Crash-Severity Modeling

Part 1

By Xiao Qin

Introduction

- The crash severity model is to use statistical methods for identifying factors that are significantly associated with the consequence of a traffic crash, and their relationships.
- The response variable is the person who sustains the most severe injury in a crash in the KABCO scale (i.e., killed, incapacitating injury, non-incapacitating injury, possible injury, and no injury).
- A variety of models have been developed to account for data issues and methodological limitations.
- The common modeling approaches include logistic, probit and their variations.
- The impact of a factor on the injury severity levels can be estimated through its marginal effect or odds ratio.

Objectives

- Learn the characteristics of crash injury severity data.
- Gain the knowledge about data limitations and modeling challenges.
- Understand the assumptions, property and limitations of models for crash injury severity levels.
- Develop crash injury severity models and perform analysis.
- Interpret the modeling results.

Recall: Definition (most common)

TABLE 2.3 Comprehensive crash costs (per person) (2018 dollars).

Severity of injuries	Costs
Fatal (K)	$10,855,000
Incapacitating (type A)	$1,187,000
Nonincapacitating (type B)	$327,000
Possible (type C)	$151,000
No injury (property damage only or PDO)	$50,000

Source: NSC.[1]

Values shown above can be used to evaluate highway safety interventions in terms of lives/injuries saved.

Characteristics of Injury Severity Data

- Researchers and safety professionals rely heavily on crash data as they are the most relevant and informative resource for analyzing traffic injuries.
- However, the causes of an injury are very complicated because they involve a sequence of events and several factors (i.e., driver, vehicle, environment), as discussed in Chapter 2. (see next slide)
- Crash injury severity modeling helps describe, identify, and evaluate the factors contributing to various levels of injury severity.

Characteristics of Injury Severity Data

In a crash report, there are different methods/codes to measure or define injuries (still subjective as it is governed by the opinion of the police officer). In some reports, they classify the injury as the first injury outcome in the sequence of events ("first harmful event"), whereas, in other reports, they defined it as the "most harmful event." For example,

Vehicle swerved to avoid an animal (driver not injured)

↓

Vehicle runs off the traveled-way (driver not injured)

↓

Vehicle hits a break-away pole (speed limit sign) (driver slightly injured by glass, say Type B)

↓

Vehicle goes down the embankment/sideslope and hit the bottom ditch hard (driver is severely injured, say Type A, from the external forces)

↓

Vehicle goes up the backslope and hits a tree (at low speed), the final resting place (driver does not sustain additional injuries)

Characteristics of Injury Severity Data

In a crash report, there are different methods/codes to measure or define injuries (still subjective as it is governed by the opinion of the police officer). In some reports, they classify the injury as the first injury outcome in the sequence of events ("**first harmful event**"), whereas, in other reports, they defined it as the "most harmful event." (**will be classified as 'Type B'**)

Vehicle swerved to avoid an animal (driver not injured)

↓

Vehicle runs off the traveled-way (driver not injured)

↓

Vehicle hits a break-away pole (speed limit sign) (driver slightly injured by glass, say Type B)

↓

Vehicle goes down the embankment/sideslope and hit the bottom ditch hard (driver is severely injured, say Type A, from the external forces)

↓

Vehicle goes up the backslope and hits a tree (at low speed), the final resting place (driver does not sustain additional injuries)

Characteristics of Injury Severity Data

In a crash report, there are different methods/codes to measure or define injuries (still subjective as it is governed by the opinion of the police officer). In some reports, they classify the injury as the first injury outcome in the sequence of events ("first harmful event"), whereas, in other reports, they defined it as the "**most harmful event**." (**will be classified as 'Type A'**)

Vehicle swerved to avoid an animal (driver not injured)

↓

Vehicle runs off the traveled-way (driver not injured)

↓

Vehicle hits a break-away pole (speed limit sign) (driver slightly injured by glass, say Type B)

↓

Vehicle goes down the embankment/sideslope and hit the bottom ditch hard (driver is severely injured, say Type A, from the external forces)

↓

Vehicle goes up the backslope and hits a tree (at low speed), the final resting place (driver does not sustain additional injuries)

Characteristics of Injury Severity Data

- Unlike crash count, which is a nonnegative integer, injury severity has a **finite number of outcomes** (e.g., killed, injury type A, injury type B, injury type C, no injury) that are categorized on the KABCO scale.
- Crash injury severity data usually are **imbalanced** on the KABCO scale, where the number of fatal or severe injuries is substantially fewer than the number of less severe and no injury crashes.
- Crash-severity models can be classified as **nonordinal** (e.g., multinomial logit (MNL) and multinomial probit) or **ordered** probabilistic (e.g., ordered probit and order logistic) if an ordinal structure for the response variable is assumed.
- Crash severity models can also be categorized as **fixed** or **random** parameter models according to the parameter assumptions.
- Model variations are available if **restrictions** such as irrelevant and independent alternatives (IIA), proportional odds, or homogeneity are relaxed.

Imbalanced Data Between Injury Severity Levels

- This imbalance of data in each injury category presents a challenge for classification algorithms. In predictive modeling, imbalanced data introduce a bias toward the majority that causes less accurate predictions of severe crashes (minority).
- A common method of treating imbalanced data is to combine similar injury types (i.e., K, A, B, and C) into one category on a new scale (i.e., injury and noninjury) to gain more balanced data.
- Other methods for handling imbalanced data include resampling techniques that aim to create a balanced injury scale data.
- Resampling involves oversampling less-representative classes (the Synthetic Minority Over-sampling Technique (SMOTE)), undersampling overly-representative classes, or using ensemble methods (e.g., Bootstrap aggregating).

Ordinal Nature of Crash Injury Severity Data

- An ordinal scale quantitatively categorizes crashes from the highest to lowest levels of injury severity (i.e., KABCO).
- Recognizing this ordinal structure within data is important because it aids in the selection of an appropriate methodology.
- Utilizing the intrinsic ordinal information preserved in the data may lead to the estimation of fewer parameters.
- Additionally, the potential dependency between adjacent categories may share unobserved effects. If such a correlation exists but is not accounted for, it can lead to biased parameter estimates and incorrect inferences.
- Nevertheless, the ordinality assumption should be exercised with caution, as it can be overly restrictive for models under certain circumstances, such as when lower severity crashes are underreported.

Unobserved Heterogeneity

- Differences in drivers' risk-taking behaviors, physiological attributes, and other factors lead to unobserved heterogeneity among road users involved in crashes.
- Data heterogeneity affects the model parameters among injury observations. Large effects, when unaccounted for, could lead to biased parameter estimates and incorrect statistical inferences.

Underreporting

- It has been well-documented that crashes with lower severity levels are less likely to be reported to governmental authorities compared to more severe crashes.
- For example, people involved in a reportable property damage only collision (above the minimum reportable threshold) may not be interested in seeing their vehicle insurance premiums go up and would therefore directly pay for the damages themselves or worse, flee from the crash scene (which is more common than we think).
- There is a lot of variation in the extent of underreporting, which can depend on the study location and severity levels.

Underreporting Varies by Location/time

- For instance, about three decades ago, Hauer and Hakkert (1988) stated that approximately 20% of severe injuries, 50% of minor injuries, and up to 60% of no-injury crashes were not reported.
- Elvik and Mysen (1999) reported underreporting rates of 30%, 75%, and 90% for serious, slight, and very slight injuries, respectively.
- According to Blincoe et al. (2002), up to 25% of all minor injuries and almost 50% of no-injury crashes were likely to be nonreported.
- The underreporting is a more significant issue in low and middle-income countries than in high-income countries.
- Some studies have proposed methods to minimize this bias even if the underreporting rate is unknown (see Kumara and Chin (2005); Yamamoto et al. (2008); Ma (2009) Ye and Lord (2011); Patil et al. (2012)). (see references in textbook)

Inconsistent Classification -KABCO

- Furthermore, there is a possibility of inconsistency in the classification of a crash outcome into no injury or possible injury levels; and/or an arbitrary crash threshold for the vehicle or property damages exceeding a certain amount.
- Developed by the National Safety Council (NSC) in 1966, the KABCO scale was adopted by the states to report injury severity at the scene of a crash; but KABCO Injury classification scale and definitions vary by state (FHWA: https://safety.fhwa.dot.gov/hsip/spm/conversion_tbl/pdfs/kabco_ctable_by_state.pdf)
- The 4th edition of the Model Minimum Uniform Crash Criteria (MMUCC) was the first major change to the KABCO scale since its inception and states were required to adopt this new definition for serious injury reporting by 04/15/2019.

MMUCC 4TH Edition: KABCO

- This edition not only changed injury severity names but also provided clear examples of specific injuries for each severity level. The new edition also brought significant clarity to a serious injury 'A', with the following guidance:
- "A suspected serious injury is any injury other than fatal which results in one or more of the following:
  - Severe laceration resulting in exposure of underlying tissues/muscle/organs or resulting in significant loss of blood,
  - Broken or distorted extremity (arm or leg),
  - Crush injuries,
  - Suspected skull, chest or abdominal injury other than bruises or minor lacerations,
  - Significant burns (second and third degree burns over 10% or more of the body),
  - Unconsciousness when taken from the crash scene, and
  - Paralysis."

KABCO	MMUCC 3 (1994-2016)	MMUCC 4/5 (2017-)
K	Fatal Injury	Fatal Injury
A	Incapacitating Injury	Suspected Serious Injury
B	Non-Incapacitating Injury	Suspected Minor Injury
C	Possible Injury	Possible Injury
O	No Injury	No Apparent Injury

Inconsistent Classification: KABCO vs. AIS

- KABCO or AIS (Abbreviated Injury Scale)
  - The Abbreviated Injury Scale (AIS) is an anatomically-based injury severity scoring system that classifies each injury by body region on a 6-point scale. AIS is the system used to determine the Injury Severity Score (ISS) of the multiply injured patient.
  - The AIS is an internationally accepted standard developed by the Association for the Advancement of Automotive Medicine (AAAM) in 1969 and most recently updated in 2015.
  - The AIS classifies individual injuries by body region (e.g., head, face, neck, abdomen, spine) as follows:

    AIS 1 – Minor

    AIS 2 – Moderate

    AIS 3 – Serious

    AIS 4 – Severe

    AIS 5 – Critical

    AIS 6 – Maximal (unsurvivable)

- The distribution of both KABCO and Maximum Abbreviated Injury Scale (MAIS) varies. MAIS was found to be more consistent between states than KABCO.

Other Data Issues

- Small Sample size
  - Will affect the proportion (see unbalanced data above)
    - See next slide for minimum values.
- Endogeneity
  - An endogenous variable is an explanatory variable whose value is determined or influenced by one or more variables in the model.
  - Carson and Mannering (2001) studied the endogeneity problem by exploring the effectiveness of ice-warning signs in reducing the frequency of ice-related crashes.
  - An indicator variable for the presence of an ice warning sign is typically used when developing a crash-frequency model.
  - As ice-warning signs are more likely to be placed at locations with high numbers of ice-related crashes, this indicator variable may be endogenous (the explanatory variable will change as the dependent variable changes).

Minimum Sample Size

Table 4
Three evaluation criteria by sample size for the three models*.

Sample size	Mean of absolute-percentage-bias (APB)			Max of absolute-percentage-bias (APB)			Total root-mean-square-error (RMSE)		
	MNL	ML	OP	MNL	ML	OP	MNL	ML	OP
100	5.50E+13	2.10E+11	143%	9.70E+14	2.90E+12	2.10E+01	7.40E+15	1.60E+13	20.7
500	2.00E+14	1.10E+04	25%	4.50E+15	1.10E+05	94%	1.30E+16	1.20E+06	4.5
2000	16%	26%	11%	45%	167%	40%	12.9	28.7	2.2
5000	9%	13%	5%	27%	52%	20%	7.6	13.7	1.2
10,000	4%	5%	4%	13%	13%	14%	4.7	8.7	0.7
20,000	2%	3%	2%	9%	21%	9%	1.9	3.4	0.4

* MNL: multinomial logit model; OP: ordered probit model; ML: mixed logit model.

In terms of the values of all three criteria, the multinomial logit and mixed logit are more sensitive to small sample sizes than the ordered probit model. This is especially noticeable for the sample sizes equal to 100 and 500. Nonetheless, for a sample size below 500, all models perform poorly.

According to the three criteria, the minimum sample size for the ordered probit, multinomial logit, and mixed logit models should be 2000, 5000 and 10,000, respectively.

Utility and Utility Function

- **Utility** is a measure of relative satisfaction such as a consumer will choose a product with the combination of quality and price to achieve the maximum utility; or a traveler will choose the combination of mode and destination that provides the most utility.
- In the context of safety, we are looking for a combination of factors that lead to the worst injuries.
- The **utility function** usually favors the maximum utility (e.g., high injury severity levels) and is usually a linear form of covariates as follows:

$$U_{ni} = \beta_{0i} + \beta_{1i} x_{n1i} + \beta_{2i} x_{n2i} + \ldots + \beta_{ki} x_{nki} = \mathbf{x}'_{ni} \boldsymbol{\beta}_i \qquad (4.1)$$

Where U_{ni} is the utility value of crash n with injury severity level i; x_{nki} is the kth variable related to injury level i; β_{i0} is the constant for injury level i; and, β_{ki} is the estimable coefficients for the covariates.

