A Unimodal Ordered Logit model for ranked choices

Melvin Wong

José Ángel Martín-Baos

Michel Bierlaire

Transport and Mobility Laboratory (EPFL)

September 2021

STRC | **21st Swiss Transport Research Conference**
Monte Verità / Ascona, September 12 – 14, 2021

Transport and Mobility Laboratory (EPFL)

A Unimodal Ordered Logit model for ranked choices

Melvin Wong, Michel Bierlaire
Transport and Mobility Laboratory
École polytechnique fédérale de Lausanne
TRANSP-OR, Station 18, CH-1015 Lausanne
phone:
fax:
{melvin.wong,michel.bierlaire}@epfl.ch

José Ángel Martín-Baos
Department of Mathematics
University of Castilla-La Mancha
Paseo de la Universidad, 4, Ciudad Real, Spain
phone:
fax:
JoseAngel.Martin@uclm.es

September 2021

Abstract

Ordinal scale responses capture qualitative user feedback which can be used to model individual choice preference, or are employed in traffic accident analysis to evaluate accident severity. We present a new choice model for ordinal scale responses in choice tasks that combines a Multinomial Logit model with a Poisson probability mass function. The Poisson distribution, which is suitable for modelling the occurrence of the number of events in a fixed time frame, independent of previous events, can be adapted into the unobserved error distribution of a standard MNL model to capture the natural ordering of the choices by imposing a unimodal constraint on the *a posteriori* choice probability. In this paper we describe the theoretical framework and the specification of the Unimodal Logit model. We apply our model to evaluate accident severity concerning road collisions. Our results are compared against the traditional ordered logit model and the MNL model.

Keywords

Discrete choice modelling; Ordered logit; Unimodality; Accident analysis

1 Introduction

Ordinal scale responses such as public transport satisfaction or driver star-rating in ride-hailing services, capture qualitative user feedback which is usually used to model individual choice preference among a set of ranked options (Krueger *et al.*, 2019, Tirachini and del Río, 2019, Fu, 2020, Loa and Habib, 2021). In transportation ordinal scale responses are also employed in traffic accident analysis to evaluate crash severity (Kaplan and Prato, 2012, Zeng *et al.*, 2019). A distinguishing feature of such choice sets is that the responses have an inherent correlation with other alternatives in the set based its proximate covariance (Small, 1987).

The earliest example of using ordering information for regression is the Proportional Odds Model (McCullagh, 1980). The choice categories are described as contiguous intervals on a continuous scale and the points of division between intervals are assumed to be unknown. Ordered Logit or Generalized Ordered Logit models are derived based on defining an unobserved latent variable that varies across the contiguous intervals that depends on exogenous features of the choice and the choice is specified as a probability function of lying in any one of the contiguous intervals.

Alternatively, using a classical multinomial logit (MNL) model is also possible. The MNL model is computationally tractable, simpler to model and interpret, but is generally not suitable for ordered choice modelling tasks for a number of reasons. In particular, although the MNL model might produce relatively good model fit, the MNL model and its variants such as the Nested Logit (NL), Probit or Mixed Logit (ML) model do not conform to the specification of the ordered nature of ranking data (Train, 2003). Although the Ordered Logit is widely used to this day to model data with a natural ordering of choices, it relies on the basic Proportional Odds concept of a continuous scale and contiguous interval method.

Our proposed approach to modelling ordered choice data uses a familiar perspective of a utility error correction term. This study develops a new choice model – the Unimodal Logit Model which employs a Poisson distribution function in the utility error correction term to account for proximity correlation between alternatives. Our proposed model focuses on the idea of constrained unimodality (Hall *et al.*, 2001). Imposing this constraint accounts for the proximity correlation between alternatives and non-correlated attributes in discrete choice data without fixed thresholds. We hypothesize that imposing such constraint leads to better model fit without increasing the number of parameters of the model.

Unimodality in artificial neural networks for ordered objects classification have been studied previously (da Costa *et al.*, 2008, Beckham and Pal, 2017). In this study, we adopt similar approaches for discrete choice modelling. The unimodal constraint is the ordering nature of the choices where the further a choice is from the desired option, the less probable it is to be selected, with respect to the selected choice. The Poisson distribution, which is suitable for modelling the occurrence of the number of events in a fixed time frame, independent of previous events, can be adapted into the unobserved error distribution of a standard MNL model to account for the proximity of the ranking order of choices. The expected value of the Poisson distribution is expressed as an unconstrained variable in the model that captures the latent response variable and the alternative specific observed utility component captures the attractiveness of the alternative. Our assumption is that the distribution of ordered choices should follow a unimodal distribution. The unimodality constraint imposed by the new error correction term in the utility would penalize the probabilities of non-chosen alternatives based on the ordering proximity.

2 Background

2.1 Modelling ordered choice

Discrete choice analysis aims to model the determinants of preferences of individuals over a set of choices. We assume that in a choice decision, there are J alternatives ($i = 1, ..., J$) and the utilities for individual $n = 1, ..., N$ are given by $U_{n1}, ..., U_{nJ}$. Let $y_{ni} = 1$ denote that the individual n prefers alternative i and $y_{ni} = 0$ otherwise, implying that $U_{ni} \geq \max\{U_{n1}, ..., U_{nJ}\}$ when $y_{ni} = 1$.

The random utility theory assumption is that the decision maker knows all utilities U_{n1} to U_{nJ} but the analyst does not observe those utilities, therefore the utilities are defined as $U_{ni} = V_{ni} + \varepsilon_{ni}$, where V_{ni} are the observed characteristics and ε_{ni} are the unobserved components of the utility, assumed to be independently and identically distributed (i.i.d.), $\varepsilon_{ni} \sim \text{Gumbel}(0, 1)$. This leads to the MNL model expression for the probability of the decision maker selecting alternative i (McFadden, 1974):

$$P(y_{ni} = 1) = P(U_{ni} \geq \max\{U_{n1}, ..., U_{nJ}\}) = \frac{\exp(V_{ni})}{\sum_{j=1}^{J} \exp(V_{nj})} \tag{1}$$

where V_{ni} is modelled as a linear function of the characteristics of the individual n and alternative i. For example: $V_{ni} = \sum_m \beta_{im} x_{nm}$ with m characteristics and β_{im} is an alternative specific parameter associated with alternative i.

For choices with natural ordering, the i.i.d. assumption does not hold, and the unobserved error component depends on some proximity measure between the alternatives. For example, in a vehicle ownership model where the choice set is the number of household vehicles, any decision maker choosing alternative j would prefer j to $j-1$, $j-1$ to $j-2$, $j-2$ to $j-3$, and so on (Sheffi, 1979). When the alternatives are ordered along a dimension which provides a natural rank relative to other alternatives, the unobserved error components close to each other should be more correlated than the error components further away (Small, 1987).

