

EVALUATING ALTERNATE DISCRETE CHOICE FRAMEWORKS FOR MODELING ORDINAL DISCRETE VARIABLES

Naveen Eluru*

Assistant Professor

Department of Civil Engineering & Applied Mechanics

McGill University

Suite 483, 817 Sherbrooke St. W.

Montréal, Québec, H3A 2K6

Canada

Ph: 514 398 6823, Fax: 514 398 7361

Email:naveen.eluru@mcgill.ca

Submission Date: January 7 2013

* Corresponding author

ABSTRACT

There is considerable debate on the appropriate discrete choice framework for examining injury severity. Researchers in the safety field have employed ordered and unordered frameworks for examining the various factors influencing injury severity. The objective of the current study is to investigate the performance of the ordered and unordered response frameworks at a fundamental level. Towards this end, we undertake a comparison of the alternative frameworks by estimating ordered and unordered response models using data generated through ordered, unordered data and a combination of ordered and unordered data generation processes. We also examine the influence of aggregate sample shares on the appropriateness of the modeling framework. Rather than be limited by the aggregate sample shares in an empirical dataset, simulation allows us to explore the influence of a broad spectrum of sample shares on the performance of ordered and unordered frameworks. We also extend the data generation process based analysis to under reported data and compare the performance of the ordered and unordered response frameworks. Finally, based on these simulation exercises, we provide a discussion of the merits of the different approaches. The results clearly highlight the emergence of the generalized ordered logit model as a true equivalent ordered response model to the multinomial logit model for ordinal discrete variables.

Keywords: Ordered and unordered discrete choice models for injury severity, ordinal discrete variables, generalized ordered logit, comparison

1 INTRODUCTION

Discrete choice models and their variants are employed extensively for analyzing decision processes in various fields including transportation, marketing, social sciences, bio-statistics and epidemiology. Discrete choice models in their broadest sense can be characterized as ordered and unordered response frameworks. The ordered response frameworks are suited for examining discrete variables that are ordinal in nature while the unordered response frameworks are applicable to analyzing all discrete variables. The ordered response models represent the decision process under consideration using a single latent propensity. The choice probabilities are determined by partitioning the uni-dimensional propensity into as many categories as the dependent variable alternatives through a set of thresholds. Examples of ordered discrete variables in the field of transportation include: (1) driver and passenger injury severity in traffic collisions, (2) household vehicle (automobile and bicycle) ownership, and (3) activity participation indicators (such as number of tours, number of stops, activity episode participation frequency and activity duration). The prevalent mechanism to analyze ordered discrete variables is to employ the ordered response models such as ordered logit and ordered probit models depending on the distributional assumptions of the unobserved component of the latent propensity.

Unordered discrete choice frameworks offer a potential alternative to the analysis of ordered discrete variables. These models are characterized, usually, by a latent variable per alternative and an associated decision rule. The most commonly employed unordered discrete models – the multinomial logit (MNL) model and its extensions – have their origin in the random utility domain. The latent variable per alternative is referred to as the alternative utility and the alternative with the highest utility is designated as the chosen alternative. There are a number of studies that have considered the multinomial logit (and its extensions) for examining ordinal discrete choice variables. For example, (1) injury severity (see Yasmin and Eluru, 2012, Savolainen et al., 2011, Eluru and Bhat, 2007 for detailed literature reviews on severity models), and (2) vehicle ownership (see Anowar et al., 2012 for a list of studies).

The applicability of the two frameworks for analyzing ordinal discrete variables has evoked considerable debate on using the appropriate choice model for analysis. There has been considerable debate more recently in the safety community in adopting the appropriate framework for analysis in the injury severity context. There are many strengths and weaknesses for the ordered framework vis-à-vis the unordered framework. The ordered response models explicitly recognize the inherent ordering within the decision variable whereas the unordered response models neglect the ordering or require artificial constructs to consider the ordering (for example the ordered generalized extreme value model (OGEV)). On the other hand, the traditional ordered response models restrict the impact of exogenous variables on the choice process to be same across all alternatives while the unordered response models allow the model parameters to vary across alternatives (see Eluru et al., 2008 for a discussion). The restricted number of parameters ensures that ordered response models have a parsimonious specification. The unordered response models might not be as parsimonious but offer greater explanatory power because of the additional exogenous effects that can be explored. In fact, several studies highlight the advantages of multinomial logit models over the ordered response models (see for example Bhat and Pulugurta, 1998).

Another concern with the ordered response framework is in the context of modeling datasets that might be affected by under reporting¹ - an aspect of great relevance to safety literature. In fact, unordered response framework is considered to be more effective compared to the ordered response framework in this context. In the case of an under reported decision variable, the traditional multinomial logit model provides estimates that are unbiased i.e. the elasticity effects of the variables are not affected by the under reported data. This is quite critical in terms of examining exogenous variable impacts on the decision process. Further, the unordered response model can be applied by altering the constants if the true population shares are available. In the case of an ordered response model, the parameter estimates are expected to be biased and hence might lead to erroneous policy implications². In summary, in the context of accident literature there are two important aspects that need to be examined.

- 1) Which model framework offers superior statistical fit (and thereby behavioral interpretability) to the dataset under consideration?
- 2) How do these frameworks perform in the presence of under reported data?

It is in this light that we undertake the current research effort. The objective of the current study is to investigate the performance of the ordered and unordered response frameworks at a fundamental level. Towards this end, we undertake a comparison of the alternative frameworks by estimating ordered and unordered response models using data generated through ordered, unordered data and a combination of ordered and unordered data generation processes (more on this in Section 2.1.1). Subsequently, we examine the influence of aggregate sample shares on the appropriateness of the modeling framework i.e. are ordered response frameworks more suitable to examine ordered discrete variable with a particular share. Rather than be limited by the aggregate sample shares in an empirical dataset³, simulation allows us to explore the influence of a broad spectrum of sample shares on the performance of ordered and unordered frameworks. Third, we extend the data generation process based analysis to under reported data and compare the performance of the ordered and unordered response frameworks (in the context of different data generation processes and varying sample shares). Finally, based on these simulation exercises, we provide a summary of the strengths and weaknesses of the two frameworks for analyzing ordered discrete variables in general and for injury severity modeling in particular.

The remainder of the paper is organized as follows. Section 2 provides a background for the proposed research methodology; highlights the motivation for our research and discusses the experimental setup of our study. Section 3 briefly outlines the econometric frameworks of the three alternative frameworks considered. Section 4 presents the results from the simulation exercise for model comparisons. In Section 5, we discuss the simulation results in the context of datasets with under reporting. Section 6 provides a discussion of findings from our analysis while simultaneously providing guidelines on the appropriateness of the modeling frameworks for ordinal discrete variables. We conclude the paper in Section 7 with a discussion of the limitations of the current study and directions for future research.

¹ Injury severity reporting is considered to be substantially affected by under reporting (see Elvik and Mysen 1999, Yamamoto et al., 2008)

² Of course, the *true* advantage of the multinomial logit model in the context of under reporting is slightly reduced because the availability of “true” population level measures to the analyst is quite often rare.

³ An empirical dataset provides a single realization of the aggregate sample share; thus limiting us in exploring the performance difference of the two frameworks as a function of aggregate sample share.

