers, in general, drive more aggressively than other drivers; see Paleti *et al.*, 2010).[15] Interestingly, the results also indicate that crashes involving

---

[15] Many previous studies have also observed injury severity risk differences (for the most severely injured person) between crashes involving teenagers and crashes that do not involve teenagers. Some of these studies suggest a higher injury severity in crashes involving teenage drivers (Chen and Chen, 2011, Mujalli and de Oña, 2011), while other studies have found the opposite effect (Yamamoto *et al.*, 2008, Schneider IV *et al.*, 2009) or no effect at all (Dissanayake and Ratnayake, 2006, Christoforou *et al.*, 2010). Our study suggests a non-monotonic effect of the involvement of teenage drivers on injury severity levels, underscoring the need to use modeling structures more flexible than the SOR structure. Indeed, the inconsistent results in the effects of the teenage driver variable in earlier studies may be partially attributable to the use of the restrictive SOR structure.

drivers aged 55 years or older are less likely to result in IF injuries, perhaps because of the driving experience of these individuals that helps them (and their vehicle occupants) avoid severe injuries as a crash starts to develop.

Compared to crashes that involve only male drivers, crashes with both a male and a female driver lead to a reduced probability of no injury as well as, in general, a reduced probability of IF injury (with a corresponding increase in the combined probability of the intermediate injury severity levels). The flexibility of our GORP framework allows such double-edged reductions at the extremes to be captured. However, if a crash involves only women drivers, there is a reduced probability of no injuries and a higher probability of IF injuries (see Yamamoto *et al.*, 2008, Xie *et al.*, 2009, Chen and Chen, 2011, and Liu and Donmez, 2011 for a similar result). Given that the drivers of vehicles are the ones who typically sustain the most severe injury, a plausible reason for this effect is that women are more vulnerable to injuries due to their generally smaller physical frames. Alternatively, the result may be reflecting differences in driving styles and behaviors between men and women. As expected, crashes involving one or more drivers DUI of alcohol have an increased probability of IF injuries, possibly due to generally reckless driving behavior and inability to take quick evasive actions. This last result highlights the importance of continued investment in awareness campaigns and public policies to reduce alcohol consumption while driving.

4.2.4 Characteristics of vehicles involved in the crash

As mentioned earlier, among all the characteristics of vehicles involved in the crash explored in this study, only vehicle type had a statistically significant effect on injury severity. Specifically, a crash involving a commercial truck has a high and statistically significant positive impact on the probability of the IF injury level. This is a clear consequence of the huge mass of a commercial truck relative to typical motorized passenger vehicles.

4.2.5 Environmental factors

Rainy or foggy weather conditions increase the probability of the "no injury" level as well as substantially reduce the probability of the IF injury level (relative to normal conditions). This weather-related finding may be a reflection of more cautious driving in adverse weather conditions (see also Eluru *et al.*, 2008, Malyshkina and Mannering, 2009, and Quddus *et al.*,

2010 for a similar result). The results related to lighting conditions are rather complex, because these lighting-related variables appear in the latent risk propensity as well one or both thresholds. Crashes that occur at nighttime (under lighted or unlighted conditions) clearly increase the probability of the "no injury" level. Beyond that, the specific effects on the other categories are context-dependent, though, in general, the results also suggest an increase in the probability of the IF severity level. This bipolar effect is perhaps a reflection of more careful driving during night times, but also the inability to react in dark conditions in ways that can reduce the consequences of a crash as it starts to develop.

The effect of the average daily traffic volume variable was tested using several different functional forms. But, the best data fit was obtained using a squared transformation. Lower daily average traffic volume leads, in general, to higher severity injuries. This finding is consistent with a number of previous studies that have shown that crash severity is higher on low-traffic volume roads, and significantly decreases with increasing volumes, presumably due to better safety design on highways with high traffic volumes (see Das *et al.*, 2009, Christoforou *et al.*, 2010, and Chen and Chen, 2011). Lower traffic volume could be related with higher speeds that more often lead to severe crashes.[16]

4.2.6 Spatial dependence

A unique feature of our formulation is that it enables the accommodation of spatial effects through the long-term propensity variable in the GORP model. The spatial dependency parameter δ in the final S-GORP model is moderate in magnitude (about 0.22) but highly statistically significant, supporting the hypothesis that unobserved factors at each highway location have spatial effects at other highway locations. However, the spatial dependency effect fades relatively rapidly with distance as indicated by the inverse distance cubed-based weight matrix specification. Ignoring the spatial parameter is tantamount to ignoring heteroscedasticity in the error term (since the spatial error specification also leads to error heteroscedasticity in addition to spatial autocorrelation), rendering parameter estimation in an aspatial model both inconsistent and inefficient. The extent of the inconsistency and inefficiency will vary depending

---

[16] We also tested the variable "daily average percentage of trucks", but did not find any statistically significant effect of this variable on crash severity, after controlling for whether a commercial truck was involved in the crash or not. This is not surprising, since the percentage of truck traffic should impact crash occurrence more so than injury severity given a crash.

on the empirical context; however, testing and accommodating for any spatial effects is an important consideration that has not been addressed in the injury severity literature.