2.2 Ordered Logit model

To deal with ordinal choices, the analyst can treat the problem as a linear regression and use a squared error loss function to model the choice as count-data. Although this takes into account the proximity of the choices, it is not an ideal approach as it assumes that choices are continuous real-valued instead of discrete classes. The Ordered Logit model is a type of ordered response model based on the cumulative probabilities of the response variable and the logit is assumed to be a linear function of the attributes and regression coefficients which is constant across the responses. This approach is used to account for the proximity between the responses while retaining the discrete nature of the response. The most prominent method is to transform the responses into a cumulative response inequality:

$$y_{ni} = 1 \Leftrightarrow y_n^* \leq \tau_1 \quad \text{for} \quad i = 1 \tag{2}$$

$$y_{ni} = 1 \Leftrightarrow \tau_{i-1} < y_n^* \leq \tau_i \quad \text{for} \quad i = 2, ..., J-1, \text{ and} \tag{3}$$

$$y_{ni} = 1 \Leftrightarrow \tau_J < y_n^* \quad \text{for} \quad i = J \tag{4}$$

The response variable (ratings, severity, agreeableness, etc.), are assumed to be derived from a latent variable y^*. For instance, in the *Presidential questionnaire* example given in Train (2003), a latent variable is used to represent the opinion of the respondent, where a higher or lower level of y^* means the President is doing a better or poorer job respectively, and the choice task is to select from 5 possible ranked responses ($i = 1$ to $i = 5$) where 1 is "very poor job" and 5 is "very good job". This latent variable is defined as a linear function of exogenous variables:

$$y_n^* = \sum_m \beta_m X_{mn} + \varepsilon_n, \quad \text{for} \quad n = 1, 2, ..., N \tag{5}$$

where X_{mn} represents the attributes that enter the latent variable. The parameters β_m are the constant across all alternatives. The way the attributes X_{mn} enter the latent variable is defined by the analyst, but it was suggested by convention to use a linear form, which produces a familiar random utility function (Greene and Hensher, 2010). The motivation to use this method is that there is an unobserved component that is a linear function of exogenous X_{mn} variables and a random variable ε_n. The probability of y_n^* lying in each range of thresholds (τ_1 to τ_4) varies with $\beta_m X_{mn}$. The decision maker chooses the alternative from the reference point of y_n^* given the set of thresholds between each alternative. Existing methods for modelling ordinal data are based on post-hoc cut-off points, and only after estimation that these points can be evaluated. The ordered logit can be interpreted as a censoring of the underlying latent variable which ranges over all real numbers: $-\infty < y*_n < \infty$, the thresholds τ_j are in increasing order ($\tau_1 < \tau_2 < ... < \tau_{J-1}$), individual specific, with $J - 1$ number of thresholds, where J being the number of possible choices. The choices 1 through 5 can be represented as:

1 if $y_n^* < \tau_1$,

2 if $\tau_1 \leq y_n^* < \tau_2$,

3 if $\tau_2 \leq y_n^* < \tau_3$,

4 if $\tau_3 \leq y_n^* < \tau_4$, and

5 if $\tau_4 \leq y_n^*$

The difference between two adjacent thresholds (e.g. between τ_2 and τ_3) are assumed to be the same for all individuals and the thresholds are defined as an estimable parameter in the choice model (Greene and Hensher, 2010). The parameters β_m in the latent variable y_n^* are not indexed by the alternatives and so the effects of the parameters are constant across the responses. If we add a constant to the latent variable, the model would not be identified, as the operation can be cancelled by subtracting the same value from each threshold, and it would not change the probabilities. This can be resolved by fixing the first threshold to $\tau_1 = 0$.

For the ordered logit model to be estimated, the choice probability is written as:

$$P(y_n > i) = \frac{\exp(\tau_i - \sum_m \beta_m X_{mn})}{1 + \exp(\tau_i - \sum_m \beta_m X_{mn})}, \quad \text{for} \quad i = 1, 2, ..., J - 1 \tag{6}$$

The above equation corresponds to taking the sigmoid function: $sigmoid(x) = 1/(1 + \exp(-\phi))$ where $\phi = \sum_m \beta_m X_{mn} - \tau_i$. The probability expression for individual n choosing alternative i is simply the difference between $P(y_n > i)$ and $P(y_n > (i - 1))$:

$$P(y_{n1} = 1) = 1 - \frac{\exp(\tau_1 - \sum_m \beta_m X_{mn})}{1 + \exp(\tau_1 - \sum_m \beta_m X_{mn})} \tag{7}$$

$$P(y_{ni} = 1) = P(y_n > (i-1)) - P(y_n > i) \tag{8}$$

$$= \frac{\exp(\tau_{i-1} - \sum_m \beta_m X_{mn})}{1 + \exp(\tau_{i-1} - \sum_m \beta_m X_{mn})} - \frac{\exp(\tau_i - \sum_m \beta_m X_{mn})}{1 + \exp(\tau_i - \sum_m \beta_m X_{mn})}, \quad \text{for} \quad i = 2, ..., J-1 \tag{9}$$

$$P(y_{nJ} = 1) = \frac{\exp(\tau_{J-1} - \sum_m \beta_m X_{mn})}{1 + \exp(\tau_{J-1} - \sum_m \beta_m X_{mn})} \tag{10}$$

2.3 Generalized Ordered Logit model

A limitation of the ordered logit model is the proportional odds assumption – the ratios between each pair of choices are equal. The Generalized Ordered Logit model relaxes this assumption by modifying the latent variable into a vector, expressed as a function of alternative specific β parameters (Maddala, 1986, Eluru *et al.*, 2008) and the expression is formulated as:

$$P(y_n > i) = \frac{\exp(\tau_i - \sum_m \beta_{im} X_{mn})}{1 + \exp(\tau_i - \sum_m \beta_{im} X_{mn})}, \quad \text{for} \quad i = 1, 2, ..., J-1 \tag{11}$$

Wang and Abdel-Aty (2008) and Quddus *et al.* (2010) further combined alternative specific and generic β parameters in the latent variable function: $y_n^* = \sum_m \beta_m X_{mn} + \sum_m \beta_{im} X_{mn} + \varepsilon_n$. To ensure that the cut-offs retain the increasing order condition ($\tau_1 < \tau_2 < ... < \tau_{J-1}$), Eluru *et al.* (2008) described the following linear function for the threshold:

$$\tau_i = \tau_{i-1} + \exp(\sum_m \delta_{im} Z_{imn}) \tag{12}$$

where δ_{im} are the estimated parameters and Z_{imn} are the exogenous variables associated with the i-th cut-off. Studies have used the Generalized Ordered Logit model to model the underlying risk factors of bus accident severity in the United States (Kaplan and Prato, 2012), effects of driver concentration while driving on crash injury severity in Ethiopia (Abegaz *et al.*, 2014), and sustained injuries in an e-bicycle crash model in China (Wang *et al.*, 2018).

