4. Ordered response models

4.1 General model approaches

Ordered (ordinal) dependent variables in a microeconometric analysis:
These qualitative variables have also more than two possible mutually exclusive categories which are (in contrast to multinomial variables), however, naturally ordered (in the case of two categories, the variables are binary)

Examples for microeconometric analyses with ordered response models:
- Analysis of the individual satisfaction of a person with life (e.g. on a eleven-point scale of integers from zero for “completely dissatisfied” to ten for “completely satisfied”)
- Analysis of the personal strength of agreement to a political program (e.g. with the categories “strong disagreement”, “weak disagreement”, weak agreement”, “strong agreement”)
- Analysis of the years of education of a person (e.g. with the categories “less than nine years”, “between nine and 12 years”, “at least 13 years”)
- Analysis of the credit rating of firms (e.g. on a scale from D to AAA)
- Analysis of the stated importance of equity issues in international climate negotiations (e.g. on a five-point scale from “no importance” to “very high importance”)
Preliminary remarks to ordered dependent variables:

- The values for the categories of the ordered dependent variables are completely arbitrary if they preserve the order so that e.g. the sequences 1, 2, 3, 4 or 10, 20, 30, 40 or -20, -10, 0, 10 are possible and all reveal the same information for an ordinal variable with four categories (e.g. “strong agreement”, “weak agreement”, weak disagreement”, “strong disagreement”). As a consequence, expectations, variances, or covariances for values of ordinal variables have no meaning.

- For notational simplicity, the dependent ordered variables \( y_i \) of an observation \( i \) take the values 1, 2, ..., \( J \) for the categories where “1” < “2” < ... < “\( J \)”

- In general, ordered dependent variables could therefore be analyzed with multinomial response models as discussed before. However, this ignores the ordering information so that this would lead to inefficient ML estimators of the parameters compared with the use of ordered response models.

- This coding 1, 2, ..., \( J \) should not mislead to the application of linear regression models which would necessarily imply that the difference between “1” and “2” (e.g. “strong disagreement” and “weak disagreement”) is the same as the difference between “2” and “3” (e.g. “weak disagreement” and “weak agreement”) or “3” and “4” (e.g. “weak agreement” and “strong agreement”). In addition, linear regression models would imply that “2” is twice as high as the value “1” (e.g. persons with a “weak disagreement” would agree twice as strong to a political program as persons with a “strong disagreement”).
Continuous latent variable (which sometimes can be interpreted as varying utility) in ordered response models \((i = 1, \ldots, n)\):

\[
y_i^* = \beta' x_i + \varepsilon_i
\]

As in binary response models, \(x_i = (x_{i1}, \ldots, x_{ik})'\) is again a vector of \(k\) explanatory variables, \(\beta = (\beta_1, \ldots, \beta_k)'\) is the corresponding \(k\)-dimensional parameter vector, and \(\varepsilon_i\) is an error term. These unobservable latent variables can be related to the observed variables \(y_i\) or \(y_{ij}\) \((i = 1, \ldots, n; j = 1, \ldots, J)\):

\[
y_i = j \quad \text{if} \quad \kappa_{j-1} < y_i^* \leq \kappa_j
\]

\[
y_{ij} = \begin{cases} 1 & \text{if } y_i = j \\ 0 & \text{otherwise} \end{cases}
\]

This threshold mechanism divides the latent variable \(y_i^*\) in \(J\) intervals by using \(J + 1\) threshold parameters \(\kappa_0, \kappa_1, \ldots, \kappa_J\) with \(\kappa_0 < \kappa_1 < \cdots < \kappa_J\). According to this mechanism, higher values of the latent variable \(y_i^*\) lead to higher values of the ordered dependent variable \(y_i\) with the values or intervals \(j = 1, \ldots, J\). It follows:

\[
y_i = 1 \quad \text{if} \quad \kappa_0 < y_i^* \leq \kappa_1 \iff \kappa_0 - \beta' x_i < \varepsilon_i \leq \kappa_1 - \beta' x_i
\]

\[
y_i = 2 \quad \text{if} \quad \kappa_1 < y_i^* \leq \kappa_2 \iff \kappa_1 - \beta' x_i < \varepsilon_i \leq \kappa_2 - \beta' x_i
\]

\[
y_i = 3 \quad \text{if} \quad \kappa_2 < y_i^* \leq \kappa_3 \iff \kappa_2 - \beta' x_i < \varepsilon_i \leq \kappa_3 - \beta' x_i
\]

\[\vdots \quad \vdots \]