4.3 Measures of Fit

The performance of the S-GORP model structure proposed here may be compared to that of the standard ordered response probit (SORP) model using statistical and other measures of fit.[17]

The composite log-likelihood value for the S-GORP model (with 44 parameters) is -2,381,395, while the corresponding value for the SORP model (with 24 parameters) is -2,437,428. The two models (note that the S-GORP model nests the SORP model; see Section 2.1) may then be compared using the adjusted composite likelihood ratio test (*ADCLRT*) statistic that is approximately chi-squared distributed (the *ADCLRT* statistic is similar to the likelihood ratio test statistic used in ordinary maximum likelihood estimation, though its construction is not as simple as the likelihood ratio statistic; see Bhat (2011) for a detailed discussion). The *ADCLRT* statistic value is 349.5, which is larger than the chi-squared table value with 20 degrees of freedom at any reasonable level of significance. This result clearly illustrates the superior data fit offered by the S-GORP model.

To ensure that the superior data fit of the S-GORP model is not simply an artifact of overfitting on the overall estimation sample, we also evaluated the performance of the S-GORP model and SORP model on various market segments of the estimation sample. At an aggregate level we compared the predicted and actual (observed) shares for each injury severity level for each market segment, using the root mean squared error (RMSE) and the absolute percentage error measures. The predicted shares for the SORP model were obtained in the usual way, while the procedure for the S-GORP model is discussed in Appendix A. At a disaggregate level, we computed the implied predictive log-likelihood (after estimating the predicted probability of the observed injury severity level for each crash) and compared the two models using a chi-squared predictive log-likelihood ratio test with 20 degrees of freedom. In all the market segments we tested, the S-GORP model provided a better data fit at both the aggregate and disaggregate levels. To focus the discussion and conserve on space, Table 3 presents these data fit statistics for the full sample and for five market segments based on selected variables. For each selected variable, the data fit for the market segment with the most number of observations is provided

---

[17] The SORP model estimation results are suppressed here to conserve on space, but are available from the authors.

(thus, for example, for the dummy variable that indicates whether or not a teenage driver was involved in the crash, Table 3 presents the data fit for the market segment of crashes in which no teenage driver was involved, because this segment represents 85% of the total sample of crashes). The results clearly show that the predicted shares from the S-GORP model are closer to the true shares than the predicted shares from the SORP model for the full sample and each market segment. The results of the predictive log-likelihood ratio test also reject the SORP model in preference for the S-GORP model, again for the full sample as well as each market segment. In summary, all the aggregate and disaggregate measures of fit point to the superior performance of the S-GORP model over the SORP model.

In the next section, we focus on the elasticity effects as implied by the S-GORP model structure obtained in our paper.

4.4 Elasticity Effects and Implications

Section 4.2 discussed the effects of variables on crash injury severity. However, the coefficients do not directly provide a sense of the magnitude and direction of effects of each variable on each injury severity level (this is not specific to the S-GORP structure adopted here, but is also an issue in the SOR and UR model structures). Of course, the magnitude and direction of effects vary across crashes based on crash context, but one can compute aggregate-level effects to characterize the overall impacts of each variable. To do so, we compute the aggregate-level "pseudo-elasticity effects" of variables, as discussed in the subsequent paragraphs.

For dummy variables, we first predict the probabilities of each injury severity level for each crash, assigning the base value of "0" for all dummy variables characterizing each single exogenous discrete variable. For ease, we will focus here on the precise procedure for one of the discrete variables in the model, with the same procedure being adopted for other discrete variables. Thus, consider the crash location variables. We first compute the crash-level probability of each injury severity level after assigning zero values for both the "off-roadway" and "shoulder or median" variables for each crash, with the "on-roadway" variable already assigned a value of zero because it is the base category; all other variables are at their values in the original data. The detailed procedure to compute the crash-level probabilities in our spatial model (for one realization of the estimated parameters from their sampling distributions) is discussed in Appendix A. Then, the crash-level probabilities are added to obtain the expected

value of the number of crashes at each injury severity level in the base case (label the resulting vector of four values in this base case as BASE). Subsequently, the same procedure as above is undertaken but after changing the value of the "off-roadway" dummy variable for each crash from the value of zero to the value of one, and obtaining the expected value of the number of crashes at each injury severity level in this new case (label the resulting vector of four values as OFF). Next, the same procedure as above is implemented, but now starting with the base data and changing the value of the "shoulder or median" dummy variable for each crash from the value of zero to the value of one (label the resulting vector of four expected values as SHOULDER). Subsequently, to obtain an aggregate-level elasticity of the "off-roadway" dummy variable, we compute the change between the OFF and BASE vectors as a percentage of the BASE vector, yielding four elasticity values (one for each injury severity level). Similarly, to obtain an aggregate-level elasticity of the SHOULDER dummy variable, we compute the change between the SHOULDER and BASE vectors as a percentage of the BASE vector, once again yielding four elasticity values. Finally, we compute the mean and standard errors of the aggregate-level elasticity effects as computed above across 200 bootstrap draws taken from the sampling distributions of the estimated parameters.

For count variables, the procedure is simpler. We simply change the count variable (number of vehicles, number of passengers and number of lanes in our empirical analysis) for each crash by the value of one, and compute the percentage change in the expected number of crashes at each injury severity level. For continuous variables (average daily traffic volume, shoulder width, roadway width and median width), we increase the value of the variable by 10% for each observation.

Table 4 presents the aggregate-level elasticity values. To keep the discussion focused and the presentation uncluttered, only the elasticity effects of the IF injury level for the S-GORP model are presented. The numbers in the table may be interpreted as the percentage change in the probability of the IF injury level due to a change in the exogenous variable. For example, the first entry in the table indicates that the probability of an IF injury is 174.38% (with a standard deviation of 38.13%) higher for off-roadway crashes compared to on-roadway crashes, other characteristics being equal (alternatively, another way to interpret this result is that off-roadway crashes are 2.274 times more likely to result in IF injury than on-roadway crashes). Other entries may be similarly interpreted. The directions of the elasticity effects of the model are consistent

with the discussions in the previous section. The table suggests that the most important variables affecting IF injuries are (1) at least one occupant is ejected, (2) collision type is head-on, (3) at least one vehicle in the crash overturned, (4) at least one occupant unrestrained, and (5) at least one vehicle involved in the crash is a commercial truck. Crash location (crash occurs off-roadway or on shoulder/median) and DUI of alcohol also have significant impacts, but not as much as the factors identified above.

