# Multinomial Response Models

(GB: Chapter 15.9-15.10)

Isabel Casas

SYDDANSK UNIVERSITET

# Multinomial data

- Conditional logit model
- Nested logit model
- Ordered probit model
- Ordered logit model

# Conditional Logit Model (CLM)

- MNL from an underlying utility comparison
- Interpretation of $\beta$
  - Marginal effects of $\mathbf{x}_j$
- Maximum likelihood estimation
- Example

# Multinomial response models

$$
\begin{aligned}
P(y_i = j | \mathbf{x}_i, \mathbf{z}_i, w_i) =& G(\alpha_j + \beta \mathbf{x}_{ij} + \gamma_j \mathbf{z}_i + \delta_j w_{ij}) \\
i =& 1, \dots, n, \quad j = 0, 1, \dots, J
\end{aligned}
$$

- $i$ is the individual, $j$ is the alternative
- $\mathbf{x}_{ij}$ alternative specific variables with generic coefficient $\beta$,
- $\mathbf{z}_i$ individual specific variable with an alternative specific coefficient $\gamma_j$
- $w_{ij}$ alternative specific variables with an alternative specific coefficient $\delta_j$

# Data for MNL

Multinomial logit model:

$$P(y_i = j|\mathbf{z}_i) = G(\alpha_j + \gamma_j \mathbf{z}_i)$$

| id | status | educ | exper | expersq | black |
|---|---|---|---|---|---|
| 1 | 2 | 10 | 0 | 0 | 1 |
| 2 | 3 | 16 | 4 | 16 | 0 |
| 3 | 2 | 10 | 0 | 0 | 1 |
| 4 | 1 | 10 | 0 | 0 | 1 |
| 5 | 2 | 11 | 0 | 0 | 1 |
| 171 | 3 | 12 | 7 | 49 | 0 |

All variables are individual specific and alternative specific coefficients

# MNL from underlying utility comparison

status: school(1), home(2), work(3)

$$
\begin{aligned}
y_{i1}^* =& \alpha_1 + \mathbf{z}_i\gamma_1 + \epsilon_i = \alpha_1 + \gamma_{11}\ educ_i + \gamma_{12}\ exper_i + \ldots \\
y_{i2}^* =& \alpha_2 + \mathbf{z}_i\gamma_2 + \epsilon_i = \alpha_2 + \gamma_{21}\ educ_i + \gamma_{22}\ exper_i + \ldots \\
y_{i3}^* =& \alpha_3 + \mathbf{z}_i\gamma_3 + \epsilon_i = \alpha_3 + \gamma_{31}\ educ_i + \gamma_{32}\ exper_i + \ldots
\end{aligned}
$$

SYDDANSK UNIVERSITET

# Conditional logit model (CLM)

$$P(y_i = j|\mathbf{x}_i, \mathbf{w}_i) = G(\beta\mathbf{x}_{ij} + \delta_j w_{ij})$$

| id | mode | choice | wait | vcost | travel | gcost |
|---|---|---|---|---|---|---|
| 1 | air | no | 69 | 59 | 100 | 70 |
| 1 | train | no | 34 | 31 | 372 | 71 |
| 1 | bus | no | 35 | 25 | 417 | 70 |
| 1 | car | yes | 0 | 10 | 180 | 30 |
| 2 | air | no | 64 | 58 | 68 | 68 |

- All variables are alternative specific, the coefficients can be constant or alternative specific
- The simplest model assume constant coefficients.

# CLM from an underlying utility comparison

Interurban trips between Sydney and Melbourne.

| | |
|---|---|
| individual: | Factor indicating individual with levels 1 to 200. |
| mode: | Factor indicating travel mode with levels "car", "air", "train", or "bus". |
| choice: | Factor indicating choice with levels "no" and "yes". |
| wait: | Terminal waiting time, 0 for car. |
| vcost: | Vehicle cost component. |
| travel: | Travel time in the vehicle. |
| gcost: | Generalized cost measure. |
| income: | Household income. |
| size: | Party size. |

Interest: Estimate the probability for each alternative or mode.