2 BACKGROUND AND CURRENT STUDY IN CONTEXT

2.1 Earlier research

To be sure, some of the aspects highlighted above have been examined in earlier research. For example, Bhat and Pulugurta, 1998 undertook a comparison exercise of vehicle ownership decisions through the ordered logit and the multinomial logit models. In their study, the authors estimated the two models on four datasets and confirmed that the multinomial logit model offers superior data fit and validation capabilities. The study highlight how an ordered response model offers a parsimonious specification while the unordered response model offers enhanced behavioral interpretability through the addition of exogenous variable effects at the alternative level. Yamamoto et al. 2008 conducted an analysis of potential under reported data by comparing the performance of ordered probit and sequential binary probit models. The authors found that the sequential probit models outperform the ordered probit model in terms of bias values in the parameters. Ye and Lord, 2011 compared the ordered probit, multinomial logit and mixed logit model in terms of under reported data. The authors concluded that all the three models considered in the study perform poorly in the presence of under reported data. The exact impact of under reporting on these model frameworks needs further investigation. The study employed data simulation; however, the models were estimated with just one parameter and for a particular aggregate sample share.

More recently Patil et al. 2012 demonstrated the application of a conditional maximum likelihood estimation approach to address under reporting in the context of crash severity analysis using nested logit model. Anowar et al. 2012 undertook a comparison of the ordered and unordered response models in the context of vehicle ownership. In their study, they found that the unordered response models outperform the ordered response models. Yasmin and Eluru, 2012 undertook a comprehensive comparison of various modeling frameworks including ordered logit, generalized ordered logit, multinomial logit, nested logit and ordered generalized extreme value logit for analyzing driver injury severity. In their study, the results clearly establish the superiority of the generalized ordered logit in the context of driver injury severity. In the study the authors also explored the issue of how different frameworks perform in the presence of under reporting in the data. The authors computed elasticity effects for the “true” and under reported datasets and concluded that the error in elasticity effects estimated from the unordered systems is not any better than the error in elasticity effects estimated from the ordered systems in their empirical context.

Interestingly, none of these studies establish that the comparison relationship identified will hold for all possible datasets and sample shares. For example, in the context of accident injury severity it is possible that the additional flexibility of the MNL model can enhance the predictive capabilities. However, it is completely possible that for another decision process (or other injury severity datasets from other geographical regions) the additional flexibility offered by MNL might not yield similar benefits.

2.2 Limitations of earlier research

2.2.1 Influence of Data Generation Process

An important reason for the inability to generalize the results from earlier analysis can be attributed to ignoring the underlying data generation process. The aforementioned studies do not examine the comparison of ordered and unordered frameworks from a fundamental data generation process perspective. The implicit assumption while estimating a discrete choice model for analyzing a decision process is that the model framework considered, reasonably represents the *true* data generation process of the decision maker. To elaborate, if the underlying data generation process (dgp) of the decision considered is ordered in reality an ordered model might be appropriate for analyzing the decision framework. At the same time, if the dgp for the decision maker is unordered in nature then the unordered response model might be a more plausible alternative. Unfortunately, the dgp is latent to the analyst and is seldom known. In any empirical comparison of the ordered and unordered response framework, the impact of the underlying dgp is hard to capture. For example, in the context of injury severity (as an ordinal variable) the underlying dgp could be ordered or unordered or more likely a combination of both. The earlier research efforts on comparison of the various frameworks have not explicitly examined the influence of the inherent dgp on the appropriateness of the framework. To conclusively establish the superiority of a particular framework, one can simulate data and examine how each framework performs based on the dgp under consideration. The comparison exercises undertaken so far have not tackled the comparison from this pure data generation process. In our current study, we attempt to examine the role of dgp on model estimation.

2.2.2 Generalized Ordered Logit model

Even if we confine ourselves to empirical datasets, the earlier literature has compared the traditional ordered logit/probit models with the multinomial logit model. However, the generalized logit model developed in the 1980s (Terza, 1985) and recently employed in traffic safety literature (Castro et al., 2012a; Eluru et al., 2008; Srinivasan, 2002) and modified for count models (Narayanamorrthy et al., 2012; Paleti et al., 2012; Castro et al., 2012b) offers a representative comparison between the ordered and unordered response mechanisms. Researchers have been very proactive in employing the more advanced variants of the multinomial logit model; however, recent advances in ordered regime have not been considered in the comparison (except for Yasmin and Eluru, 2012). It is important that an equivalent ordered response framework for the multinomial logit model is considered for the comparison exercise. The generalized ordered logit (GOL) model allows for the impact of exogenous variables to affect the threshold parameters thus relaxing the restrictive assumption of the traditional ordered response structure on limiting the parameters to be the same for all alternatives. In fact, the generalized ordered response model theoretically can estimate the same number of parameters as the multinomial logit for an ordinal discrete variable. Hence, an exercise comparing the alternative frameworks is incomplete without considering the generalized ordered logit.

2.2.3 Influence of Aggregate Population Sample Share

While undertaking the comparison exercise of alternative frameworks on empirical datasets (be it vehicle ownership or injury severity analysis), the analyst is often restricted to one dataset sample (or a very small number of datasets) with fixed aggregate shares. Thus it is unlikely that the analyst can examine the influence of sample shares on the appropriateness of the alternative

decision frameworks. To investigate the effect of aggregate population sample shares on the appropriateness of the modeling approach, a feasible alternative is to undertake a comparison with simulated databases with varying aggregate sample shares. The current study undertakes this exercise to confirm (or invalidate) the hypothesis that aggregate sample shares of the dataset affects the appropriateness of the modeling framework.

2.2.4 Under reporting

A commonly stated disadvantage of the traditional ordered models is the inability to obtain unbiased estimates in the presence of under reporting (Yamamoto et al., 2008). The recent study by Yasmin and Eluru, 2012 however provides counter evidence to the hypothesis. It is to be expected that an under reported sample will provide incorrect prediction when used on a population sample. However, prior to using a model framework, we need to explore for the presence of systematic biases in alternative model frameworks. The issue is particularly relevant to crash severity modeling as most police compiled datasets are affected by underreporting (see Elvik and Mysen, 1999).

2.2.5 Mixed data generation process

In reality the dgp for the empirical comparison for models is probably a mix of ordered and unordered processes; i.e. for some proportion of the population the underlying decision process might represent an ordered response framework while the remainder of the population might follow an unordered decision process. Depending on the nature of the empirical dataset and aggregate sample shares it is possible that the proportion might have a significant influence on the compatibility of the modeling framework. Even if it is nearly impossible to realize the true proportion value, there is a need to evaluate the performance of alternative frameworks under different proportion levels. The effect of the potential mixing in the dgp has never been studied in the context of ordered and unordered models.

2.3 **Current Study**

The preceding discussion clearly highlights the different aspects of alternative framework comparison that have not been considered in previous research efforts. The main focus of our study is to augment the literature on comparison of ordered and unordered frameworks by focussing on the aforementioned aspects. To achieve this, we propose a four pronged approach. First, we consider the ordered logit, generalized ordered logit and the multinomial logit models for the comparison exercise. Second, we resort to a simulation exercise to understand the influence of underlying dgp on the modeling frameworks. We examine the performance of three models identified in various dgps. Third, we consider the influence of aggregate sample shares on the appropriateness of the model frameworks by considering a range of aggregate shares for the data generation. Finally, we repeat the comparison exercise (second and third steps) in the context of “*artificially*” generated under reported data.

3 **Methodology**

Prior to discussing the experimental design employed in the study we briefly provide details of the three model frameworks employed in our study.

