Phd Program in Transportation

Transport Demand Modeling

Filipe Moura

Ordered Discrete Choice Models (ODCM)

Uses of ordered discrete choice models

- In many transportation applications discrete data are ordered, for instance when respondents are asked for:
  - quantitative ratings
    - for example, on a scale from 1 to 10 rate the following)
  - Ordered opinions / satisfaction levels
    - for example, you disagree, are neutral, or agree
  - Categorical frequency data
    - for example, property damage only crash, light injury crash, serious injury crash and fatal crash.
- An application of the standard or nested multinomial discrete models presented earlier d**oes not account for the ordinal nature of the discrete data** and thus the ordering information is lost.

Basic structure of an OCDM

- To address the problem of ordered discrete data, ordered probability models have been developed.
- OCDM have the following structure
  - **Dependent variable** => Ordered outcome Y (“*lhs*”)
  - **Latent (unobserved) variable** => Used to determine thresholds (or cut off levels) between possible Y outcomes (*intermediate calculation*)
  - **Independent variables** => Observable variables that explain the likelihood of belonging to higher or lower categories of outcome (“rhs” ⇔ *utility function*)
- OCDM can either be Probit or Logit
  - Probit => error terms follow a standard normal CDF

$$f(\varepsilon_i) = \frac{\exp(-\varepsilon_i^2/2)}{\sqrt{2\pi}}, \quad -\infty < \varepsilon_i < +\infty$$

  - Logit => error terms follow a logistic CDF

$$f(\varepsilon_i) = \frac{\exp(\varepsilon_i)}{[1+\exp(\varepsilon_i)]^2}, \quad -\infty < \varepsilon_i < +\infty$$

Latent (or unobserved) variable - z

- The **latent (unobserved) variable** z is used as a basis for modeling the ordinal ranking of data (also denoted as ***Y****)
- The value on the observed variable Y depends on whether or not you **have crossed a particular threshold**
- For example, it might be that:
  - if your score on the unobserved latent variable z was 37 or less, your score on Y would be 1;
  - if your z score was between 37 and 53, Y would equal 2; and
  - if your z score was above 53, Y would equal 3.
- Put another way, you can think of Y as being a collapsed version of z
  - z can take on an infinite range of values which might then be collapsed into j categories of Y.

Latent (or unobserved) variable - z

- z is typically specified as a linear function for each observation:

$$z = \beta X + \varepsilon$$

where ***X*** is a vector of variables determining the discrete ordering for observation *n*

β is a vector of estimable parameters

ε is a random disturbance

- Observed ordinal data, *y*, for each observation *n* are defined as:

$$
\begin{array}{ll}
y = 1 & \text{if } z \leq \mu_0 \\
y = 2 & \text{if } \mu_0 < z \leq \mu_1 \\
y = 3 & \text{if } \mu_1 < z \leq \mu_2 \\
y = \ldots & \\
y = I & \text{if } z \geq \mu_{I-2}
\end{array}
$$

where μ_ι **are:**

- **estimable parameters** (referred to as thresholds/cut offs) that define *y* (i.e., integer ordering)
- **estimated jointly with the model parameters** (β)
- ***I*** **is the highest integer ordered response** (e.g., 'very satisfied' or 'very unsatisfied').

Estimation of Probit OCDM (I)

- The estimation problem becomes determining the probability of *I* specific ordered responses for each observation *n*
- The determination depends on the CDF function
  - If ε is assumed to be **normally** distributed across observations with mean = 0 and variance = 1 then it is a **Probit OCDM**, where

$$\Phi(\mu) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} EXP\left[-\frac{1}{2}w^2\right] dw$$

  - The probability of belonging to each level is determined by

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

, where μ_0 is set equal to zero

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

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

..

..

..