\[
y_i = J \quad \text{if} \quad \kappa_{J-1} < y_i^* \leq \kappa_J \iff \kappa_{J-1} - \beta' x_i < \varepsilon_i \leq \kappa_J - \beta' x_i
\]
In order to comprise all real values of the latent variable \( y_i^* \), it follows \( \kappa_0 = -\infty \) and \( \kappa_J = \infty \) so that the number of threshold parameters decreases to \( J - 1 \):

\[
y_i = 1 \text{ if } -\infty < y_i^* \leq \kappa_1 \Leftrightarrow -\infty < \varepsilon_i \leq \kappa_1 - \beta'x_i \\
\vdots \text{ } \vdots \text{ } \vdots \\
y_i = J \text{ if } \kappa_{J-1} < y_i^* \leq \infty \Leftrightarrow \kappa_{J-1} - \beta'x_i < \varepsilon_i \leq \infty
\]

Remarks:

• The resulting \( J - 1 \) threshold parameters are required to divide the range of the unobservable latent variable \( y_i^* \) into \( J \) cells which correspond to the \( J \) categories of the observed ordered dependent variable \( y_i \). Therefore, the categories of \( y_i \) represent a censored version of the true underlying values of the latent variable \( y_i^* \) (e.g. preferences)

• It follows that very low values of \( y_i^* \) are linked with \( y_i = 1 \) (e.g. “strong disagreement”), for \( y_i^* > \kappa_1 \) the value of \( y_i \) increases to 2 (e.g. “weak disagreement”), \( y_i \) increases to the value 3 for \( y_i^* > \kappa_2 \) (e.g. “weak agreement”) etc.

• In line with the discussion before, it can be seen that the difference between two levels of the ordered dependent variables \( y_i \) (e.g. “2” and “3” for “weak disagreement” and “weak agreement” or “3” and “4” for “weak agreement” and “strong agreement”) are not the same as on the scale of the latent variable \( y_i^* \) so that the threshold parameters capture a nonlinear transformation (and can be estimated within the ordered response models)
Different ordered response models result from different types of the density function of the latent variables $y_i^*$ on the basis of a distribution assumption about $\varepsilon_i$ with the distribution function $F_i(\varepsilon_i)$ as discussed later. If the $J - 1$ threshold parameters $\kappa_1, \kappa_2, \ldots, \kappa_{J-1}$ and the parameters in $\beta$ are summarized in the vector $\theta$, it follows for the probability that $y_i$ takes the value $j$ ($j = 1, \ldots, J$):

$$p_{ij}(x_i, \theta) = P(y_i = j|x_i, \theta) = P(y_{ij} = 1|x_i, \theta) = F_i(\kappa_j - \beta'x_i) - F_i(\kappa_{j-1} - \beta'x_i)$$

On the basis of $\kappa_0 = -\infty$ and $\kappa_J = \infty$, it follows $F(-\infty) = 0$ and $F(\infty) = 1$. However, these probabilities comprise too many parameters so that not all threshold parameters are identified if a constant is included in the ordered response model. Therefore, one parameter has to be normalized. Common approaches are to set the first threshold parameter $\kappa_1 = 0$ or to drop the constant term from $x_i$. In the following, we consider the second approach.

Based on the aforementioned probabilities and the binary variables $y_{ij}$, the ML estimation of the $k + J - 1$ parameters $\beta_1, \ldots, \beta_k$ and $\kappa_1, \ldots, \kappa_{J-1}$ in ordered response models (instead of $k \cdot (J-1)$ slope parameters and constants in pure multinomial logit and probit models) is identical to the ML estimation in multinomial discrete choice models. Therefore, it follows for the log-likelihood function:

$$\log L(\theta) = \sum_{i=1}^{n} \sum_{j=1}^{J} y_{ij} \log p_{ij}(x_i, \theta)$$
4.2 Ordered probit and logit models

Ordered probit models:
These ordered response models assume that the error terms $\varepsilon_i$ are standard normally distributed (as in binary probit models which are special cases of ordered probit models with $J = 2$)

Choice probabilities in ordered probit models $(i = 1, \ldots, n; j = 1, \ldots, J)$:
$$p_{ij}(x_i, \theta) = P(y_i = j|x_i, \theta) = \Phi_i (\kappa_j - \beta'x_i) - \Phi_i (\kappa_{j-1} - \beta'x_i)$$

For the specific probabilities this means:
$$P(y_i = 1|x_i, \theta) = \Phi_i (\kappa_1 - \beta'x_i)$$
$$P(y_i = 2|x_i, \theta) = \Phi_i (\kappa_2 - \beta'x_i) - \Phi_i (\kappa_1 - \beta'x_i)$$
$$P(y_i = 3|x_i, \theta) = \Phi_i (\kappa_3 - \beta'x_i) - \Phi_i (\kappa_2 - \beta'x_i)$$
$$\vdots \quad \vdots$$
$$P(y_i = J|x_i, \theta) = 1 - \Phi_i (\kappa_{J-1} - \beta'x_i)$$