3.1 Ordered Logit Model

In the traditional ordered response model, the discrete injury severity levels (y_i) are assumed to be associated with an underlying continuous, latent variable (y_i^*). This latent variable is typically specified as a linear function as follows

$$y_i^* = \mathbf{X}_i \boldsymbol{\beta} + \varepsilon_i, \quad \text{for } i = 1, 2, \dots, N \quad (1)$$

where,

i ($i = 1, 2, \dots, N$) represents the drivers

\mathbf{X}_i is a vector of exogenous variables (excluding a constant)

$\boldsymbol{\beta}$ is a vector of unknown parameters to be estimated

ε is the random disturbance term assumed to be standard logistic

Let j ($j = 1, 2, \dots, J$) and τ_j denote the injury severity levels and the thresholds associated with these severity levels, respectively. These unknown thresholds are assumed to partition the propensity into $J - 1$ intervals. The unobservable latent variable y_i^* is related to the observable ordinal variable y_i by the τ_s with a response mechanism of the following form:

$$y_i = j, \text{ if } \tau_{j-1} < y_i^* < \tau_j, \quad \text{for } j = 1, 2, \dots, J \quad (2)$$

In order to ensure the well-defined intervals and natural ordering of observed severity, the thresholds are assumed to be ascending in order, such that $\tau_0 < \tau_1 < \dots < \tau_J$ where $\tau_0 = -\infty$ and $\tau_J = +\infty$. The probability expressions take the form:

$$\pi_{ij} = Pr(y_i = j | X_i) = \Lambda(\tau_j - X_i \boldsymbol{\beta}) - \Lambda(\tau_{j-1} - X_i \boldsymbol{\beta}) \quad (3)$$

where Λ represents the standard logistic cumulative distribution function.

3.2 Generalized Ordered Logit Model

The generalized ordered response model relaxes the constant threshold across population restriction to provide a flexible form of the traditional OL model. The basic idea of the GOL is to represent the threshold parameters as a linear function of exogenous variables (Maddala 1983, Terza 1985, Srinivasan 2002, Eluru *et al.* 2008). Thus the thresholds are expressed as:

$$\tau_{ij} = \text{function of } (Z_{ij}) \quad (4)$$

where, Z_{ij} is a set of exogenous variable (including a constant) associated with j th threshold.

Further, to ensure the accepted ordering of observed discrete severity ($-\infty < \tau_{i,1} < \tau_{i,2} < \dots < \tau_{i,J-1} < +\infty$). We employ the parametric form employed by Eluru *et al.* (2008):

$$\tau_{i,j} = \tau_{i,j-1} + \exp(\delta_{ij} Z_{ij}) \quad (5)$$

δ_{ij} is a vector of parameters to be estimated. The remaining structure and probability expressions are similar to the OL model. For identification reasons, we need to either suppress the latent propensity of one of the δ_{ij} vectors.

3.3 Multinomial Logit Model

Consider the probability of an accident i ending in a specific injury-severity level j . The alternative specific latent variables take the form of:

$$U_{ij} = \boldsymbol{\beta}_j \mathbf{X}_{ij} + \varepsilon_{ij} \quad (6)$$

where

$\boldsymbol{\beta}_j$ is a vector of coefficients to be estimated for outcome j

X_{ij} is a vector of exogenous variables

U_{ij} is a function of covariates determining the severity

ε_{ij} is the random component assumed to follow a gumber type 1 distribution.

Thus, the MNL probability expression is as follows:

$$P_i(j) = \frac{\exp[\beta_j X_{ij}]}{\sum_{j=1}^J \exp[\beta_j X_{ij}]} \quad (7)$$

4 Experimental Design

The experimental design considered for the comparison exercise is outlined in detail in this section. The objective is to evaluate the performance of the three frameworks on simulated data. In particular, we focus on data generated from a true ordered process, a true unordered process and a wide spectrum of mixed processes i.e. a proportion (p) of the population follows an ordered decision making process and the remainder of the population (1-p) follows an unordered decision making process. Specifically, we consider the range of p from 0 to 1 in intervals of 0.1, thus traversing the pure ordered dgp to pure unordered dgp in 11 steps⁴. While generating nied dgp, we ensure that the ordered and unordered dgp originate from the same aggregate shares. Within each dgp, we estimate the three models and compare the performance of the frameworks in terms of data fit.

The simulation exercise is undertaken for a four alternative ordered dependent variable – typical injury severity categories employed in safety literature (see Eluru et al., 2008, Yasmin and Eluru 2012). We consider three independent variables (standard univariate normal random variables) to influence the decision process (see Ferdous et al., 2010 and Bhat et al., 2010 for similar data generation examples). The same three variables are used in the three model systems. In the OL model these variable are incorporated in the latent propensity. Hence, the OL model requires 3 parameters and 3 thresholds (a total of 6 parameters). In the GOL model the independent variables are incorporated in the latent propensity and thresholds 2 and 3 (threshold 1 is set to 0 for identification) resulting in 9 parameters, 1 propensity constant, 1 constant in threshold 2 and 1 constant in threshold 3, yielding a total of 12 parameters. The MNL model is estimated considering the first alternative as base. The MNL model has 9 parameters for independent variables and 3 constants (again a total of 12 parameters).

4.1 Data generation

4.1.1 Data Generation for model estimation comparison

For ordered logit data, we generate the latent propensity for an ordered logit model using the three independent variables and a standard logistic error term. The discrete dependent variable value is determined based on the threshold parameters assumed. For generalized logit data, we generate the latent propensity and threshold vectors using the three independent variables and a standard logistic error term. The discrete dependent variable value is determined based on the

⁴ To be sure, the idea is to examine, if there is any effect of mixed dgp on the appropriateness of the framework for analysis. There is very little information available to the analyst in terms of realizing the extent of mixture in the dgp. However, we believe the process will shed light on the strength and weakness of the alternative frameworks as the value of p changes.

location of the propensity with regard to the individual specific threshold. For multinomial logit data, we generate the systematic alternative utilities based on the three independent variables. The error components for the alternatives are generated using standard Type 1 extreme value distribution. The chosen alternative is determined based on the alternative with the highest utility.

To cover a wide range of possible dgp's we focus on two major configurations:

- (1) Ordered logit dgp and Multinomial logit dgp mixed data generation
- (2) Generalized ordered logit and Multinomial logit mixed data generation

The two dgps for the above configurations are generated by combining the independent dgps from ordered and unordered systems using a uniform random variable (U). For an individual, the ordered process generated dependent variable is the chosen alternative if $U < p$ and unordered process generated dependent variable is the chosen alternative if $U \geq p$. In this manner we can easily generate a mixed dgp with varying levels of mixture. The above process is repeated for 6 aggregate sample shares. The parameter vectors and the aggregate sample shares for the three regimes are provided in Table 1. The sample shares selected in our analysis were targeted at covering the most probable combination of a four alternative dependent variable. Sample 1 is the equal share alternative. Sample 2 has proportions starting at a high value for the first alternative and gradually reducing for the last alternative. Sample 3 is the mirror image of Sample 2. Sample 4 considers high proportions allocated to interior alternatives. Sample 5 reflects extreme loading on the last alternative with diminishing shares for the other alternatives. Sample 6 is the mirror image of Sample 5. The sample shares are chosen so as to reflect the commonly observed sample shares in safety literature for driver injury severity (sample 2 and 6) and pedestrian injury severity (sample 3 and sample 5).

The reader would note that it is not straight forward to generate parameters that impact the dependent variable in a similar fashion across the different model frameworks. Hence, we resort to assuming parameters that provides the same aggregate shares in the population for the three frameworks. The parameter set for each sample for the three model frameworks are also presented in table 1. To elaborate on the parameters provided in Table 1, we discuss the estimates for Sample 1 for each of the model frameworks. The OL the independent variable parameters are 1.25, 0.25, and 0.50 and thresholds parameters are -1.50, 0.00 and 1.50. The GOL model constant parameters are 1.90, 0.50, 0.70 and the independent variable parameters: (1) in the propensity equation are 1.5, 0.75, 1.50, (2) threshold 2 are 0.75, 0.50, 1.25, and (3) threshold 3 are -1.50, 0.75, and 0.50. For the MNL model the first alternative is considered the base; the constant parameters are -0.75, -0.25, and -0.50 and and the parameters for independent variables: (1) alternative 2 are 1.50, -0.25, and 0.50, (2) alternative 3 are -0.50, 0.75, and 0.25 and (3) alternative 4 are -1.25, -0.50, and -0.25. The remainder of the parameter set also follows the same structure. With the parameter set presented in Table 1, 50 realizations of the data with 5000 observations each are generated for each proportion value (p). We generate a total of 6 aggregate sample shares for the two configurations identified above. Within each configuration, we estimate the OL, GOL and MNL. In all the simulation exercises the three independent variables are retained to be the same across models.