Flexible forms of the Generalized Extreme Value (GEV) class of models (Ben-Akiva *et al.*, 2002) have also been suggested as a way to handle ordered responses due to its consistency with random utility theory. One of which is the Ordered GEV (OGEV) model, which accounts for the correlation between choices of close proximity (Small, 1987). The OGEV model is an extension of the GEV model where each ordered choice alternative is a member of a nest with one or more adjacent alternatives based on their ordering structure. This specification allows for outcomes to be ordered, while still providing the flexibility of the MNL model. By changing the number of alternatives and the nest parameters in each nest, the OGEV can capture different levels of cross-elasticity. The OGEV is described with the following choice probability (Wen and Koppelman, 2001):

$$P(y_{ni} = 1) = \sum_{r=i}^{i+R} \left[\frac{w_{r-i} \exp(V_{ni})^{1/\mu_r}}{\sum_{j\in B_r} w_{r-j} \exp(V_{nj})^{1/\mu_r}} \times \frac{(\sum_{j\in B_r} w_{r-j} \exp(V_{nj})^{1/\mu_r})^{\mu_r}}{\sum_{s=1}^{J+L}(\sum_{j\in B_s} w_{s-j} \exp(V_{nj})^{1/\mu_s})^{\mu_s}} \right] \tag{13}$$

where $j\in B_r$ is the set of overlapping alternatives in nest r. L is a positive integer that defines the maximum number of contiguous alternatives in the nest. $w\geq 0$ is the allocation weight of the alternative to the nest and the sum of all weights is equal to one. $\exp(V_{ni}/\mu) = 0$ for $i < 1$ and $i > J$, and $0 < \mu_r \leq 1$. This model reduces to an MNL model when $\mu_r = 1 \ \forall \ r$ (Small, 1987).

The main structural difference between the OGEV and the Ordered Logit is that the former does not require the estimation of a set of cut-off points. The lack of the need to specify a latent variable makes the OGEV model more flexible in setting different correlation patterns.

Although the Generalized Ordered Logit and OGEV models are suitable for handling potential ordering of the choice preference, neither fully account for heterogeneity of outcomes. For instance, in ride-hailing services, the platform is designed to encourage drivers and passengers to stay on the platform, hence using a rating system to incentivize the experience (Fielbaum and Tirachini, 2020). Therefore, there might be unobserved heterogeneity between a 5 star-rating choice compared to a 4 or 3 star-rating which might not be captured in the ordered logit model specification.

2.4 Other extensions to Ordered Logit

The Dogit OGEV (DOGEV) extension model developed by Fry and Harris (2005) combines both item preferences and ordering of outcomes as a two-part choice generating process. The Dogit model allows some subsets of relative probabilities to be determined consistently within the irrelevant alternatives (IIA) axiom, while other subsets are not (Gaundry and Dagenais, 1979). It simultaneously accounts for correlation among pairs of alternatives, yet flexible enough to allow for IIA. The Dogit model is defined by the following (see Gaundry and Dagenais (1979)):

$$P(y_{ni} = 1) = \frac{\exp(V_{ni}) + \theta_i \sum_j \exp(V_{nj})}{(1 + \sum_j \theta_j) \sum_j \exp(V_{nj})} \tag{14}$$

The property of the Dogit model is that the probability ratio of any two alternatives can be dependent on several or all alternatives. Gaundry and Dagenais (1979) showed that the Dogit specification captures additional parameters to the MNL model, interpreted as a "captivity" coefficient. Therefore it allows for respondents to be drawn towards a specific ranked choice if the choice has a certain level of attractiveness.

Despite the strong theoretical foundation of the DOGEV model and being able to capture the latent effects in the choice process consistently within the general random utility theory framework, few applications of this model in transportation science literature can be found. Habib and Weiss (2014) suggest a lack in commercial estimation software has limited Dogit type models in use for modelling behavior.

3 The Unimodal Logit

3.1 Unimodality in ordered choices

In the Ordered Logit model, the posterior probabilities are not guaranteed to be unimodal, although the thresholds are assumed to be in ascending order. To illustrate how a non-unimodal probability mass function (pmf) generates undesirable properties in an ordinal choice problem, we show two different choice scenarios (scenario 1, s_1 and scenario 2, s_2) in Figure 1. Both scenarios are assumed to be modelling the same ordinal data and the selected choice for an observation in both scenarios is 2 ("disagree").

We assume that maximum likelihood estimation is used in both scenarios. Although the estimates in both s_1 and s_2 might have identical probability mass for choice 2, and therefore, similar log likelihood value, $\ln(P(y_{s_1} = 2)) = \ln(P(y_{s_2} = 2))$, the posterior probability mass of s_1 does not make sense. This is because maximum likelihood estimation only takes into account the logsum of the selected choice. The unimodal probability distribution of s_2 is preferred as the probability mass decreases to the left and right of the desired choice.

3.2 Model specification

When a natural ordering of alternatives appears in the choice set, we can impose a unimodal constraint on the pmf. Specifically, the *a posteriori* choice probabilities are unimodal. This mild

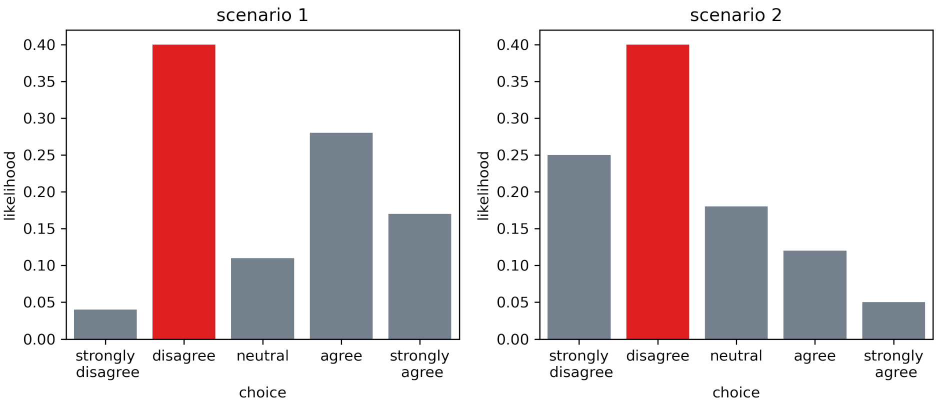

Figure 1: Maximum likelihood estimation ignores the relationships and ordering nature between classes. In this example, both scenario estimates an identical likelihood function, but the unimodal distribution on the right is more preferable for ordinal data than the distribution on the left.

assumption, in conjunction with the alternative specific utility function, enables the ordering of the alternatives to be captured in the unobserved error terms. Using the Presidential question example again with five options: [“very poor job”, “poor job”, “neutral”, “good job”, “very good job”] labeled 1 through 5 respectively, assume of the highest *a posteriori* probability is $P(y_{n4} = 1|X)$, then the next highest *a posteriori* probability should be adjacent to y_{n4} and the *a posteriori* probabilities should be monotonically decreasing to the left and right of $P(y_{n4} = 1|X)$. It would not be logical if the second highest likely option is $P(y_{n1} = 1|X)$ after $P(y_{n4} = 1|X)$, as it would not make sense if a decision maker’s opinion jumps from “good job” to “very poor job” to “very good job”.