Overall, the results suggest that continued information campaigns regarding the importance of seat-belt usage and driving sober, as well as stricter enforcement of seat-belt and DUI laws, are important countermeasures to reduce injury severity in crashes. Further, informing motorists that shoulders are meant only for emergency stopping, and not for casual stopping, would be helpful as a countermeasure. The results also point to the dangerous implications of a crash involving commercial trucks, suggesting strict enforcement of "rest and sleep laws" for commercial drivers as well as instruction to non-commercial drivers on safe driving procedures around trucks (such as keeping away from the blind spot of trucks, and assessing truck passing opportunities carefully before following through) to avoid collisions with trucks in the first place. On roadways with a substantial percentage of commercial truck vehicles, posting appropriate safety advisory/warning signs may make motorists more alert. The results also point to the need for continued investment in vehicle design (to reduce the occurrence of overturning in crashes) and in vehicle safety and protective technology (to decrease injuries sustained in overturning and head-on crashes). Finally, the cumulative effects of the "shoulder or median" and "outside shoulder width" variables suggest a careful investigation into the advantages and problems of reducing outer shoulder widths, especially near exit and entrance ramps, as a means to discourage stopping in these areas.

5. CONCLUSIONS

This paper has proposed an econometric structure for injury severity analysis that simultaneously (a) recognizes the ordinal nature of the categories in which injury severity are recorded, but also allows flexibility in an efficient manner in capturing the effects of explanatory variables on each injury severity category, (b) accommodates unobserved heterogeneity in the effects of contributing factors, and (c) explicitly recognizes spatial dependencies in the injury severity levels experienced in crashes that occur close to one another in space. The resulting spatial

random coefficients generalized ordered response probit (SRC-GORP) model is estimated using a relatively straightforward-to-implement composite marginal likelihood (CML) inference approach. To our knowledge, the SRC-GORP model is the first formulation of its kind to be proposed and applied in the econometric literature in general, and in the safety analysis literature in particular.

The proposed SRC-GORP model is applied to model injury severity levels on highway segments in Austin, Texas, using the crash incident files maintained by the Texas Department of Transportation. The empirical results clearly reveal the benefits, both in terms of capturing flexibility in variable effects and data fit, to adopting a generalized ordered-response structure rather than the traditional standard ordered-response structure. However, in the current empirical context, we did not find statistically significant unobserved heterogeneity effects in the impact of factors influencing injury severity. But such a determination itself requires first accommodating and testing for such potential heterogeneity, as we have done here, rather than summarily dismissing its presence *a priori*. The results reveal the presence of spatial effects arising from common unobserved factors that affect the latent injury propensity at spatially proximal crash locations. From a substantive standpoint, the results underscore the important effects of crash characteristics, highway design attributes, and driver, vehicle, and environmental factors in determining injury severity levels in crashes on highways.

In summary, it is our hope that the method developed and applied in this paper will serve as a catalyst for the more extensive use of spatial models and flexible structure models for various crash contexts in the injury severity literature.

REFERENCES

Anselin, L., 1988. *Spatial Econometrics: Methods and models*. Kluwer Academic, Dordrecht, The Netherlands.

Amoros, E., Martin, J.-L., Laumon, B., 2006. Under-reporting of road crash casualties in France. *Accident Analysis and Prevention* 38(4), 627-635.

Anastasopoulos, P.C., Mannering, F.L., 2011. An empirical assessment of fixed and random parameter logit models using crash- and non-crash-specific injury data. *Accident Analysis and Prevention* 43(3), 1140-1147.

Bagdadi, O., 2012. Estimation of the severity of safety critical events. *Accident Analysis and Prevention* 50, 167-174.

Bhat, C.R., 2011. The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B* 45(7), 923-939.

Bhat, C.R., Varin, C., Ferdous, N., 2010. A comparison of the maximum simulated likelihood and composite marginal likelihood estimation approaches in the context of the multivariate ordered response model. In: Greene, W., Hill, R.C. (Eds.), *Advances in Econometrics: Maximum Simulated Likelihood Methods and Applications*, 65-106, Emerald Group Publishing Limited, Bingley, UK.

Chen, F., Chen, S., 2011. Injury severities of truck drivers in single- and multi-vehicle accidents on rural highways. *Accident Analysis and Prevention* 43(5), 1677-1688.

Christoforou, Z., Cohen, S., Karlaftis, M.G., 2010. Vehicle occupant injury severity on highways: an empirical investigation. *Accident Analysis and Prevention* 42(6), 1606-1620.

Das, A., Abdel-Aty, M., Pande, A., 2009. Using conditional inference forests to identify the factors affecting crash severity on arterial corridors. *Journal of Safety Research* 40(4), 317-327.

Dissanayake S., Ratnayake, I., 2006. Identification of factors leading to high severity of crashes in rural areas using ordered probit modeling. *Journal of the Transportation Research Forum* 45(2), 87-101.

Dupont, E., Martensen, H., Papadimitriou, E., Yannis, G., 2010. Risk and protection factors in fatal accidents. *Accident Analysis and Prevention* 42(2), 645-653.