SYDDANSK UNIVERSITET

# Data for CLM

| individual | mode | choice | wait | vcost | travel | gcost | income | size |
|---|---|---|---|---|---|---|---|---|
| 1 | air | no | 69 | 59 | 100 | 70 | 35 | 1 |
| 1 | train | no | 34 | 31 | 372 | 71 | 35 | 1 |
| 1 | bus | no | 35 | 25 | 417 | 70 | 35 | 1 |
| 1 | car | yes | 0 | 10 | 180 | 30 | 35 | 1 |
| 2 | air | no | 64 | 58 | 68 | 68 | 30 | 2 |
| 2 | train | no | 44 | 31 | 354 | 84 | 30 | 2 |
| 2 | bus | no | 53 | 25 | 399 | 85 | 30 | 2 |
| 2 | car | yes | 0 | 11 | 255 | 50 | 30 | 2 |
| 3 | air | no | 69 | 115 | 125 | 129 | 40 | 1 |
| 3 | train | no | 34 | 98 | 892 | 195 | 40 | 1 |
| 3 | bus | no | 35 | 53 | 882 | 149 | 40 | 1 |
| 3 | car | yes | 0 | 23 | 720 | 101 | 40 | 1 |

Syddansk Universitet

# CLM from an underlying utility comparison

Example: car(0), plain(1), train(2) and bus (3). The utility of each choice, assuming constant coefficients:

$$
\begin{aligned}
y_{0i}^* =& \mathbf{x}_{0i}\beta + \epsilon_{0i} = \beta_1 \ wait_{0i} + \beta_2 \ vcost_{0i} + \ldots \\
y_{1i}^* =& \mathbf{x}_{1i}\beta + \epsilon_{1i} = \beta_1 \ wait_{1i} + \beta_2 \ vcost_{1i} + \ldots \\
y_{2i}^* =& \mathbf{x}_{2i}\beta + \epsilon_{2i} = \beta_1 \ wait_{2i} + \beta_2 \ vcost_{2i} + \ldots \\
y_{3i}^* =& \mathbf{x}_{3i}\beta + \epsilon_{3i} = \beta_1 \ wait_{3i} + \beta_2 \ vcost_{3i} + \ldots
\end{aligned}
$$

# CLM from an underlying utility comparison

- $\mathbf{x}_{ji}$ differs across alternatives and possibly across individuals.
- E.g. the commute time for individual $i$ using transportation $j$
- $\mathbf{x}_{ji}$ does not contain the unity.
- $\beta$ same across alternatives $\Rightarrow$ effect of a higher $gcost$ on utility is the same for all alternatives and individuals
- $\epsilon_{ij}$ are $J+1$ unobservables (taste shifters) affecting utility, e.g. individual preferences for the different alternatives.
- $\mathbf{x}_{ji}$ is independent of $\epsilon_{ji}$

# CLM from an underlying utility comparison

The individual picks the alternative with the highest utility:

$$y_i = \arg\max(y^*_{0i}, y^*_{1i}, y^*_{2i}, y^*_{3i}) = \begin{cases} 0 & \text{if car max. utility} \\ 1 & \text{if plain max. utility} \\ 2 & \text{if train max. utility} \\ 3 & \text{if bus max. utility} \end{cases}$$

$\Rightarrow$ Probability of choosing alternative $j$:

$$P(y_i = j|\mathbf{x}_i) = P(x_{ji}\beta + \epsilon_{ji} > x_{hi}\beta + \epsilon_{hi}, \text{ for } h \neq j|\mathbf{x}_i)$$

If $(\epsilon_{0i}, \epsilon_{1i}, \dots, \epsilon_{Ji})$ follows some joint distribution $\Rightarrow$ finding the above probability requires a $J+1$ dimensional integral.

# CLM from an underlying utility comparison

If $\epsilon_{ji}$'s are independent and follow a **type-I extreme value (Gumbel)** distribution, then it can be shown that:

$$p_j = P(y_i = j|\mathbf{x}_i) = \frac{\exp(\mathbf{x}_{ji}\beta)}{\sum_{h=0}^{J}\exp(\mathbf{x}_{hi}\beta)}, \quad j = 0, 1, \ldots, J$$

What is a (standard) type-I extreme value distribution?

$$\begin{aligned} F(x) =& e^{-e^{-x}} \\ f(x) =& F'(x) = e^{-x}e^{-e^{-x}} \end{aligned}$$

# CLM from an underlying utility comparison

If $\epsilon_{ji}$'s are independent and follow a **type-I extreme value (Gumbel)** distribution, then it can be shown that:

$$p_j = P(y_i = j|\mathbf{x}_i) = \frac{\exp(\mathbf{x}_{ji}\beta)}{\sum_{h=0}^{J}\exp(\mathbf{x}_{hi}\beta)}, \quad j = 0, 1, \ldots, J$$

What is a (standard) type-I extreme value distribution?