We say that the Unimodal Logit model captures natural ordering of the choices if there exist an integer $c \in J$ such that:

$$p(y_{ni}|X) \geq p(y_{ni+1}|X), \text{ for all } \; i \geq c \quad \text{and,}$$
$$p(y_{ni-1}|X) \leq p(y_{ni}|X), \text{ for all } \; i \leq c$$

In the Poisson pmf, the probability of i occurrences of an event in a set of N observations is defined as:

$$P(i) = \frac{\lambda^i \exp(-\lambda)}{i!}, \text{ for } \; i = 0, 1, 2, .. \tag{15}$$

where λ is the mean of the pmf. If we take $P(i)$ to be the a priori of an ordinal response, and the number of occurrences i to be equivalent to the ordinal choice selection, then a unimodal distribution can be enforced by including $\ln(P(i))$ as the adjustment term in the systematic part of the utility of each alternative i:

$$U_{in} = V_{in} + \ln(P(i)) + \varepsilon_{in} \tag{16}$$

$$= V_{in} + \ln\left(\frac{\lambda^i \exp(-\lambda)}{i!}\right) + \varepsilon_{in} \tag{17}$$

$$= V_{in} + i\ln(\lambda) - \lambda - \ln(i!) + \varepsilon_{in} \tag{18}$$

In the form of an MNL choice probability, we have:

$$P(y_{ni} = 1) = \frac{\exp(\mu\Phi_{in})}{\sum_{j=1}^{J} \exp(\mu\Phi_{jn})}, \tag{19}$$

and

$$\Phi_{in} = V_{in} + i\ln(\lambda) - \lambda - \ln(i!) + \beta_{i0} \tag{20}$$

where $\mu \geq 0$ is the scale parameter and β_{i0} is an alternative specific constant to capture the mean of the error term. For consistency, we include V_{in} in the utility function. Since the alternatives are associated with the same subject in question, this term is not very useful from an ordinal response choice standpoint (Small, 1987). The utility function in Equation 20 can be rearranged into:

$$U_{in} = V_{in} - \lambda + f(\lambda, i) + \varepsilon_{in} \tag{21}$$

where the proximity correction is comprised of a scalar term λ and a function $f(\lambda, i)$ that depends on λ and i. λ and $f(\lambda, i)$ control the mode (point of highest probability mass) of the unimodal distribution and the adjacent utilities are monotonically decreasing from the mode. The variance of the distribution is λ. To ensure that λ is always greater than 0, we set λ to be a function of y_n^* (Equation 5) as follows:

$$\lambda = f(y_n^*) = \ln(1 + \exp(y_n^*)) \tag{22}$$

This formulation for modelling ordered choice has several implications. First, the (log) Poisson adjustment term to the systematic utility conforms to the general formulation $U_i = \beta x + f(A_{in}) + \varepsilon_i$ described by Axhausen and Schüssler (2007). Axhausen and Schüssler suggest four behavioural processes can be expected: loosing visibility, joint risks, super-alternative, or super-visibility.

Since our adjustment term reduces the utilities of the non-chosen alternatives proportionally further away from the chosen alternative, therefore giving the non-chosen alternatives a lower probability of being selected. The Unimodal Logit model falls under the category of *loosing visibility*.

Second, since the utilities do not include attributes of competing alternatives, the relative probability of the decision maker choosing between two alternatives is independent of any new alternatives. Therefore it retains the proportional odds ratio between competing choices. For example, given a Likert scale from "strongly disagree" (1) to "strongly agree" (5), adding a 6th choice – e.g. "very strongly agree" (6) to the choice set does not violate the IIA property.

Third, the Poisson function takes into account the "zero" option, for instance when a choice set containing the number of items to buy which includes a "no purchase" alternative. In some cases, a zero-truncated Poisson (ZTP) distribution might be more appropriate. For example, a movie rating database has to have scores between 1 to 5 stars. A Zero-truncated Unimodal Logit has the following ZTP pmf.

$$P(i|i>0) = \frac{\lambda^i \exp(-\lambda)}{i!(1-\exp(-\lambda))}, \text{ for } i = 1, 2, 3, ... \tag{23}$$

and the utility of the Zero-truncated Unimodal Logit is defined by:

$$U_{in} = V_{in} + i\ln(\lambda) - \lambda - \ln(i!) - \ln(1-\exp(-\lambda)) + \varepsilon_{in} \tag{24}$$

Finally, similar to the OGEV specification, the Unimodal Logit does not require a set of thresholds, as it only deals with the correlation between the alternatives in the utility function. To incorporate heterogeneity between outcomes, we can include alternative specific parameters in V_{in}. This allows the observed utility to capture any taste heterogeneity, independent from the proximity correction error component.

The probabilities of the Unimodal Logit have a closed-form expression and we use a maximum likelihood function over the observations, maximizing the *a posteriori* probability, to estimate the parameters of the model:

$$LL(\hat{\boldsymbol{\beta}}) = \sum_{n=1}^{N}\sum_{i=1}^{J} z_{ni} \ln P(y_{ni} \mid \boldsymbol{\beta}, \mathbf{X}) \tag{25}$$

z_{ni} is the ground truth information where $z_{ni} = 1$ if the decision maker chooses alternative i and $z_{ni} = 0$ otherwise. $\boldsymbol{\beta}$ and $\mathbf{X}$ are the vectors of estimated parameters and exogenous variables respectively.

4 Case study

This case study aims to understand the suitability of modelling ranked ordered choices using the Unimodal Logit model. A crash severity analysis for identifying the contributing factors towards the severity of injury or manner of collision is used to test our model hypothesis. Crash severity is measured by the most severe injury sustained by all involved parties and the level of severity is usually defined by the amount of property damage, and the number of injured or fatalities (Mannering and Bhat, 2014).