→ As in binary probit models, the parameterization of the standard normal distribution of $\varepsilon_i$ is not restrictive. In fact, the normalization of the normal distribution with an expected value of zero and variance one is necessary for the identification of the parameters in the choice probabilities
Ordered logit models:
These ordered response models are derived in the same way as ordered probit models and thus assume that the error terms $\varepsilon_i$ are standard logistically distributed (as in binary logit models which are special cases of ordered logit models with $J = 2$)

Choice probabilities in ordered logit models ($i = 1, \ldots, n$; $j = 1, \ldots, J$):
$$p_{ij}(x_i, \theta) = P(y_i = j|x_i, \theta) = \Lambda_i(\kappa_j - \beta'x_i) - \Lambda_i(\kappa_{j-1} - \beta'x_i)$$
For the specific probabilities this means:
$$P(y_i = 1|x_i, \theta) = \Lambda_i(\kappa_1 - \beta'x_i)$$
$$P(y_i = 2|x_i, \theta) = \Lambda_i(\kappa_2 - \beta'x_i) - \Lambda_i(\kappa_1 - \beta'x_i)$$
$$P(y_i = 3|x_i, \theta) = \Lambda_i(\kappa_3 - \beta'x_i) - \Lambda_i(\kappa_2 - \beta'x_i)$$
$$\vdots$$
$$P(y_i = J|x_i, \theta) = 1 - \Lambda_i(\kappa_{J-1} - \beta'x_i)$$

→ In the same way as in the case of binary probit and logit models, the assumptions of standard normal or standard logistic distributions of $\varepsilon_i$ in ordered probit or ordered logit models usually lead to very similar estimation results in practice across these two types of ordered response models (see later)
Similar to (pure) multinomial logit models, the ML estimators $\hat{\beta}_h$ can neither be interpreted as the estimators of the effect of the respective explanatory variable $x_{ih}$ nor do they (generally, see later) indicate the direction of the estimator of marginal probability effects, i.e. a positive (negative) $\hat{\beta}_h$ does not necessarily lead to positive (negative) estimators of these effects. Instead, it follows for the estimator of the (partial) marginal probability effect of a (continuous) explanatory variable $x_{ih}$ in general ordered response models as well as in ordered probit and logit models ($i = 1,\ldots, n; j = 1,\ldots, J$):

$$\frac{\partial \hat{p}_{ij}(x_i, \hat{\theta})}{\partial x_{ih}} = \left[ f_i(\hat{\kappa}_{j-1} - \hat{\beta}'x_i) - f_i(\hat{\kappa}_j - \hat{\beta}'x_i) \right] \hat{\beta}_h$$

$$\frac{\partial \hat{p}_{ij}(x_i, \hat{\theta})}{\partial x_{ih}} = \left[ \varphi_i(\hat{\kappa}_{j-1} - \hat{\beta}'x_i) - \varphi_i(\hat{\kappa}_j - \hat{\beta}'x_i) \right] \hat{\beta}_h$$

$$\frac{\partial \hat{p}_{ij}(x_i, \hat{\theta})}{\partial x_{ih}} = \left\{ \Lambda_i(\hat{\kappa}_{j-1} - \hat{\beta}'x_i) \left[ 1-\Lambda_i(\hat{\kappa}_{j-1} - \hat{\beta}'x_i) \right] - \Lambda_i(\hat{\kappa}_j - \hat{\beta}'x_i) \left[ 1-\Lambda_i(\hat{\kappa}_j - \hat{\beta}'x_i) \right] \right\} \hat{\beta}_h$$

In these equations $f_i(\cdot)$, $\varphi_i(\cdot)$, and $\Lambda_i(\cdot)[1-\Lambda_i(\cdot)]$ again symbolize general density functions of $\varepsilon_i$ as well as the corresponding density functions of the standard normal and standard logistic distributions.
The estimators of a discrete change of $p_{ij}(x_i, \theta)$ due to a discrete change $\Delta x_{ih}$ of a (particularly discrete) explanatory variable $x_{ih}$ in general ordered response models as well as in ordered probit and logit models are ($i = 1, \ldots, n; j = 1, \ldots, J$):

$$\Delta \hat{p}_{ij}(x_i, \hat{\theta}) = P(y_i = j|x_i + \Delta x_{ih}, \hat{\theta}) - P(y_i = j|x_i, \hat{\theta}) =$$