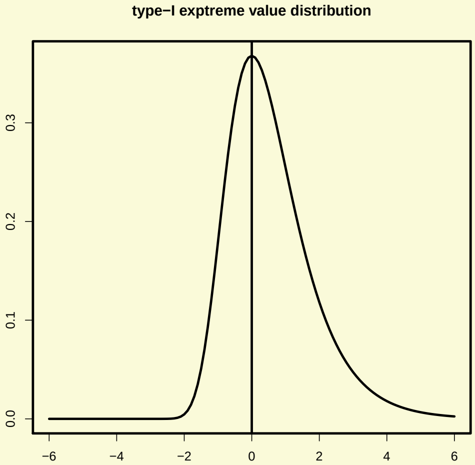
$$\begin{aligned} F(x) =& e^{-e^{-x}} \\ f(x) =& F'(x) = e^{-x}e^{-e^{-x}} \end{aligned}$$

Remember the MNL probabilities

$$P(y_i = j|\mathbf{z}_i) = \frac{\exp(\mathbf{z}_i\gamma_j)}{1 + \sum_{h=1}^{J}\exp(\mathbf{z}_i\gamma_h)}, \quad j = 1, \ldots, J$$

# Type-I extreme value distribution

Looks a bit like the normal but it is a bit skewed.



## Marginal effects of $\mathbf{x}_{jk}$

The effect of the $k$th element of $\mathbf{X}$ (e.g. $wait$) on the probability of alternative $j$ (e.g. bus)

Two types of effects:

1. Effect on the probability of taking alternative $j$ (i.e. the bus)
   - E.g. effect of increase of bus waiting time on probability of choosing the bus
2. Effect on the probability of taking another alternative, $p$ (e.g. the car)
   - E.g. effect of increase of bus waiting time on probability of choosing the car

SYDDANSK UNIVERSITET

# Marginal effects of $\mathbf{x}_{jk}$

1. Effect of change of $\mathbf{x}_{jk}$ on prob. of alternative $j$: E.g. effect of bus waiting time on prob. of taking the bus

$$\frac{\partial p_j(\mathbf{X})}{\partial \mathbf{x}_{jk}} = \frac{\beta_k \exp(\mathbf{x}_j\beta)\sum_{h=0}^{J}\exp(\mathbf{x}_h\beta) - \beta_k\left[\exp(\mathbf{x}_j\beta)\right]^2}{\left[\sum_{h=0}^{J}\exp(\mathbf{x}_h\beta)\right]^2}$$

$$= \beta_k p_j(\mathbf{X}) - \beta_k[p_j(\mathbf{X})]^2$$

$$= \beta_k p_j(\mathbf{X})(1 - p_j(\mathbf{X}))$$

- The sign of $\beta_k$ is the sign of the marginal effect
- $p_j(\mathbf{X})(1 - p_j(\mathbf{X})) \leq 0.25$
- Rule of thumb: divide the coefficient by 4 in order to have an upper bound of the marginal effect
- For individual specific variables, the sign of the coefficient is not necessary the sign of the effect

# Marginal effects of $\mathbf{x}_{jk}$

2. Effect of change of $X_{jk}$ on prob. of alternative $p$: E.g. effect of bus waiting time on prob. of taking the car

$$\frac{\partial p_p(X)}{\partial X_{jk}} = \frac{-\beta_k \exp(X_j\beta)\exp(X_p\beta)}{\left[\sum_{h=0}^{J}\exp(X_h\beta)\right]^2}$$

$$= -\beta_k p_j(X) p_p(X)$$

The sign of $\beta_k$ is the - sign of the marginal effect

# Estimation CML

The log–likelihood function is:

$$\ell(\beta) = \sum_{i=1}^{n} \ell_i(\beta) = \sum_{i=1}^{n} \sum_{j=0}^{J} \mathbf{1}[y_i = j] \log p_j(X_{ij}, \beta)$$

- Usual properties of the ML estimators: consistency, asymptotic efficiency and asymptotic normality
- Same tests as previously
- Same three estimators for the asymptotic variance
- Same consequences for the model misspecifications