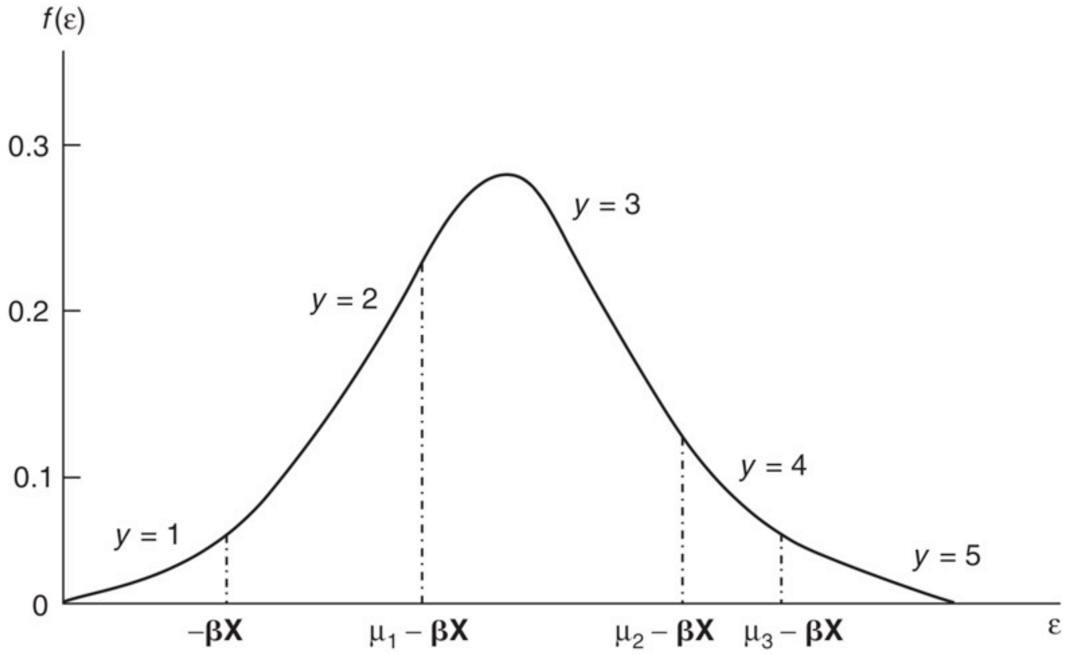
$$P(y=I) = 1 - \Phi(\mu_{I-2} - \beta X),$$

Estimation of Probit OCDM (II)

- The generic equation for the estimation of the probability of belonging to each level is

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

, where μ_i and μ_{i+1} represent the upper and lower thresholds for outcome *i.*



Estimation of a Probit OCDM (III)

- The likelihood function is (over the population of *N* observations)

$$L(y \mid \boldsymbol{\beta}, \mu) = \prod_{n=1}^{N} \prod_{i=1}^{I} \left[\Phi\left(\mu_i - \boldsymbol{\beta} \mathbf{X}_n\right) - \Phi\left(\mu_{i+1} - \boldsymbol{\beta} \mathbf{X}_n\right) \right]^{\delta_{in}}$$

, where δ_{in} is equal to 1 if the observed discrete outcome for observation n is *I*, and zero otherwise.

- This equation leads to a log-likelihood of

$$LL = \sum_{n=1}^{N} \sum_{i=1}^{I} \delta_{in} LN\left[\Phi\left(\mu_i - \boldsymbol{\beta} \mathbf{X}_n\right) - \Phi\left(\mu_{i+1} - \boldsymbol{\beta} \mathbf{X}_n\right) \right]$$

- Maximizing this log likelihood function is subject to the constraint

$$0 \leq \mu_1 \leq \mu_2 \ldots \leq \mu_{I-2}$$

Estimation of Logit OCDM (II)

- If ε is assumed to be **logistically** distributed across observations with mean = 0 and variance = 1 then it is a **Logit OCDM**

$$\Lambda(\mu) = \frac{e^{\varepsilon}}{1 + e^{\varepsilon}}$$

- Formulas for probability estimates are

$$P(Y_i > j) = \frac{\exp(X_i\beta - \kappa_j)}{1 + [\exp(X_i\beta - \kappa_j)]},\ \mathrm{j} = 1, 2, \ldots, \mathrm{M} - 1, \text{ which implies}$$

$$P(Y_i = 1) = 1 - \frac{\exp(X_i\beta - \kappa_1)}{1 + [\exp(X_i\beta - \kappa_1)]}$$

$$P(Y_i = j) = \frac{\exp(X_i\beta - \kappa_{j-1})}{1 + [\exp(X_i\beta - \kappa_{j-1})]} - \frac{\exp(X_i\beta - \kappa_j)}{1 + [\exp(X_i\beta - \kappa_j)]} \quad j = 2, \ldots, M - 1$$

$$P(Y_i = M) = \frac{\exp(X_i\beta - \kappa_{M-1})}{1 + [\exp(X_i\beta - \kappa_{M-1})]}$$

Estimation of Logit OCDM (II)