Making A (Deterministic) Choice

- Utility maximization is the process of choosing the alternative with the maximum utility value.
- In a binary outcome model with injury and no injury, if $U(injury)>U(no\ injury)$, then the probability of injury $P_r\ (injury)=1$; and if $U(injury)<U(no\ injury)$, then $P_r\ (injury)=0$. This is a deterministic choice that can be depicted in Fig. 4.1.

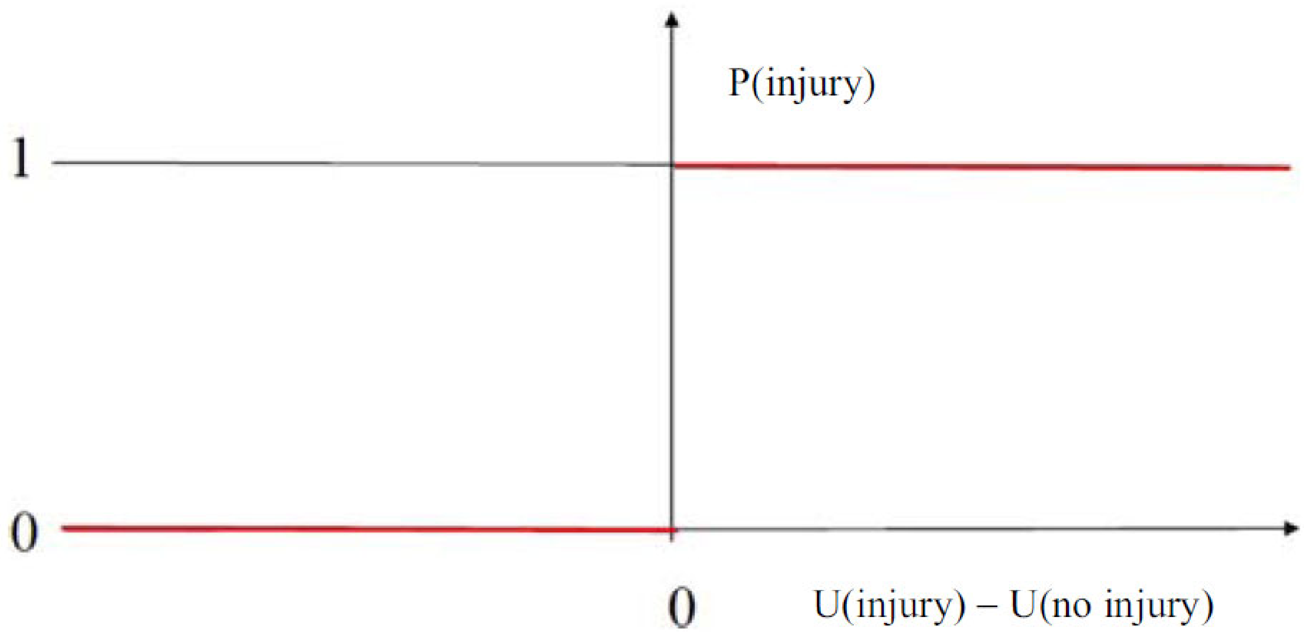


FIGURE 4.1 Deterministic choice of a binary variable.

Making A (Deterministic) Choice

- Utility maximization is the process of choosing the alternative with the maximum utility value.
- In a binary outcome model with injury and no injury, if $U(injury)>U(no\ injury)$, then the probability of injury $P_r\ (injury)=1$; and if $U(injury)<U(no\ injury)$, then $P_r\ (injury)=0$. This is a deterministic choice that can be depicted in Fig. 4.1.

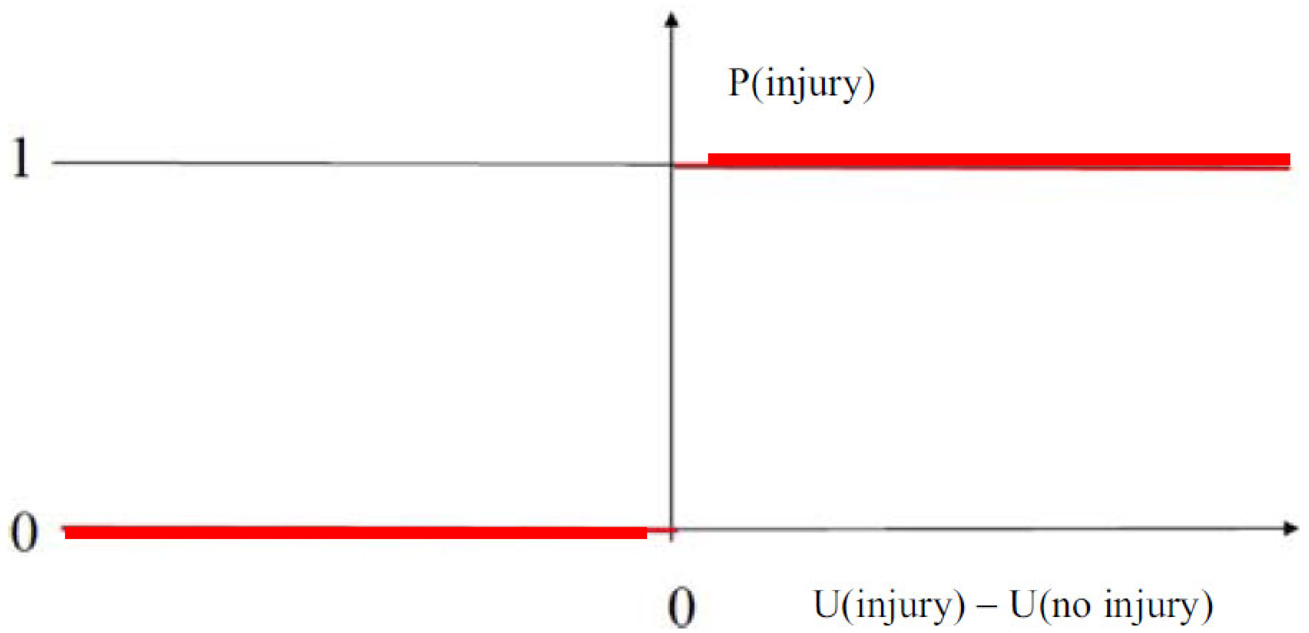


FIGURE 4.1 Deterministic choice of a binary variable.

Random Utility Function

- Because it is impossible that the utility of each crash injury outcome can be specified with certain, a random un-specificable error term is added to the end.

$$U_{ni} = \beta_{0i} + \beta_{1i}x_{n1i} + \beta_{2i}x_{n2i} + \cdots + \beta_{ki}x_{nki} + \varepsilon_{ni} = V_{ni} + \varepsilon_{ni} \quad (4.2)$$

where V_{ni} represents the deterministic portion of U_{ni}.

- **Reasons for adding a disturbance term:**
  - Variables have been omitted from the function (some important data may not be available),
  - The functional form may be incorrectly specified (it may not be linear),
  - Proxy variables may be used (variables that approximate missing variables in the database),
  - Variations in β_i that are not accounted for (β_i may vary across observations).

Random Utility Model

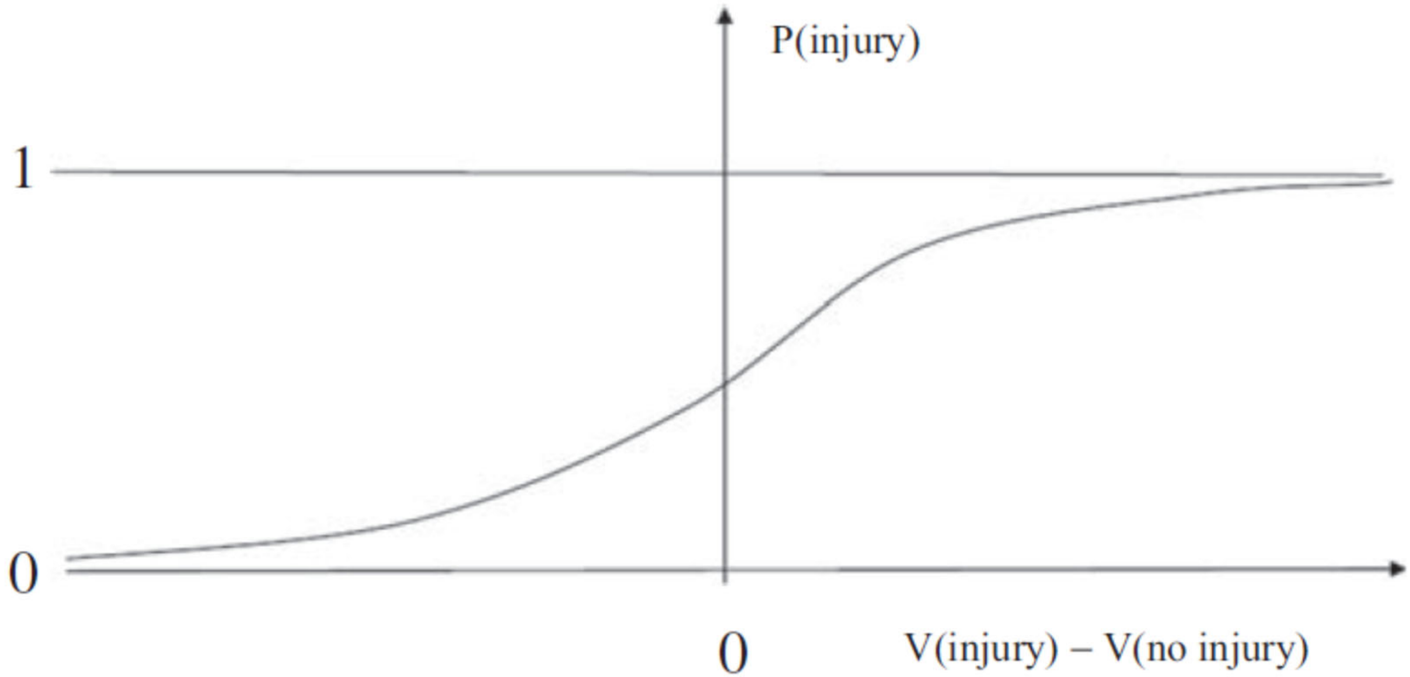


FIGURE 4.2 Stochastic choice of a binary variable.

Models are estimated by assuming a distribution for the random error term, ε's. Now, instead of being a deterministic outcome, the probability of each outcome alternative is determined by the distributional form (Fig. 4.2).

Random Utility Model

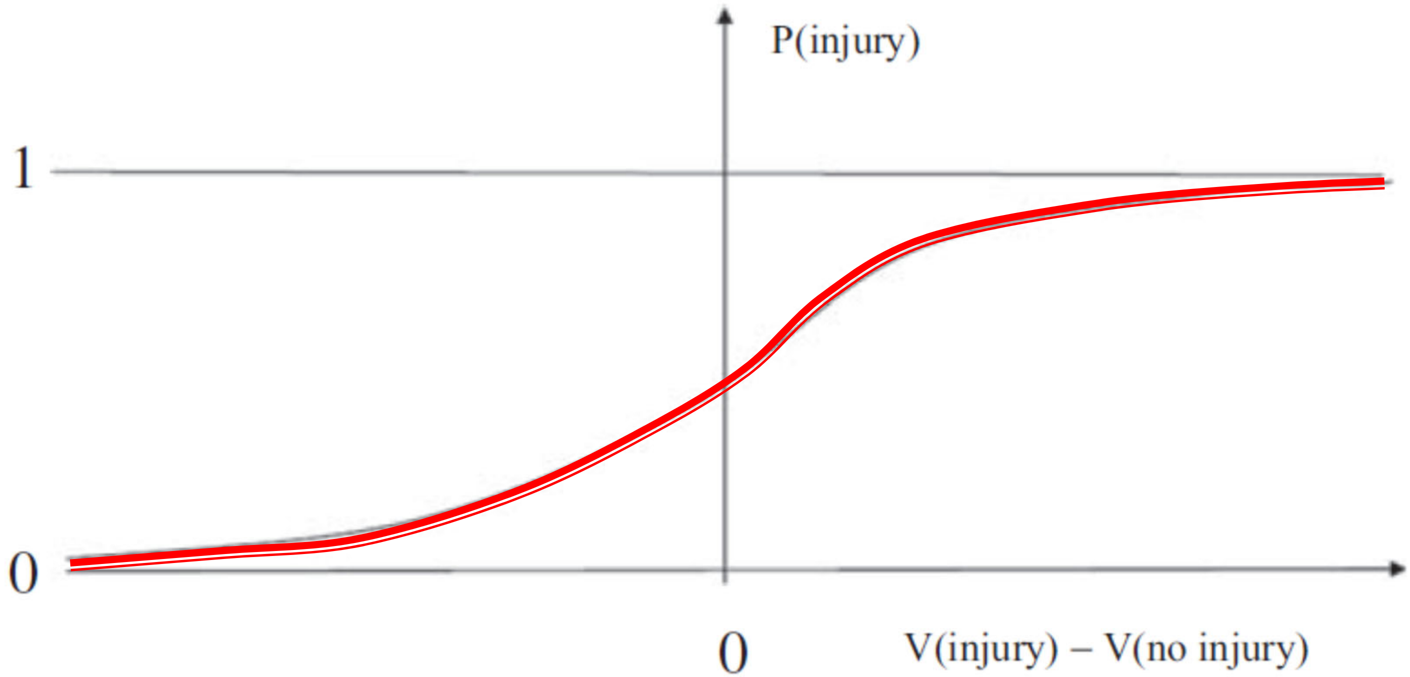


FIGURE 4.2 Stochastic choice of a binary variable.