Eluru, N., Bhat, C.R., 2007. A joint econometric analysis of seat belt use and crash-related injury severity. *Accident Analysis and Prevention* 39(5), 1037-1049.

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

Eluru, N., Paleti, R., Pendyala, R.M., Bhat, C.R., 2010. Modeling injury severity of multiple occupants of vehicles: Copula-based multivariate approach. *Transportation Research Record* 2165, 1-11.

Franzese R.J., Hays J.C., 2008. Empirical models of spatial interdependence. In: Box-Steffensmeier J., Brady H., Collier D. (eds), *Oxford Handbook of Political Methodology*, 570-604, Oxford University Press, New York.

Fu, V., 1998. Estimating generalized ordered logit models. *Stata Technical Bulletin (STB)* 44, 27-30.

Greene, W.H., Hensher D.A., 2009. *Modeling Ordered Choices: A Primer*. Cambridge University Press, Cambridge, UK.

Godambe, V.P., 1960. An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics* 31(4), 1208-1211.

Haleem, K., Abdel-Aty, M., 2010. Examining traffic crash injury severity at unsignalized intersections. *Journal of Safety Research* 41(4), 347-357.

Helai, H., Chor, C., Haque, M., 2008. Severity of driver injury and vehicle damage in traffic crashes at intersections: A Bayesian hierarchical analysis. *Accident Analysis and Prevention* 40(1), 45-54.

Hu, W., Donnell, E.T., 2010. Median barrier crash severity: Some new insights. *Accident Analysis and Prevention* 42(6), 1697-1704.

Hu, J., Lee, J.B., Yang, K.H., King, A.I, 2005. Injury patterns and sources of non-ejected occupants in trip-over crashes: a survey of NASS-CDS database from 1997 to 2002. *Annual Proceedings / Association for the Advancement of Automotive Medicine* 49, 119-132.

Huang, H., Chin, H.C., Haque, M.M., 2008. Severity of driver injury and vehicle damage in traffic crashes at intersections: a Bayesian hierarchical analysis. *Accident Analysis and Prevention* 40(1), 45-54.

Janssen, W., 1994. Seat-belt wearing and driving behavior: an instrumented vehicle study. *Accident Analysis and Prevention* 26(2), 249-261.

Jung, S., Qin, X., Noyce, D., 2010. Rainfall effect on single-vehicle crash severities using polychotomous response models. *Accident Analysis and Prevention* 42(1), 213-224.

Kermanshah, M., Najaf, P., Jahromi, H.N., 2011. Developing nested logit probability models for rollover accident occurrence and driver injury severity. *Transportation Research Journal* 1(1), 35-45.

Khorashadi, A., Niemeier, D., Shankar, V., Mannering, F.L, 2005. Differences in rural and urban driver-injury severities in accidents involving large trucks: An exploratory analysis. *Accident Analysis and Prevention* 37(5), 910-921.

Lee, J., Mannering, F.L., 2002. Impact of roadside features on the frequency and severity of run-off-road accidents: an empirical analysis. *Accident Analysis and Prevention* 34(2), 349-361.

LeSage, J.P., Pace, R., 2009. *Introduction to Spatial Econometrics*. CRC Press, Taylor & Francis Group, Boca Raton, FL.

Lindsay, B.G., 1988. Composite likelihood methods. *Contemporary Mathematics* 80, 221-239.

Liu, Z., Donmez, B., 2011. Effects of distractions on injury severity in police-involved crashes. Proceedings of the Transportation Research Board 90th Annual Meeting, Washington, D.C.

Lord, D., Mannering, F.L., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transportation Research Part A* 44(5), 291-305.

Ma, Z., Shao, C., Yue, H., Ma, S., 2009. Analysis of the logistic model for accident severity on urban road environment. *Proceedings IEEE Intelligent Vehicles Symposium*, 983-987.

Malyshkina, N., Mannering, F.L., 2009. Markov switching multinomial logit model: an application to accident-injury severities. *Accident Analysis and Prevention* 41(4), 829-838.

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

Molenberghs, G., Verbeke, G., 2005. *Models for Discrete Longitudinal Data*, Springer Series in Statistics. Springer Science + Business Media, Inc., New York.

Moore, D.N., Schneider IV, W.H., Savolainen, P.T., Farzaneh, M., 2011. Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations. *Accident Analysis and Prevention* 43(3), 621-630.

Mujalli, R.O., de Oña, J., 2011. A method for simplifying the analysis of traffic accidents injury severity on two-lane highways using Bayesian networks. *Journal of Safety Research* 42(5), 317-326.

Müller, G., Czado, C., 2005. An autoregressive ordered probit model with application to high frequency financial data. *Journal of Computational and Graphical Statistics* 14(2), 320-338.

Murphy, S.L., Xu, J.Q., Kochanek, K.D., 2012. Deaths: Preliminary Data for 2010. National Vital Statistics Reports 60(4), National Center for Health Statistics, Hyattsville, MD. Available at: http://www.cdc.gov/nchs/data/nvsr/nvsr60/nvsr60_04.pdf

National Highway Traffic Safety Administration (NHTSA), 2011. Traffic safety facts 2009. DOT HS 811 402, National Center for Statistics and Analysis, U.S. Department of Transportation. Available at: http://www-nrd.nhtsa.dot.gov/Pubs/811402.pdf

National Highway Traffic Safety Administration (NHTSA), 2012. 2010 Motor vehicle crashes: overview. DOT HS 811 552, National Center for Statistics and Analysis, U.S. Department of Transportation. Available at: http://www-nrd.nhtsa.dot.gov/pubs/811552.pdf

Paleti, R., Eluru, N., Bhat, C.R., 2010. Examining the influence of aggressive behavior on driver injury severity in traffic crashes. *Accident Analysis and Prevention* 42(6), 1839-1854.