4.1.2 Data Generation for under reporting

The same process described in Section 4.1.1 is repeated for the under reported data with one minor change. The entire data generated is not used in the estimation exercise. Towards generating an under reported dataset, – a dataset that has few individuals choosing the first alternative - we remove a fraction of the records that are assigned to the first level for the ordinal variable. Specifically, we only sample about 70% (i.e. remove 30 %) of the data points that are in the first category. The resulting overrepresentation of data is different for different aggregate shares. The OL, GOL and MNL models are estimated using the under reported sample.

4.2 Performance Evaluation Metrics

The emphasis of the research effort is on examining the performance of the OL, GOL and MNL models in retrieving the behavior in the simulated data. It is important to note that parameter retrieval for comparison is not a possible evaluation measure. Because, based on how we generated the mixed data it is quite complex to generate the “true” model parameters for the system unless one finds the equivalent a latent segmentation based combination. The discovery of such parameter space is far from trivial and is not essential for our research effort. Hence we resort to comparing the model fit measures for the estimation sample for the three models. In terms of model fit comparison, the GOL model is a generalized version of OL; we can compare these two models by using likelihood ratio test for selecting the preferred model. However, to compare the ordered approaches with the unordered approach the likelihood ratio test is not appropriate.

Towards this exercise, we employ the Bayesian Information Criterion (BIC) measure. The BIC for a given empirical model is equal to $-2\ln(L) + K \ln(Q)$, where $\ln(L)$ is the log-likelihood value at convergence, K is the number of parameters, and Q is the number of observations. The model with the *lower* BIC values is the preferred model. BIC measure comparison for the GOL and MNL models actually collapses to a comparison of the log-likelihood because both these models have the same number of parameters. However, we still maintain the BIC comparison because BIC imposes higher penalty for additional parameters (as is common when we estimate GOL/MNL). The BIC parameter allows us to ensure we are not over fitting that data, a relevant criterion for accident datasets that usually have a large set of exogenous variables.

We compute two measures of BIC: (1) the proportion of datasets for which a framework has the lowest BIC and (2) average BIC. The first measure is computed as the proportion a framework has the lowest BIC by checking which framework provides the lowest BIC for each dataset. This measure is computed once for every aggregate sample. The proportion will allow us to identify the superior model over the entire analysis for each aggregate share. The second measure - average BIC - is computed by averaging the BIC value for each p value i.e. for each aggregate sample we have 11 measures.

In the simulation exercise for the under-reported sample, we report similar measures computed in a slight different manner. Specifically, in the case of under reporting, we examine the performance of the three models in prediction i.e. estimate models with the underreported data and then employ the estimated parameters to predict the dependent variable in the “original” sample. The process allows us to see how the models perform when estimates are affected by under reporting. The predictive ability of these under reported estimates in the context of true

sample offers the most realistic comparison. The two measures computed earlier are computed for these predicted samples. The exercise is particularly relevant in the context of safety literature.

5 RESULTS

5.1 Model Estimation Comparison

The model estimation comparison results are discussed in this section. The OL, GOL and MNL models are estimated for 6600 datasets of 5000 observations each (2 (data configurations) * 6 (aggregate shares))* 11 (p values) * 50 samples). For all these model estimations, we compute the BIC measures. The two BIC measures computed are reported in Tables 2 and 3.

5.1.1 Proportion measure

In Table 2, the proportion of a particular model has the lowest BIC is reported for all 6 aggregate shares. In terms of the first data configuration (OL and MNL mixed dgp), we observe that the MNL model performs really well for 4 aggregate shares while GOL model performs really well for 2 aggregate shares. The MNL model clearly outperforms the other models in aggregate samples that are left skewed (samples 2 and 6). The MNL model also performs better than the GOL model for the equal share alternative sample. The GOL model outperforms the other two systems when we have aggregate shares that are right skewed (samples 3 and 5). In terms of the sample with high proportions of interior alternatives (sample 4) the MNL model is slightly better than the combined performance of OL and GOL model. The results clearly show that even for an OL and MNL mixture process the OL model rarely offers a good fit. This clearly establishes that the use of traditional ordered logit models for analyzing ordered discrete variables is quite restrictive. This supports the hypothesis of earlier safety research advocating the use of multinomial logit models. However, the results also indicate that employing the GOL model in this context reduces the difference substantially and in some cases even improves model fit.

For the second data configuration (GOL and MNL mixed dgp) the results for the proportion measure are quite different. The results clearly show that in the presence of GOL based ordered process the OL model is clearly inefficient. In terms of the comparison between GOL and MNL; the result is an equal split across the 6 aggregate shares. Similar to what is observed earlier the MNL model outperforms the GOL model in Sample 1 (equal shares), Sample 2 (right skewed) and Sample 6 (extremely right skewed). The GOL model outperforms the MNL model for Sample 3 (left skewed), Sample 5 (extremely left skewed) and Sample 4 (high proportion of interior shares). The major difference in the current configuration is the switch for sample 4 from an MNL preference to a GOL preference. The results support the hypothesis that GOL offers a credible alternative to MNL for modeling ordinal discrete variables. Of course, the most striking result from the above discussion is that distinct aggregate shares provide variation in model preference clearly highlighting that the aggregate share has an influence on how the alternative model frameworks perform.

5.1.2 Average BIC measure

The results corresponding to the average BIC measure are presented for all aggregate sample shares for all p values. The results are reported in Table 3 for the two data configurations. Table 3 provides information on the sample, the p value, the average BIC value for a constants only model as reference, and the average BIC values for the three models for data configuration 1 and 2. The lowest average BIC values for each model system and data configuration are underlined. The reader would note that as the p value increases the model shifts from a purely ordered (OL/GOL) process to a purely MNL process. In fact, the reader would note that the estimates for $p=1$ for data configurations 1 and 2 yield the same results because at $p=1$ we have a purely MNL process which is not influenced by the change in the ordered process from OL to GOL.

The average BIC measure is a more representative metric of comparison because the metric in addition to identifying the superior framework also provides a sense of the difference between the model frameworks. The difference will allow us to see how different the best model framework performs relative to the other frameworks. For the first data configuration, the results closely resemble the trends described in the proportion measures. The OL model performs competitively for p values of 0-0.1 usually indicating that unless the data is purely or very close to purely ordered logit ($p \leq 0.1$) the ordered system cannot perform better than GOL/MNL even using a stringent penalty for excess parameters. The result clearly highlights the inadequate explanatory power offered by the traditional ordered logit model. The GOL model usually takes over from the OL model once p values are >0.1 . The GOL model performs as well or better than the MNL model for p -values of up to 0.3 for Samples 1, and 2. On the other hand, for samples 3 and 5, the GOL model outperforms the MNL model for p values ranging from 0.2 through 0.8/0.9. A particularly important aspect to notice is even when the MNL model outperforms the GOL model the difference in LL is well below 2% in all cases (and even below 0.1% in many cases). The measure clearly indicates for a particular sample when GOL is inferior to the MNL the difference is a marginal one.