Models are estimated by assuming a distribution for the random error term, ε's. Now, instead of being a deterministic outcome, the probability of each outcome alternative is determined by the distributional form (Fig. 4.2).

Unordered Response Models

Modeling Crash Severity as an Unordered Discrete Outcome

- In a probabilistic model, we are looking for a combination of factors that lead to the worst injuries, or highest utility.
- Treating the dependent variable with multiple responses as ordinal or as nominal significantly impacts which methodologies should be considered.

Binary Probit Models

- The name comes from probability and unit. The purpose of the model is to estimate the probability that an observation t with particular characteristics will fall into a specific category (1 or 0).

$$y_t = \begin{cases} 1 & if\ U_t^* > 0 \\ 0 & otherwise \end{cases}$$ Where $$U_t^* = x_t'\beta + \varepsilon_t$$

$$P_r(y_t = 1|x_t) = P_r(U_t^* > 0) = P_r(x_t'\beta + \varepsilon_n > 0) = \Phi(x_t'\beta)$$

Where $\varepsilon_t \sim N(0,1)$

- Likelihood function: $\mathcal{L}(\beta) = \Phi(x_t'\beta)^{y_t}[1 - \Phi(x_t'\beta)]^{1-y_t}$
- Maximum likelihood estimate:

$$\hat{\beta} = \arg\max_{\beta} [\ln \mathcal{L}(\beta)] = \arg\max_{\beta} \left[\sum_t (y_t \ln \Phi(x_t'\beta) + (1 - y_t) \ln(1 - \Phi(x_t'\beta))) \right]$$

Multinomial Probit Models

- The problem with the multinomial probit is that the outcome probabilities are **not closed form** and estimation of the likelihood functions requires numerical integration.
- The difficulties of extending the probit formulation to more than two discrete outcomes have lead researchers to consider other disturbance term distributions.

Modeling Choices

- From a model estimation perspective, a desirable property of an assumed distribution of disturbances (ε's) is that the maximums of randomly drawn values from the distribution have the same distribution as the values from which they were drawn. Because of this property, the highest utility value of all other options in a multinomial case can be defined as $x'_{nj}\beta_j(\forall j \neq i)$.
- The normal distribution does not posses this property (the maximums of randomly drawn values from the normal distribution are not normally distributed).
- Distributions of the maximums of randomly drawn values from some underlying distribution are referred to as extreme value distributions (Gumbel, 1958).
- Extreme value distributions are categorized into three families: Type 1, Type 2, and Type 3.
- The most common extreme value distribution is Type 1, or the Gumbel distribution. Based on the error distributional assumption of the Gumbel distribution (Type 1 extreme value), the most known discrete choice model is the MNL model.

Making (Probabilistic) Choices

- To derive an estimable model of discrete outcomes with I denoting all possible outcomes for observation *n*, and $P_n(i)$ being the probability of observation *n* having discrete outcome *i* ($i \in I$)

$$P_n(i) = P(U_{in} \geq U_{In}) \ \forall \ I \neq i \ .$$

$$P_{ni} = \Pr\left(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}, \forall j \neq i\right) = \Pr\left(\varepsilon_{nj} < V_{ni} + \varepsilon_{ni} - V_{nj}, \forall j \neq i\right) \qquad (4.3)$$

- Estimable models are developed by assuming a distribution of the random disturbance term, ε's.

Multinomial Logit Model

- The most common extreme value distribution is the Type 1 distribution (sometimes referred to as the Gumbel distribution). It has the desirable property that maximums of randomly drawn values from the extreme value Type 1 distribution are also extreme value Type 1 distributed.
- The probability density function of Gumbel distribution is,

$$f(x) = e^{-x} e^{-\exp(-x)} \qquad F(x) = e^{-\exp(-x)}$$

- When Gumbel is assumed and ε_{nis} are independent, the cumulative distribution over all j≠i is the product of individual cumulative distributions as:

$$P_{ni} \mid \varepsilon_{ni} = \prod_{j \neq i} e^{-\exp\left[-\left(V_{ni} + \varepsilon_{ni} - V_{nj}\right)\right]} \qquad (4.4)$$

- Since ε_{nis} is not given, the choice probability is the integral of $P_{ni} \mid \varepsilon_{nis}$ over all values of ε_{nis} weighted by its density as:

$$P_{ni} = \int \left\{ \prod_{j \neq i} e^{-\exp\left[-\left(V_{ni} + \varepsilon_{ni} - V_{nj}\right)\right]} \right\} e^{-\varepsilon_{ni}} e^{-\exp(-\varepsilon_{ni})} d\varepsilon_{ni} \qquad (4.5)$$

Multinomial logit model

This results in a closed-form expression known as the MNL model, formulated as:

$$P_{ni} = \Pr(y_n = i) = \frac{\exp(\mathbf{x}'_{ni}\boldsymbol{\beta}_i)}{\sum_{i=1}^{I}\exp(\mathbf{x}'_{ni}\boldsymbol{\beta}_i)} \qquad (4.6)$$

If level I is the reference level, the model becomes

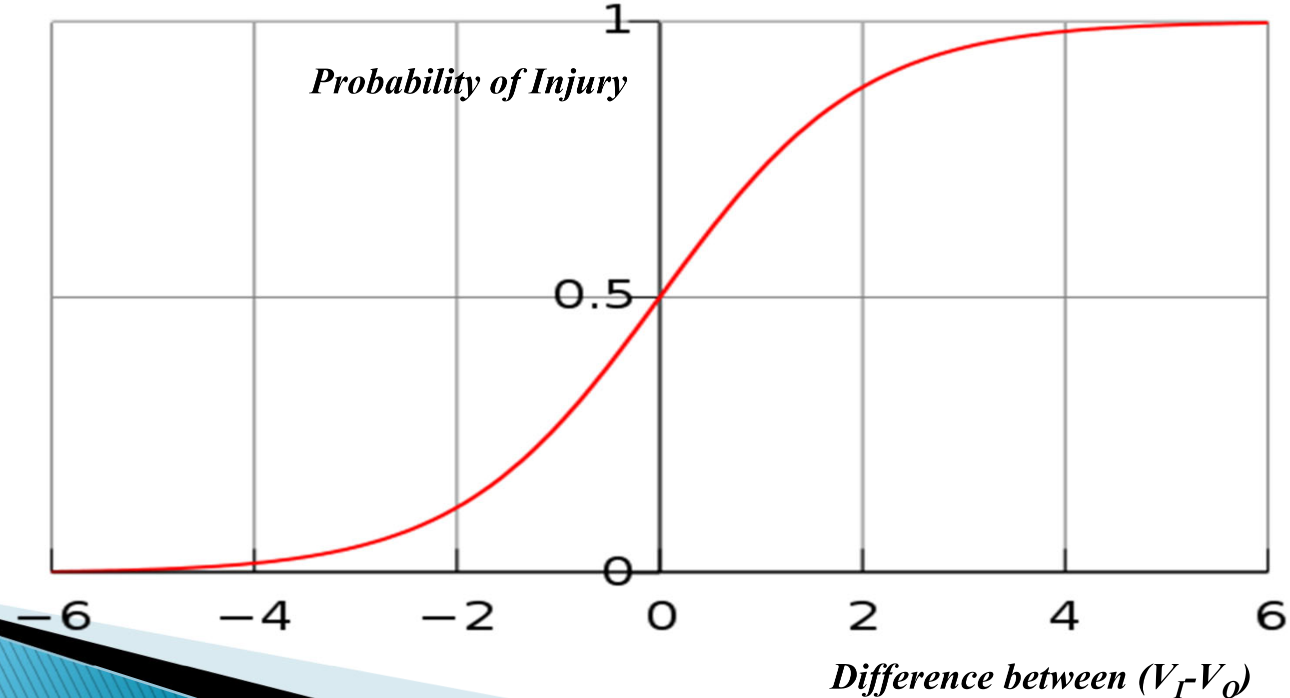
$$\log\left[\frac{P_n(i)}{P_n(I)}\right] = \mathbf{x}'_{ni}\boldsymbol{\beta}_i \qquad (4.7)$$

Note that in crash severity modeling, the lowest injury severity level (i.e., "no injuries" or "property damage only" (PDO)) is usually set to be the reference level instead of level I (in R code, it's "reflevel" argument in "mlogit").

Binomial Logit Model

If there are two injury outcome (injury (I) and no injury (O)), the probability of injury) can be written as a Logit (logistic) Model.

$$P(I) = \frac{e^{V_I}}{e^{V_I} + e^{V_O}} = \frac{1}{1 + e^{(V_O - V_I)}}$$



Parameter Estimate

- Binary Logit Example
  - $logitP = log\left(\frac{p}{1-p}\right) = \boldsymbol{\beta x}$
  - The parameter vector $\boldsymbol{\beta}$ is readily estimated using standard maximum likelihood methods and the steps are as follows:

1. $L = \mathrm{P_r}(y_1, y_2, \dots yn) = \prod_i {P_i}^{y_i}(1-Pi)^{1-yi} = \prod_i \left(\frac{P_i}{1-Pi}\right)^{y_i}(1-Pi)$
2. $LL = \sum_i log\left(\frac{P_i}{1-Pi}\right) + \sum_i log\ (1-Pi) = \sum_i \beta xiyi - \sum_i \log(1+\exp(\boldsymbol{\beta} X_i)$
3. $\frac{\partial LL}{\partial \boldsymbol{\beta}} = 0$; solve a series of k equations for k estimable parameter vector $\boldsymbol{\beta}$.

- A more general form: If δ_{in} is defined as being equal to 1 if the observed discrete outcome for observation n is i and zero otherwise, the likelihood function is:

$$L = \prod_{n=1}^{N} \prod_{i=1}^{I} P(i)^{\delta_{in}}$$

Probability, Odds, Odds Ratio and Marginal Effect

- P is probability of an event
- $\frac{P}{1-P} = \frac{Probability\ of\ event}{Probability\ of\ no\ event}$ is the odds (O) of such an event,
- $\mathrm{P} = \frac{O}{O+1}$; an odds of 3 means P (event)=0.75
- Odds ratio (OR) is defined as the ratio of the odds of event A in the presence of B and the odds of A in the absence of B. An odd ratio is used to compared two dichotomous (binary) variables (the variables have only two categories or levels), say odds (injury crash|x=1) vs. odds (injury crash|x=0) if x is seatbelt use.
- $\exp(\beta x)$ is the odds ratios if x is an indicator variable
- Marginal effects $\frac{\partial p}{\partial x_1} = \frac{\beta_1 e^{\beta x}}{\left(1+e^{-\beta x}\right)^2}$ is a way of presenting results as differences in probabilities, which is more informative than odds ratios. It is about derivatives.
- Marginal effects for continuous variables apply to a small change in *x* when effects are non-linear. They are not changes by 1 unit, strictly speaking.

Elasticity

- **Elasticity** is computed from the partial derivative for each observation *n* (*n* subscripting omitted):

$$E_{x_{ki}}^{P(i)} = \frac{\partial P(i)}{\partial x_{ki}} \times \frac{x_{ki}}{P(i)}$$

  Where $P(i)$ is the probability of outcome *i* ; and x_{ki} is the value of variable *k* for outcome *i*.

- **Elasticity values** are interpreted as the percent effect that a 1% change in x_{ki} has on the outcome probability $P(i)$: $E_{x_{ik}}^{P(i)} = \left[1 - P(i)\right] \beta_{ki} x_{ki}$
  - If the computed elasticity value is less than one, the variable x_{ki} is said to be inelastic and a 1% change in x_{ki} will have less than a 1% change in outcome *i*'s selection probability.
  - If the computed elasticity is greater than one it is said to be elastic and a 1% change in x_{ki} will have more than a 1% change in outcome *i*'s selection probability.

Pseudo-Elasticity

1. The values are point elasticity's and as such are valid only for small changes x_{ik} and considerable error may be introduced when an elasticity is used to estimate the probability change caused by a doubling of x_{ki}.
2. Elasticities are not applicable to indicator variables
   - Some measure of the sensitivity of indicator variables is made by computing a pseudo-elasticity. The equation is

$$E_{x_{ki}}^{P(i)} = \frac{EXP\left[\Delta\left(\beta_i X_i\right)\right]\sum_{\forall I} EXP\left(\beta_{kI} x_{kI}\right)}{EXP\left[\Delta\left(\beta_i X_i\right)\right]\sum_{\forall I_n} EXP\left(\beta_{kI} x_{kI}\right) + \sum_{\forall I \neq I_n} EXP\left(\beta_{kI} x_{kI}\right)} - 1$$

where I_n is the set of alternate outcomes with x_k in the function determining the outcome, and I is the set of all possible outcomes.

mlogit: Multinomial Logit Models

- Yves Croissant. **mlogit: Multinomial Logit Models**. https://cran.r-project.org/web/packages/mlogit/
  - Random utility model and the multinomial logit model. https://cran.r-project.org/web/packages/mlogit/vignettes/c3.rum.html
  - Logit models relaxing the iid hypothesis (including The nested logit model). https://cran.r-project.org/web/packages/mlogit/vignettes/c4.relaxiid.html
- Estimation of multinomial logit models in R : The mlogit Packages by Yves Croissant (71 pages) (http://www2.uaem.mx/r-mirror/web/packages/mlogit/vignettes/mlogit.pdf)
- mlogit: Multinomial logit model (Estimation by maximum likelihood of the multinomial logit model, with alternative-specific and/or individual specific variables) https://www.rdocumentation.org/packages/mlogit/versions/1.1-1/topics/mlogit

mlogit: Multinomial Logit Models

Maximum likelihood estimation of random utility discrete choice models. The software is described in Croissant (2020) <doi:10.18637/jss.v095.i11> and the underlying methods in Train (2009) <doi:10.1017/CBO9780511805271>.