In the case of M = 3, these equations simplify to

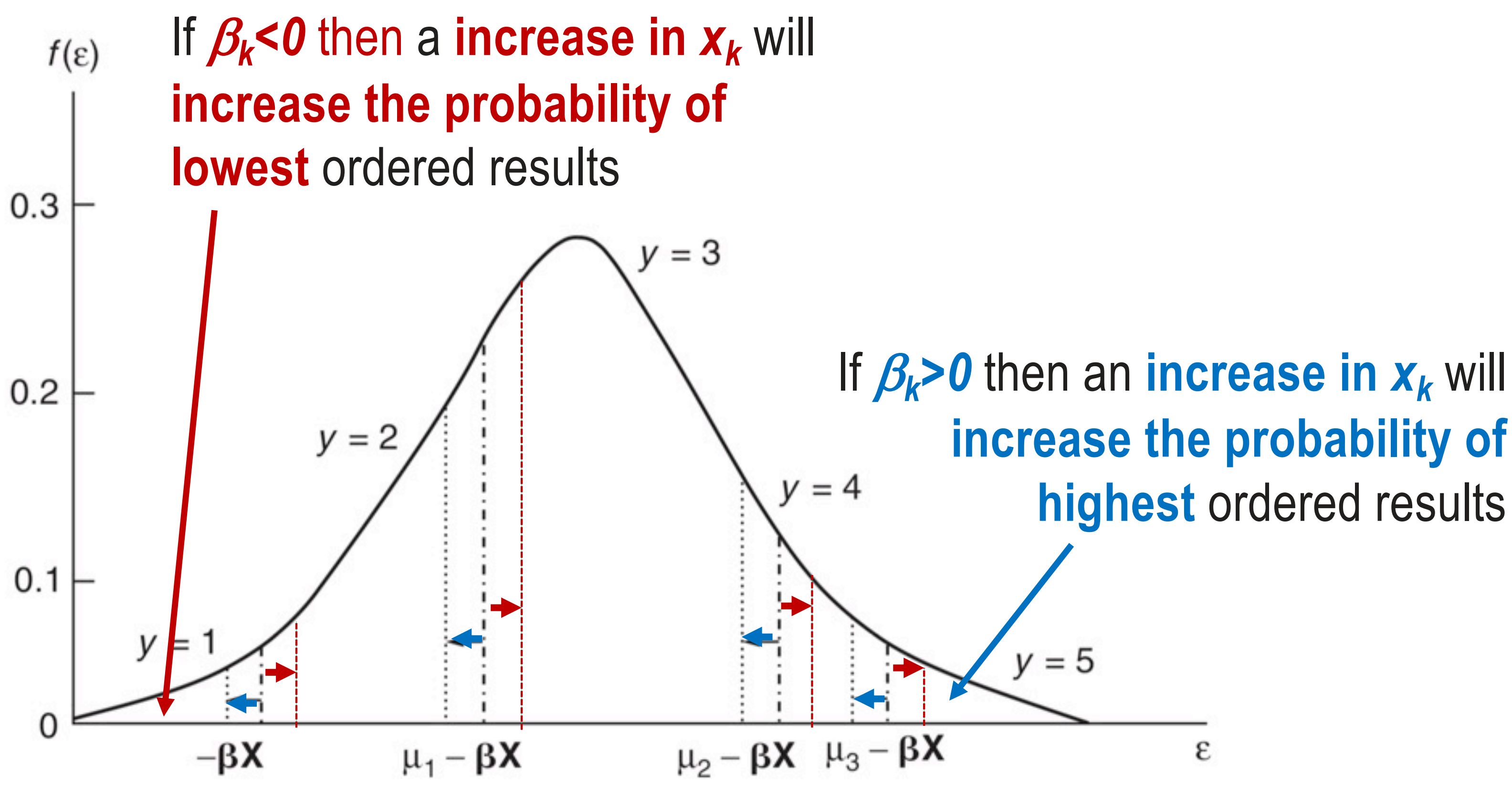
$$P(Y=1)=\frac{1}{1+\exp(Z_i-\kappa_1)}$$

$$P(Y=2)=\frac{1}{1+\exp(Z_i-\kappa_2)}-\frac{1}{1+\exp(Z_i-\kappa_1)}$$

$$P(Y=3)=1-\frac{1}{1+\exp(Z_i-\kappa_2)}$$

Effect of βX ordered probability models (I)

- How do independent variables X impact outcome results in ordered Probit or Logit models?