We used a publicly available open dataset of the High Severity Traffic Crash Data Report[1] from the City of Tempe, Arizona (City of Tempe, 2018). The data contains 39,793 records, with information on the collision, time, location, condition of the road, weather, lighting, age, gender, type of violation and action taken by the driver. Data records were taken between 2012 and 2019.

Table 1 describes the variables X_n used in the estimation. Five accident severity levels $j = \{1, 2, 3, 4, 5\}$ are defined: 1: no injury, 2: possible injury, 3: non-incapacitating (minor) injury, 4: incapacitating (major) injury, and 5: fatal injury. The severity of the accident is given by the highest severity of injury of all persons involved. Among the observations, 27,473 were classified as "no injury", 7,251 were "possible injury", 4,411 were "minor injury", 559 were "major injury" and 99 were "fatal injury". The attributes are classified into i) environmental attributes and ii) collision related attributes. Environmental attributes refer to variables such as road condition, weather, lighting, and time of day. Collision related attributes refer to the variables associated with the accident instance: cause of accident and behaviour of driver, vehicle type, alcohol level, victim's age, total number of injured parties and location of incident (relative to a road intersection).

In this study, we estimated three models: 1) Ordered Logit model, 2) Unimodal Logit model, and 4) Zero-truncated Unimodal Logit model. In the Ordered Logit model, the latent propensity function is specified as $y_n^* = \sum_m \beta_m X_{mn} + \varepsilon_n$ of the crash n, where β_m is the parameter to be estimated, ε_n is a stochastic error term and X_n is the crash attribute. the latent propensity function

[1] `https://data.tempe.gov/datasets/tempegov::1-08-crash-data-report-detail/about`

Name	Variable	Summary statistics	mean
Environmental			
Time	hour_afternoon	Time of accident between 12pm to 6pm	0.471
	hour_night	Time of accident between 6pm to 12am	0.215
Lighting	light_darklighted	Lighting condition: dark, lighted	0.222
	light_darknotlighted	Lighting condition: dark, not lighted	0.007
Road cond.	surf_wet	Wet surface at site of accident	0.032
Weather	meteo_cloudy	Weather: Cloudy	0.054
	meteo_rain	Weather: Rain	0.023
Collision characteristics			
Age	age	Defines the age of the driver	30.2
Total injuries	total_injuries	Total number of injured involved	0.441
Alcohol use	alcohol	Indicates whether alcohol was a factor	0.044
Location	nonintersection	Collision not near an intersection	0.429
Type	type_cyclist	Car-cyclist collision	0.022
	type_pedestrian	Car-pedestrian collision	0.007
	type_driverless	Car-driverless vehicle collision	0.001
Action	action_turn	Driving behaviour: Making a turn	0.306
	action_lanes	Driving behaviour: Changing lanes	0.077
	action_straight	Driving behaviour: Going straight ahead	0.493
	action_slowing	Driving behaviour: Slowing in traffic	0.029
Cause	cause_speeding	Violation: Speed to fast for conditions	0.312
	cause_yield	Violation: Failed To yield right of way	0.199
	cause_unsafe	Violation: Unsafe lane change	0.078
	cause_signal	Violation: Disregarded traffic signal	0.066
	cause_following	Violation: Followed too closely	0.066
	cause_distraction	Violation: Inattention distraction	0.039
	cause_turn	Violation: Made improper turn	0.042
Collision	acc_rearend	Collision type: Rear end	0.377
	acc_left	Collision type: Left turn	0.19
	acc_side	Collision type: Sideswipe	0.131

Table 1: Summary statistics of the Tempe High Severity Traffic Crash Data

is mapped to a logistic distribution defined by a set of fixed thresholds defined by τ_1, τ_2, τ_3, and τ_4 where $\tau_1 = 0$, $\tau_2 = \tau_1 + \delta_2$, $\tau_3 = \tau_2 + \delta_3$, and $\tau_4 = \tau_3 + \delta_4$. We estimate the differences δ_2, δ_3, and δ_4 in the Ordered Logit model. In the Unimodal and Zero-truncated Unimodal Logit model, the linear propensity function is limited to $\lambda_n > 0$ by using the transformation: $\lambda_n = \ln(1 + \exp(y_n^*))$, since the Poisson distribution requires a positive λ component.

We set the first alternative (no injury) as the reference choice. The estimated parameters are interpreted as the impact of the attribute on the severity of the outcome. However, caution must be taken when interpreting the signs of the β estimates as it does not always determine the direction of the impact (Wooldridge, 2002).

5 Results

5.1 Model evaluation

Two goodness-of-fit measures are chosen for model evaluation: pseudo R-squared measure (ρ^2) (McFadden, 1974) and Bayesian Information Criterion (BIC). The former tests how well the model fits the data and the latter tests the efficiency of the model by considering the number of parameters used (complexity). The pseudo R-squared measure is given by:

$$\rho^2 = 1 - \frac{\ln LL(\hat{\beta})}{\ln LL(\bar{\beta})} \tag{26}$$

where $LL(\hat{\beta})$ denotes the maximized likelihood of the estimated model, and $\bar{\beta}$ denotes the null model (β parameters associated with the exogenous variables are zero). ρ^2 is bounded by 0 and 1 and a ρ^2 measure closer to 1 indicates a better model fit. The BIC of a model is equal to:

$$BIC = -2LL(\beta) + M \ln(Q) \tag{27}$$

where M is the number of estimated parameters and Q is the number of observations. The model with the lower BIC is generally the preferred model.

5.2 Out-of-sample evaluation

In order to evaluate model performance on out-of-sample forecasting, we use three quantitative metric: Discrete Classification Accuracy, Geometric Mean Probability of Correct Assignment (GMPCA) and Quadratic Weighted Kappa (QWK).

The Discrete Classification Accuracy shows the number of correct assignments if each prediction is assigned to the highest probability class. It is the most commonly used metric in the literature, however, this metric have several problems. The most important one is that if the data is imbalance, i.e. there are more observations of some classes than others (such as in this case study), by assigning each prediction to the class with highest probability, the less frequent classes will be under-represented in the predicted outcomes, whereas the more frequent ones will be over-represented. Therefore, a model that always predicts the majority classes might achieve a high accuracy score.