For the second data configuration, the results indicate that in the presence of GOL dgp, OL model is consistently inferior to GOL and MNL. In fact, the comparison here is a comparison of GOL and MNL models only. The trends observed here are similar to the trends observed for the proportion measure. For Samples 1, 2 and 6, the MNL model outperforms the GOL model for majority of the p value segments. For samples 3, 4 and 5 GOL model outperforms the MNL model. Again these results clearly highlight the potential of GOL model for examining ordinal discrete models. Further, the GOL model consistently performs well when compared to the MNL model irrespective of aggregate share. The results indicate that the performance of the GOL model for some empirical datasets (Eluru et al., 2008, Yasmin and Eluru, 2012) is not a lucky realization; but a realization of its strength in modeling of ordinal discrete variables. Again, the reader would note that in the event of a superior performance of the MNL model the GOL performance is well within 2% in all cases.

In summary the proportion measure and the average BIC highlight how GOL model offers a true ordered equivalent to the MNL model for a wide range of aggregate shares and inherently latent underlying dgps.

5.2 Performance in the Presence of Under-Reporting

The entire data generation and model estimation process was repeated to examine the performance of alternative frameworks under the presence of under reported data. The process involved estimating the alternative frameworks for the under reported data and employing the model parameters for predicting the outcome for the original dataset. To compare the alternative frameworks we employ the same measures generated previously. The results for the under reporting exercise are presented in Tables 4 and 5.

5.2.1 Proportion measure

In Table 4, the proportion of a particular model that has the lowest BIC is reported for all 6 aggregate shares. The proportion measure of prediction indicates that the overall performance trends for the various frameworks are not altered in the presence of under reporting. The result concludes that there is no evidence to suggest, at least for model prediction, one model is superior to another. Again, the result provides credence to the finding from Yasmin and Eluru, 2012 that the GOL model performed better than the MNL model. If the GOL model performs better than MNL model in estimation for a particular aggregate share; the performance in the presence of under reporting (given that the overall share structure is marginally affected) is likely to be very similar. The result is comforting in the sense that analysts can choose their preferred model solely on model estimation metrics and proceed with post processing analysis.

5.2.2 Average BIC measure

The results corresponding to the average BIC measure are presented for all aggregate sample shares for all p values. The results are reported in Table 5 for the two data configurations. The lowest average BIC values for each model system and data configuration are underlined.

The results for the average BIC measure also very closely resemble the results observed for the model estimation. The average BIC measures examining the BIC for every p value for all the 6 aggregate shares indicates that the trends obtained in model estimation continue to be followed even in the context of under reported data providing further evidence to the lack of systematic bias in terms of the model frameworks for prediction. The reader would also note that in the event of a superior performance of the MNL model the GOL performance is not inferior by a huge margin. Another aspect to note in the under reported case is the higher BIC values compared to the corresponding values in Table 3 indicating that there is a minor loss in efficiency when under reported sample based parameters are used to predict for the “true” sample. The result is in line with intuitive expectations.

5.3 Implications of the Findings

The two significant reasons for employing unordered models for ordered decision variables were: (1) substantial improvement in data fit offered by the unordered response models and (2) improved performance of unordered response models in presence of underreporting in the data. The current simulation exercise clearly indicates that it is possible to address these limitations with an improved ordered model, the Generalized Ordered Logit. The findings from our study have significant implications for accident analysis research. There is growing recognition within the safety community that modeling injury severity exogenous to seat belt use, alcohol

consumption, or collision type is not realistic. For instance, the common unobserved factors that influence seat belt usage might also influence injury severity (See Eluru and Bhat, 2007). Accommodating these unobserved correlations when we consider injury severity as an unordered process increases the number of error terms to be handled in the model formulation and estimation resulting in computationally intensive model structures. The estimation approaches require either simulation or some form of approximation to handle the complexity. However, with a GOL structure the number of error terms to be handled can be restricted to one without any loss in model explicative power. In fact, employing GOL might allow us to simultaneously examine multiple ordered decisions in a straight forward manner. Moreover, the simpler error structure of the GOL model will lend itself to closed form couplings such as copulas (Bhat and Eluru, 2009). The corresponding model for the unordered response model would require K couplings where K corresponds to the number of alternatives in the injury severity model.

The advantages of considering GOL model further increase when we intend to correlate two ordinal variables. In choice scenarios that require us to capture spatial correlation, the unordered systems are particularly cumbersome whereas GOL model while retaining the simplicity of the traditional ordered systems offers a competitive alternative to capture spatial correlations (see for example Castro et al., 2012, Narayanmoorthy et al., 2012). Further, it is important to note that the GOL framework explicitly recognizes the inherent ordering in the discrete variable while also remaining unaffected by the Independence from Irrelevant Alternatives (IIA) property of the MNL model. In summary, the GOL model avoids the shortcomings of the MNL model while maintaining reasonably close statistical fit in comparison to the MNL model.

6 CONCLUSIONS

The applicability of the ordered and unordered frameworks for analyzing ordinal discrete variables has evoked considerable debate on using the appropriate choice model for analysis. The ordered response models explicitly recognize the inherent ordering within the decision variable whereas the unordered response models neglect the ordering or require artificial constructs to consider the ordering (for example the ordered generalized extreme value model. On the other hand, the traditional ordered response models restrict the impact of exogenous variables on the choice process to be same across all alternatives while the unordered response models allow the model parameters to vary across alternatives. The unordered response models might not be as parsimonious as the ordered response models but offer greater explanatory power because of the additional exogenous effects that can be explored. Another concern with the ordered response framework is in the context of modeling datasets that might be affected by under reporting.

It is in this background that we undertake the current research effort. The objective of the current study is to investigate the performance of the ordered and unordered response frameworks at a fundamental level. Towards this end, we undertake a comparison of the alternative frameworks by estimating ordered (OL and GOL) and unordered response models (MNL) using data generated through ordered, unordered data and a combination of ordered and unordered data generation processes. Subsequently, we examine the influence of aggregate sample shares on the appropriateness of the modeling framework. Rather than be limited by the aggregate sample shares in an empirical dataset, simulation allows us to explore the influence of a broad spectrum of sample shares on the performance of ordered and unordered frameworks. Third, we extend the

data generation process based analysis to under reported data and compare the performance of the ordered and unordered response frameworks. Finally, based on these simulation exercises, we provide a summary of the strengths and weaknesses of the two frameworks for analyzing ordered discrete variables.

The simulation exercise is undertaken for a four alternative ordered dependent variable. To cover a wide range of possible dgps we focus on two major data generation configurations: (1) Ordered logit dgp and Multinomial logit dgp mixed data generation and (2) Generalized ordered logit and Multinomial logit mixed data generation. For each of these configurations 6 different aggregate sample share based dgps are produced. The study also generates under reported datasets by removing a fraction of the records that are assigned to the first level for the ordinal variable. Specifically, we only sample about 70% (i.e. remove 30 %) of the data points that are in the first category.

The emphasis of the research effort is on examining the performance of the OL, GOL and MNL models in retrieving the behavior in the simulated data. For this purpose, we employ the Bayesian Information Criterion (BIC) measures. We compute two measures of BIC: (1) the proportion of datasets for which a framework has the lowest BIC and (2) the average BIC. The first measure is computed by checking which framework provides the lowest BIC for each dataset. This measure is computed once for every aggregate sample. The proportion will allow us to identify the superior model over the entire analysis for each aggregate share. The average BIC is computed by averaging the BIC value for each proportion (p) value i.e. for each aggregate sample we have 11 measures. The most striking result from the comparison exercise is that distinct aggregate shares provide variation in the model preference clearly highlighting that the aggregate share has an influence on how the alternative model frameworks perform. Further, the GOL model consistently performs well when compared to the MNL model irrespective of aggregate share. Even in cases where the MNL outperforms the GOL, the difference between the two frameworks is marginal. The results clearly indicate the emergence of the GOL model as the true ordered equivalent vis-a-vis the MNL model for examining ordinal discrete variables. Moreover, the result is of great significance for econometric modellers in general and safety researchers in particular for developing joint models and spatial effect models involving ordinal discrete variables. The application of GOL model frameworks for ordinal data reduces the computational burden of formulating joint models and spatial effect models compared to using the unordered response models.