# Estimation CLM in R

My original data is called $TravelMode$. We need to format it first so $mlogit$ understands it. So the new data set is $TM$:

```
> TM<-mlogit.data(TravelMode, choice="choice", shape="long", alt.var="mode")
> head(TM, 12)
```

| | individual | mode | choice | wait | vcost | travel | gcost | income | size |
|---|---|---|---|---|---|---|---|---|---|
| 1.air | 1 | air | FALSE | 69 | 59 | 100 | 70 | 35 | 1 |
| 1.train | 1 | train | FALSE | 34 | 31 | 372 | 71 | 35 | 1 |
| 1.bus | 1 | bus | FALSE | 35 | 25 | 417 | 70 | 35 | 1 |
| 1.car | 1 | car | TRUE | 0 | 10 | 180 | 30 | 35 | 1 |
| 2.air | 2 | air | FALSE | 64 | 58 | 68 | 68 | 30 | 2 |
| 2.train | 2 | train | FALSE | 44 | 31 | 354 | 84 | 30 | 2 |
| 2.bus | 2 | bus | FALSE | 53 | 25 | 399 | 85 | 30 | 2 |
| 2.car | 2 | car | TRUE | 0 | 11 | 255 | 50 | 30 | 2 |

# Estimation CLM in R

- Variable $Y$ : $choice$, the alternative chosen by the individual
- Variable $mode$ includes all the possible anternatives
- The reference alternative: $car$
- Alternative–specific variables: $wait, vcost, travel, gcost$
- Individual–specific variables: $income, size$

# Estimation CLM in R

Model without individual-specific effects:

- formula: $choice \sim wait + vcost + travel + gcost| - 1$
- No include $reflevel$

```
> CML.1<-mlogit(choice~wait+vcost+travel+gcost|-1,data=TM)
> summary(CML.1)

Coefficients :
         Estimate Std. Error t-value  Pr(>|t|)
wait   -0.0348066  0.0046940 -7.4152 1.215e-13 ***
vcost  -0.0224295  0.0143541 -1.5626 0.1181505
travel -0.0063447  0.0018417 -3.4451 0.0005709 ***
gcost   0.0318293  0.0137286  2.3185 0.0204238 *
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
```

What does $\beta_{vcost} = -0.02$ mean?

# Estimation CLM in R

```
> library(mlogit)
> data("Fishing", package="mlogit")

     mode price.beach price.pier price.boat price.charter catch.beach catch.pier catch.boat catch.charter
1 charter     157.930    157.930    157.930       182.930      0.0678     0.0503     0.2601        0.5391
2 charter      15.114     15.114     10.534        34.534      0.1049     0.0451     0.1574        0.4671
3    boat     161.874    161.874     24.334        59.334      0.5333     0.4522     0.2413        1.0266
4    pier      15.134     15.134     55.930        84.930      0.0678     0.0789     0.1643        0.5391
5    boat     106.930    106.930     41.514        71.014      0.0678     0.0503     0.1082        0.3240
6 charter     192.474    192.474     28.934        63.934      0.5333     0.4522     0.1665        0.3975
```

# Estimation CLM in R

```
> Fish<-mlogit.data(Fishing,choice="mode", varying=2:9, shape="wide")
> head(Fish,8)
           mode   income     alt   price  catch chid
1.beach   FALSE 7083.332   beach 157.930 0.0678    1
1.boat    FALSE 7083.332    boat 157.930 0.2601    1
1.charter  TRUE 7083.332 charter 182.930 0.5391    1
1.pier    FALSE 7083.332    pier 157.930 0.0503    1
2.beach   FALSE 1250.000   beach  15.114 0.1049    2
2.boat    FALSE 1250.000    boat  10.534 0.1574    2
2.charter  TRUE 1250.000 charter  34.534 0.4671    2
2.pier    FALSE 1250.000    pier  15.114 0.0451    2
```

- choice: the variable with alternative choice
- shape: wide (one row for each choice situation)
- varying: columns 2:9 contain the alternative specific variables

Run the model and find the marginal effects.