Patil, S., Geedipally, S.R., Lord, D., 2012. Analysis of crash severities using nested logit model-Accounting for the underreporting of crashes. *Accident Analysis and Prevention* 45, 646-653.

Petridou, E., Moustaki, M., 2000. Human factors in the causation of road traffic crashes. *European Journal of Epidemiology* 16(9), 819-826.

Quddus, M.A., Wang, C., Ison, S., 2010. Road traffic congestion and crash severity: an econometric analysis using ordered response models. *Journal of Transportation Engineering* 136(5), 424-435.

Rana, T.A., Sikder, S., Pinjari, A.R., 2010. Copula-based method for addressing endogenity in models of severity of traffic crash injuries: Application to two-vehicle crashes. *Transportation Research Record* 2147, 75-87.

Rifaat, S.M., Tay, R., de Barros, A., 2011. Effect of street pattern on the severity of crashes involving vulnerable road users. *Accident Analysis and Prevention* 43(1), 276-283.

Savolainen, P.T., Mannering, F.L., 2007. Probabilistic models of motorcyclists' injury severities in single- and multi-vehicle crashes. *Accident Analysis and Prevention* 39(5), 955-963.

Savolainen, P.T., Mannering, F.L., Lord, D., Quddus, M.A., 2011. The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. *Accident Analysis and Prevention* 43(5), 1666-1676.

Schneider IV, W.H., Savolainen, P.T., Zimmerman, K., 2009. Driver injury severity resulting from single-vehicle crashes along horizontal curves on rural two-lane highways. *Transportation Research Record* 2102, 85-92.

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

Srinivasan, K., 2002. Injury severity analysis with variable and correlated thresholds: ordered mixed logit formulation. *Transportation Research Record* 1784, 132-142.

Ulfarsson, G., Mannering, F.L., 2004. Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents. *Accident Analysis and Prevention* 36(2), 135-147.

Varin, C., Czado, C., 2010. A mixed autoregressive probit model for ordinal longitudinal data. *Biostatistics* 11(1), 127-138.

Varin, C., Vidoni, P., 2009. Pairwise likelihood inference for general state space models. *Econometric Reviews* 28(1-3), 170-185.

Varin, C., Reid, N., Firth, D., 2011. An overview of composite likelihood methods. *Statistica Sinica* 21(1), 5-42.

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

Wang, Z., Chen, H., Lu, J.J., 2009. Exploring impacts of factors contributing to injury severity at freeway diverge areas. *Transportation Research Record* 2102, 43-52.

Wang, Y.G., Peng, H., Xiang W.S., Xiang, W., Pei, Y.L., 2011. Examining injury severity in left turning crashes at intersections. *Periodica Polytechnica Civil Engineering* 55(2), 191-197.

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

Winston, C., Maheshri, V., Mannering, F.L., 2006. An exploration of the offset hypothesis using disaggregate data: the case of airbags and antilock brakes. *Journal of Risk and Uncertainty* 32(2), 83-99.

Xie, Y., Zhang, Y., Liang, F., 2009. Crash injury severity analysis using Bayesian ordered probit model. *Journal of Transportation Engineering* 135(1), 18-25.

Xu, X., Reid, N., 2011. On the robustness of maximum composite likelihood estimate. *Journal of Statistical Planning and Inference* 141(9), 3047-3054.

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

Yan, X., Ma, M., Huang, H., Abdel-Aty, M., Wu, C., 2011. Motor vehicle-bicycle crashes in Beijing: Irregular maneuvers, crash patterns, and injury severity. *Accident Analysis and Prevention* 43(5), 1751-1758.

Ye, F., Lord, D., 2011. Investigating the effects of underreporting of crash data on three commonly used traffic crash severity models: multinomial logit, ordered probit and mixed logit models. *Transportation Research Record* 2241, 51-58.

Zhang Y., Li, Z., Liu P., Zha, L., 2011. Exploring contributing factors to crash injury severity at freeway diverge areas using ordered probit model. *Procedia Engineering* 21, 178-185.

Zhao, Y., Joe, H., 2005. Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics* 33(3), 335-356.

Zhu, X., Srinivasan, S., 2011. A comprehensive analysis of factors influencing the injury severity of large-truck crashes. *Accident Analysis and Prevention* 43(1), 49-57.

Appendix A:
Computation of crash-level shares and probabilities for each injury severity level

Let $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{b}}', \hat{\overline{\boldsymbol{\Omega}}}', \hat{\delta}, \hat{\boldsymbol{\gamma}}', \hat{\boldsymbol{\alpha}}')'$ be the estimated counterpart of the parameter vector $\boldsymbol{\theta} = (\boldsymbol{b}', \overline{\boldsymbol{\Omega}}', \delta, \boldsymbol{\gamma}', \boldsymbol{\alpha}')'$. The procedure to compute the probabilities of the different injury severity levels for each crash is then as follows:

(1) Random coefficients are drawn from a multivariate normal distribution $\hat{\boldsymbol{\beta}}_q \sim MVN_L(\hat{\boldsymbol{b}}, \hat{\boldsymbol{\Omega}})$, where $\hat{\boldsymbol{\Omega}}$ is the covariance matrix formed from the elements of $\hat{\overline{\boldsymbol{\Omega}}}$ (if there are no statistically significant unobserved heterogeneity effects, $\hat{\boldsymbol{\beta}}_q = \hat{\boldsymbol{b}}$ for all q).