$$\left[ F_i(\kappa_j - \hat{\beta}'x_i - \hat{\beta}_h\Delta x_{ih}) - F_i(\kappa_{j-1} - \hat{\beta}'x_i - \hat{\beta}_h\Delta x_{ih}) \right] - \left[ F_i(\kappa_j - \hat{\beta}'x_i) - F_i(\kappa_{j-1} - \hat{\beta}'x_i) \right]$$

$$\Delta \hat{p}_{ij}(x_i, \hat{\theta}) = \left[ \Phi_i(\kappa_j - \hat{\beta}'x_i - \hat{\beta}_h\Delta x_{ih}) - \Phi_i(\kappa_{j-1} - \hat{\beta}'x_i - \hat{\beta}_h\Delta x_{ih}) \right] -$$

$$\left[ \Phi_i(\kappa_j - \hat{\beta}'x_i) - \Phi_i(\kappa_{j-1} - \hat{\beta}'x_i) \right]$$

$$\Delta \hat{p}_{ij}(x_i, \hat{\theta}) = \left[ \Lambda_i(\kappa_j - \hat{\beta}'x_i - \hat{\beta}_h\Delta x_{ih}) - \Lambda_i(\kappa_{j-1} - \hat{\beta}'x_i - \hat{\beta}_h\Delta x_{ih}) \right] -$$

$$\left[ \Lambda_i(\kappa_j - \hat{\beta}'x_i) - \Lambda_i(\kappa_{j-1} - \hat{\beta}'x_i) \right]$$

→ On the basis of the estimators of (partial) marginal probability effects and of discrete probability effects, it is again possible to estimate average marginal and discrete probability effects of an explanatory variable $x_{ih}$ across all i as well as marginal and discrete probability effects of $x_{ih}$ at the mean of the explanatory variables (the procedure for this estimation with STATA by considering differences of estimated probabilities is identical to the case of multinomial logit models)
Interpretation:

- The ML estimators $\hat{\beta}_h$ determine the estimated effect of the explanatory variable $x_{ih}$ on the (indeed not interesting) latent variable $y_i^*$. Since the values of $y_i^*$ are directly connected to the values of the ordered dependent variable $y_i$, the sign of $\hat{\beta}_h$ gives information about the direction of the estimated effect of $x_{ih}$ on increasing values of $y_i$ (but not on the single probabilities).

- Furthermore, the direction of the estimators of marginal and discrete probability effects for the categories $j = 1$ and $j = J$ is clear. For the estimators e.g. of marginal probability effects (with $\kappa_0 = -\infty$ and $\kappa_J = \infty$) it follows:

\[
\frac{\partial \hat{p}_{i1}(x_i, \hat{\theta})}{\partial x_{ih}} = \left[ f_i(-\infty) - f_i(\hat{\kappa}_1 - \hat{\beta}'x_i) \right] \hat{\beta}_h = -f_i(\hat{\kappa}_1 - \hat{\beta}'x_i) \hat{\beta}_h
\]

\[
\frac{\partial \hat{p}_{ij}(x_i, \hat{\theta})}{\partial x_{ih}} = \left[ f_i(\hat{\kappa}_{j-1} - \hat{\beta}'x_i) - f_i(\infty) \right] \hat{\beta}_h = f_i(\hat{\kappa}_{j-1} - \hat{\beta}'x_i) \hat{\beta}_h
\]

Therefore, a positive (negative) $\hat{\beta}_h$ implies that an increasing $x_{ih}$ leads to an increase (decrease) of $\hat{p}_{ij}(x_i, \hat{\theta})$ and a decrease (increase) of $\hat{p}_{i1}(x_i, \hat{\theta})$.

- In contrast, the direction of the estimators of marginal and discrete probability effects for $j = 2, \ldots, J - 1$ are ambiguous, even when positive (negative) $\hat{\beta}_h$ imply positive (negative) estimators of probability effects for high values of $y_i$ and negative (positive) estimators for small values of $y_i$. The sign of the estimators of probability effects changes exactly once at different values of $y_i$ when moving from small to high values of $y_i$. 
Example: Determinants of secondary school choice (I)

The determinants of the choice of 675 pupils in Germany between the three secondary school types Hauptschule, Realschule, and Gymnasium are again analyzed. In contrast to the previous application of a (pure) multinomial logit model, however, the natural ordering of the three categories is now used so that the coding of 1 for “Hauptschule”, the coding of 2 for “Realschule”, and the coding of 3 for “Gymnasium” of the dependent variable secondary school type (schooltype) indicates this ordering. The explanatory variables are the same as in the previous multinomial logit model analysis:

- Years of education of the mother (motheduc) as mainly interesting explanatory variable
- Dummy variable for labor force participation of the mother (mothinlf) that takes the value one if the mother is employed
- Logarithm of household income (loghhincome)
- Logarithm of household size (loghhsize)
- Rank by age among the siblings (birthorder)
- Year dummies for 1995-2002

The ML estimation of the ordered probit and logit models with STATA leads to the following results:
Example: Determinants of secondary school choice (II)

oprobit schooltype motheduc mothinlf loghhincome loghhsize birthorder year1995 year1996

Ordered probit regression

Number of obs = 675
LR chi2(13) = 203.31
Prob > chi2 = 0.0000
Log likelihood = -631.18653
Pseudo R2 = 0.1387

| schooltype | Coef. | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|------------|-------|-----------|-------|-------|---------------------|
| motheduc   | .2731466 | .0279933  | 9.76  | 0.000 | .2182808            |
|            |        |           |       |       | .3280124            |
| mothinlf   | -.1649158 | .0993593 | -1.66 | 0.097 | -.3596565           |
|            |        |           |       |       | .0298249            |
| loghhincome| .6492508 | .1049847  | 6.18  | 0.000 | .4434845            |
|            |        |           |       |       | .8550172            |
| loghhsize  | -.6145965 | .2016538 | -3.05 | 0.002 | -.1.009831          |
|            |        |           |       |       | -.2193624           |
| birthorder | -.1049289 | .0576033 | -1.82 | 0.069 | -.2178293           |
|            |        |           |       |       | .0079714            |
| year1995   | .0457511 | .1877388  | 0.24  | 0.807 | -.3222101           |
|            |        |           |       |       | .4137124            |
| year1996   | .0898004 | .1952066  | 0.46  | 0.645 | -.2.927974          |
|            |        |           |       |       | .4723982            |
| year1997   | -.2803249 | .1982527 | -1.41 | 0.157 | -.668893            |
|            |        |           |       |       | .1082432            |
| year1998   | .0750879 | .2120345  | 0.35  | 0.723 | -.3404922           |
|            |        |           |       |       | .4906679            |
| year1999   | -.1605177 | .2046103 | -0.78 | 0.433 | -.5615465           |
|            |        |           |       |       | .240511             |
| year2000   | .0009375 | .2037127  | 0.00  | 0.996 | -.3983321           |
|            |        |           |       |       | .4002071            |
| year2001   | .0330622 | .1972731  | 0.17  | 0.867 | -.3535859           |
|            |        |           |       |       | .4197104            |
| year2002   | -.105459 | .1960686  | -0.54 | 0.591 | -.4897463           |
|            |        |           |       |       | .2788283            |
|         /cut1 | 8.439737 | 1.061008  |       |       | 6.360201            |
|         /cut2 | 9.369887 | 1.068849  |       |       | 7.274981            |
|             |        |           |       |       | 11.46479            |
Example: Determinants of secondary school choice (III)


Ordered logistic regression

Number of obs = 675
LR chi2(13) = 205.38
Prob > chi2 = 0.0000
Log likelihood = -630.1549
Pseudo R2 = 0.1401

| schooltype | Coef. | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|------------|-------|-----------|-------|------|----------------------|
| motheduc   | .4678007 | .0514731 | 9.09  | 0.000 | .3669152 -- 0.5686862 |
| mothinlf   | -.3025754 | .1661858 | -1.82 | 0.069 | -.6282935 -- 0.0231427 |
| loghhincome| 1.149409 | .1933411 | 5.94  | 0.000 | .7704671 -- 1.52835   |
| loghhsize  | -1.063832 | .3429954 | -3.10 | 0.002 | -1.736091 -- 0.3915736 |
| birthorder | -.1734154 | .0975637 | -1.78 | 0.075 | -.3646368 -- 0.0178061 |
| year1995   | .0767835  | .3175815 | 0.24  | 0.809 | -.5728199 -- 0.623632 |
| year1996   | .1352987  | .3263589 | 0.41  | 0.678 | -.504353 -- 0.7749504 |
| year1997   | -.4836479 | .3314329 | -1.46 | 0.144 | -1.133245 -- 0.1659487 |
| year1998   | .1242217  | .35564   | 0.35  | 0.727 | -.5728199 -- 0.8212632 |
| year1999   | -.2312418 | .3402744 | -0.68 | 0.497 | -.8981673 -- 0.4356837 |
| year2000   | -.0389321 | .3385088 | -0.12 | 0.908 | -.7023971 -- 0.624533 |
| year2001   | .0343839  | .3297169 | 0.10  | 0.917 | -.6118494 -- 0.6806172 |
| year2002   | -.2241305 | .3284821 | -0.68 | 0.495 | -.8679436 -- 0.4196826 |

/cut1 | 14.8607 | 1.967535 | 11.00441 | 18.717 |
/cut2 | 16.41517 | 1.987395 | 12.51994 | 20.31039 |
Example: Determinants of secondary school choice (IV)

Interpretation:

- The estimation results in the ordered probit and ordered logit models are very similar.