Version:	1.1-1
Depends:	R (≥ 2.10), dfidx
Imports:	Formula, zoo, lmtest, statmod, MASS, Rdpack
Suggests:	knitr, car, nnet, lattice, AER, ggplot2, texreg, rmarkdown
Published:	2020-10-02
Author:	Yves Croissant [aut, cre]
Maintainer:	Yves Croissant <yves.croissant at univ-reunion.fr>
License:	GPL-2 | GPL-3 [expanded from: GPL (≥ 2)]
URL:	https://cran.r-project.org/package=mlogit, https://r-forge.r-project.org/projects/mlogit/
NeedsCompilation:	no
Citation:	mlogit citation info
Materials:	NEWS
In views:	Econometrics, SocialSciences
CRAN checks:	mlogit results

Documentation:

Reference manual:	mlogit.pdf
Vignettes:	2. Data management, model description and testing
	3. Random utility model and the multinomial logit model
	4. Logit models relaxing the iid hypothesis
	5. The random parameters (or mixed) logit model
	6. The multinomial probit model
	7. Miscellaneous models
	Exercise 1: Multinomial logit model
	Exercise 2: Nested logit model
	Exercise 3: Mixed logit model
	Exercise 4: Multinomial probit
	mlogit

MNL Model Coefficient Estimates

Exercise 4.1: Coefficient Estimates for MNL*

*: Only partial variables are displayed due to the space. Total number of variables is *k*=36 (1 intercept variable and 17 variables for each of the two levels: B/C and K/A).

	B or C			K or A		
Variable	Estimate	Std. Error	Pr(>∣z∣)	Estimate	Std. Error	Pr(>∣z∣)
Rule violation	0.3192	0.0611	0.0000	0.9368	0.0888	0.0000
Reckless behavior	0.2263	0.0507	0.0000	0.3559	0.0767	0.0000
Signal	0.6930	0.1471	0.0000	0.6216	0.2727	0.0227
Two-way	0.7419	0.1555	0.0000	1.3280	0.2709	0.0000
None	0.4295	0.1379	0.0018	0.8710	0.2580	0.0007
Total units	0.3264	0.0269	0.0000	0.3849	0.0358	0.0000
Snow	−0.6935	0.0753	0.0000	−1.0676	0.1305	0.0000
Ice	−0.5375	0.1080	0.0000	−0.7336	0.1836	0.0001
Wet	0.0467	0.0675	0.4891	−0.3037	0.1113	0.0064
Dark	0.0991	0.0613	0.1059	0.3775	0.0901	0.0000

AIC: 17,869.32; Log-Likelihood: −8898.7, McFadden R^2: 0.062873.

Akaike information criterion (AIC) = $2k-2LL$ = 2*36+2*8897.7=72+17797.4=17,869.4
McFadden's $R^2 = 1-LL_{mod}/LL_0$;
Where LL_{mod} log likelihood value for the fitted model and LL_0 for the null model which includes only an intercept as predictor.

Solution

1. Determine the functional form: $P_{ni} = \Pr(y_n = i) = \frac{\exp(\mathbf{x}'_{ni}\boldsymbol{\beta}_i)}{\sum_{i=1}^{I}\exp(\mathbf{x}'_{ni}\boldsymbol{\beta}_i)}$

In this functional form, y_n is the crash injury severity with three levels: PDO (i = 1), B or C (i = 2); and K or A (i = 3). Xβ is a vector of explanatory variables that determines the severity of crash observation n (n=1,10000), and β_i is a vector of coefficients for injury severity level i.

2. Estimate the coefficients using the R "mlogit" package:

"crash_mnl <- mlogit.data(data_model_ch4, shape = "wide", choice = "INJSVR")

multi_logit <- mlogit(INJSVR ~ 0 | YOUNG + OLD + FEMALE + ALCFLAG + DRUGFLAG + SAFETY + DRVRPC_SPD + DRVRPC_RULEVIO + DRVRPC_RECK + TRFCONT_SIGNAL + TRFCONT_2WAY + TRFCONT_NONE + TOTUNIT + ROADCOND_SNOW + ROADCOND_ICE + ROADCOND_WET + LGTCOND_DARK, data = crash_mnl)"

3. summarize your findings. In the MNL model, the coefficient estimates are explained as the comparison between injury level i and the base level PDO (i=1). As can be seen in the table, if a driver was influenced by drugs, his or her chance of getting injured increases drastically, with respective probabilities of level B or C and level K or A being 4.49 ($e^{1.5013}$) times and 12.95 (or $e^{2.561}$) times that of PDO. The exponentiated value of the logit coefficients is also called odds ratio.

Statistical Evaluation

- To determine if the estimated parameter is significantly different from zero, the *t*-statistic is: $t = \dfrac{\beta\ \text{-}\ 0}{S.E.(\beta)}$
- The likelihood ratio test: -2[$LL(\beta_R)$ – $LL(\beta_U)$]
  - where $LL(\beta_R)$ is the log-likelihood at convergence of the "restricted" model and $LL(\beta_U)$ is the log-likelihood at convergence of the "unrestricted" model.
  - This statistic is χ^2 distributed with degrees of freedom equal to the difference in the numbers of parameters between the restricted and unrestricted model (the difference in the number of parameters in the β_R and the β_U parameter vectors).
- Overall model fit is the ρ^2 statistic or McFadden's R^2 (it is similar to R^2 in regression models in terms of purpose but should not be explained as R^2). The ρ^2 statistic is: $\rho^2 = 1 - \dfrac{LL(\beta)}{LL(0)}$
  - where $LL(\beta)$ is the log-likelihood at convergence with parameter vector β and $LL(0)$ is the initial log-likelihood (with all parameters set to zero, only keep the intercept).
  - To account for the estimation of potentially insignificant parameters a corrected ρ^2 is estimated as:

$$\text{corrected } \rho^2 = 1 - \frac{LL(\beta)\ \text{-}\ K}{LL(0)}$$

where K is the number of parameters estimated in the model.

Specification Errors

- **Independence of Irrelevant Alternatives (IIA) property**
  - This problem arises when only some of the functions, which determine possible outcomes, share unobserved elements (that show up in the disturbances).
  - If all outcomes shared the same unobserved effects, the problem would self correct because in the differencing of outcome functions common unobserved effects would cancel out.
  - Small-Hsaio test.
- **Omitted variables**
  - the omitted variable is correlated with other variables included in the model,
  - the mean values of the omitted variable vary across alternate outcomes and outcome specific constants are not included in the model, or
  - the omitted variable is correlated across alternate outcomes or has a different variance in different outcomes.

Because one or more of these conditions are likely to hold, omitting relevant variables is a serious specification problem.

- **Presence of an irrelevant variable**.
  - Estimates of parameter and choice probabilities remain consistent in the presence of an irrelevant variable but the standard errors of the parameter estimates will increase (loss of efficiency).
- **Disturbances that are not independently and identically distributed (IID)**.
  - Dependence among a subset of possible outcomes causes the IIA problem resulting in inconsistent parameter estimates and outcome probabilities.
  - Having disturbances with different variances (not identically distributed) also results in inconsistent parameter estimates and outcome probabilities.

Independence Of Irrelevant Alternatives (IIA)

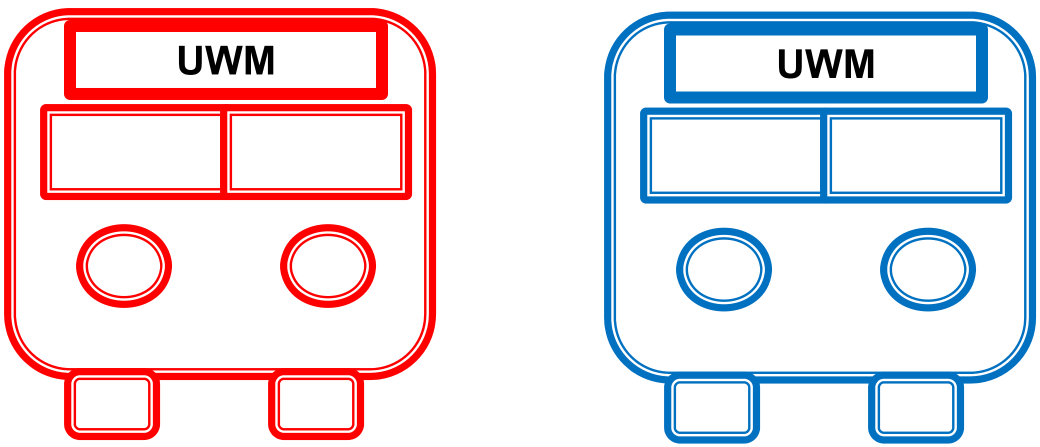
- It is a property of MNL models, but not for all discrete choice or discrete outcome models.
- If adding a new mode, mode shares will be taken from the other available modes.
- However, the ratio of the probabilities of the two alternatives is not affected by the utility of any other alternative in the choice set. In other words, this ratio is not influenced by any change in the utility of a third ("irrelevant") alternative.
- Because the ratio $P(i)/P(j)=exp(U_i-U_j)$ is unaffected by the third alternative.

$$P_{ni} = \Pr(y_n = i) = \frac{\exp(\mathbf{x}'_{ni}\boldsymbol{\beta}_i)}{\sum_{i=1}^{I}\exp(\mathbf{x}'_{ni}\boldsymbol{\beta}_i)}$$

The Nested Logit Model (Generalized Extreme Value Models)

Blue Bus, Red Bus

Can we increase the bus ridership by painting the bus with a different color?



Issues with Logit Model: Red Bus Blue Bus

- Clearly, the bus share should not have been changed. What is wrong?
- The problem is with the underlying assumption in the logit model. The logit model requires that alternatives be independent (i.e. ε_{red} and ε_{blue} be independent). This is not the case in this example.
- Obviously, the errors of the perceived utility from alternative of the red bus is dependent on the error from the blue bus, and vice versa. This does not justify the use of logit model.
- Note: when the alternatives are distinctly different and independent, the logit model shall work well.

Nested Logit Model

- To overcome the IIA problem, the idea behind a nested logit model is to group alternate outcomes suspected of sharing unobserved effects into nests (this sharing sets up the disturbance term correlation that violates the derivation assumption).
- Because the outcome probabilities are determined by differences in the functions determining these probabilities (both observed and unobserved), shared unobserved effects will cancel out in each nest providing that all alternatives in the nest share the same unobserved effects.