# MNL vs CLM

MNL:

- Variables $z_i$ are the same across alternatives for the same individual
- But effects of $z$ vary across alternatives $\Rightarrow$ different $\alpha_j$ that I call $\gamma_j 0$ and different $\gamma_j$ for each alternative $j$

CLM:

- $x_{ij}$ varies across alternatives (contains characteristics of each alternative) and possibly across individuals (as characteristics may depend on the user),
- But $\beta$ (the effect of $x_{ij}$) is fixed across alternatives

# Mixed Logit model

Missing both types of variables. Sometime, the mixed is still called CML:

$$
\begin{aligned}
y_{0i}^* =& x_{0i}\beta + z_i\gamma_0 + \epsilon_{0i} \\
y_{1i}^* =& x_{1i}\beta + z_i\gamma_1 + \epsilon_{1i} \\
y_{2i}^* =& x_{2i}\beta + z_i\gamma_2 + \epsilon_{2i}
\end{aligned}
$$

- $x_{ji}$ are alternative–specific characteristics
- $\mathbf{z}_i$ are individual–specific characteristics
- $\gamma_j$ are alternative–specific coefficient

# Estimation CLM in R

Model with effects of $income$ and $size$:

- formula:
  $choice \sim wait + vcost + travel + gcost | income + size$
- Include $reflevel = "car"$

```
> CML.2<-mlogit(choice~wait+vcost+travel+gcost|income+size,data=TM, reflevel="car")
Coefficients :
                  Estimate Std. Error t-value  Pr(>|t|)
altair           5.2865000  1.2026299  4.3958 1.104e-05 ***
alttrain         5.7082954  0.7266516  7.8556 3.997e-15 ***
altbus           4.7163288  0.8238724  5.7246 1.037e-08 ***
wait            -0.1025547  0.0113755 -9.0154 < 2.2e-16 ***
vcost           -0.0533528  0.0252825 -2.1103   0.03484 *
travel          -0.0102498  0.0034179 -2.9988   0.00271 **
gcost            0.0464263  0.0249000  1.8645   0.06225 .
altair:income    0.0080781  0.0134186  0.6020   0.54717
alttrain:income -0.0594981  0.0148925 -3.9952 6.465e-05 ***
altbus:income   -0.0199412  0.0163648 -1.2185   0.22302
altair:size     -0.5307014  0.3210340 -1.6531   0.09831 .
alttrain:size    0.1628226  0.2394588  0.6800   0.49653
altbus:size     -0.2399000  0.3580260 -0.6701   0.50282
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
```

# Independence of Irrelevant Alternatives (IIA)

An unpleasant characteristic of the CLM is the IIA assumption: The relative probability (odd ratio) for alternatives $j$ and $p$:

$$\frac{P(y=j|\mathbf{X})}{P(y=p|\mathbf{X})} = \exp((\mathbf{x}_j - \mathbf{x}_p)\beta))$$

- It only depends on $j$ and $p$, not other alternatives
- i.e. adding or changing a third alternative does not change this relative probability
- Example: Car, blue bus where $P(y=car|\mathbf{X})/P(y=bluebus|\mathbf{X}) = 1 \Rightarrow P(car|X) = 0.5$
- Add red bus then $P(y=car|X)/P(y=bluebus|\mathbf{X}) = 1$ but $P(car|\mathbf{X}) = 1/3$
- Absurd: IIA
- http://www.statisticalhorizons.com/iia

# Relaxing the IIA

- If we assume that the $\epsilon_{ji}$ are correlated with $\epsilon_{pi}$ such that $(\epsilon_j, \epsilon_p) \sim N(0, R)$ with $R$ a correlation matrix, then:
- The Conditional Probit Model
  - Relative probabilities of $j$ and $p$ are no longer independent of the characteristics of other alternatives
  - However, it is very complicated to estimate this model

# Relaxing the IIA

**The Nested Logit Model**:

- Divide alternatives into groups (*similar* alternatives in the same group)
- Two-step modelling approach

  Step 1 choose between groups
  - E.g. $ground$ nest: $bus$, $train$ and $car$ modes
  - $fly$ nest: $air$ mode

  Step 2 choose between alternatives within group (modelled as a CLM)
- The CLM is nested within a more general choice model (Step 1) that does not have the IIA property.
- Choice probability = product of the probabilities at the two stages.
- The ML estimator is consistent and asymptotically normal under the classical assumptions

# Nested Logit in R

```
> NL<-mlogit(choice~wait+vcost+travel+gcost,data=TM, reflevel="car",
nests=list(fly="air", ground=c("train", "bus", "car")), unscaled=T)
> summary(NL)

Coefficients :
           Estimate Std. Error t-value  Pr(>|t|)
altair     5.398425   1.054053  5.1216 3.030e-07 ***
alttrain   4.618518   0.772444  5.9791 2.244e-09 ***
altbus     3.967942   0.709512  5.5925 2.238e-08 ***
wait      -0.100622   0.012464 -8.0732 6.661e-16 ***
vcost     -0.421429   0.096475 -4.3683 1.252e-05 ***
travel    -0.070754   0.014151 -5.0000 5.733e-07 ***
gcost      0.411450   0.095204  4.3218 1.548e-05 ***
iv.fly     0.868599   0.155595  5.5824 2.372e-08 ***
iv.ground  0.252502   0.052095  4.8469 1.254e-06 ***
```

- nests contains the list of different groups
- unscaled= T because we have only one choice in the nest fly

# Ordered response models

- Ordered Probit Model (OPM)
- Ordered Logit Model (OLM)

# Ordered responses (Chapter 5.10)

When outcomes can be ranked!