- The values of 203.31 and 205.39 of the likelihood ratio test statistic imply that the null hypothesis that all 13 parameters of the explanatory variables are zero (which would imply that the ordered response models only comprise the two thresholds) can be rejected at any common significance levels.

- The parameter estimates for motheduc are positive and highly significantly different from zero due to the z statistics of 9.76 and 9.09 in the ordered probit and logit models, respectively.

- These parameter estimates therefore imply that the years of education of the mother have a strong significantly positive effect on the choice of Gymnasium and a strong significantly negative effect on the choice of Hauptschule, whereas the effect on the choice of Realschule is ambiguous from these estimation results.

- Similarly, loghhincome has a strong significantly positive effect on the choice of Gymnasium, whereas loghhsize has a strong significantly negative effect on the choice of Gymnasium.
Example: Determinants of secondary school choice (V)

Wald and likelihood ratio tests:
As an example, the null hypothesis that neither motheduc nor mothinlf has any effect on the secondary school choice, i.e. that the two corresponding parameters are zero, is tested after the ML estimation of the ordered probit model. The command for the Wald test in STATA is:

```
test motheduc mothinlf
  ( 1) [schooltype]motheduc = 0
  ( 2) [schooltype]mothinlf = 0
  ch2(  2) =   95.50
  Prob > ch2 =    0.0000
```

The corresponding commands for the likelihood ratio test in STATA are then:

```
estimates store unrestricted
                      year1999 year2000 year2001 year2002
estimates store restricted
lrtest unrestricted restricted
  Likelihood-ratio test      LR ch2(2) =  114.19
    (Assumption: restricted nested in unrestricted)  Prob > ch2 =  0.0000
```
Example: Determinants of secondary school choice (VI)

The estimation of the average marginal probability effects of motheduc across all 675 pupils on the choice of Hauptschule, Realschule, and Gymnasium leads to the following (shortened) STATA results in the ordered probit model:

```stata
margins, dydx(motheduc) predict(outcome(1))

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
</tr>
<tr>
<td>motheduc</td>
<td>- .0786704</td>
<td>.0075726</td>
<td>-10.39</td>
<td>0.000</td>
<td>-.0935124 -.0638283</td>
</tr>
</tbody>
</table>
```

```stata
margins, dydx(motheduc) predict(outcome(2))

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
</tr>
<tr>
<td>motheduc</td>
<td>- .007464</td>
<td>.0025874</td>
<td>-2.88</td>
<td>0.004</td>
<td>-.0125352 -.0023928</td>
</tr>
</tbody>
</table>
```

```stata
margins, dydx(motheduc) predict(outcome(3))

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
</tr>
<tr>
<td>motheduc</td>
<td>.0861344</td>
<td>.0073071</td>
<td>11.79</td>
<td>0.000</td>
<td>.0718128  .100456</td>
</tr>
</tbody>
</table>
```
Example: Determinants of secondary school choice (VII)

The estimation of corresponding marginal probability effects of motheduc at the means of the explanatory variables across all 675 pupils leads to the following (shortened) STATA results in the ordered probit model:

```
margins, dydx(motheduc) atmeans predict(outcome(1))
```

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
</tr>
<tr>
<td>motheduc</td>
<td>-.0859161</td>
<td>.0086732</td>
<td>-9.91</td>
<td>0.000</td>
<td>-.1029152</td>
</tr>
</tbody>
</table>

```
margins, dydx(motheduc) atmeans predict(outcome(2))
```

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
</tr>
<tr>
<td>motheduc</td>
<td>-.0199433</td>
<td>.0049997</td>
<td>-3.99</td>
<td>0.000</td>
<td>-.0297426</td>
</tr>
</tbody>
</table>

```
margins, dydx(motheduc) atmeans predict(outcome(3))
```

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
</tr>
<tr>
<td>motheduc</td>
<td>.1058594</td>
<td>.0110476</td>
<td>9.58</td>
<td>0.000</td>
<td>.0842065</td>
</tr>
</tbody>
</table>
Example: Determinants of secondary school choice (VIII)