Nested Logit Model Structure

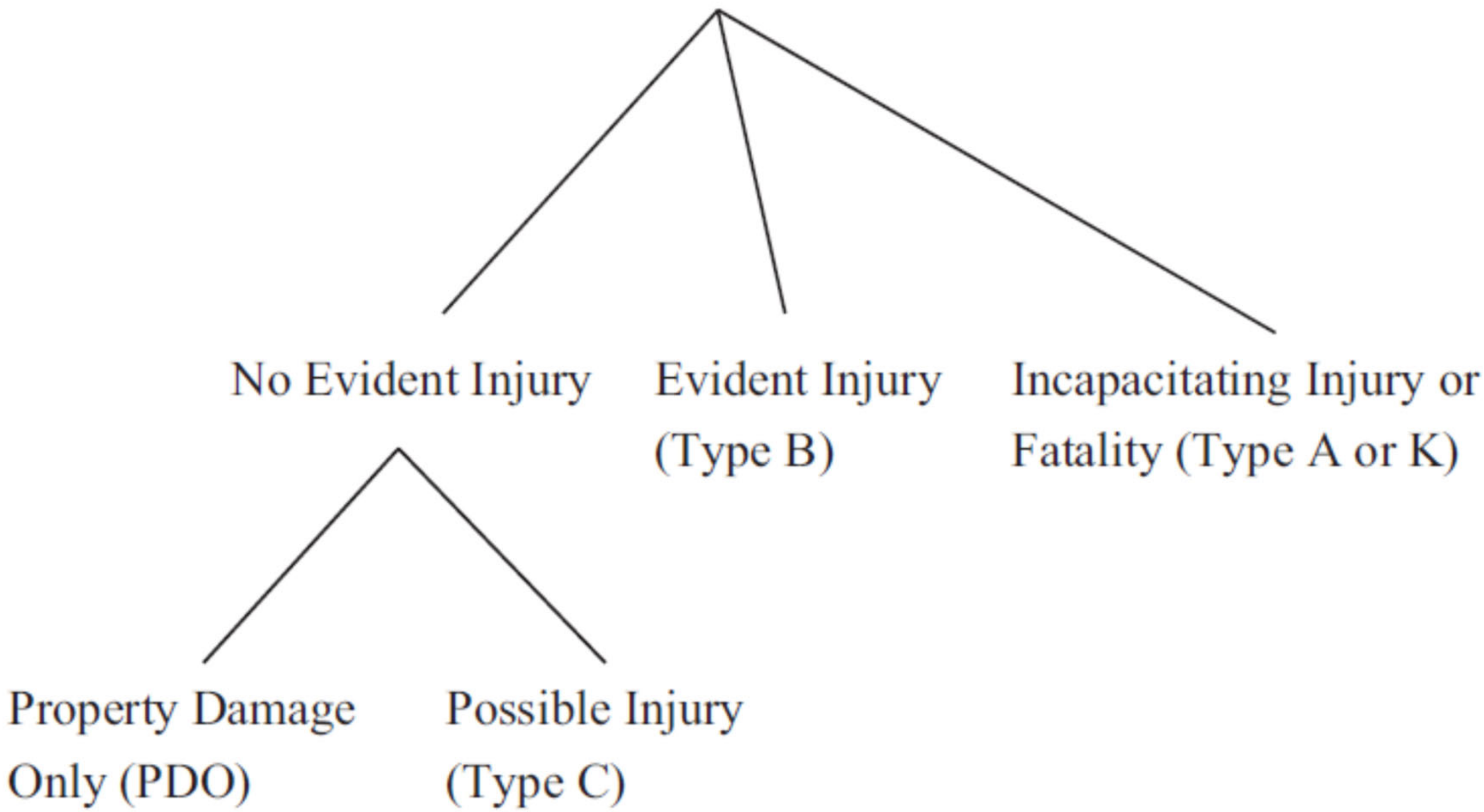


FIGURE 4.3 Nested structure of accident severities.

A General Formation

- Mathematically, McFadden (1981) has shown the GEV disturbance assumption leads to the following model structure for observation *n* choosing outcome *i*

$$P_n(i) = \frac{\exp[\boldsymbol{\beta}_i \mathbf{X}_{in} + \phi_i LS_{in}]}{\sum_{\forall I} \exp[\boldsymbol{\beta}_I \mathbf{X}_{In} + \phi_I LS_{In}]} \quad (4.8a)$$

$$P_n\left(j|i\right) = \frac{\exp[\boldsymbol{\beta}_{j|i} \mathbf{X}_n]}{\sum_{\forall J} \exp[\boldsymbol{\beta}_{J|i} \mathbf{X}_{Jn}]} \quad (4.8b)$$

$$LS_{in} = LN\left[\sum_{\forall J} \exp(\boldsymbol{\beta}_{J|i} \mathbf{X}_{Jn})\right], \quad (4.8c)$$

Where

$P_n(i)$ is the unconditional probability of observation n having discrete outcome i,

X's are vectors of characteristics that determine the probability of discrete outcomes,

β's are vectors of estimable parameters,

$P_n(j|i)$ is the probability of observation n having discrete outcome j conditioned on the outcome being in outcome category i,

J is the conditional set of outcomes (conditioned on i), I is the unconditional set of outcome categories, LS_{in} is the inclusive value (logsum), and ϕ_i is an estimable parameter.

ϕ

- In order to be consistent with McFadden's generalized extreme value derivation of the model, the parameter estimate for ϕ_i in the nested logit model must be between zero and one.
- If ϕ_i equals to one or is not significantly different from one, there is no correlation between the severity levels in the nest, meaning the model reduces to the multinomial logit model.
- If ϕ_i equals to zero, a perfect correlation is implied among the severity levels in the nest, indicating a deterministic process by which crashes result in particular severity levels.
- The t test can be used to test if ϕ_i is significantly different from 1. Because ϕ_i is less than or equal to one, this is a one-tailed t test (half of the two-tailed t-test).
- It is important to note that the typical t-test implemented in many commercial software packages are against zero instead of one. Thus, the t value must be calculated manually. The IIA assumption for a MNL model can also be tested with the Hausman-McFadden (1984) test which has been widely implemented in commercial statistical software.

Estimation of a Nested Model

Usually done in a sequential fashion.

1. Estimate the conditional model using only the observations in the sample that are observed having discrete outcomes *J*. In the example illustrated in the Figure this is a binary model of commuters observed taking the arterial or the freeway.
2. Once these estimation results are obtained, the logsum is calculated (this is the denominator of one or more of the conditional models) for all observations, both those selecting *J* and those not (for all commuters in our example case).
3. These computed logsums (in our example there is just one logsum) are used as independent variables in the functions. Note that not all unconditional outcomes need to have a logsum in their respective functions (the example shown in the Figure would only have a logsum present in the function for the non-freeway choice).

NL Model Coefficient Estimates

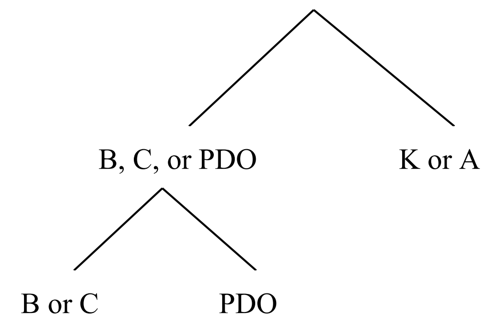
Exercise 4.2: Coefficient Estimates for NL

	B or C			K or A		
Variable	Estimate	Std. Error	Pr(>|z|)	Estimate	Std. Error	Pr(>|z|)
Rule violation	0.3192	0.0611	0.0000	0.9368	0.0888	0.0000
Reckless behavior	0.2263	0.0507	0.0000	0.3559	0.0767	0.0000
Signal	0.6930	0.1471	0.0000	0.6216	0.2727	0.0227
Two-way	0.7419	0.1555	0.0000	1.3280	0.2709	0.0000
None	0.4295	0.1379	0.0018	0.8710	0.2580	0.0007
Total units	0.3264	0.0269	0.0000	0.3849	0.0358	0.0000
Snow	−0.6935	0.0753	0.0000	−1.0676	0.1305	0.0000
Ice	−0.5375	0.1080	0.0000	−0.7336	0.1836	0.0001
Wet	0.0467	0.0675	0.4891	−0.3037	0.1113	0.0064
Dark	0.0991	0.0613	0.1059	0.3775	0.0901	0.0000

AIC: 17,869.32; Log-Likelihood: −8898.7, McFadden R^2: 0.062873.

Solution

1. Establish the nested structure of crash severities. :



2. Determine the functional form based on Eq. 4.8 (a), (b) and (c). For example, $P_n(j|i)$ is the probability of crash n having injury outcome B or C conditioned on the injury outcome being in category not a K or A injury. I is the unconditional set of outcome categories (for example, the upper three branches in the figure: no K/A injury and K/A injury). LSni is the inclusive value (logsum).

3. Estimate the coefficients using the R "mlogit" package:

nested_logit <- mlogit(INJSVR ~ 0|YOUNG + OLD + FEMALE + ALCFLAG + DRUGFLAG + SAFETY + DRVRPC_SPD + DRVRPC_RULEVIO + DRVRPC_RECK + TRFCONT_SIGNAL + TRFCONT_2WAY + TRFCONT_NONE + TOTUNIT + ROADCOND_SNOW + ROADCOND_ICE + ROADCOND_WET + LGTCOND_DARK, data = crash_mnl, nests = list(KA = c("3"), non_KA = c("1", "2")), un.nest.el = TRUE).

4. summarize your findings. The AIC value of the NL model (17871.11) is greater than that of MNL model (17869.32), indicating inferior performance. The inclusive value is 0.7161 and its t-value is -0.474. Apparently, the log-sum coefficient is not significantly different from 1. When the inclusive value is equal to one or not significantly different from 1, there is no correlation between the severity levels in the nest. We can conclude that for this dataset, the MNL model is more appropriate.

Mixed Logit: Motivation

- More aggregate – cannot include specific accident characteristics (driver characteristics, vehicle characteristics, restraint usage, alcohol consumption, etc.).
- Without detailed accident information, the approach potentially introduces a heterogeneity problem.
- Heterogeneity could result in varying effects of X that could be captured with random parameters.
- Mixed logit may be appropriate.
  - Relaxes possible IIA problems with a more general error-term structure.
  - Can test a variety of distribution options for β .
  - Estimated with simulation based maximum likelihood.

Mixed Logit Model

This model is similar to the random parameter model for the crash-frequency model. This means that the coefficients are allowed to vary across observations.

$$P_{ni}(i) = \int \frac{\exp(\mathbf{x}'_{ni}\boldsymbol{\beta}_i)}{\sum_J \exp(\mathbf{x}'_{nJ}\boldsymbol{\beta}_J)} f(\boldsymbol{\beta}|\boldsymbol{\phi})d\boldsymbol{\beta} \tag{4.9}$$

where $f(\boldsymbol{\beta}|\boldsymbol{\varphi})$ is a density function of $\boldsymbol{\beta}$ and $\boldsymbol{\varphi}$ is a vector of parameters which specify the density function, with all other terms as previously defined.

In a statistics term, the weighted average of several functions is called a mixed function, and the density that provides the weights is called the mixing distribution. Mixed logit is a mixture of the standard logit function evaluated at different $\boldsymbol{\beta}$ with $f(\boldsymbol{\beta})$ being the mixing distribution.

Model Specifications

- The injury severity level probability is a mixture of logits. When all parameters β are fixed, the model reduces to the multinomial logit model.
- When β is allowed to vary, the model is not in a closed form, and the probability of crash observation n having a particular injury outcome i can be calculated through integration.
- Simulation-based maximum likelihood methods such as Halton draws are usually used.
- The choice of the density function of β depends on the nature of the coefficient and the statistical goodness of fit.
  - The lognormal distribution is useful when the coefficient is known to have the same sign for each observation.
  - Triangular and uniform distributions have the advantage of being bounded on both sides.
  - Furthermore, triangular assumes that the probability increases linearly from the beginning to the mid-range and then decreases linearly to the end.
  - A uniform distribution assumes the same probability for any value within the range.

Random Coefficient or Error Components?

- Random coefficient: $U_{nj} = \beta_n' x_{nj} + \varepsilon_{nj}$ where β_n can be decomposed into mean α and deviations μ_n such as $(\alpha' x_{nj} + \mu_n' x_{nj})$ and ε_{nj} is a random term that is iid extreme value.
- Error components: $U_{nj} = \alpha' x_{nj} + \mu_n' z_{nj} + \varepsilon_{nj}$ where x_{nj} and z_{nj} are vectors of observable variables relating to alternative j. α is a vector of fixed parameters and μ is random with zero mean, and ε_{nj} is iid extreme value. So, the random portion of utility is $(\mu_n' z_{nj} + \varepsilon_{nj})$ which can be correlated over alternatives depending on z.
- Error-component and random-coefficient specifications are formally equivalent; but a researcher thinks about the model affects the specification of the mixed logit.
- It is important to know that the mixing distribution, whether driven by random parameters or by error components, captures variance and correlations in unobserved factors. But there is a limit on how much one can learn about things that are not seen.

ML Model Coefficient Estimates

Exercise 4.3: Coefficient Estimates for ML

	B or C			K or A		
Variable	Estimate	Std. Error	Pr(>|z|)	Estimate	Std. Error	Pr(>|z|)
Intercept	−0.4882	0.0346	0.0000	−1.9324	0.0637	0.0000
Old	−0.0200	0.0587	0.7340	0.6171	0.0985	0.0000
Female	0.9975	0.0551	0.0000	0.1470	0.6700	0.8264
Alcohol	0.2647	0.1272	0.0375	−3.6985	8.5099	0.6638
Speed	0.3429	0.0513	0.0000	−0.0065	0.4962	0.9896
Snow	−0.8892	0.2158	0.0000	−10.8309	11.2638	0.3363
Dark	−0.0737	0.0616	0.2318	0.2168	0.3050	0.4771
sd. Female_K or A				1.5142	0.8969	0.0914
sd. Alcohol_K or A				−8.7659	12.5946	0.4864
sd. Speed_K/A				−1.0809	0.7895	0.1710
sd. Snow_B or C	1.8159	0.6711	0.0068			
sd. Snow_K or A				8.6369	7.9126	0.2750
sd. Dark_K or A				−0.3950	1.1620	0.7339

AIC: 18,659.75, Log-Likelihood: −9309.9, McFadden R^2: 0.030,777.