The Geometric Mean Probability of Correct Assignment (GMPCA) (Hillel, 2019) provides a robust measure of the model's average accuracy, taking into account relative variances in decision probabilities. The advantage of this metric is that it has a clear physical interpretation, the geometric average correctness of the output model. A GMPCA of one indicates a perfect (totally correct) classifier, while a GMPCA of zero indicates a completely erroneous classifier. It can be calculated as:

$$G_{\text{GMPCA}} = \left(\prod_{n=1}^{N} P\left(i_n \mid x_n\right) \right)^{\frac{1}{N}} \tag{28}$$

The Quadratic Weighted Kappa (QWK) ($-1 \leq \kappa \geq 1$) evaluates whether a model does better than random chance ($\kappa > 0$) or worse than random chance ($\kappa < 0$), where a score of 1 denotes the best score (Cohen, 1968). The weighting allows the evaluation against random chance to be scaled by how far the it is from the predicted choice. The result is a chance-corrected proportion of weighted agreement which is appropriate for ordinal choice problems. The Quadratic Weighted Kappa can be defined as:

$$\kappa = 1 - \frac{\sum_{i=1}^{J} \sum_{j=1}^{J} w_{ij} \hat{y}_{ij}}{\sum_{i=1}^{J} \sum_{j=1}^{J} w_{ij} y_{ij}} \tag{29}$$

where i and j represents the observed and predicted choice indices respectively, $w_{ij} \in \mathbb{R}_{\geq 0}$ and $w_{ii} = 0$. $\hat{y}_{ij}$ is the proportion of observed agreement from the forecast and y_{ij} is the proportion of agreement by chance.

5.3 Model results

The models were estimated using Biogeme software (Bierlaire, 2020). Table 2 presents the estimates results of each of the three models, the robust standard error and the t-test score for each parameter respectively. As aforementioned, the Unimodal Logit model specification examines the effects of the interaction between the attributes and the crash severity. Given that the crash severity follows a unimodal distribution, we expect that the Unimodal specification would capture a better model fit and more accurate estimates of the attribute coefficients. All three models present significant estimates for most of the coefficients, the alternative specific constants (Unimodal Ordered) and the threshold values.

The zero-truncated version of the Unimodal Logit model performs poorly compared to the Unimodal Logit, but still slightly better than the Ordered Logit model. This might indicate that that the crash severity may not follow a ZTP distribution.

In particular, the Ordered Logit and the Unimodal Logit model exhibit similar signs for all estimates except for `cause_distraction`, which is also shown to be significant in the latter but not in the former. The `cause_speeding` parameter is shown to be *not significant* (-1.96 < rob_tTest < 0) in the Ordered Logit, whereas it is a *significant* factor in the Unimodal Logit (rob_tTest<-1.96). A study by Christoforou *et al.* (2011) predicted the type of accidents from real-time data using prevailing traffic conditions. Their findings showed that driving at *inappropriate speeds* does not always allow for evasive actions to be taken in order to avoid an accident, thus influencing the severity of the crash. Similarly, we observed that the Unimodal Logit captures this phenomena better than the Ordered Logit model. Another interesting finding is that `type_driverless` parameter is significant in the Unimodal Logit, but not in the Ordered Logit model. This might indicate a positive relationship between driverless vehicles and reduction in collision severity.

We used a 20% out of sample data to evaluate model performance. Table 3 shows the comparison between the three models on the evaluation metrics. Figure 2 illustrates the choice probabilities generated by the Ordered and Unimodal Logit models. With discrete classification accuracy, the models are not significantly different, in addition, the Ordered Logit (0.839) performs better than the zero-truncated Unimodal Logit model (0.826). However, this accuracy is not reliable as it only considered the "highest proability" output. As mentioned in Train (2003), such metric does not give a reliable measure of predictive power. Comparing the models on GMPCA and QWK, we see that the Unimodal Logit (0.653, 0.805) provides greater improvement from the Ordered Logit (0.581, 0.758) and the Zero-truncated Unimodal Logit (0.59, 0.787). The Zero-truncated Unimodal Logit performs similarly to the Ordered Logit. From our case study, our results show that the Unimodal approach is more reliable method of modelling ordinal data.

	Ordered Logit			Unimodal Logit			Zero-trunc Unimodal Logit		
Variables	values	rob_stdErr	rob_tTest	values	rob_stdErr	rob_tTest	values	rob_stdErr	rob_tTest
ASC_noinjury (1)				ref.			ref.		
ASC_possinjury (2)				3.673	0.039	93.809	2.452	0.036	68.248
ASC_nonincap (3)				4.117	0.039	104.798	3.446	0.035	97.33
ASC_incap (4)				2.449	0.056	43.533	1.907	0.054	35.544
ASC_fatal (5)				0.788	0.109	7.193	0.319	0.111	2.889
acc_leftturn	-0.276	0.048	-5.784	-0.318	0.038	-8.269	-0.383	0.052	-7.369
acc_rearend	-0.785	0.049	-15.929	-0.851	0.039	-21.986	-1.13	0.056	-20.1
acc_sides	-1.294	0.063	-20.629	-1.209	0.041	-29.721	-1.745	0.067	-25.958
action_lanes	-0.618	0.102	-6.082	-0.926	0.073	-12.645	-1.39	0.108	-12.868
action_slowing	-0.915	0.102	-9.004	-1.18	0.08	-14.726	-1.829	0.113	-16.181
action_straight	-0.83	0.048	-17.341	-1.265	0.034	-37.643	-1.897	0.054	-35.375
action_turn	-1.041	0.055	-18.912	-1.48	0.037	-39.706	-2.176	0.059	-36.981
age	-0.008	0.001	-10.184	-0.015	0.001	-22.834	-0.017	0.001	-18.351
alcohol	0.384	0.076	5.02	0.379	0.082	4.625	0.524	0.107	4.918
cause_distraction	0.08	0.079	1.013	-0.287	0.06	-4.749	-0.249	0.088	-2.83
cause_following	-0.185	0.07	-2.647	-0.39	0.054	-7.237	-0.455	0.079	-5.79
cause_signal	-0.412	0.07	-5.869	-0.799	0.062	-12.943	-0.948	0.08	-11.885
cause_speeding	-0.027	0.049	-0.543	-0.271	0.04	-6.78	-0.28	0.058	-4.824
cause_turn	-0.153	0.084	-1.832	-0.355	0.059	-6.05	-0.411	0.089	-4.611
cause_unsafe	-0.492	0.102	-4.811	-0.473	0.068	-6.985	-0.686	0.103	-6.676
cause_yield	-0.108	0.053	-2.038	-0.341	0.042	-8.166	-0.4	0.059	-6.784