ACKNOWLEDGMENTS

The corresponding author would like to acknowledge discussions with Abdul Pinjari and Rajesh Paleti in the manuscript preparation process. Three anonymous reviewers provided useful comments on an earlier version of the paper. The corresponding author would like to acknowledge financial support from Natural Sciences and Engineering Research Council (NSERC) of Canada.

REFERENCES

- Anowar S., S. Yasmin, N. Eluru, and L. Miranda-Moreno (2012). Analyzing Car Ownership in Two Quebec Metropolitan Regions: A Comparison of Latent Ordered and Unordered Response Models. Technical Paper, Department of Civil Engineering and Applied Mechanics, McGill University.
- Bhat, C.R., and V. Pulugurta (1998). A Comparison of Two Alternative Behavioral Choice Mechanisms for Household Auto Ownership Decisions. *Transportation Research Part B*, Vol. 32, No. 1, pp. 61-75.
- Bhat, C.R., Eluru, N., (2009). A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. *Transportation Research Part B: Methodological* 43 (7), 749-765.
- Bhat, C.R., I.N. Sener, and N. Eluru (2010), "A Flexible Spatially Dependent Discrete Choice Model: Formulation and Application to Teenagers' Weekday Recreational Activity Participation," *Transportation Research Part B*, Vol. 44, No. 8-9, pp. 903-921
- Castro, M., R. Paleti, and C.R. Bhat (2012a), "A Spatial Generalized Ordered Response Model to Examine Highway Crash Injury Severity," Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin
- Castro, M., R. Paleti, and C.R. Bhat (2012b), "A Latent Variable Representation of Count Data Models to Accommodate Spatial and Temporal Dependence: Application to Predicting Crash Frequency at Intersections," *Transportation Research Part B*, Vol. 46, No. 1, pp. 253-272
- Eluru, N., and C. R. Bhat (2007). A Joint Econometric Analysis of Seat Belt Use and Crash-Related Injury Severity. *Accident Analysis and Prevention*, Vol. 39, No. 5, pp.1037-1049.
- Eluru, N., C. R. Bhat, and D. A. Hensher. A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes. *Accident Analysis and Prevention*, Vol. 40, No. 3, 2008, pp.1033-1054.
- Elvik, R., and A. B. Mysen (1999). incomplete Accident Reporting: Meta-analysis of Studies Made in 13 Countries. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1665, Transportation Research Board of the National Academies, Washington, D.C., 1999, pp. 133-140.
- Ferdous, N., N. Eluru, C.R. Bhat, and I. Meloni (2010), "A Multivariate Ordered Response Model System for Adults' Weekday Activity Episode Generation by Activity Purpose and Social Context," *Transportation Research Part B*, Vol. 44, No. 8-9, pp. 922-943
- Maddala, G. S. Limited-dependent and Qualitative Variables in Econometrics. Cambridge University Press, New York, 1983.
- Narayanamoorthy, S., R. Paleti, and C.R. Bhat, "A Spatial Model for Examining Bicycle and Pedestrian Injuries," Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, July 2012
- Paleti, R., C.R. Bhat, and P. Singh, "A Spatial Model for Examining Firm Location Decisions," Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, July 2012
- Patil, S., Geedipally, S.R., Lord, D., (2011). Analysis of crash severities using nested logit model—accounting for the under reporting of crashes. *Accident Analysis & Prevention*.

- Savolainen, P. T., F. L. Mannering, D. Lord, and M. A. Quddus (2011). The Statistical Analysis of Highway Crash-Injury Severities: A Review and Assessment of Methodological Alternatives. *Accident Analysis and Prevention*, Vol. 43, No. 5, pp. 1666-1676.
- Srinivasan, K. K (2002). Injury Severity Analysis with Variable and Correlated Thresholds: Ordered Mixed Logit Formulation. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1784, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 132-142.
- Terza, J.V (1985). Ordinal Probit: A Generalization. *Communications in Statistics: Theory and Methods*, Vol. 14, No. 1, 1985, pp. 1-11.
- Yamamoto, T., J. Hashiji, and V. N. Shankar. Under reporting in Traffic Accident Data, Bias in Parameters and the Structure of Injury Severity Models. *Accident Analysis and Prevention*, Vol. 40, No. 4, 2008, pp. 1320-1329.
- Yasmin S., and N. Eluru, "Evaluating Alternate Discrete Choice Frameworks for Modeling Crash Injury Severity," Technical Paper, Department of Civil Engineering and Applied Mechanics, McGill University, Original version August 2012
- Ye, F., and D. Lord. Investigating Effects of Under reporting of Crash Data on Three Commonly Used Traffic Crash Severity Models: Multinomial Logit, Ordered Probit, and Mixed Logit Models. In *Transportation Research Board 90th Annual Meeting*. Washington DC: Transportation Research Board (TRB), 2011.

LIST OF TABLES

Table 1: OL, GOL and MNL estimates and the corresponding Aggregate Shares for a 4 level ordinal discrete variable

Table 2: Comparison of BIC Proportion Measures for Model Estimation

Table 3: Comparison of Average BIC value for the Estimation Sample

Table 4: Comparison of BIC Proportion Measures for Under reporting Data Prediction

Table 5: Predictive BIC for OL, GOL, and MNL based on under reported sample estimates

Table 1: OL, GOL and MNL estimates and the corresponding Aggregate Shares for a 4 level ordinal discrete variable

Sample	Sample Shares		OL parameters		GOL parameters			MNL parameters	
1	Alternative 1	25.0	Propensity	1.25 0.25 0.50	Propensity	1.90 1.50 0.75 1.50	Alternative 1	---	
	Alternative 2	25.0	Threshold 1	-1.50	Threshold 1	---	Alternative 2	-0.75 1.50 -0.25 0.50	
	Alternative 3	25.0	Threshold 2	0.00	Threshold 2	0.50 0.75 0.50 1.25	Alternative 3	-0.25 -0.50 0.75 0.25	
	Alternative 4	25.0	Threshold 3	1.50	Threshold 3	0.70 -1.50 0.75 0.50	Alternative 4	-0.50 -1.25 -0.50 -0.25	
2	Alternative 1	62.6	Propensity	1.00 0.25 0.50	Propensity	-0.75 1.5 -0.75 0.5	Alternative 1	---	
	Alternative 2	25.1	Threshold 1	0.65	Threshold 1	---	Alternative 2	-1.55 1.50 -0.75 0.50	
	Alternative 3	7.8	Threshold 2	2.40	Threshold 2	0.25 -0.50 0.75 0.25	Alternative 3	-2.55 -0.50 0.75 0.25	
	Alternative 4	4.5	Threshold 3	3.60	Threshold 3	-0.50 -1.25 0.25 -0.25	Alternative 4	-3.50 -1.25 0.25 -0.25	
3	Alternative 1	4.5	Propensity	2.50 0.75 1.50	Propensity	4.25 1.50 -0.75 0.5	Alternative 1	---	
	Alternative 2	9.8	Threshold 1	-6.00	Threshold 1	---	Alternative 2	-1.00 1.50 -0.75 0.50	
	Alternative 3	24.5	Threshold 2	-3.75	Threshold 2	0.19 -0.50 0.75 0.25	Alternative 3	1.75 -0.50 0.75 0.25	
	Alternative 4	61.2	Threshold 3	1.00	Threshold 3	0.20 -1.25 0.25 -0.25	Alternative 4	2.75 -1.25 0.25 -0.25	
4	Alternative 1	5.9	Propensity	1.50 0.75 1.50	Propensity	4.65 1.50 -1.75 0.50	Alternative 1	---	
	Alternative 2	44.2	Threshold 1	-4.50	Threshold 1	---	Alternative 2	1.45 1.50 -1.75 0.50	
	Alternative 3	44.1	Threshold 2	0.00	Threshold 2	1.32 -0.5 0.75 0.25	Alternative 3	1.65 -0.50 0.75 0.25	
	Alternative 4	5.8	Threshold 3	4.50	Threshold 3	2.00 -1.25 0.25 -0.25	Alternative 4	-0.85 -1.25 0.25 -0.25	
5	Alternative 1	2.1	Propensity	1.50 0.75 1.50	Propensity	4.30 0.50 -0.50 0.75	Alternative 1	---	
	Alternative 2	2.4	Threshold 1	-6.00	Threshold 1	---	Alternative 2	-0.50 0.50 -0.50 0.75	
	Alternative 3	4.3	Threshold 2	-5.00	Threshold 2	-1.65 0.75 -0.50 0.50	Alternative 3	0.25 0.75 -0.50 0.50	
	Alternative 4	91.2	Threshold 3	-4.00	Threshold 3	-0.50 0.50 0.50 -0.50	Alternative 4	4.00 0.50 0.50 -0.50	
6	Alternative 1	90.6	Propensity	1.50 0.75 1.50	Propensity	-3.25 1.50 -0.75 0.50	Alternative 1	---	
	Alternative 2	4.6	Threshold 1	3.75	Threshold 1	---	Alternative 2	-4.20 1.50 -0.75 0.50	
	Alternative 3	2.5	Threshold 2	4.75	Threshold 2	-1.00 -0.50 0.75 0.25	Alternative 3	-4.00 -0.50 0.75 0.25	
	Alternative 4	2.3	Threshold 3	5.75	Threshold 3	-1.65 -1.25 0.25 -0.25	Alternative 4	-4.50 -1.25 0.25 -0.25	