Solution

1. **Determine the density function in the R "mlogit" package**, random parameter object "rpar" contains all the relevant information about the distribution of random parameters. Currently, the normal ("n"), log-normal ("ln"), zero-censored normal ("cn"), uniform ("u") and triangular ("t") distributions are available. For illustration, normal distribution is chosen as the density function of random parameter β.
2. **Estimate the coefficients using the R "mlogit" package:**

   crash_data_mixed <- mlogit.data(data_mixed_ch4, shape = "long", choice = "INJSVR", chid.var = "ID", alt.var = "OUTCOME")

   mixed_logit <- mlogit(INJSVR ~ OLD_2 + OLD_3 + FEMALE_2 + FEMALE_3 + ALCFLAG_2 + ALCFLAG_3 + DRVRPC_SPD_2 + DRVRPC_SPD_3 + ROADCOND_SNOW_2 + ROADCOND_SNOW_3 + LGTCOND_DARK_2 + LGTCOND_DARK_3, data = crash_data_mixed, rpar = c(FEMALE_2 = 'n', FEMALE_3 = 'n', ALCFLAG_3 = 'n', DRVRPC_SPD_3 = 'n', ROADCOND_SNOW_2 = 'n', ROADCOND_SNOW_3 = 'n', LGTCOND_DARK_3 = 'n'),panel = FALSE, correlation = FALSE, R = 100, halton = NA).

   Note: rpar argument names random coefficients ('n' for a normal distribution); halton=NA means default halton draws are applied. (if interested, read "Halton Sequences for Mixed Logit" by Kenneth Train at https://eml.berkeley.edu/wp/train0899.pdf)
3. **Summarize the findings**. The ML model can account for the data heterogeneity by treating coefficients as random variables. the snowy surface parameter for truck K or A injuries is fixed (-10.830); and for severity B or C, it is normally distributed with a mean of -0.8892 and a standard deviation of 1.8159, meaning that 31% of truck crashes occurring on snowy pavement have an increased possibility of B or C injuries. It is plausible that people often drive more slowly and cautiously on snowy roads but that the slick conditions still have a tendency to cause accidents.

- In Milton et al. (2008), the application of the mixed logit model (also called the random parameters logit model) is undertaken by considering injury-severity proportions for individual roadway segments.
- For all of the random parameters, the normal distribution was found to provide the best statistical fit (among normal, lognormal, triangular and uniform).

"*the constant for the property-damage only proportion is normally distributed with mean −0.355 and standard deviation 1.776. ….This variability is likely capturing the unobserved heterogeneity in the roadway segments that could include factors such as visual noise and other physical and environmental factors. ….The average daily traffic (ADT) per lane is normally distributed with a mean 0.0403 and standard deviation 0.515. … 46.9% of the distribution is less than 0 and 53.1% is greater than 0…. a complex interaction among traffic volume, driver behavior and accident-injury severity*."

Table 2
Mixed logit estimation results for annual accident-severity proportions on roadway segments

Variable	Parameter estimate	Standard error	*t*-Statistic
Property damage only			
Constant (standard error of parameter distribution)	−0.355 (1.776)	0.182 (0.694)	−1.95 (2.56)
Average daily traffic per lane in thousands (standard error of parameter distribution)	0.0403 (0.515)	0.0190 (0.122)	2.12 (4.23)
Average annual snowfall in inches (standard error of parameter distribution)	0.0974 (0.335)	0.0418 (0.173)	2.33 (1.93)
Possible injury			
Pavement friction (scaled 0–100), fixed parameter	−0.0124	0.00293	−4.21
Percentage of trucks (standard error of parameter distribution)	−0.129 (0.1143)	0.0309 (0.0298)	−4.18 (3.84)
Injury			
Average daily truck traffic in thousands (standard error of parameter distribution)	−0.302 (0.433)	0.0716 (0.111)	−4.22 (3.90)
Number of horizontal curves per mile, fixed parameter	−0.267	0.0547	−4.89
Number of grade breaks per mile, fixed parameter	−0.0712	0.0284	−2.51
Number of interchanges per mile (standard error of parameter distribution)	−0.601 (1.441)	0.190 (0.450)	−3.17 (3.20)
Number of observations		1,280	
Restricted log-likelihood (constant only)		−24,849.51	
Log-likelihood at convergence		−21,980.66	

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

- In Milton et al. (2008), the application of the mixed logit model (also called the random parameters logit model) is undertaken by considering injury-severity proportions for individual roadway segments.

"*The percentage of trucks…had a mean of−0.129 and standard deviation 0.1143, being less than 0 for 87.1% of the roadway segments and greater than 0 for 12.9% of the segments…in a small proportion of roadway segments, the truck percentage increases the proportion of possible injury accidents, while in a majority of roadway segments, the proportion tends to decrease. Note that this variable implies that for 87.1% of roadway segments increasing truck percentages make the severity proportions more likely to be minor (property damage only) or major (injury)... 75.2% of the roadway segments negative values (an increasing number of trucks decreases the likelihood of accidents resulting in injury) and 24.8% positive values (an increasing number of trucks increases the likelihood of accidents resulting in injury).The net effect of these two truck variables points to a fairly complex picture of the effect of trucks on accident-injury severities.*"

Table 2
Mixed logit estimation results for annual accident-severity proportions on roadway segments

Variable	Parameter estimate	Standard error	*t*-Statistic
Property damage only			
Constant (standard error of parameter distribution)	−0.355 (1.776)	0.182 (0.694)	−1.95 (2.56)
Average daily traffic per lane in thousands (standard error of parameter distribution)	0.0403 (0.515)	0.0190 (0.122)	2.12 (4.23)
Average annual snowfall in inches (standard error of parameter distribution)	0.0974 (0.335)	0.0418 (0.173)	2.33 (1.93)
Possible injury			
Pavement friction (scaled 0–100), fixed parameter	−0.0124	0.00293	−4.21
Percentage of trucks (standard error of parameter distribution)	−0.129 (0.1143)	0.0309 (0.0298)	−4.18 (3.84)
Injury			
Average daily truck traffic in thousands (standard error of parameter distribution)	−0.302 (0.433)	0.0716 (0.111)	−4.22 (3.90)
Number of horizontal curves per mile, fixed parameter	−0.267	0.0547	−4.89
Number of grade breaks per mile, fixed parameter	−0.0712	0.0284	−2.51
Number of interchanges per mile (standard error of parameter distribution)	−0.601 (1.441)	0.190 (0.450)	−3.17 (3.20)
Number of observations		1,280	
Restricted log-likelihood (constant only)		−24,849.51	
Log-likelihood at convergence		−21,980.66	

Ordered Response Models

Modeling Crash Severity as an Ordered Discrete Outcome

- The primary rationale for using ordered discrete choice models for modeling crash severity is that there is an intrinsic order among injury severities, with fatality being the highest order and property damage being the lowest. Including the ordinal nature of the data in the statistical model defends the data integrity and preserves the information.
- Second, the consideration of ordered response models avoids the undesirable properties of the multinomial model such as the independence of irrelevant alternatives in the case of a multinomial logit model or a lack of closed-form likelihood in the case of a multinomial probit model.
- Third, ignoring the ordinality of the variable may cause a lack of efficiency (i.e., more parameters may be estimated than are necessary if the order is ignored).
- Although there are many positives to the ordered model, the disadvantage is that imposing restrictions on the data may not be appropriate despite the appearance of a rank. Therefore, it is important to test the validity of the ordered restriction.

Model Structure

- Ordered probability models are derived by defining an unobserved variable, Z, ($Z = \beta X + \varepsilon$) that is used as a basis for modeling the ordinal ranking of data.
- Observed ordinal data, *y*, for each observation are defined as,

$$y_n = \begin{cases} 1, if\ z_n \le \mu_1 \left(\text{PDO or no injury}\right) \\ 2, if\ \mu_1 < z_n \le \mu_2 \left(\text{injury C}\right) \\ 3, if\ \mu_2 < z_n \le \mu_3 \left(\text{injury B}\right) \\ 4, if\ \mu_3 < z_n \le \mu_4 \left(\text{injury A}\right) \\ 5, if\ \mu_4 < z_n \left(\text{K or fatal injury}\right) \end{cases} \quad (4.11)$$

where the μs are estimable thresholds, along with the parameter vector β. The model is estimated using maximum likelihood estimation

Ordered Logit/Probit

The ordinal logit/probit model applies a latent continuous variable, z_n, as a basis for modeling the ordinal nature of crash severity data, and z_n is specified as a linear function of $\boldsymbol{X_n}$:

$$z_n = \boldsymbol{\beta}'\boldsymbol{X_n} + \varepsilon_n \qquad (4.10)$$

Where $\boldsymbol{X_n}$ is a vector of explanatory variables determining the discrete ordering (i.e., injury severity) for n th crash observation, $\boldsymbol{\beta}$ is a vector of estimable parameters, and ε_n is an error term that accounts for unobserved factors influencing injury severity.

Ordered Probit Model

If ε is assumed to be **normally distributed** across observations with N(0,1), an **ordered probit model** results with the ordered selection probabilities being

$$P(y=1) = \Phi(-\beta X)$$

$$P(y=2) = \Phi(\mu_1 - \beta X) - \Phi(\beta X)$$

$$P(y=3) = \Phi(\mu_2 - \beta X) - \Phi(\mu_1 - \beta X)$$

..

$$P(y=I) = 1 - \Phi(\mu_{I-1} - \beta X) \qquad (4.12)$$

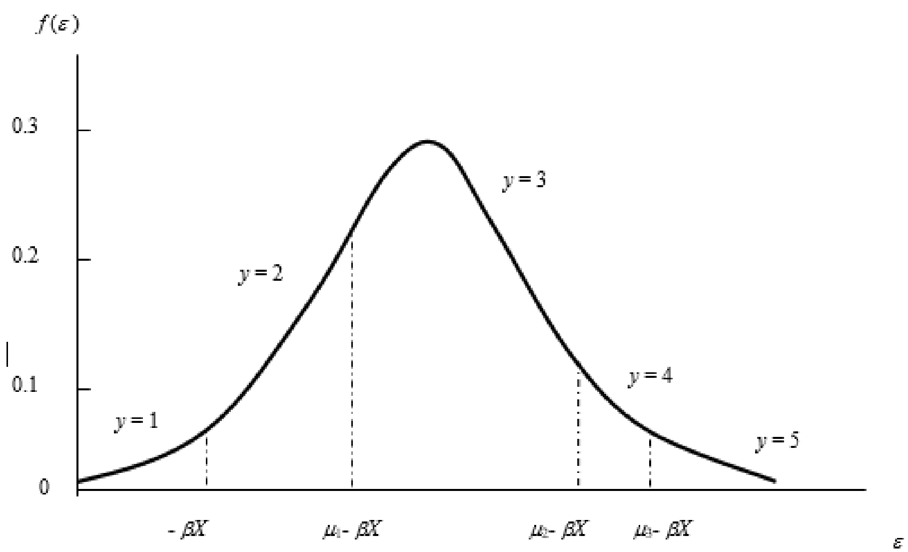


Illustration of an ordered probability model with $\mu_0 = 0$.

Where Φ (.) is the cumulative standard normal distribution. Note threshold μ_0 is set equal to 0 without loss of generality (this implies that one need only estimate I-2 thresholds.

Limitation

- The difficulty arises because the areas between the shifted thresholds may yield increasing or decreasing probabilities after shifts to the left or right, especially for the intermediate categories (i.e., y=2, y=3, and y=4).

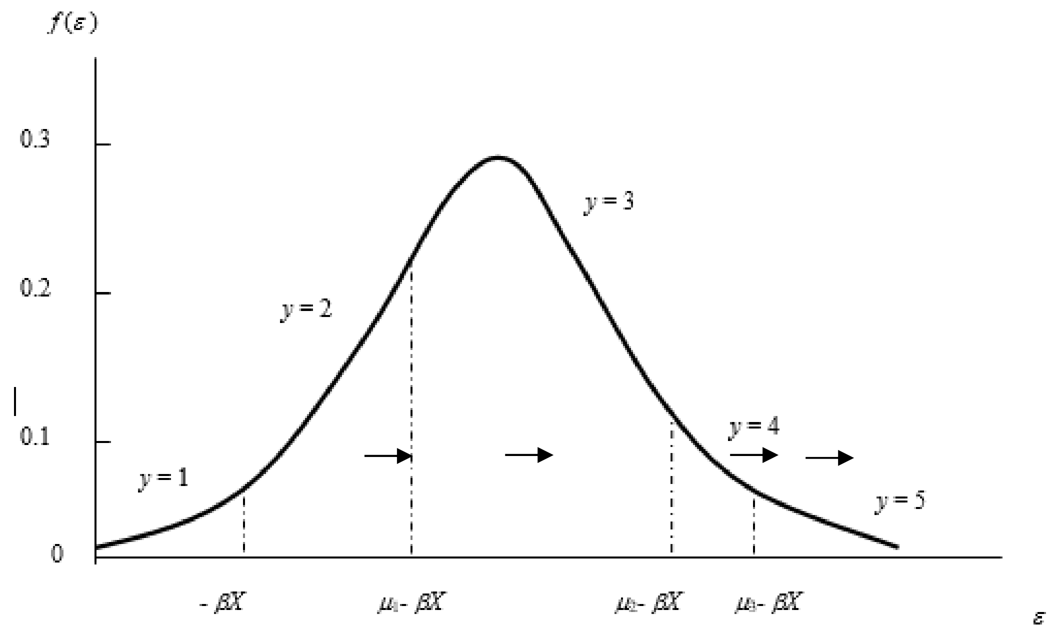


Illustration of an ordered probability model with $\mu_0 = 0$.