- Grades in school: -3, 0, 2, 4, 7, 10, 12
- Attitudes to various issues:
  - agree, neutral, disagree
  - many, few, none
- But *distance* between choices still does not (necessarily) make sense
- E.g. moving from -3 to 0 may take more or less effort than moving from 10 to 12.

# Ordered probit model (OPM)

Model set-up:

- $\mathbf{y}$ takes a value in $\{0, 1, 2, \ldots, J\}$
- $\mathbf{X}$ is $n \times k$ (not unity)
- $(\mathbf{x}_i, y_i)$ is a random draw from the population
- Value of $\mathbf{y}$ is determined by a latent variable, $\mathbf{y}^*$:

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon}|\mathbf{X} \sim N(0, 1)$$

- $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$, no intercept.
- Does it remind you of any model?

# Ordered probit model (OPM)

$$
\begin{aligned}
y &= 0 & Y^* &\leq \alpha_1 \\
y &= 1 & \alpha_1 < Y^* &\leq \alpha_2 \\
&\vdots & &\vdots \\
y &= J & \alpha_J &< Y^*
\end{aligned}
$$

- We have to estimate the coefficients $\beta$ and
- the critical values $\alpha_j$

# Ordered probit model (OPM)

$$P(y=0|X) = P(Y^* \leq \alpha_1|X) = P(\epsilon \leq \alpha_1 - \mathbf{X}\boldsymbol{\beta}|X) = \Phi(\alpha_1 - \mathbf{X}\boldsymbol{\beta})$$

$$\begin{aligned} P(y=1|X) =& P(\alpha_1 < Y^* \leq \alpha_2|X) = P(Y^* \leq \alpha_2|X) - P(Y^* < \alpha_1|X) \\ =& P(\epsilon \leq \alpha_2 - \mathbf{X}\boldsymbol{\beta}|X) - P(\epsilon < \alpha_1 - \mathbf{X}\boldsymbol{\beta}|X) \\ =& \Phi(\alpha_2 - \mathbf{X}\boldsymbol{\beta}) - \Phi(\alpha_1 - \mathbf{X}\boldsymbol{\beta}) \end{aligned}$$

$P(y=J|X)$?

# Ordered probit model (OPM)

We have the standard binary probit model:

$$P(y = 0|X) = \Phi(\alpha_1 - \mathbf{X}\boldsymbol{\beta}) = 1 - \Phi(\mathbf{X}\boldsymbol{\beta} - \alpha_1)$$
$$P(y = 1|X) = 1 - \Phi(\alpha_1 - \mathbf{X}\boldsymbol{\beta}) = \Phi(\mathbf{X}\boldsymbol{\beta} - \alpha_1)$$

If

- $J = 1$ and
- $\alpha_1$ is then the intercept $(-\alpha_1 = \beta_0)$
- Remember there is no $\beta_0$ in OPM

# Ordered probit model (OPM)

Estimation: the Log-lik for observation $i$:

$$
\begin{aligned}
\ell_i(\alpha,\beta) = & \mathbf{1}[y_i = 0] \log(\Phi(\alpha_1 - \mathbf{X}\boldsymbol{\beta})) \\
& + \mathbf{1}[y_i = 1] \log(\Phi(\alpha_2 - \mathbf{X}\boldsymbol{\beta}) - \Phi(\alpha_1 - \mathbf{X}\boldsymbol{\beta})) \\
& \dots \\
& + \mathbf{1}[y_i = J] \log(1 - \Phi(\alpha_J - \mathbf{X}\boldsymbol{\beta}))
\end{aligned}
$$

The estimator maximises $\ell(\alpha,\beta) = \sum_i \ell_i(\alpha,\beta)$ over $\beta$ and $\alpha$.