Table 2: Comparison of BIC Proportion Measures for Model Estimation

Sample	Proportion of lowest BIC					
	Mixed DGP: OL and MNL			Mixed DGP: GOL and MNL		
	OL	GOL	MNL	OL	GOL	MNL
1	0.18	0.19	0.63	0.00	0.36	0.64
2	0.18	0.19	0.63	0.00	0.14	0.86
3	0.14	0.67	0.19	0.00	0.57	0.43
4	0.19	0.28	0.53	0.00	0.65	0.35
5	0.25	0.64	0.11	0.00	0.75	0.25
6	0.33	0.04	0.63	0.00	0.14	0.86

Table 3: Comparison of Average BIC value for the Estimation Sample

Sample	p	BIC for constants only Model	Mixed DGP: OL and MNL			Mixed DGP: GOL and MNL		
			OL	GOL	MNL	OL	GOL	MNL
1	0.0	13887	<u>11837</u>	11883	11909	9629	<u>6736</u>	7323
	0.1	13887	<u>12516</u>	12538	12591	10394	<u>8533</u>	8738
	0.2	13887	13083	<u>13038</u>	13077	11092	<u>9700</u>	9782
	0.3	13888	13430	<u>13297</u>	13313	11572	<u>10383</u>	10412
	0.4	13888	13692	13440	<u>13425</u>	11998	10921	<u>10901</u>
	0.5	13888	13848	13434	<u>13386</u>	12352	11275	<u>11216</u>
	0.6	13888	13887	13302	<u>13219</u>	12596	11446	<u>11359</u>
	0.7	13888	13818	13002	<u>12872</u>	12802	11521	<u>11390</u>
	0.8	13887	13682	12631	<u>12446</u>	12919	11498	<u>11334</u>
	0.9	13888	13377	11871	<u>11579</u>	13036	11284	<u>11026</u>
	1.0	13887	13054	11055	<u>10710</u>	13054	11055	<u>10710</u>
2	0.0	9811	<u>8620</u>	8665	8675	8326	<u>7172</u>	7277
	0.1	9805	<u>8777</u>	8807	8834	8700	<u>8092</u>	8095
	0.2	9815	8970	<u>8946</u>	8976	9059	8762	<u>8733</u>
	0.3	9798	9075	<u>8983</u>	8999	9291	9132	<u>9095</u>
	0.4	9826	9225	9040	<u>9024</u>	9506	9389	<u>9351</u>
	0.5	9840	9371	9061	<u>8989</u>	9689	9528	<u>9482</u>
	0.6	9813	9437	8999	<u>8858</u>	9751	9485	<u>9419</u>
	0.7	9821	9526	8894	<u>8656</u>	9830	9356	<u>9251</u>
	0.8	9816	9585	8780	<u>8456</u>	9829	9144	<u>8982</u>
	0.9	9809	9646	8505	<u>8059</u>	9763	8656	<u>8362</u>
	1.0	9812	9692	8175	<u>7726</u>	9692	8175	<u>7726</u>
3	0.0	10134	<u>6077</u>	6122	6150	9323	<u>6398</u>	7097
	0.1	10145	7605	<u>7603</u>	7791	9554	<u>7460</u>	7888
	0.2	10135	8700	<u>8650</u>	8811	9781	<u>8235</u>	8548
	0.3	10138	9335	<u>9236</u>	9353	9860	<u>8675</u>	8897
	0.4	10125	9782	<u>9439</u>	9701	9913	<u>9029</u>	9169
	0.5	10161	10074	<u>9874</u>	9916	9920	<u>9291</u>	9343
	0.6	10181	10163	<u>9885</u>	9915	9856	9366	<u>9353</u>
	0.7	10153	10036	<u>9678</u>	9707	9668	9209	<u>9183</u>
	0.8	10171	9854	<u>9418</u>	9439	9522	9044	<u>9026</u>
	0.9	10161	9350	8749	<u>8728</u>	9159	8541	<u>8498</u>
	1.0	10153	8828	8084	<u>7978</u>	8828	8084	<u>7978</u>

4	0.0	10584	<u>7087</u>	7133	7144	9512	<u>6383</u>	6998
	0.1	10589	<u>8027</u>	8055	8158	9818	<u>8146</u>	8231
	0.2	10589	8778	<u>8763</u>	8871	10043	9138	<u>9095</u>
	0.3	10576	9229	<u>9166</u>	9246	10125	9561	<u>9560</u>
	0.4	10590	9584	<u>9454</u>	9494	10203	<u>9894</u>	9906
	0.5	10576	9778	9556	<u>9547</u>	10129	10002	<u>10001</u>
	0.6	10564	9819	9494	<u>9435</u>	10017	<u>9906</u>	9925
	0.7	10559	9731	9252	<u>9140</u>	9794	<u>9545</u>	9604
	0.8	10562	9561	8904	<u>8741</u>	9553	<u>9088</u>	9151
	0.9	10548	9133	8106	<u>7869</u>	9118	8168	<u>8142</u>
	1.0	10571	8751	7173	<u>6945</u>	8751	7173	<u>6945</u>
5	0.0	3718	<u>2652</u>	2697	2707	3267	<u>2887</u>	2947
	0.1	3769	<u>2983</u>	3008	3031	3372	<u>3064</u>	3135
	0.2	3817	3263	<u>3258</u>	3302	3493	<u>3238</u>	3314
	0.3	3807	3384	<u>3356</u>	3413	3492	<u>3285</u>	3355
	0.4	3875	3559	<u>3519</u>	3577	3590	<u>3424</u>	3485
	0.5	3926	3676	<u>3636</u>	3685	3652	<u>3537</u>	3580
	0.6	3913	3684	<u>3643</u>	3687	3630	<u>3548</u>	3576
	0.7	3955	3710	<u>3660</u>	3694	3630	<u>3579</u>	3587
	0.8	4020	3734	<u>3687</u>	3709	3659	<u>3615</u>	3613
	0.9	4040	3649	<u>3601</u>	3604	3610	3567	<u>3557</u>
	1.0	4074	3594	3536	<u>3521</u>	3594	3536	<u>3521</u>
6	0.0	4160	<u>2959</u>	3005	3017	3393	<u>3157</u>	3183
	0.1	4153	<u>3183</u>	3224	3251	3570	<u>3464</u>	3465
	0.2	4144	<u>3397</u>	3422	3446	3732	<u>3687</u>	3677
	0.3	4141	<u>3541</u>	3543	3555	3840	3824	<u>3804</u>
	0.4	4148	3686	3659	<u>3649</u>	3938	3927	<u>3895</u>
	0.5	4153	3832	3770	<u>3729</u>	4038	4010	<u>3957</u>
	0.6	4130	3907	3808	<u>3729</u>	4070	4006	<u>3925</u>
	0.7	4115	3967	3803	<u>3671</u>	4110	3989	<u>3867</u>
	0.8	4118	4023	3816	<u>3635</u>	4127	3947	<u>3787</u>
	0.9	4132	4093	3787	<u>3546</u>	4139	3838	<u>3620</u>
	1.0	4123	4114	3697	<u>3428</u>	4114	3697	<u>3428</u>