Limitation

- The difficulty arises because the areas between the shifted thresholds may yield increasing or decreasing probabilities after shifts to the left or right, especially for the intermediate categories (i.e., y=2, y=3, and y=4).
- **The change depends on the location of the thresholds.**
- A trade-off is inherently being made between recognizing the ordering of responses and losing the flexibility in specification offered by unordered probability models.

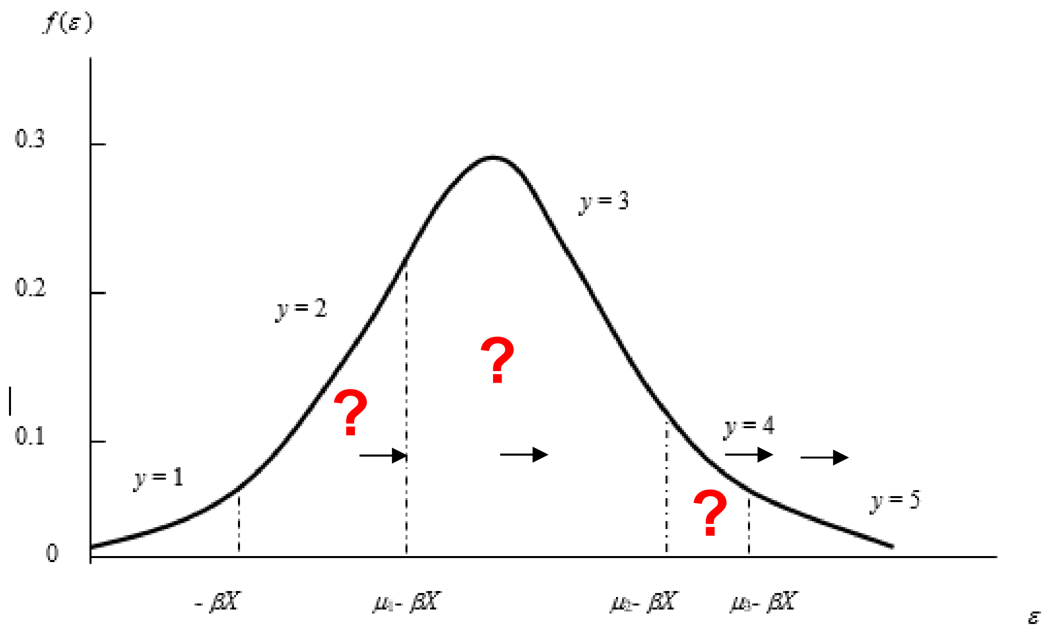


Illustration of an ordered probability model with $\mu_0 = 0$.

Estimation

$P(u = i) = \Phi(\mu_i - \beta X) - \Phi(\mu_{i-1} - \beta X)$

Where μ_i and μ_{i-1} represent the upper and lower thresholds for outcome i.

The likelihood function is:

$$L(u|\beta,\mu) = \prod_{n=1}^{N}\prod_{i=1}^{I}[\Phi(\mu_i - \beta X_n) - \Phi(\mu_{i-1} - \beta X_n)]^{\delta_{in}}$$

$$LL(u|\beta,\mu) = \sum_{n}\sum_{i}\delta_{in}LN[\Phi(\mu_i - \beta X_n) - \Phi(\mu_{i-1} - \beta X_n)]$$

where $\delta_{in} = 1$ if the observed discrete outcome for observation n is i, and zero otherwise.

Maximize the LL is subject to the constraint $0 \leq \mu_1 \leq \mu_2 \ldots \leq \mu_{I-2}$

Order Logit Model

- Ordered logit can also be conceptualized as a latent variable model.
- Let Z be a continuous random variable that depends on a set of explanatory variables X, $Z = \beta X + \varepsilon$, that is used as a basis for modeling the ordinal ranking of data.
- Similarly, we do not observe Z directly. Instead, there is a set of cut points or thresholds *us* that are used to transform Z into Y. If we assume that ε follows **a standard logistic distribution**, it follows the cumulative logit, also known as ordered or ordinal logit model.

- Recall 4.11, $y_n = \begin{cases} 1, if\ z_n \le \mu_1 \left(\text{PDO or no injury}\right) \\ 2, if\ \mu_1 < z_n \le \mu_2 \left(\text{injury C}\right) \\ 3, if\ \mu_2 < z_n \le \mu_3 \left(\text{injury B}\right) \\ 4, if\ \mu_3 < z_n \le \mu_4 \left(\text{injury A}\right) \\ 5, if\ \mu_4 < z_n \left(\text{K or fatal injury}\right) \end{cases}$

$$\Pr\left(y_n > i\right) = \Pr\left(Z_n > \mu_i\right) = \Pr\left(\varepsilon_n > \mu_i - \mathbf{x}_n'\boldsymbol{\beta}\right)$$

$$= \frac{1}{1+\exp\left(\mu_i - \mathbf{x}_n'\boldsymbol{\beta}\right)} = \frac{\exp\left(\mathbf{x}_n'\boldsymbol{\beta} - \mu_i\right)}{1+\exp\left(\mathbf{x}_n'\boldsymbol{\beta} - \mu_i\right)} \qquad (4.15)$$

Where ε_n follows logistic distribution whose CDF is: $F\left(\varepsilon_n\right) = \dfrac{\exp\left(\varepsilon_n\right)}{1+\exp\left(\varepsilon_n\right)}$

So, $Pr(\varepsilon_n > \mu_i - \boldsymbol{x}_n'\boldsymbol{\beta}) = 1 - F(\varepsilon_n) = \frac{1}{1+exp(\mu_i - \boldsymbol{x}_n'\boldsymbol{\beta})}$.

- Model can be specified as a set of I-1 equations

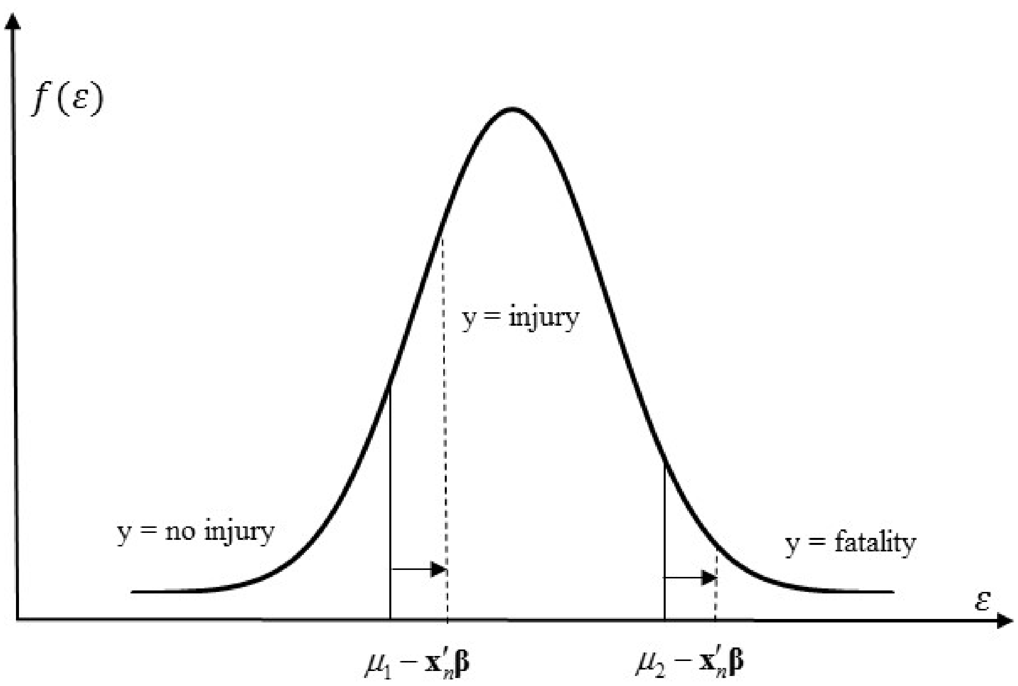
$$log\left(\frac{P_{ni}}{1-P_{ni}}\right) = \boldsymbol{x}_{\boldsymbol{n}}'\boldsymbol{\beta} - \mu_i \quad \text{i=1, …, I-1} \qquad (4.13)$$

Proportional Odds Assumption

- The fact that you can calculate odds ratios highlights a key assumption of ordered logit:
  - "Proportional odds assumption"
  - Also known as the "parallel regression assumption"
    - Which also applies to ordered probit
- Model assumes that variable effects on the odds of lower vs. higher outcomes are consistent; or regression parameters have to be the same for different response outcomes.
- If this assumption doesn't seem reasonable, consider multinomial logit, generalized ordered logistic and proportional odds model.

Airbag Example for Proportional Odds

Consider a model of three injury levels - no injury, injury, and fatality. Suppose that one of the factors is airbag. A negative parameter of the airbag indicator (1 if it was deployed and zero otherwise) becomes greater and hence, shifts values to the right on the X-axle. Thus, the model constrains the effect of the seatbelt to simultaneously decrease the probability of a fatality and increase the no injury probability. But we know for a fact that the activation of an airbag may cause injury and/or decrease no injury; but unfortunately, ordered models cannot account for this bi-directional possibility because the shift in thresholds is constrained to move in the same direction.



OP Model Coefficient Estimates

Exercise 4.4: Coefficient Estimates for OP

	Ordinal probit model			Ordinal logit model		
Variable	Estimate	Std. Error	Pr(>|z|)	Estimate	Std. Error	Pr(>|z|)
Young	0.0963	0.0309	0.0019	0.1514	0.0521	0.0037
Old	0.1285	0.0307	0.0000	0.1961	0.0520	0.0002
Female	0.3398	0.0270	0.0000	0.6116	0.0454	0.0000
Alcohol	0.2977	0.0623	0.0000	0.4945	0.1082	0.0000
Drugs	1.0187	0.1459	0.0000	1.7663	0.2455	0.0000
Safety constraints	−0.4321	0.1642	0.0085	−0.7798	0.2881	0.0068
Speed	0.3090	0.0288	0.0000	0.5299	0.0488	0.0000
Rule violation	0.3329	0.0315	0.0000	0.5416	0.0534	0.0000
Reckless behavior	0.1569	0.0264	0.0000	0.2644	0.0446	0.0000
Signal	0.3353	0.0799	0.0000	0.5744	0.1342	0.0000
Two-way	0.5364	0.0830	0.0000	0.9019	0.1403	0.0000
None	0.3190	0.0752	0.0000	0.5144	0.1265	0.0000
Total units	0.1635	0.0120	0.0000	0.2867	0.0210	0.0000
Snow	−0.4450	0.0400	0.0000	−0.7666	0.0682	0.0000
Ice	−0.3358	0.0577	0.0000	−0.5727	0.0974	0.0000
Wet	−0.0615	0.0357	0.0844	−0.0906	0.0596	0.1288
Dark	0.1237	0.0319	0.0001	0.2108	0.0540	0.0001
Threshold coefficients						
1|2	0.5518	0.1811		0.8867	0.3154	
2|3	1.8791	0.1818		3.1692	0.3171	
AIC	18,072.64			18,036.83		

Solution

- This exercise uses the same dataset as Exercise 4.1. In this exercise, an ordinal probit and an ordinal logistic regression model are respectively applied in order to recognize the ordinality of injury level, the dependent variable.

- First, determine the functional form: Eq. 4.12 for the ordinal probit model and Eq. 4.15 for the ordinal logistic model. In both equations, the μs are estimable thresholds, along with the parameter vector β.

- Second, estimate the coefficients using the R "ordinal" package:

```
crash_data_ordinal <- data_model_ch5
op_model <- clm(as.factor(INJSVR) ~ YOUNG + OLD + FEMALE + ALCFLAG + DRUGFLAG + SAFETY +
DRVRPC_SPD + DRVRPC_RULEVIO + DRVRPC_RECK
        + TRFCONT_SIGNAL + TRFCONT_2WAY + TRFCONT_NONE
        + TOTUNIT + ROADCOND_SNOW + ROADCOND_ICE + ROADCOND_WET + LGTCOND_DARK,
        data = crash_data_ordinal, link = "probit")
```

- Note that the response (INJSVR) should be a factor, which will be interpreted as an ordinal response with levels ordered as in the factor. Replace "probit" with "logit" if you want to run an ordinal logit model. Other distribution options are: "cloglog", "loglog", "cauchit", "Aranda-Ordaz", "log-gamma".

- Third, present the model results of the coefficients and finally, summarize the findings.

Generalized Ordered Logistic and Proportional Odds Model

- A generalized ordered logistic model (gologit) provides results similar to those that result from running a series of binary logistic regressions/ cumulative logit models.
- The ordered logit model is a special case of the gologit model where the coefficients β are the same for each category.
- A gologit model and an MNL model, whose variables are freed from the proportional odds constraint, both generate many more parameters than an ordered logit model.
- The partial proportional odds model (PPO) is in between, as some of the coefficients β are the same for all categories and others may differ.
- A PPO model allows for the parallel lines/ proportional odds assumption to be relaxed for those variables that violate the assumption.