- The problem with ordered probability models is associated with the **interpretation of intermediate categories**

Effect of βX ordered probability models (II)

- Example
  - Suppose the following five categories in the previous slide
    - y = 1 ⇔ disagree strongly
    - y = 2 ⇔ disagree
    - y = 3 ⇔ neutral
    - y = 4 ⇔ agree
    - y = 5 ⇔ agree strongly
  - It is tempting to interpret a positive β_k as implying that an **increase x_k** will **increase the likelihood of agreeing.**
  - This conclusion is incorrect because of the ambiguity in the interior category probabilities.
  - The correct interpretation is that an **increase in x_k increases the likelihood of agreeing strongly** and **decreases the likelihood of disagreeing strongly**.

Marginal effects

- **Marginal effects** are computed for **each category**
- These marginal effects provide the **direction of the probability** for each category
- It is calculated as

$$P(y = i)/\partial X = [\phi(\mu_{i-1} - \boldsymbol{\beta}\mathbf{X}) - \phi(\mu_i - \boldsymbol{\beta}\mathbf{X})]\boldsymbol{\beta}$$

, where $\Phi(.)$ is the standard normal density

Example of marginal effects: Health status (fair, good, excelente) – I

Health status	Ordered logit coefficients	Ordered probit coefficients
Age	-0.03*	-0.01*
Income	0.28*	0.16*
Number of diseases	-0.05*	-0.03*
Intercept cut1	-1.39	-0.79
Intercept cut2	0.95	0.54

- Coefficient interpretation: the health status is better (from fair to good to excellent categories) with lower age, higher income and lower number of diseases.
- The thresholds/intercept parameters are significantly different from each other so the three categories should not be combined into one (or less than 3)
- The logit and probit ordered model coefficients differ by a scale factor (and therefore we cannot interpret the magnitude of the coefficients – marginal effects should be calculated).

Example of marginal effects: Health status (fair, good, excelente) – II

Health status	Ordered logit marginal effects for fair health status	Ordered logit marginal effects for good health status	Ordered logit marginal effects for excellent health status
Age	0.002*	0.005*	-0.007*
Income	-0.02*	-0.05*	0.07*
Number of diseases	0.003*	0.009*	-0.01*

- Marginal effects interpretation: one unit increase in income is associated with being **2% less likely** to be in “fair” health status, **5% less likely** to be “good health status” and **7% more likely** to be in an “Excellent” health status.
- The marginal effects sum up to zero.
- The marginal effects for the Probit model (not reported here) are like those of the Logit model.

Ordered versus unordered models (I)

- Example:
  - Severity of vehicle accidents: property damage only, injury, and fatality
  - Suppose that one key factor X explaining the severity level is the **deployment of airbag**
  - **Ordered model's estimates:**
    - Increase the probability of a fatality (and decrease the probability of property damage only); or
    - Decrease the probability of fatality (and increase the probability of property damage only)
  - **However,** the deployment of an air bag decreases fatalities but increases the number of minor injuries (from air bag deployment)

Ordered versus unordered models (II)

- Example:
  - **Unordered model's estimates** (e.g., MNL):
    - The estimation would result in the parameter for the air bag deployment variable having a **negative value** in the severity function for the **fatality** outcome; and
    - **Also a negative value** for the **property damage only** outcome (with the injury outcome having an increased probability in the presence of an air bag deployment).
  - In this example, an ordered probability model is not appropriate because it **does not have the flexibility to explicitly control interior category probabilities**.
  - A trade-off is inherently being made between **recognizing the ordering of responses** and **losing the flexibility** in specification **offered by unordered** probability models.

Elasticities

- Elasticities are common in transportation studies

$$\varepsilon_{i,k} = \frac{\partial \ln \text{Prob}(y_i = 1|\mathbf{x}_i)}{\partial \ln x_{i,k}}$$
$$= \frac{\partial \text{Prob}(y_i = 1|\mathbf{x}_i)}{\partial x_{i,k}} \frac{x_{i,k}}{\text{Prob}(y_i = 1|\mathbf{x}_i)}$$

- Because they are a ratio of percentage change, elasticities cannot be computed for dummy variables.
  - For these semi-elasticities (or pseudo-elasticities) should be computed (the denominator is equal to 1).

$$e_{i,k} = \frac{\text{Prob}(y_i = j|\mathbf{x}_i, d_i = 1) - \text{Prob}(y_i = j|\mathbf{x}_i, d_i = 0)}{\frac{1}{2}\left[\text{Prob}(y_i = j|\mathbf{x}_i, d_i = 1) + \text{Prob}(y_i = j|\mathbf{x}_i, d_i = 0)\right]}$$

  - The denominator computation removes the asymmetry in the computation (not dependent on whether the change is from $d_i = 1$ to *0* or from *0* to *1*).