Table 4: Comparison of BIC Proportion Measures for Under reporting Data Prediction

Sample	Proportion of lowest BIC					
	Mixed DGP: OL and MNL			Mixed DGP: GOL and MNL		
	OL	GOL	MNL	OL	GOL	MNL
1	0.18	0.19	0.63	0.01	0.37	0.62
2	0.19	0.18	0.63	0.00	0.14	0.86
3	0.14	0.67	0.19	0.01	0.56	0.43
4	0.19	0.29	0.52	0.06	0.64	0.30
5	0.25	0.64	0.11	0.00	0.75	0.25
6	0.33	0.04	0.63	0.00	0.14	0.86

Table 5: Predictive BIC for OL, GOL, and MNL based on under reported sample estimates

Sample	p	Mixed DGP: OL and MNL			Mixed DGP: GOL and MNL		
		OL	GOL	MNL	OL	GOL	MNL
1	0.0	<u>11925</u>	11972	11998	9718	<u>6807</u>	7402
	0.1	<u>12609</u>	12630	12684	10486	<u>8612</u>	8826
	0.2	13185	<u>13137</u>	13178	11188	<u>9789</u>	9876
	0.3	13540	<u>13402</u>	13419	11677	<u>10483</u>	10517
	0.4	13811	13549	<u>13534</u>	12107	11026	<u>11010</u>
	0.5	13973	13546	<u>13497</u>	12469	11385	<u>11329</u>
	0.6	14015	13414	<u>13329</u>	12714	11556	<u>11471</u>
	0.7	13955	13121	<u>12988</u>	12933	11637	<u>11508</u>
	0.8	13822	12748	<u>12559</u>	13053	11612	<u>11448</u>
	0.9	13524	11988	<u>11690</u>	13181	11400	<u>11141</u>
	1.0	13205	11168	<u>10820</u>	13205	11168	<u>10820</u>
2	0.0	<u>8744</u>	8791	8800	8436	<u>7276</u>	7382
	0.1	<u>8901</u>	8933	8961	8823	<u>8211</u>	<u>8214</u>
	0.2	9100	<u>9077</u>	9108	9196	8896	<u>8865</u>
	0.3	9205	<u>9114</u>	9133	9437	9277	<u>9239</u>
	0.4	9359	9174	<u>9160</u>	9658	9539	<u>9500</u>
	0.5	9511	9197	<u>9127</u>	9851	9686	<u>9638</u>
	0.6	9581	9139	<u>8998</u>	9914	9643	<u>9575</u>
	0.7	9674	9037	<u>8795</u>	9998	9517	<u>9407</u>
	0.8	9736	8923	<u>8595</u>	9995	9301	<u>9134</u>
	0.9	9803	8648	<u>8194</u>	9927	8808	<u>8504</u>
	1.0	9854	8322	<u>7863</u>	9854	8322	<u>7863</u>
3	0.0	<u>6094</u>	6139	6167	9377	<u>6419</u>	7121
	0.1	7625	<u>7624</u>	7812	9589	<u>7474</u>	7904
	0.2	8721	<u>8670</u>	8833	9821	<u>8264</u>	8572
	0.3	9359	<u>9260</u>	9378	9896	<u>8703</u>	8923
	0.4	9808	<u>9654</u>	9726	9947	<u>9058</u>	9196
	0.5	10101	<u>9900</u>	9942	9950	<u>9320</u>	9368
	0.6	10192	<u>9913</u>	9943	9885	9397	<u>9380</u>
	0.7	10064	<u>9706</u>	9736	9697	9239	<u>9211</u>
	0.8	9882	<u>9444</u>	9466	9548	9071	<u>9052</u>
	0.9	9378	8775	<u>8756</u>	9187	8567	<u>8525</u>
	1.0	8858	8110	<u>8005</u>	8858	8110	<u>8005</u>

4	0.0	<u>7112</u>	7158	7170	9584	<u>6406</u>	7024
	0.1	<u>8053</u>	8081	8185	9880	<u>8174</u>	8262
	0.2	8806	<u>8791</u>	8900	10095	9143	<u>9133</u>
	0.3	9259	<u>9195</u>	9276	10162	9600	<u>9594</u>
	0.4	9618	<u>9486</u>	9526	10240	<u>9931</u>	9937
	0.5	9814	9589	<u>9580</u>	10163	10037	<u>10034</u>
	0.6	9858	9530	<u>9470</u>	10048	<u>9939</u>	9957
	0.7	9773	9289	<u>9174</u>	9829	<u>9580</u>	9639
	0.8	9604	<u>8202</u>	8774	9590	<u>9121</u>	9184
	0.9	9182	8145	<u>7903</u>	9164	8205	<u>8178</u>
	1.0	8805	7209	<u>6979</u>	8805	7209	<u>6979</u>
5	0.0	<u>2662</u>	2708	2718	3282	<u>2900</u>	2961
	0.1	<u>2994</u>	3019	3043	3386	<u>3076</u>	3148
	0.2	3275	<u>3271</u>	3315	3509	<u>3252</u>	3329
	0.3	3396	<u>3369</u>	3425	3506	<u>3299</u>	3370
	0.4	3573	<u>3533</u>	3591	3605	<u>3439</u>	3500
	0.5	3690	<u>3650</u>	3699	3667	<u>3552</u>	3595
	0.6	3698	<u>3657</u>	3702	3645	<u>3563</u>	3591
	0.7	3725	<u>3675</u>	3709	3645	<u>3595</u>	3601
	0.8	3749	<u>3702</u>	3723	3674	3630	<u>3627</u>
	0.9	3664	<u>3615</u>	3619	3625	3581	<u>3571</u>
	1.0	3609	3550	<u>3535</u>	3609	3550	<u>3535</u>
6	0.0	<u>2998</u>	3045	3058	3436	<u>3204</u>	3230
	0.1	<u>3226</u>	3268	3295	3617	<u>3513</u>	3514
	0.2	<u>3443</u>	3469	3493	3784	3741	<u>3731</u>
	0.3	<u>3590</u>	3592	3605	3895	3880	<u>3859</u>
	0.4	3737	3711	<u>3701</u>	3996	3985	<u>3952</u>
	0.5	3886	3825	<u>3783</u>	4098	4069	<u>4016</u>
	0.6	3963	3865	<u>3785</u>	4132	4067	<u>3984</u>
	0.7	4024	3862	<u>3727</u>	4173	4051	<u>3926</u>
	0.8	4082	3873	<u>3692</u>	4190	4009	<u>3846</u>
	0.9	4153	3847	<u>3603</u>	4201	3900	<u>3677</u>
	1.0	4175	3746	<u>3485</u>	4175	3746	<u>3485</u>