Ordinal Dependent Variables

Philip Leifeld

GV903: Advanced Research Methods, 28 January 2020, Week 18

University of Essex

1. Ordinal Data

Ordinal Dependent Variables

Ordinal data are a frequent data type in political science:

Ordinal Dependent Variables

Ordinal data are a frequent data type in political science:

- Surveys with Likert scales, for example:
  "disagree strongly", . . . , "neutral", . . . , "agree strongly".

Ordinal Dependent Variables

Ordinal data are a frequent data type in political science:

- Surveys with Likert scales, for example:
  "disagree strongly", ..., "neutral", ..., "agree strongly".
- Expert judgments and classifications of institutions. For example: quality of democracy – very high, high, medium, low, very low.

Ordinal Dependent Variables

Ordinal data are a frequent data type in political science:

- Surveys with Likert scales, for example:
  "disagree strongly", . . . , "neutral", . . . , "agree strongly".
- Expert judgments and classifications of institutions. For example: quality of democracy – very high, high, medium, low, very low.
- Military, bureaucratic, or party ranks.

Ordinal Dependent Variables

Ordinal data are a frequent data type in political science:

- Surveys with Likert scales, for example:
  "disagree strongly", . . . , "neutral", . . . , "agree strongly".
- Expert judgments and classifications of institutions. For example: quality of democracy – very high, high, medium, low, very low.
- Military, bureaucratic, or party ranks.
- Interval-scaled variables that were measured only in ordinal intervals, like income categories, age bands etc.

Ordinal Dependent Variables

Ordinal data are a frequent data type in political science:

- Surveys with Likert scales, for example:
  "disagree strongly", . . . , "neutral", . . . , "agree strongly".
- Expert judgments and classifications of institutions. For example: quality of democracy – very high, high, medium, low, very low.
- Military, bureaucratic, or party ranks.
- Interval-scaled variables that were measured only in ordinal intervals, like income categories, age bands etc.

Some people would argue that *all* ordinal data have underlying continuous variables and were measured imperfectly.

Ordinal Dependent Variables

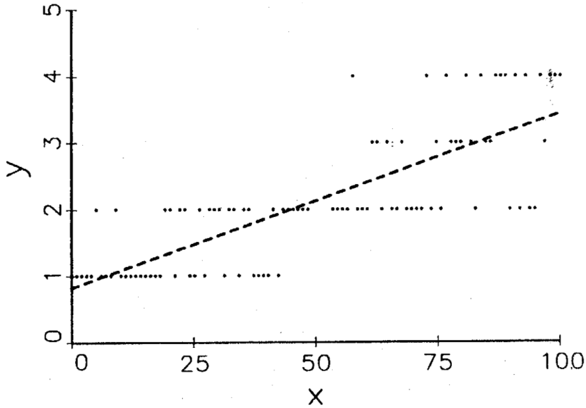
Ordinal data are a frequent data type in political science:

- Surveys with Likert scales, for example:
  "disagree strongly", . . . , "neutral", . . . , "agree strongly".
- Expert judgments and classifications of institutions. For example: quality of democracy – very high, high, medium, low, very low.
- Military, bureaucratic, or party ranks.
- Interval-scaled variables that were measured only in ordinal intervals, like income categories, age bands etc.

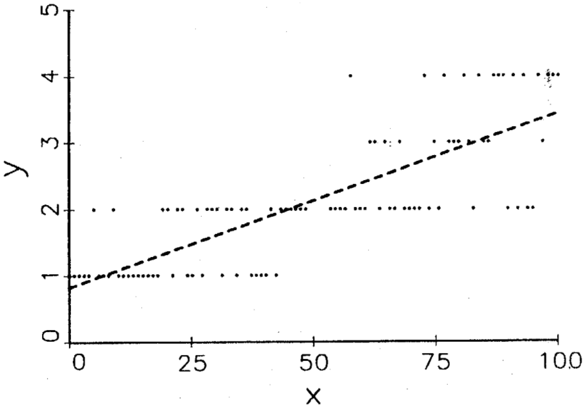
Some people would argue that *all* ordinal data have underlying continuous variables and were measured imperfectly.