Likelihood ratio test

- The likelihood ratio test is used to test if the model is a statistical improvement over a base model (null or restricted model).
- The test statistic is simply twice the difference between the log likelihoods for the null and alternative models.

$$LR = 2(LL_{null} - LL_{estimated}) \sim \chi^2_{(n^{\underline{o}}\ new\ parameters\ in\ the\ estimated\ model)}$$

Likelihood ratio test

- The likelihood ratio test provides a more convenient approach for testing homogeneity of strata in the data.
  - e.g. Does it make sense to build different models for both genders, separately?

$$LR = 2[\Sigma_{g=groups} \log L_g - \log L_{pooled}]$$

- The statistic has a limiting chi squared distribution with degrees of freedom equal to g-1 times the number of parameters in the model (where *g* refers to the number of groups).
- The null hypothesis states that the same ordered choice model applies to both strata.

Fit measures

- Pseudo R^2

$$R_{Pseudo}{}^{2} = 1 - \log L_{Model} / \log L_{No\ Model}$$

- Adjusted Pseudo R^2

$$Adjusted\ R_{Pseudo}{}^{2} = 1 - [\log L_{Model} - M] / \log L_{No\ Model}$$

, where M is the number of parameters in the model

Measures of fit based on predictions

$$Count\ R^2 = \frac{Number\ of\ Correct\ Predictions}{n}$$

, where n is the number of observations

$$Adjusted\ Count\ R^2 = \frac{Number\ of\ Correct\ Predictions - n_j^*}{n - n_j^*}$$

, where n_j^* is the count of the most frequent outcome

Fit measures

- Fit measures could be used to compare models to each other, not only to baseline (null) models.
- The following measures are not normalized to the unit interval, but are based on the log likelihood function

$$\textit{Log Akaike Information Criterion} = AIC = (-2\log L + 2M)/n,$$

$$\textit{Finite Sample AIC} = AIC_{FS} = AIC + 2M(M+1)/(n-M-1),$$

$$\textit{Bayes Information Criterion} = BIC = (-2\log L + M/\log n)/n,$$

$$\textit{Hannan-Quinn IC} = HQIC = (-2\log L + 2M\log\log n)/n.$$

, where *M* is the number of parameters,

n is the sample size

- They reward parsimony and small samples, where a better model is one with a smaller information criterion.

Parallel slopes assumption

- Important assumption in an Ordered Model

$$P(y \leq i) = F(\mu_i - \beta X)$$

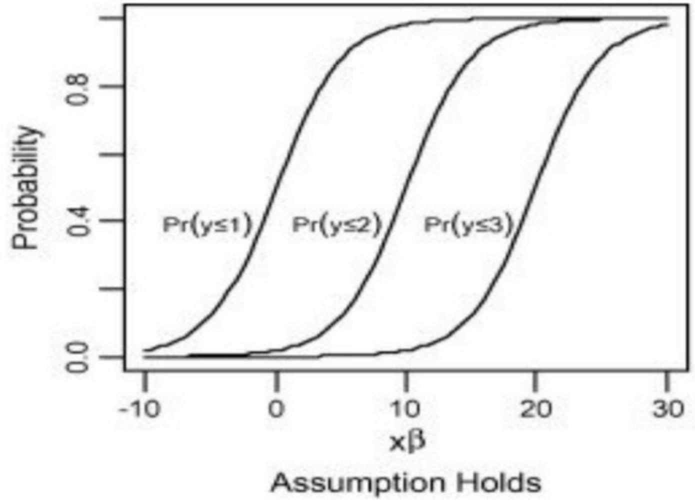
- This is the definition of a set of binary models with different intercepts

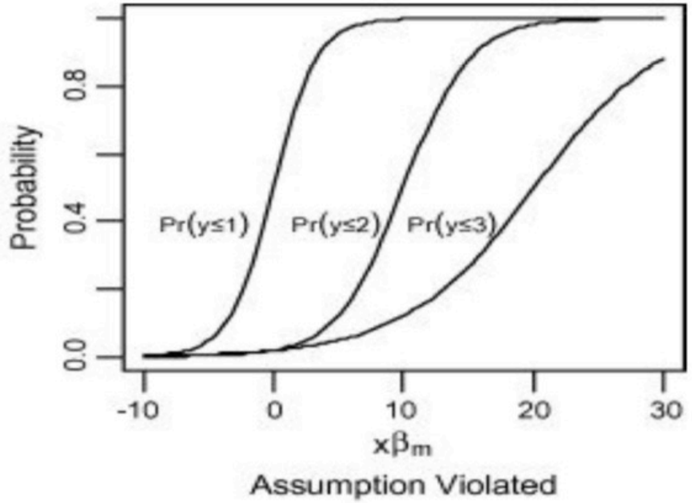
$$\mu_i - \beta X = (\mu_i - \beta_0) - \sum_{k=1}^{K} \beta_k x_k$$