The estimation of the average marginal probability effects of motheduc across all 675 pupils on the choice of Hauptschule, Realschule, and Gymnasium leads to the following (shortened) STATA results in the ordered logit model:

```
margins, dydx(motheduc) predict(outcome(1))

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
<th></th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
</tr>
<tr>
<td>motheduc</td>
<td>-.0795201</td>
<td>.0081198</td>
<td>-9.79</td>
<td>0.000</td>
<td>-.0954346 -.0636056</td>
</tr>
</tbody>
</table>
```

```
margins, dydx(motheduc) predict(outcome(2))

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
<th></th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
</tr>
<tr>
<td>motheduc</td>
<td>-.0082689</td>
<td>.0030382</td>
<td>-2.72</td>
<td>0.006</td>
<td>-.0142235 -.0023142</td>
</tr>
</tbody>
</table>
```

```
margins, dydx(motheduc) predict(outcome(3))

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
<th></th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
</tr>
<tr>
<td>motheduc</td>
<td>.087789</td>
<td>.0079214</td>
<td>11.08</td>
<td>0.000</td>
<td>.0722633 .1033146</td>
</tr>
</tbody>
</table>
```
Example: Determinants of secondary school choice (IX)

The estimation of corresponding marginal probability effects of motheduc at the means of the explanatory variables across all 675 pupils leads to the following (shortened) STATA results in the ordered logit model:

```
margins, dydx(motheduc) atmeans predict(outcome(1))
```

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
</tr>
<tr>
<td>motheduc</td>
<td>-.0849093</td>
<td>.0089859</td>
<td>-9.45</td>
<td>0.000</td>
<td>-.1025214 -.0672972</td>
</tr>
</tbody>
</table>

```
margins, dydx(motheduc) atmeans predict(outcome(2))
```

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
</tr>
<tr>
<td>motheduc</td>
<td>-.0276524</td>
<td>.007027</td>
<td>-3.94</td>
<td>0.000</td>
<td>-.0414252 -.0138797</td>
</tr>
</tbody>
</table>

```
margins, dydx(motheduc) atmeans predict(outcome(3))
```

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
</tr>
<tr>
<td>motheduc</td>
<td>.1125617</td>
<td>.0127688</td>
<td>8.82</td>
<td>0.000</td>
<td>.0875354  .1375881</td>
</tr>
</tbody>
</table>
Example: Determinants of secondary school choice (X)

Interpretation:

• The estimated average marginal probability effects or marginal probability effects at the means of the explanatory variables are very similar in the ordered probit and logit models, respectively, and strengthen the significantly positive effect of the years of education of the mother on the choice of Gymnasium and the significantly negative effect on the choice of Hauptschule.

• The estimation results particularly imply that motheduc has a significantly negative effect on the choice of Realschule. This estimated negative effect is stronger on the basis of the marginal probability effects at the means of the explanatory variables than of the average marginal probability effects.

• The estimated average marginal probability effect of motheduc of -0.0075 in the ordered probit model implies that an increase of the years of education of the mother by one (unit) leads to an approximately estimated decrease of the choice probability for Realschule by 0.75 percentage points, whereas the estimated value of -0.0200 at the means of the explanatory variables implies an approximately estimated decrease by 2.00 percentage points.

• While these estimated effects are very similar to those in the multinomial logit model, the estimated standard deviations are higher in the latter model which points to efficiency losses compared to ordered response models.
Example: Determinants of secondary school choice (XI)

The estimation of the average probability of the choice of Hauptschule across all 675 pupils leads to the following STATA results in the ordered probit model:

```
margins, predict(outcome(1))
Predictive margins                                    Number of obs  =  675
Model VCE : OIM
Expression : Pr(schooltype==1), predict(outcome(1))

|            Delta-method |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+---------------------+------------------+-+--------+------------------+-+--------+------------------+-+--------+------------------+
_cons       |  .2931277  .015854    18.49   0.000     .2620545     .324201
-------------+---------------------+------------------+-+--------+------------------+-+--------+------------------+
```

The estimation of (average) discrete changes of probabilities due to a discrete change of an explanatory variable requires the estimation of (average) probabilities for specific values of the explanatory variable. For example, the average change of the probability of the choice of Hauptschule across all 675 pupils due to an increase of motheduc from the minimum value of seven years to the maximum value of 18 years of education can be estimated on the basis of the estimated average probabilities at these specific values. The corresponding STATA commands after the ML estimation of the ordered probit model are:
Example: Determinants of secondary school choice (XII)

margins, at(motheduc=7) predict(outcome(1))

|          | Delta-method | Margin | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|----------|--------------|--------|-----------|-------|------|---------------------|
| _cons    | .6897138     | .0423121 | 16.30     | 0.000 |      | .6067836             | .772644 |  

margins, at(motheduc=18) predict(outcome(1))

|          | Delta-method | Margin | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|----------|--------------|--------|-----------|-------|------|---------------------|
| _cons    | .0100681     | .0052127 | 1.93      | 0.053 |      | -.0001485             | .0202848 |  