Continued on next page

Variables	Ordered Logit values	rob_stdErr	rob_tTest	Unimodal Logit values	rob_stdErr	rob_tTest	Zero-trunc Unimodal Logit values	rob_stdErr	rob_tTest
hour_afternoon	-0.363	0.031	-11.677	-0.703	0.027	-26.343	-0.753	0.035	-21.437
hour_night	-0.333	0.047	-7.082	-0.428	0.041	-10.338	-0.485	0.056	-8.678
light_darklighted	-0.347	0.048	-7.31	-0.638	0.04	-15.883	-0.709	0.056	-12.713
light_darknotlighted	-0.663	0.194	-3.427	-0.906	0.155	-5.856	-1.13	0.243	-4.654
meteo_cloudy	-0.172	0.069	-2.485	-0.177	0.068	-2.616	-0.211	0.091	-2.331
meteo_rain	-0.054	0.152	-0.356	-0.167	0.133	-1.253	-0.124	0.19	-0.655
nonintersection	-0.316	0.03	-10.653	-0.67	0.026	-25.694	-0.691	0.034	-20.073
surf_wet	-0.183	0.134	-1.364	-0.199	0.119	-1.671	-0.305	0.169	-1.808
total_injuries	2.659	0.042	62.862	3.551	0.051	69.493	3.789	0.065	58.207
type_cyclist	1.46	0.082	17.722	0.619	0.117	5.289	0.804	0.128	6.265
type_driverless	-0.52	0.355	-1.465	-1.478	0.194	-7.631	-1.744	0.331	-5.267
type_pedestrian	1.596	0.208	7.657	3.838	0.473	8.122	3.066	0.506	6.065
tau1	0.0	0.053	0.0						
delta2	2.611	0.038	68.111						
delta3	3.31	0.084	39.596						
delta4	2.303	0.154	14.98						
Log likelihood	-17148.44			-13471.31			-16731.04		
BIC	34628.6			27274.4			33793.8		
ρ^2	0.665			0.737			0.673		
Optimization time	0:01:02.27			0:06:26.2			0:07:40.4		

Table 2: Crash severity analysis: model estimates

Model	Discrete Classification Acc.	GMPCA	QWK
Ordered Logit	0.839	0.581	0.758
Unimodal Logit	0.842	0.653	0.805
Zero-truncated Unimodal Logit	0.826	0.59	0.787

Table 3: 20% out-of-sample predictive evaluation

6 Conclusion

The current study proposes a new form of choice model for ordered choices. This new choice model, called the Unimodal Logit model, uses a Poisson distribution function to account for choice proximity in the error correction term in the utility. This idea is based on the fact that ordered choices exhibits unimodality, specifically, it is the *a posteriori* distribution which is unimodal. Imposing this constraint in our model allows for a Logit model to correct for this underlying correlation between neighbouring choices, which retaining all the traits of a Ordered choice model, including the interpretation of the β parameters.

The results show several key differences on a crash severity analysis and the Unimodal Logit model was shown to capture the influence of driving speed, distracted driving and driverless vehicles in the model. The Unimodal Logit model also exhibits better model fit when comparing BIC and ρ^2 measures. This study provides new evidence on using a unimodal modelling approach on ordered choices. In terms of predictive performance the Unimodal Logit model shows improvement over the Ordered Logit when accounting for choice proximity. So far, we have looked at Poisson distribution, but we would be investigating other forms of unimodal distribution that have a closed form, such as a negative binomial distribution in future studies. Furthermore, the Unimodal approach could also be combined with other error correction logit models such as Path Size Logit (Sobhani *et al.*, 2019), or Residual Logit (Wong and Farooq, 2021), which could be an interesting avenue for research.

7 Supplementary code and data

Biogeme code and data used in analysis and modelling are available on: `https://github.com/mwong009/unimodal-logit`

8 References

Abegaz, T., Y. Berhane, A. Worku, A. Assrat and A. Assefa (2014) Road traffic deaths and injuries are under-reported in ethiopia: a capture-recapture method, *PloS one*, **9** (7) e103001.

Axhausen, K. W. and N. Schüssler (2007) Similarity, visibility and similar concepts: A note, *Arbeitsberichte Verkehrs-und Raumplanung*, **458**.

Beckham, C. and C. Pal (2017) Unimodal probability distributions for deep ordinal classification, paper presented at the *International Conference on Machine Learning*, 411–419.

Ben-Akiva, M., D. McFadden, K. Train, J. Walker, C. Bhat, M. Bierlaire, D. Bolduc, A. Boersch-Supan, D. Brownstone, D. S. Bunch *et al.* (2002) Hybrid choice models: Progress and challenges, *Marketing Letters*, **13** (3) 163–175.

Bierlaire, M. (2020) A short introduction to pandasbiogeme, *Technical Report*, **TRANSP-OR 200605**, Transport and Mobility Laboratory, ENAC, EPFL.

Christoforou, Z., S. Cohen and M. G. Karlaftis (2011) Identifying crash type propensity using real-time traffic data on freeways, *Journal of Safety research*, **42** (1) 43–50.

City of Tempe (2018) Crash data report (detail), https://open.tempe.gov/datasets/1-08-crash-data-report-detail.

Cohen, J. (1968) Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit., *Psychological bulletin*, **70** (4) 213.

da Costa, J. F. P., H. Alonso and J. S. Cardoso (2008) The unimodal model for the classification of ordinal data, *Neural Networks*, **21** (1) 78–91.

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

Fielbaum, A. and A. Tirachini (2020) The sharing economy and the job market: the case of ride-hailing drivers in chile, *Transportation*, 1–27.

Fry, T. R. and M. N. Harris (2005) The dogit ordered generalized extreme value model, *Australian & New Zealand Journal of Statistics*, **47** (4) 531–542.

Fu, X.-m. (2020) Does heavy ict usage contribute to the adoption of ride-hailing app?, *Travel Behaviour and Society*, **21**, 101–108.

Gaundry, M. J. and M. G. Dagenais (1979) The dogit model, *Transportation Research Part B: Methodological*, **13** (2) 105–111.

Greene, W. H. and D. A. Hensher (2010) Ordered choices and heterogeneity in attribute processing, *Journal of Transport Economics and Policy (JTEP)*, **44** (3) 331–364.

Habib, K. N. and A. Weiss (2014) Evolution of latent modal captivity and mode choice patterns for commuting trips: A longitudinal analysis using repeated cross-sectional datasets, *Transportation Research Part A: Policy and Practice*, **66**, 39–51.

Hall, P., L.-S. Huang, J. A. Gifford and I. Gijbels (2001) Nonparametric estimation of hazard rate under the constraint of monotonicity, *Journal of Computational and Graphical Statistics*, **10** (3) 592–614.

Hillel, T. (2019) Understanding travel mode choice: A new approach for city scale simulation, Thesis, University of Cambridge.

Kaplan, S. and C. G. Prato (2012) Risk factors associated with bus accident severity in the united states: A generalized ordered logit model, *Journal of safety research*, **43** (3) 171–180.

Krueger, R., T. H. Rashidi and J. Auld (2019) Preferences for travel-based multitasking: Evidence from a survey among public transit users in the chicago metropolitan area, *Transportation Research Part F: Traffic Psychology and Behaviour*, **65**, 334–343.