(2) Using the estimated $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\alpha}}$ parameters, the thresholds $\psi_{q,k}$ are computed for each injury severity level k ($k = 1,2,...,K-1$) and for each crash q using Equation (2).

(3) The values $\hat{\boldsymbol{\beta}}_q$ and the spatial parameter $\hat{\delta}$ are used to obtain an estimate of the mean vector (say $\hat{\boldsymbol{B}}_{ERROR}$) and the covariance matrix (say $\hat{\boldsymbol{\Sigma}}_{ERROR}$) based on Equation (9) in the main text.

(4) A ($Q \times 1$)-vector realization of the latent variable $\hat{\boldsymbol{y}}^*$ is drawn from a multivariate normal distribution with mean $\hat{\boldsymbol{B}}_{ERROR}$ and covariance matrix $\hat{\boldsymbol{\Sigma}}_{ERROR}$.

(5) The generated latent variables are translated to "observed" severity injury levels $\hat{\boldsymbol{y}}$ using the estimated threshold values. This "observed" vector is converted into a set of dummy variables representing whether or not the crash is observed in each of the K injury severity levels (using the threshold estimated in step (2)).

The above procedure is repeated N times. Finally, the probability of a crash resulting in injury severity level k for each crash q is computed as the average across the N dummy variable value realizations for injury severity level k for crash q. The elasticity effects presented in Section 4.4 were obtained with $N = 200$ repetitions.
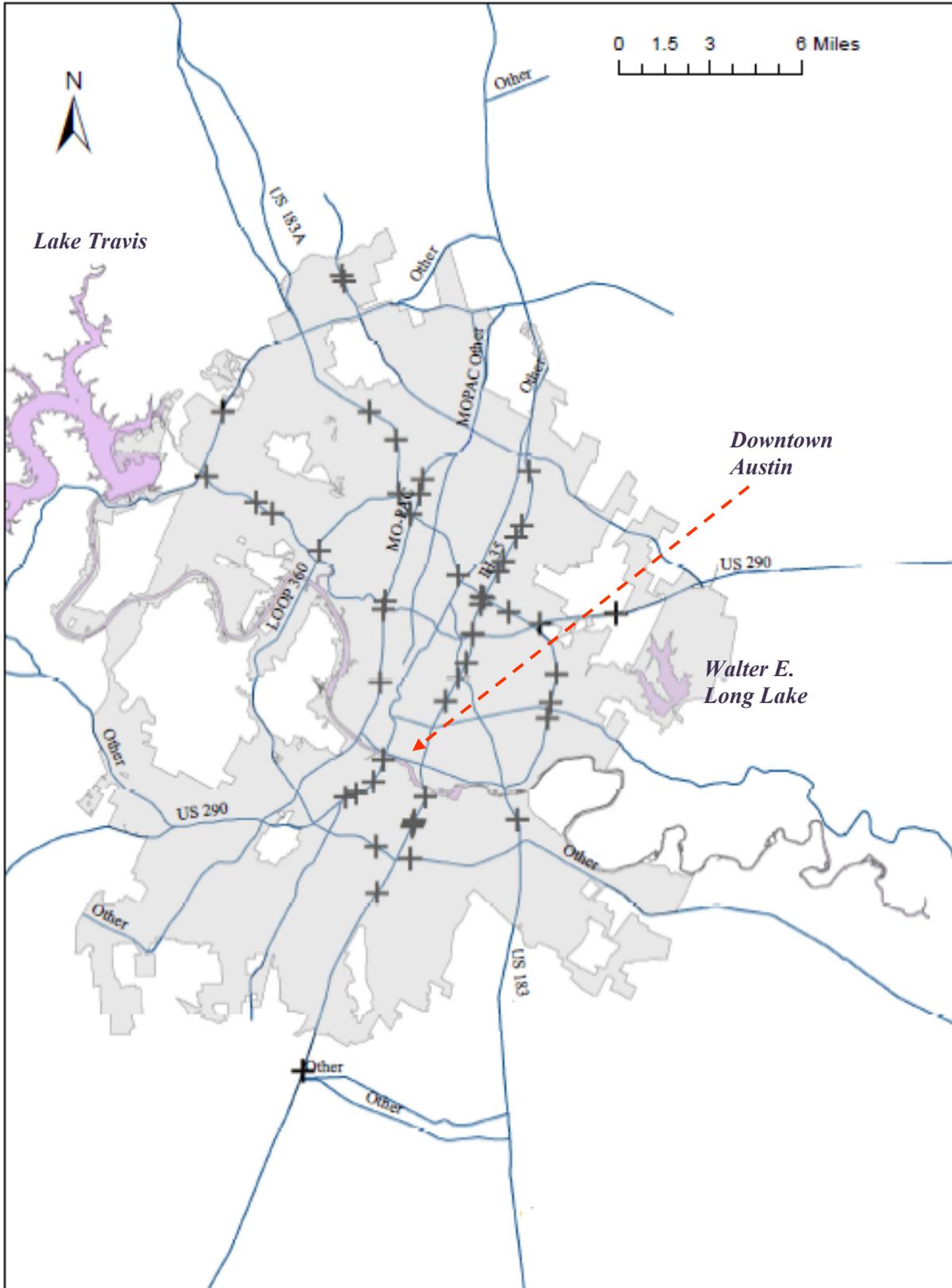
Figure 1: Crash locations for incapacitating/fatal injuries in Austin, Texas