Use ordinal logit, ordinal probit, or other ordinal models for these data!

What Happens if we Apply a Linear Model?

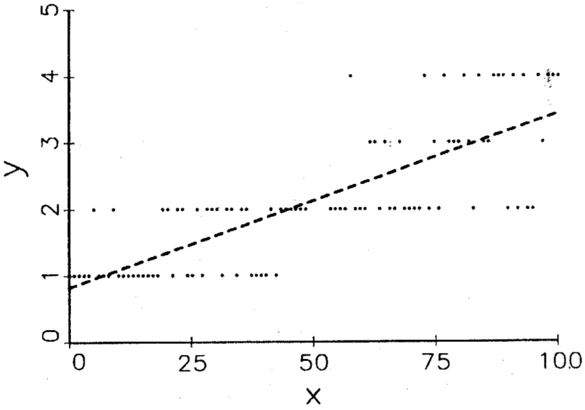


What Happens if we Apply a Linear Model?



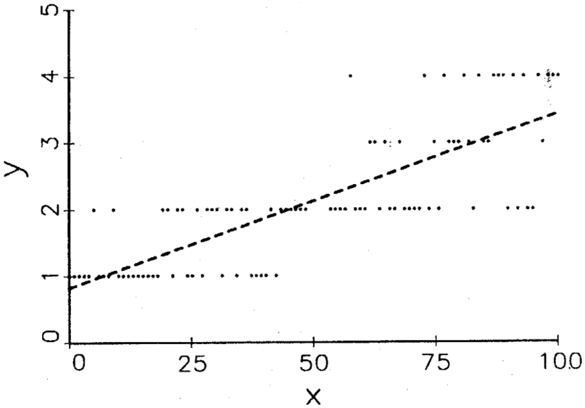
- Heteroskedasticity.

What Happens if we Apply a Linear Model?



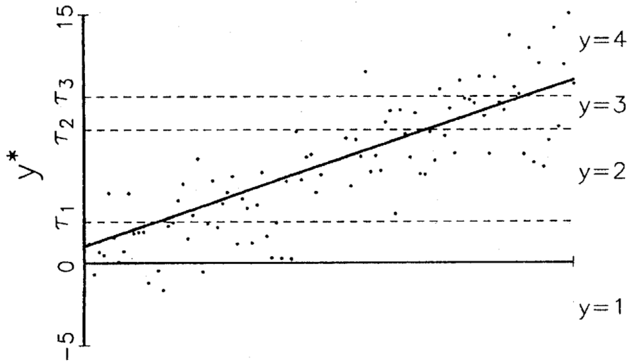
- Heteroskedasticity.
- Predictions outside of the range of the DV.

What Happens if we Apply a Linear Model?

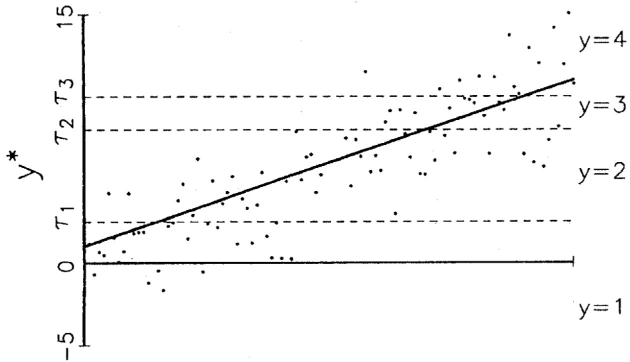


- Heteroskedasticity.
- Predictions outside of the range of the DV.
- Bias because we do not know where the cut points were drawn to discretise the continuous values.