Loa, P. and K. N. Habib (2021) Examining the influence of attitudinal factors on the use of ride-hailing services in toronto, *Transportation Research Part A: Policy and Practice*, **146**, 13–28.

Maddala, G. S. (1986) *Limited-dependent and qualitative variables in econometrics*, Cambridge university press.

Mannering, F. L. and C. R. Bhat (2014) Analytic methods in accident research: Methodological frontier and future directions, *Analytic methods in accident research*, **1**, 1–22.

McCullagh, P. (1980) Regression models for ordinal data, *Journal of the Royal Statistical Society: Series B (Methodological)*, **42** (2) 109–127.

McFadden, D. (1974) Conditional logit analysis of qualitative choice-behaviour, in Zarembka (ed.) *Frontiers in Econometrics*, 105–142, Academic Press.

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

Sheffi, Y. (1979) Estimating choice probabilities among nested alternatives, *Transportation Research Part B: Methodological*, **13** (3) 189–205.

Small, K. A. (1987) A discrete choice model for ordered alternatives, *Econometrica: Journal of the Econometric Society*, 409–424.

Sobhani, A., H. A. Aliabadi and B. Farooq (2019) Metropolis-hasting based expanded path size logit model for cyclists' route choice using gps data, *International journal of transportation science and technology*, **8** (2) 161–175.

Tirachini, A. and M. del Río (2019) Ride-hailing in santiago de chile: Users' characterisation and effects on travel behaviour, *Transport Policy*, **82**, 46–57.

Train, K. E. (2003) *Discrete Choice Methods with Simulation*, Cambridge university press.

Wang, T., J. Chen, C. Wang and X. Ye (2018) Understand e-bicyclist safety in china: Crash severity modeling using a generalized ordered logit model, *Advances in mechanical engineering*, **10** (6) 1687814018781625.

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

Wen, C.-H. and F. S. Koppelman (2001) The generalized nested logit model, *Transportation Research Part B: Methodological*, **35** (7) 627–641.

Wong, M. and B. Farooq (2021) Reslogit: A residual neural network logit model for data-driven choice modelling, *Transportation Research Part C: Emerging Technologies*, **126**, 103050.

Wooldridge, J. M. (2002) Econometric analysis of cross section and panel data mit press, *Cambridge, MA*, **108**.

Zeng, Q., W. Gu, X. Zhang, H. Wen, J. Lee and W. Hao (2019) Analyzing freeway crash severity using a bayesian spatial generalized ordered logit model with conditional autoregressive priors, *Accident Analysis & Prevention*, **127**, 87–95.

Appendix A: Out-of-sample results

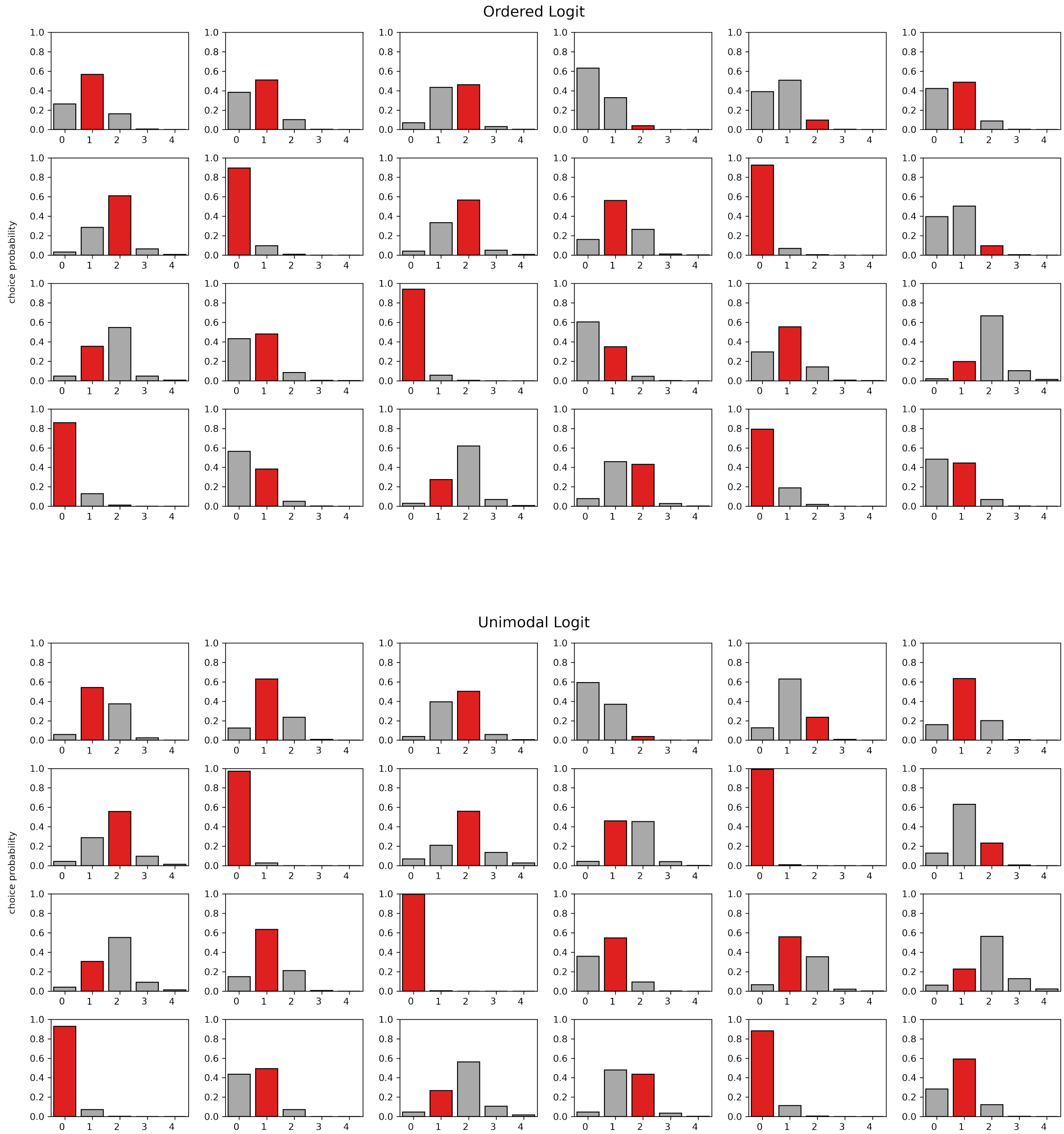

Figure 2: Choice probability output generation on 20% out-of-sample data. 24 random samples shown. Top: Ordered Logit, bottom: Unimodal Logit. Red bar indicates actual choice. Highest bar indicates selected choice.