The estimated average decrease of the probability of the choice of Hauptschule is therefore 0.6897-0.0101=0.6796 or 67.96 percentage points for an increase from seven to 18 years.
Example: Determinants of secondary school choice (XIII)

In contrast, the estimation of e.g. the probability of the choice of Hauptschule for the maximum value of motheduc = 18 years at the means of the other individual characteristics with STATA leads to the following results:

```
margins, at((means)_all motheduc=18) predict(outcome(1))
```

| Delta-method |   Margins   | Std. Err.  | z    | P>|z| | [95% Conf. Interval] |
|--------------|-------------|------------|------|-----|----------------------|
| _cons        |  | .0065595   | .0037561 | 1.75 | 0.081 | -.0008023 | .0139213 |
4.3 Discussion of ordered probit and logit models

Underlying parallel regression assumption \((j = 1, \ldots, J - 1)\) with \(F_i(\cdot) = \Phi_i(\cdot)\) in the ordered probit model and \(F_i(\cdot) = \Lambda_i(\cdot)\) in the ordered logit model \((i = 1, \ldots, n)\):

\[
P(y_i \leq j|x_i, \theta) = P(y_i^* \leq \kappa_j|x_i, \theta) = F_i(\kappa_j - \beta'x_i)
\]

Consequences:

- These probabilities for several categories \(j\) only differ due to different threshold parameters, but not due to different parameter values.
- The partial derivatives of these probabilities with respect to an arbitrary explanatory variable \(x_{ih}\) are identical for all \(j\).
- With this assumption, the following dummy variables can be specified:

\[
d_{ij} = \begin{cases} 
1 & \text{if } y_i \leq j \\
0 & \text{if } y_i > j 
\end{cases}
\]

As a consequence, the slope parameters, but not the threshold parameters, in ordered probit and logit models with two categories could be estimated by binary probit or logit models with these dummy variables as dependent variables.

- This assumption is also the reason for the aforementioned property of ordered probit and logit models that the sign of the estimators of probability effects changes exactly once at different values of \(y_i\) when moving from small to high values of \(y_i\).
Latent variables in generalized ordered probit and logit models:
\[ y_i^* = \beta_j' x_i + \varepsilon_i \]
Here it is thus allowed that the parameter vector \( \beta \) changes across \( j \). It follows for the probabilities as discussed above with \( F_i(\cdot) = \Phi_i(\cdot) \) in the generalized ordered probit model and \( F_i(\cdot) = \Lambda_i(\cdot) \) in the generalized ordered logit model:
\[ P(y_i \leq j|x_i, \theta) = F_i(\kappa - \beta_j' x_i) \]
These generalized ordered probit and logit models are clearly more flexible than conventional ordered probit and logit models:
- For example, the partial derivatives of the probabilities with respect to an arbitrary explanatory variable \( x_{ih} \) can vary across the categories \( j \)
- Furthermore, generalized ordered probit and logit models do not necessarily imply that the sign of the estimators of probability effects changes only once at different values of \( y_i \) when moving from small to high values of \( y_i \)

Statistical testing of conventional ordered probit and logit models:
- The null hypothesis of the simple ordered probit and logit models is that all parameter vectors \( \beta_j \) are identical across \( j \) (which implies the corresponding single index assumption)
- This hypothesis can e.g. be tested by using a likelihood ratio test when the generalized ordered probit or logit models are estimated (where the simple ordered probit or logit models are the restricted models)
Evaluation of the use of generalized ordered probit and logit models:

- These models do not ensure that the aforementioned probabilities are restricted to the interval between zero and one. Due to the possibly varying $\beta_j$, it is also possible that the estimated probabilities $P(y_i \leq j | x_i, \hat{\theta})$ decrease in $j$ for specific values of the explanatory variables, which is, however, not logical.

- The ML estimators of the parameters (and thus e.g. estimators of marginal probability effects) in conventional ordered probit and logit models are inconsistent if the single index assumption is violated so that the ML estimation of generalized ordered probit or logit models is necessary in this case.

- An alternative is the use of multinomial logit (or probit) models which also lead to consistent ML estimators of the parameters if the single index assumption is violated, although they are then inefficient.

- To test the robustness of estimation results in empirical studies, it is certainly useful to compare the parameter estimates and the estimates of probability effects in different model approaches (e.g. binary probit and logit models, ordered probit and logit models, multinomial logit and probit models).

Interval data (e.g. income classes):

- For these data ordered probit and logit models can generally also be used.

- The main difference to the previous analysis is that the thresholds values (i.e. the interval bounds such as income bounds) are known so that these threshold parameters need not to be estimated additionally.