Underlying Continuous Variable: Why the LM is Biased

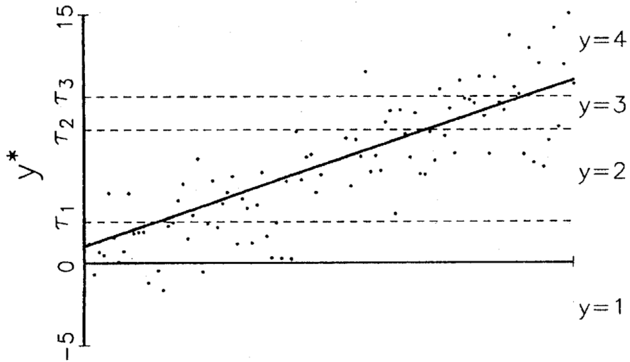


Underlying Continuous Variable: Why the LM is Biased



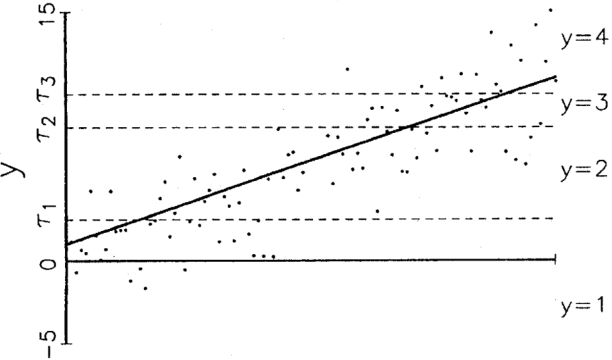
- The observed discrete values correspond to the bands between the cutpoints τ_1, τ_2, and τ_3.

Underlying Continuous Variable: Why the LM is Biased



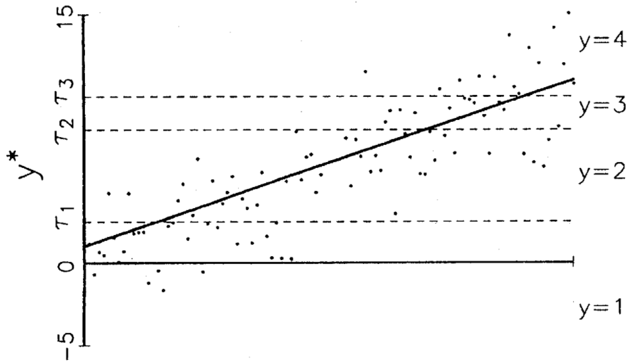
- The observed discrete values correspond to the bands between the cutpoints τ_1, τ_2, and τ_3.
- Four levels: below τ_1; between τ_1 and τ_2; between τ_2 and τ_3; above τ_3.

Underlying Continuous Variable: Why the LM is Biased



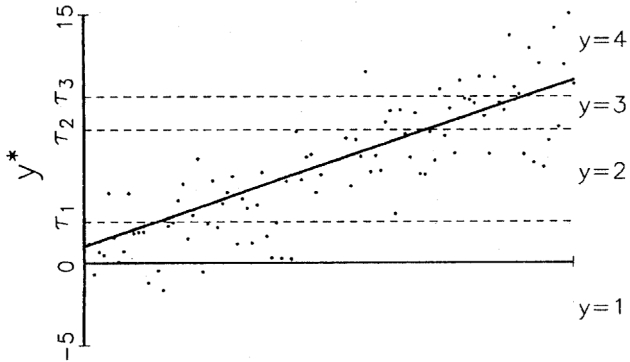
- The observed discrete values correspond to the bands between the cutpoints τ_1, τ_2, and τ_3.
- Four levels: below τ_1; between τ_1 and τ_2; between τ_2 and τ_3; above τ_3.
- The latent underlying continuous variable is unobserved.

Underlying Continuous Variable: Why the LM is Biased

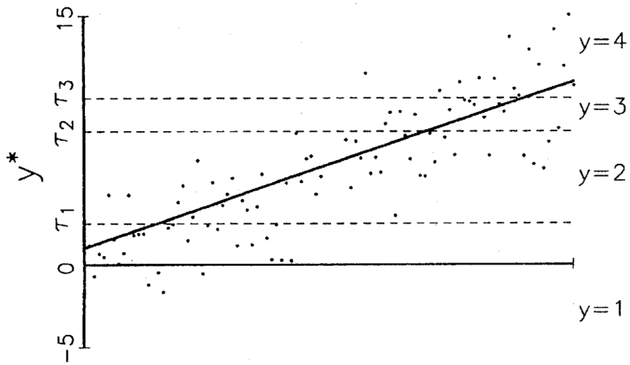


- The observed discrete values correspond to the bands between the cutpoints τ_1, τ_2, and τ_3.
- Four levels: below τ_1; between τ_1 and τ_2; between τ_2 and τ_3; above τ_3.
- The latent underlying continuous variable is unobserved.
- The cut points may not be equidistant.