In R: $polr(formula, data = mydata, method = "probit")$

- formula: $y \sim X_1 + X_2 + \ldots$
- $y$ must be a factor: $as.factor(y)$

# OPM in R (Example 15.5)

Data file: pension.txt with variables:

- prftshr: =1 if profit sharing plan
- female: =1 if female
- married: =1 if married
- age: age in years
- educ: highest grade completed
- black: =1 if black
- pctstck: 0=mostly bonds,50=mixed,100=mostly stocks

Question: What is the impact of allowing individuals to choose their own asset allocation in pension plans?

# OPM in R (Example 15.5)

```
Coefficients:
           Value Std. Error t value
choice   0.37230    0.18405  2.0228
age     -0.05124    0.02212 -2.3160
educ     0.02537    0.03513  0.7222
female   0.03947    0.20456  0.1930
black    0.10150    0.28027  0.3622
married  0.08690    0.23172  0.3750
finc25  -0.58028    0.42347 -1.3703
finc35  -0.13535    0.43088 -0.3141
finc50  -0.26930    0.42602 -0.6321
finc75  -0.58578    0.47229 -1.2403
finc100 -0.24198    0.46578 -0.5195
finc101 -0.87982    0.52564 -1.6738
prftshr  0.48392    0.21600  2.2403

Intercepts:
       Value   Std. Error t value
0|50   -3.1643  1.5957    -1.9830
50|100 -2.1308  1.5903    -1.3398
```

# Marginal effects OPM

$$\frac{\partial P(y=0|X)}{\partial X_k} = -\,\phi(\alpha_1 - \mathbf{X}\boldsymbol{\beta})\beta_k$$

$$\frac{\partial P(y=1|X)}{\partial X_k} = -\left[\phi(\alpha_2 - \mathbf{X}\boldsymbol{\beta}) - \phi(\alpha_1 - \mathbf{X}\boldsymbol{\beta})\right]\beta_k$$

$$\vdots$$

$$\frac{\partial P(y=J|X)}{\partial X_k} = \phi(\alpha_J - \mathbf{X}\boldsymbol{\beta})\beta_k$$

- The sign of $\beta_k$ determine the sign of the effect over the probability of alternative 0 and $J$
- This sign does not determine the sign of the effect over the other alternatives
- Because the alternatives are ordered: a positive $\beta_k$ implies that $Y^*$ increases with $X_k$: i.e. overall a higher chance of larger values of $Y^*$

# Reporting results of OPM

- Report the average marginal effect of variables
- Report percent correctly predicted
- Report the estimated values of critical values ($\alpha$)
- T-tests
- Testing linear restrictions: Wald, LR, LM tests
- Same issues as in standard probit:
  - Heteroscedasticity
  - Non-normality
  - Endogenous RHS variables

# OPM real example

Malchow-Møller, Munch, Schroll and Skaksen (2008): "Attitudes towards immigration–Perceived consequences and economic self-interest", Economics Letters.

- Models attitudes towards immigration
- $Y^*$: $allow\ none$ (1), $allow\ a\ few$ (2), $allow\ some$ (3) and $allow\ many$ (4)
- Create $y = AT$ has 13 values between 4 and 16.

| | Question 1 | Q2 | Q3 | Q4 | AT |
|---|---|---|---|---|---|
| poor EU-15 | 3 | 4 | 2 | 3 | 12 |
| rich EU-15 | 1 | 1 | 1 | 1 | 4 |
| poor out EU-15 | 4 | 3 | 3 | 3 | 13 |
| rich out EU-15 | 2 | 2 | 3 | 1 | 8 |

- Explanatory variables:
  - Individual characteristics: age, sex, political standing, geography, education, labour market status
  - 5 dummy in relation to people's perceptions of consequences of immigration: $wages\_down$, $take\_jobs\_away$, $bad\_for\_poor$, $take\_more\_out$ and $fill\_jobs$

# OPM real example

256 N. Malchow-Møller et al. / Economics Letters 100 (2008) 254–257

Table 2
Conditional attitudes towards immigration (ordered probit estimation)