Table 1: Sample characteristics

Variable	Share [%]	Variable	Share [%]
Crash Characteristics		***Characteristics of Drivers Involved in the Crash***	
Crash location		*Teenage driver*	
On-roadway	81.26	At least one teenage driver	15.14
Off-roadway	16.10	No teenage driver	84.86
Shoulder or median	2.64	*Driver aged more than 55 years*	
First point of impact in the crash		At least one driver aged more than 55 years	19.98
Moving vehicle	78.49	No driver aged more than 55 years	80.02
Fixed object	16.91	*Gender of drivers*	
Vehicle overturned	1.82	Only male drivers	38.76
Other point of impact	2.78	Both female and male drivers	37.76
Collision type		Only female drivers	23.48
Rear-end	33.40	*Driving under the influence(DUI) of alcohol*	
One stopped-vehicle	23.77	At least one driver DUI of alcohol	6.60
Single vehicle	21.37	No drivers DUI of alcohol	93.40
Sideswipe	17.78	***Characteristics of Vehicles Involved in the Crash***	
Head-on	1.33	*Commercial truck*	
Other collision types	2.35	At least one vehicle is a commercial truck	3.98
Highway Design Attributes		No commercial trucks	96.02
Highway-specific indicator variables		***Environmental Factors***	
Interstate 35	42.79	*Weather conditions*	
U.S Highway 183	16.72	Normal	87.21
Texas State Highway Loop 1	15.05	Rain or fog	12.79
U.S Highway 290	7.38	*Lighting conditions*	
Other highways	18.06	Daylight	69.67
Median type		Night–lighted	21.66
Barrier	59.66	Night–not lighted	6.32
No barrier	40.34	Other lighting conditions	2.35

Descriptive Statistics

Variable	Mean	Std. Dev.	Minimum	Maximum
Crash Characteristics				
Number of vehicles	2.13	0.86	1.00	10.00
Number of passengers	0.85	1.38	0.00	10.00
Number of unrestrained occupants	0.04	0.23	0.00	3.00
Number of occupants ejected	0.02	0.14	0.00	2.00
Highway Design Attributes				
Shoulders width (feet)				
Inside shoulder	12.45	7.02	0.00	30.00
Outside shoulder	15.30	8.32	0.00	40.00
Roadway width (feet)	78.67	23.71	20.00	156.00
Median width (feet)	35.39	53.55	0.00	378.00
Number of lanes	5.73	1.23	2.00	11.00
Environmental Factors				
Traffic conditions				
Average daily traffic volume (vehicles/day)	109,576	50,803	4,400	207,040
Daily average percent of single-unit trucks (%/day)	2.53	0.82	1.30	6.20
Daily average percent of combo-unit trucks (%/day)	3.97	3.82	0.20	10.00

Table 2: Estimation results S-GORP model

Variables	Latent injury risk propensity		Threshold between possible and non-incapacitating injury		Threshold between non-incapacitating and incapacitating or fatal injury	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Threshold constant between no injury and possible injury α_1	-12.9575 (t-stat: -9.877)					
Threshold constants α_k			0.0642	1.551	1.0038	4.71
Crash Characteristics						
Crash location (on-roadway)						
Off-roadway	0.2729	5.021	-0.4239	-5.849		
Shoulder or median			-0.7537	-4.201		
First point of impact in the crash (moving vehicle)						
Vehicle overturned	0.5918	5.353	-0.3048	-1.814		
Collision type (rear end)						
One stopped-vehicle					0.2292	2.226
Sideswipe			-0.1293	-2.389		
Head-on	0.8212	3.944				
Number of vehicles	0.2110	9.854				
Multiple vehicles *(one vehicle)*					0.3133	1.905
Number of passengers	-0.0742	-4.089				
0 passengers *(more than one passenger)*	-0.6397	-10.860				
1 passenger *(more than one passenger)*	-0.3044	-6.026				
1 or more unrestrained occupants *(no unrestrained occupants)*	0.7687	9.156				
1 or more occupants ejected *(no occupants ejected)*	1.2507	8.751				
Highway Design Attributes						
Highway-specific indicator variables (other highways)						
Interstate 35	0.2497*	4.085*			0.5425	6.650
U.S Highway 183	0.2497*	4.085*				
Texas State Highway Loop 1	0.2497*	4.085*	0.1779	3.501		
U.S Highway 290	0.2497*	4.085*			0.3133	1.905
Shoulders width (feet/10)						
Inside shoulder	-0.1005	-2.838				
Outside shoulder	0.0786	2.532				

*: Same parameter for all highway categories

Table 2: Estimation results S-GORP model (cont.)

Variables	Latent injury risk propensity		Threshold between possible and non-incapacitating injury		Threshold between non-incapacitating and incapacitating or fatal injury	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Highway Design Attributes						
Roadway width (feet/10)	-0.0199	-2.577				
Median width (feet/10)	-0.0095	-2.577				
Number of lanes					-0.1523	-4.588
Characteristics of Drivers Involved in the Crash						
Teenage drivers (no teenage driver)						
One or more teenager driver	-0.1892	-3.825	-0.1335	-2.269		
Drivers aged more than 55 (no driver aged more than 55 years)						
One or more driver aged more than 55 years					0.2030	2.806
Gender of drivers (only male drivers)						
Both female and male drivers	0.1484	3.813			0.1895	2.385
Only female drivers	0.2293	5.713				
Driving under the influence(DUI) of alcohol (no driver DUI)						
One or more drivers DUI of alcohol					-0.4504	-3.738
Characteristics of Vehicles Involved in the Crash						
Trucks (no trucks)						
One or more vehicles were trucks					-0.7888	-3.960
Environmental Factors						
Weather conditions (clear or cloudy)						
Rain or fog	-0.1607	-3.305			0.2890	3.082
Lighting conditions (daylight)						
Dark – lighted	-0.4112	-4.857	-0.1983	-1.493	-0.3727	-2.161
Dark – not lighted	-0.2260	-5.505	-0.4242	-7.294		
Traffic conditions						
Average daily traffic volume squared (vehicles/day/100,000)^2	-0.0715	-4.003				
Spatial parameter δ	0.2199 (t-stat: 3.078)					
Log-composite likelihood at convergence	-2,381,934.8					
Number of parameters estimated	44					
Number of observations	2087					