Vague Intuition: How Ordinal Regression Works

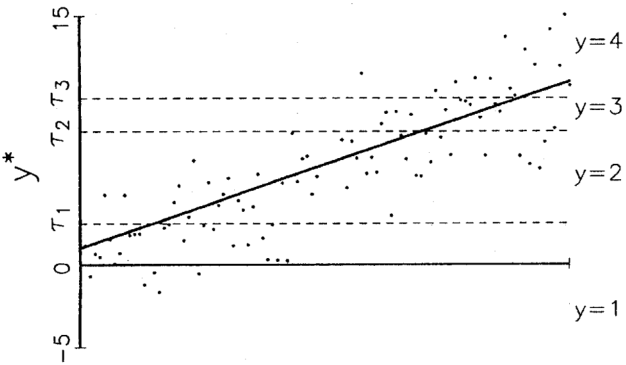


Vague Intuition: How Ordinal Regression Works



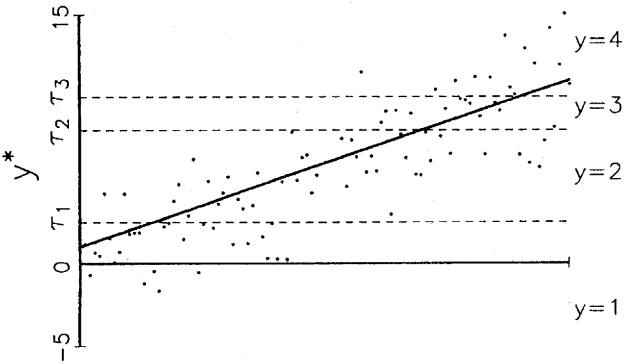
- Logit or probit models for each cut point: are we below or above the cut point?

Vague Intuition: How Ordinal Regression Works



- Logit or probit models for each cut point: are we below or above the cut point?
- The cut points are estimated from the data along with the β coefficients.

Vague Intuition: How Ordinal Regression Works



- Logit or probit models for each cut point: are we below or above the cut point?
- The cut points are estimated from the data along with the β coefficients.
- The cut points serve as intercepts for the different levels of the latent variable.

Exercise

Should the Polity IV score of a country be modelled using an ordinal regression model or a linear model?

From the Polity IV website:

> *The "Polity Score" captures [the] regime authority spectrum on a 21-point scale ranging from* -10 *(hereditary monarchy) to* $+10$ *(consolidated democracy). The Polity scores can also be converted into regime categories in a suggested three part categorization of "autocracies"* (-10 *to* -6)*, "anocracies"* (-5 *to* $+5$ *and three special values:* -66, -77 *and* -88)*, and "democracies"* ($+6$ *to* $+10$)*.*

List arguments for and against each choice.

A New Data Type: Ordinal Factors

To model ordinal data, we need a new data type: ordinal factors.

```
v <- c("Negative", "Negative", "Neutral", "Positive", "Neutral")
v
## [1] "Negative" "Negative" "Neutral"  "Positive" "Neutral"

vf <- ordered(v, levels = c("Negative", "Neutral", "Positive"))
vf
## [1] Negative Negative Neutral  Positive Neutral
## Levels: Negative < Neutral < Positive

class(vf)
## [1] "ordered" "factor"

as.numeric(vf) # saved in the right order...
## [1] 1 1 2 3 2
```

The `levels` argument tells `R` the right order of the factors.

World Values Survey Data

```
library("carData")
data(WVS)
head(WVS)
##       poverty religion degree country age gender
## 1  Too Little      yes     no     USA  44   male
## 2 About Right      yes     no     USA  40 female
## 3  Too Little      yes     no     USA  36 female
## 4    Too Much      yes    yes     USA  25 female
## 5  Too Little      yes    yes     USA  39   male
## 6 About Right      yes     no     USA  80 female
```

World Values Survey Data

```
library("carData")
data(WVS)
head(WVS)
##       poverty religion degree country age gender
## 1  Too Little      yes     no     USA  44   male
## 2 About Right      yes     no     USA  40 female
## 3  Too Little      yes     no     USA  36 female
## 4    Too Much      yes    yes     USA  25 female
## 5  Too Little      yes    yes     USA  39   male
## 6 About Right      yes     no     USA  80 female
```

- Poverty: "Do you think that what the government is doing for people in poverty in this country is about the right amount, too much, or too little?"