| | Dependent variable: AT | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Age | −0.00748 | −0.01100 | −0.00562 | −0.01240 | −0.01009 | −0.00579 |
| | (−3.27)*** | (−4.88)*** | (−2.16)** | (−5.43)*** | (−4.58)*** | (−2.09)** |
| Age^2 | 0.00001 | 0.00004 | 0.00002 | 0.00006 | 0.00003 | 0.00001 |
| | (0.41) | (1.78)* | (0.62) | (2.61)*** | (1.28) | (0.28) |
| Left | 0.26499 | 0.27507 | 0.25752 | 0.25121 | 0.26239 | 0.25341 |
| | (15.86)*** | (16.58)*** | (13.86)*** | (15.06)*** | (15.88)*** | (13.24)*** |
| Right | −0.00150 | 0.01249 | 0.01727 | 0.01812 | −0.00807 | 0.05405 |
| | (−0.09) | (0.73) | (0.90) | (1.06) | (−0.48) | (2.73)*** |
| Male | 0.10981 | 0.08719 | 0.06523 | 0.09583 | 0.08864 | 0.10492 |
| | (7.82)*** | (6.36)*** | (4.26)*** | (6.93)*** | (6.55)*** | (6.35)*** |
| Urban | 0.10545 | 0.09538 | 0.08418 | 0.08174 | 0.09593 | 0.08594 |
| | (6.90)*** | (6.27)*** | (4.99)*** | (5.35)*** | (6.33)*** | (4.95)*** |
| Immigrant | 0.26572 | 0.21890 | 0.23393 | 0.21316 | 0.24715 | 0.20432 |
| | (12.92)*** | (10.68)*** | (10.44)*** | (10.39)*** | (12.17)*** | (8.81)*** |
| Primary | −0.23129 | −0.22461 | −0.24244 | −0.24361 | −0.25370 | −0.20096 |
| | (−13.28)*** | (−13.02)*** | (−12.36)*** | (−14.10)*** | (−14.80)*** | (−9.91)*** |
| Tertiary | 0.34178 | 0.34764 | 0.35651 | 0.35877 | 0.39189 | 0.27592 |
| | (17.79)*** | (18.17)*** | (16.96)*** | (18.70)*** | (20.62)*** | (12.75)*** |
| Unemployed | −0.11626 | −0.06118 | −0.09981 | −0.15024 | −0.14065 | 0.04327 |
| | (−4.44)*** | (−1.77)* | (−3.44)*** | (−5.81)*** | (−5.48)*** | (1.09) |
| Self-employed | 0.01972 | 0.02325 | 0.07344 | 0.02541 | | |
| | (0.81) | (0.95) | (2.64)*** | (1.05) | | |
| Wages_down | −0.46354 | | | | | −0.21872 |
| | (−20.51)*** | | | | | (−7.97)*** |
| Workforce | 0.01140 | | | | | 0.00713 |
| | (0.54) | | | | | (0.25) |
| Wages_down × workforce | −0.06850 | | | | | −0.07622 |
| | (−2.39)** | | | | | (−2.27)** |
| Take_jobs_away | | −0.49945 | | | | −0.23825 |
| | | (−28.76)*** | | | | (−11.05)*** |
| Difficult_get_job | | 0.00648 | | | | 0.00599 |
| | | (0.31) | | | | (0.24) |
| Take_jobs_away × difficult_get_job | | −0.05870 | | | | −0.09779 |
| | | (−1.80)* | | | | (−2.64)*** |
| Take_jobs_away × unemployed | | −0.14474 | | | | −0.22296 |
| | | (−2.83)*** | | | | (−3.79)*** |
| Bad_for_poor | | | −0.53089 | | | −0.29943 |
| | | | (−30.14)*** | | | (−15.30)*** |
| Poor | | | −0.07497 | | | −0.06663 |
| | | | (−2.63)*** | | | (−2.24)** |

# Ordered logit model

If we assume $\epsilon|X \sim$ logistic distribution:

$$\begin{aligned} P(y=0|\mathbf{X}) =& P(\epsilon \leq \alpha_1 - \mathbf{X}\boldsymbol{\beta}|\mathbf{X}) = \Lambda(\alpha_1 - \mathbf{X}\boldsymbol{\beta}) \\ P(y=1|\mathbf{X}) =& \Lambda(\alpha_2 - \mathbf{X}\boldsymbol{\beta}) - \Lambda(\alpha_1 - \mathbf{X}\boldsymbol{\beta}) \\ P(y=J|\mathbf{X}) =& 1 - \Lambda(\alpha_J - \mathbf{X}\boldsymbol{\beta}) \end{aligned}$$

In R: $polr(formula, data = mydata, method = "logit")$

- formula: $y \sim x_1 + x_2 + ...$
- $y$ must be a factor: $as.factor(y)$
- Interactions of two factor variables A, B in R
  - A*B= A+B+ A:B
  - A:B is the interaction of the two factors