Table 3: Aggregate and disaggregate measures of fit in the estimation sample

Injury categories / Measures of fit	Entire sample			Teenage driver involved in the crash: No teenage driver			Gender of drivers involved in the crash: Only male drivers		
	Actual share	SORP predictions	S-GORP predictions	Actual share	SORP predictions	S-GORP predictions	Actual share	SORP predictions	S-GORP predictions
No Injury	52.70	52.93	52.28	52.23	52.78	51.75	57.97	56.00	57.33
Possible injury	27.70	26.86	27.63	28.29	26.95	28.22	22.62	25.61	24.05
Non-incapacitating injury	17.20	17.48	17.10	17.17	17.53	17.06	15.70	15.82	15.10
Incapacitating or fatal injury	2.40	2.73	2.99	2.32	2.74	2.97	3.71	2.57	3.52
Number of observations	2087	2087	2087	1771	1771	1771	809	809	809
Root mean square error (RMSE)	-	0.10	0.07	-	1.56	0.82	-	3.76	1.69
Mean absolute percentage error	-	1.68	1.18	-	2.69	1.31	-	6.22	2.86
Predictive likelihood ratio test	$72.31 > \chi^2_{20,0.05} = 31.41$			$54.82 > \chi^2_{20,0.05} = 31.41$			$37.19 > \chi^2_{20,0.05} = 31.41$		

Injury categories / Measures of fit	Weather conditions: Clear or cloudy			Crash location: On-roadway			Lighting conditions: Daylight		
	Actual share	SORP predictions	S-GORP predictions	Actual share	SORP predictions	S-GORP predictions	Actual share	SORP predictions	S-GORP predictions
No Injury	51.70	51.90	51.17	51.65	52.43	51.22	50.07	51.70	49.51
Possible injury	28.08	27.16	28.19	29.54	27.03	29.73	30.95	27.45	31.10
Non-incapacitating injury	17.58	18.04	17.38	16.69	17.75	16.34	17.26	18.13	16.96
Incapacitating or fatal injury	2.64	2.90	3.26	2.12	2.79	2.71	1.72	2.72	2.43
Number of observations	1820	1820	1820	1696	1696	1696	1454	1454	1454
Root mean square error (RMSE)	-	1.08	0.86	-	2.91	0.83	-	4.08	0.96
Mean absolute percentage error	-	1.83	1.48	-	5.01	1.56	-	7.00	1.71
Predictive likelihood ratio test	$62.27 > \chi^2_{20,0.05} = 31.41$			$55.14 > \chi^2_{20,0.05} = 31.41$			$48.95 > \chi^2_{20,0.05} = 31.41$		

Table 4: Elasticity effects of variables on incapacitating or fatal injuries

Variable	Estimate	Standard error
Crash Characteristics		
Crash location (on-roadway)		
Off-roadway	174.38	38.13
Shoulder or median	128.88	33.77
First point of impact in the crash (moving vehicle)		
Vehicle overturned	283.72	97.83
Collision type (rear-end)		
One stopped-vehicle	-39.15	16.37
Sideswipe	21.41	8.86
Head-on	303.52	155.47
Number of vehicles	57.44	10.06
Number of passengers	34.91	7.33
At least one unrestrained occupant *(no unrestrained occupants)*	283.46	60.90
At least one occupant ejected *(no occupants ejected)*	654.37	182.67
Highway Design Attributes		
Highway-specific indicator variables (other highways)		
Interstate 35	-47.94	16.37
U.S Highway 183	51.37	17.03
Texas State Highway Loop 1	17.27	14.29
U.S Highway 290	-11.62	29.37
Shoulders width (feet/10)		
Inside shoulder	-2.16	0.69
Outside shoulder	2.17	0.92
Roadway width (feet/10)	-2.83	1.17
Median width (feet/10)	-0.57	0.22
Number of lanes	32.21	7.45
Characteristics of Drivers Involved in the Crash		
Teenage driver (no teenage driver)		
At least one teenage driver	-14.45	9.63
Driver aged more than 55 (no driver aged more than 55 years)		
At least one driver aged more than 55 years	-35.53	11.49
Gender of drivers (only male drivers)		
Both female and male drivers	-9.32	20.67
Only female drivers	50.25	11.04
Driving under the influence(DUI) of alcohol (no driver DUI)		
At least one driver DUI of alcohol	132.37	68.97
Characteristics of Vehicles Involved in the Crash		
Commercial truck (no commercial truck)		
At least one vehicle is a commercial truck	280.27	183.89
Environmental Factors		
Weather conditions (normal)		
Rain or fog	-58.77	11.34
Lighting conditions (daylight)		
Night – lighted	16.77	11.71
Night – not lighted	29.79	49.23
Traffic conditions		
Average daily traffic volume squared (vehicles/day/100,000)^2	-3.37	0.91