Generalized Ordered Logistic Model

In the gologit model, the probability of crash injury for a given crash can be specified as (I-1) set of equations:

$$Pr(y_n > i) = \frac{exp(\mathbf{x}_n^{'}\boldsymbol{\beta}_{\boldsymbol{i}} - \mu_i)}{1 + exp(\mathbf{x}_n^{'}\boldsymbol{\beta}_{\boldsymbol{i}} - \mu_i)}, i = 1, \ldots (I - 1) \qquad (4.16)$$

Where μ_i is the cut-off point for the ith cumulative logit. Note that Equation 4.16 is different from Equation 4.14 in that β_i is a single set of coefficients that vary by category i.

Proportional Odds Model

In the PPO model formulation, it is assumed that some explanatory variables may satisfy the proportional odds assumption while some may not. The cumulative probabilities in the PPO model are calculated as follows:

$$Pr(y_n > i) = \frac{exp(\boldsymbol{x}'_n\boldsymbol{\beta}+\boldsymbol{T}'_n\boldsymbol{\gamma}_{\boldsymbol{i}}-\mu_i)}{1+exp(\boldsymbol{x}'_n\boldsymbol{\beta}+\boldsymbol{T}'_n\boldsymbol{\gamma}_{\boldsymbol{i}}-\mu_i)}, i = 1, \dots (I-1) \qquad (4.17)$$

Where x_n is a (p×1) vector of independent variables of crash n, β is a vector of regression coefficients, and each independent variable has a β coefficient. T_n is a (q×1) vector (q≤p) containing the values of crash n on the subset of p explanatory variables for which the proportional odds assumption is not assumed, and γ_i is a (q×1) vector of regression coefficients. So, γ_i represents deviation from the proportionality β_i and is an increment associated only with the ith cumulative logit, i=1,⋯,(I-1).

Proportional Odds Model (Cont'd)

An alternative but simplified way to think about the PPO model is to have two sets of explanatory variables: x_1, the coefficients of which remain the same for all injury severities and x_2, the coefficients of which vary across injury severities.

$$\Pr\left(y_n > i\right) = \frac{\exp\left(\mathbf{x}_1'\boldsymbol{\beta}_1 + \mathbf{x}_2'\boldsymbol{\beta}_{2i} - \mu_i\right)}{1 + \exp\left(\mathbf{x}_1'\boldsymbol{\beta}_1 + \mathbf{x}_2'\boldsymbol{\beta}_{2i} - \mu_i\right)} \quad (4.18)$$

Sequential Logistic/Probit Regression Model

- Although the generalized ordered logit model relaxes the proportional odds assumption by allowing some or all of the parameters to vary by severity levels, the set of explanatory variables is invariant over all severity levels.
- The sequential logit/probit regression model should be considered when the difference in the set of explanatory variables at each severity level is important.
- Sequential logit/probit regression allows different regression parameters for different severity levels. A sequential logit/probit model supposes (I-1) latent variables given as (I-1) sets of equations.
- Sequential logistic regression not only accounts for the inherent order of dependent variables but also allows **different sets of regression parameters** to be independently considered in the model specification.

Sequential logistic/probit regression model

$$
\begin{aligned}
z_{n1} &= \boldsymbol{\alpha}_1 + \mathbf{x}'_n \boldsymbol{\beta}_1 + \varepsilon_{n1} \\
z_{n2} &= \boldsymbol{\alpha}_2 + \mathbf{x}'_n \boldsymbol{\beta}_2 + \varepsilon_{n2} \\
&\vdots \\
z_{n,I-1} &= \boldsymbol{\alpha}_{I-1} + \mathbf{x}'_n \boldsymbol{\beta}_{I-1} + \varepsilon_{n,I-1}
\end{aligned} \tag{4.19}
$$

where z_{ni} is a continuous latent variable that determines whether the injury severity is observed as i or higher, β_i's are the vectors of estimated parameters, and ε_{ni}'s are error terms that are independent of x_n.

Sequential logistic/probit regression model structure

- The sequential model is a type of hierarchical model where lower stages mean lower injury severity.
- For example, stage 1 of the KABCO scale may be KABC versus O; stage 2 may be KAB versus C and stage 3 may be KA versus B. This change in definition matters when explaining the model results. Moreover, the hierarchical structure can be arranged from low to high or from high to low, which can also be called "forward" or "backward."
- It is important to know that the sequential model uses a subpopulation of the data to estimate the variant set of β_i. The subpopulation decreases as the stages progresses forward or backward. In the forward format, all data are used in the first stage to estimate β_1, but only the crashes with injury type C or higher are used in the second stage to estimate β_2. Crashes with injury type B or higher are used in the second stage to estimate β_3.

Sequential logistic/probit regression model applications

- Jung et al. (2010) applied the sequential logit model to assess the effects of rainfall on the severity of single-vehicle crashes on Wisconsin interstate highways.
  - The sequential logit regression model outperformed the ordinal logit regression model in predicting crash severity levels in rainy weather when comparing goodness of fit, parameter significance, and prediction accuracies.
  - The sequential logit model identified that stronger rainfall intensity significantly increases the likelihood of fatal and incapacitating injury crash severity, while this was not captured in the ordered logit model.
- Yamamoto et al. (2008) also reported superior performance and unbiased parameter estimates with sequential binary models as compared with traditional ordered probit models, even when underreporting was a concern.

Reference: Soyoung Jung, Xiao Qin, David Noyce (2012). Injury Severity of Multi-Vehicle Crash in Rainy Weather, ASCE, Journal of Transportation Engineering, 138(1), pp. 1-12.

Forward format:

Stage 1: $(1 - P)/P1 = \text{EXP}(\alpha_1 + \beta X1) = h1$

Stage 2: $P3/P2 = \text{EXP}(\alpha_2 + \beta X2) = h2$

Stage 1

Analysis of maximum likelihood estimates	Parameter	Estimate	Standard error
	Intercept 1	1.8666	0.7135
	Safety belt	−1.6353	0.6835
	Median-related crash	0.9213	0.4105
	DRV 2	−1.0483	0.3133
	SDV	−0.0595	0.0224
	DRV 4	−1.1587	0.3734
	SDV*DRV 4	0.0654	0.0315

Stage 2

Analysis of maximum likelihood estimates	Parameter	Estimate	Standard error
	Intercept	−0.9280	0.2959
	DCD 1	0.9487	0.3882
	OCC	−0.0506	0.0244
	Curve to the left	1.4631	0.9264

Backward format:

Stage 1: $P3/(1 - P3) = \text{EXP}(\alpha_1 + \beta X1) = I1$

Stage 2: $P2/P1 = \text{EXP}(\alpha_2 + \beta X2) = I2$

Stage 1

Analysis of maximum likelihood estimates	Parameter	Estimate	Standard error
	Intercept	−0.3052	0.1747
	DRV 2	−0.8120	0.3399
	Median-related crash	1.3261	0.4231
	Passenger car	−0.6111	0.2071
	Monday/Friday	−0.4691	0.2168

Stage 2

Analysis of maximum likelihood estimates	Parameter	Estimate	Standard error
	Intercept 1	−2.9251	0.3299
	DCD 1	0.9358	0.3321
	DRV 1	1.1142	0.3543
	DRV 3	2.0090	0.5569
	Wind speed	−0.0544	0.0246

Where
P1=probability of PDO;
P2=probability of possible injury; and
P3 = probability of fatal/incapacitating/non-incapacitating injury

DCD is defined as the minimum safe stopping distance (SSD)
OCC: Average 5-min OCC (%)

Model Interpretation

- To properly interpret model results, we need to be wary of the data formats as they can be structured differently because of different methods.
- The dependent variable can be treated as individual categories, categories higher than level i, or categories lower than level i.
- Independent variables can be continuous, indicator (1 or 0) or categorical.
- Categorical variables should be converted to dummy variables, with a dummy variable assigned to each distinct value of the original categories.
- The coefficient of a dummy variable can be interpreted as the log-odds for that particular value of dummy minus the log-odds for the base value which is 0 (e.g., the odds of being injured when drinking and driving is 10 times of someone who is sober).

Model Interpretation (Cont'd)

- The key concepts of marginal effect and elasticity are fundamental to understanding model estimates. The marginal effect is the unit-level change in y for a single-unit increase in x if x is a continuous variable.
- In a simple linear regression, the regression coefficient of x is the marginal effect, $\frac{\partial y}{\partial x_k} = \beta_k$.
- Due to the nonlinear feature of logit models, the marginal effect of any continuous independent variable is: $\frac{\partial p_i}{\partial x_{ki}} = \beta_{ki} p_i (1 - p_i)$.
- Such marginal effects are called instantaneous rates of change because they are computed for a variable while holding all other variables as constant.

Model Interpretation (Cont'd)

- Elasticity can be used to measure the magnitude of the impact of specific variables on the injury-outcome probabilities.
- For a continuous variable, elasticity is the % change in y given a 1% increase in x. It is computed from the partial derivative with respect to the continuous variable of each observation n and the formula is $E_{x_{ik}}^{P(i)} = \left[1 - P\left(i\right)\right] \beta_{ki} x_{ki}$.
- For indicator or dummy variables (those variables taking on values of 0 or 1), a pseudo elasticity of an indicator variable with respect to an injury severity category represents the percent change in the probability of that injury severity category when the variable is changed from zero to one.

References

1. Brant, R., 1990. Assessing proportionality in the proportional odds model for ordinal logistic regression. Biometrics, 46, 1171–1178. doi:10.2307/2532457.
2. Chen, F., S. Chen, 2011, Injury severities of truck drivers in single- and multi-vehicle accidents on rural highways, Accident Analysis and Prevention, vol. 43, no. 5, 1677-1688.
3. Christensen, R. H. B. 2018. Cumulative link models for ordinal regression with the R package Ordinal. Submitted in J. Stat. Software.
4. Greene, W. 2000. Econometric Analysis, 4th Edition, Prentice Hall, Upper Saddle River, NJ.
5. Greene, W. H., and D. A. Hensher. Modeling ordered choices: A primer. Cambridge University Press, 2010.
6. Gumbel, E.J. (1958) Statistics of Extremes, Columbia University Press, New York.
7. Hausman, J.A. and D. McFadden (1984), A Specification Test for the Multinomial Logit Model, Econometrica, 52, pp.1219–1240.
8. Jung, S. Y., X. Qin, and D. A. Noyce. Rainfall effect on single-vehicle crash severities using polychotomous response models. Accident Analysis and Prevention, Vol. 42, No. 1, 2010, pp. 213-224.
9. McCullagh, P. (1980). Regression models for ordinal data. Journal of the Royal Statistical Society: Series B (Methodological), 42(2), 109-127.
10. McFadden, D. (1981) Econometric Models of Probabilistic Choice. In: Manski, C. and McFadden, D., Eds., Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, 198-272.
11. McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. Journal of applied Econometrics, 15(5), 447-470.
12. Milton, J. C., V. N. Shankar, and F. Mannering. Highway accident severities and the mixed logit model: An exploratory empirical analysis. Accident Analysis and Prevention, Vol. 40, No. 1, 2008, pp. 260-266.
13. Mujalli, R. O., López, G., & Garach, L. (2016). Bayes classifiers for imbalanced traffic accidents datasets. Accident Analysis & Prevention, 88, 37-51.
14. Peterson, B., and F. E. Harrell Jr. Partial proportional odds models for ordinal response variables. Applied statistics, 1990, pp. 205-217.
15. Qin, X., Wang, K., & Cutler, C. E. (2013). Analysis of crash severity based on vehicle damage and occupant injuries. Transportation research record, 2386(1), 95-102.
16. Qin, X., Wang, K., & Cutler, C. E. (2013). Logistic regression models of the safety of large
17. trucks. Transportation research record, 2392(1), 1-10.
18. Savolainen, P. T., F. Mannering, D. Lord, and M. A. Quddus. The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. Accident Analysis and Prevention, Vol. 43, No. 5, 2011, pp. 1666-1676.
19. Savolainen, P., and F. Mannering. Probabilistic models of motorcyclists' injury severities in single- and multi-vehicle crashes. Accident Analysis and Prevention, Vol. 39, No. 5, 2007, pp. 955-963.
20. Shankar, V., F. Mannering, and W. Barfield. Statistical analysis of accident severity on rural freeways. Accident Analysis and Prevention, Vol. 28, No. 3, 1996, pp. 391-401.
21. Train, K. Discrete choice methods with simulation (Second Edition). Cambridge university press, 2009.
22. Wang, X. S., and M. Abdel-Aty. Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. Accident Analysis and Prevention, Vol. 40, No. 5, 2008, pp. 1674-1682.
23. Washington, S. P., Karlaftis, M. G., & Mannering, F. (2020). Statistical and econometric methods for transportation data analysis. 3rd Edition, Chapman and Hall/CRC.
24. Williams (2016) Understanding and interpreting generalized ordered logit models, The Journal of Mathematical Sociology, 40:1, 7-20, DOI: 10.1080/0022250X.2015.1112384.
25. Williams, R. Generalized ordered logit/partial proportional adds models for ordinal dependent variables. The Stata Journal, vol. 6, no.1, 2006, pp.58-82.).
26. Yamamoto, T., Hashiji, Junpei, Shankar, V., (2008). "Underreporting in traffic accident data, bias in parameters and the structure of injury severity models", Accident Analysis Prevention 40, 1320–1329.
27. Yasmin, S., and N. Eluru. Evaluating alternate discrete outcome frameworks for modeling crash injury severity. Accident Analysis and Prevention, Vol. 59, 2013, pp. 506-521.