$$P(y \leq 1) = F((\mu_1 - \beta_0) - \sum_{k=1}^{K} \beta_k x_k) \qquad P(y \leq 2) = F((\mu_2 - \beta_0) - \sum_{k=1}^{K} \beta_k x_k)$$

Parallel slopes assumption

- Changing the intercepts only shifts the probability curves right or left
- It doesn´t changes the slope





$$\frac{\delta P(y \leq 1)}{\delta x} = \frac{\delta P(y \leq 2)}{\delta x} = \frac{\delta P(y \leq i)}{\delta x}$$

Parallel slopes assumption

- One way to test the parallel slopes assumption is to use the Brant test, which is formulated for ordered logit

$$P(y \leq i) = F(\mu_i - \beta X)$$

- The logit formulation implies the proportional odds assumption

$$log\left(\frac{\gamma_i}{1-\gamma_i}\right) = \mu_i - \beta X$$

- This could be informally tested using i-1 binary models and compare their coefficients

Brant's wald test

- Wald test gives information about Independent variable or variables that break parallel lines assumption.

Variable	χ^2	p	s.d
All	15,53	0,004*	4
X6	1,01	0,167	1
X7	0,02	0,737	1
X8	10,65	0,001*	1
X9	4,09	0,043*	1

$(^{*}p < 0{,}05)$

Brant's wald test

- The following hypothesis are tested

$$H_0\colon \boldsymbol{\beta}_q - \boldsymbol{\beta}_1 = \mathbf{0},\ q = 2, \ldots, J-1 \qquad \text{or} \qquad H_0\colon \mathbf{R}\boldsymbol{\beta}^* = \mathbf{0}$$

$$\mathbf{R} = \begin{bmatrix} \mathbf{I} & -\mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & -\mathbf{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{I} & \mathbf{0} & \mathbf{0} & \cdots & -\mathbf{I} \end{bmatrix}, \quad \boldsymbol{\beta}^* = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \boldsymbol{\beta}_3 \\ \vdots \\ \boldsymbol{\beta}_{J-1} \end{bmatrix}$$

The Wald statistic used is:

$$\chi^2[(J-2)K] = \left(\mathbf{R}\hat{\boldsymbol{\beta}}^*\right)' \left[\mathbf{R} \times \text{Asy.Var}\left[\hat{\boldsymbol{\beta}}^*\right] \times \mathbf{R}'\right]^{-1} \left(\mathbf{R}\hat{\boldsymbol{\beta}}^*\right)$$

, where Asy.Var is the assymptotic Covariance matrix

$\hat{\boldsymbol{\beta}}^*$ is obtained by stacking the individual binary logit estimates of β

- The Brant test could be calculated both the global model and for individual coefficients

Example – Ordered Response Model

- In Portugal, injury crash severities are registered by the police and categorized as a function of the victims' stay in hospital and outcome:
  - A minor injury is registered when a victim requires hospital treatment but stays there for less than 24 h;
  - A severe injury refers to a victim who is registered as a hospital in-patient and stays there for more than a day; and
  - A fatality means a victim who dies as a consequence of crash injuries within 30 days of occurrence of the crash.
- Portuguese crash data structure allows for categorizing ROR ("run-off the road") accidents by driver injury severity and by the severity of the most seriously injured occupant.
- N = 890 observations

Source: Roque, C., Moura, F. Cardoso, J. Detecting unforgiving roadside contributors through the severity analysis of ran-off-road crashes. Accid. Anal. Prev. (2015), http://dx.doi.org/10.1016/j.aap.2015.02.012

Recommended readings

- Washington, Simon P., Karlaftis, Mathew G. e Mannering (2003) Statistical and econometric Methods for Transportation Data Analysis, CRC
- Greene, William and Hensher, David A. (2010) Modelling Ordered Choice. A primer, Cambridge University Press
- Long, J. Scott (1997) Regression Models for categorical and limited dependent variables, Sage