World Values Survey Data

```
library("carData")
data(WVS)
head(WVS)
##        poverty religion degree country age gender
## 1  Too Little      yes     no     USA  44   male
## 2 About Right      yes     no     USA  40 female
## 3  Too Little      yes     no     USA  36 female
## 4    Too Much      yes    yes     USA  25 female
## 5  Too Little      yes    yes     USA  39   male
## 6 About Right      yes     no     USA  80 female
```

- Poverty: "Do you think that what the government is doing for people in poverty in this country is about the right amount, too much, or too little?"
- Ordered factor with three levels.

World Values Survey Data

```
library("carData")
data(WVS)
head(WVS)
##        poverty religion degree country age gender
## 1  Too Little      yes     no     USA  44   male
## 2 About Right      yes     no     USA  40 female
## 3  Too Little      yes     no     USA  36 female
## 4    Too Much      yes    yes     USA  25 female
## 5  Too Little      yes    yes     USA  39   male
## 6 About Right      yes     no     USA  80 female
```

- Poverty: "Do you think that what the government is doing for people in poverty in this country is about the right amount, too much, or too little?"
- Ordered factor with three levels.
- Country: Australia; Norway; Sweden; USA.

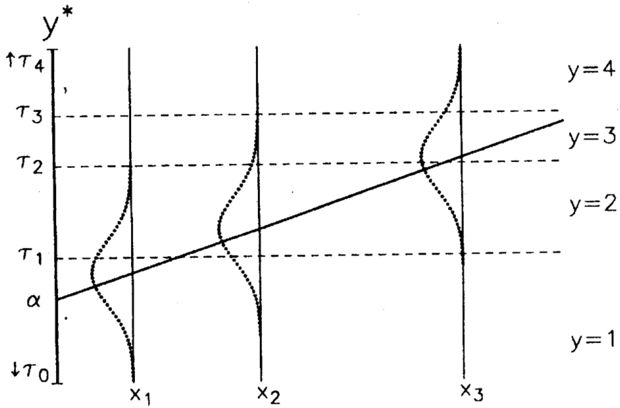
World Values Survey Data

```
library("carData")
data(WVS)
head(WVS)
##        poverty religion degree country age gender
## 1  Too Little      yes     no     USA  44   male
## 2 About Right      yes     no     USA  40 female
## 3  Too Little      yes     no     USA  36 female
## 4    Too Much      yes    yes     USA  25 female
## 5  Too Little      yes    yes     USA  39   male
## 6 About Right      yes     no     USA  80 female
```

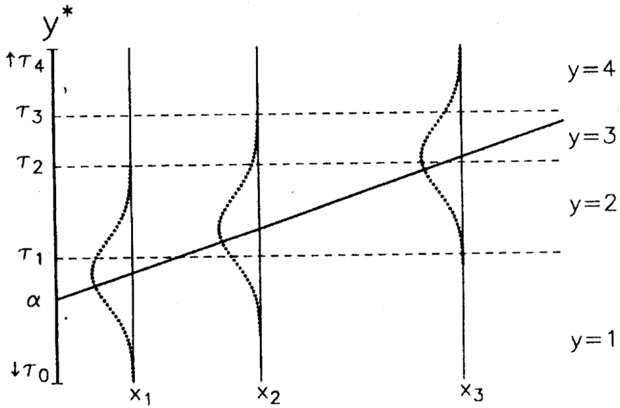
- Poverty: "Do you think that what the government is doing for people in poverty in this country is about the right amount, too much, or too little?"
- Ordered factor with three levels.
- Country: Australia; Norway; Sweden; USA.
- 5,381 observations.

2. Ordinal Logit and Probit

Distribution of y^* Given x for Ordered Regression

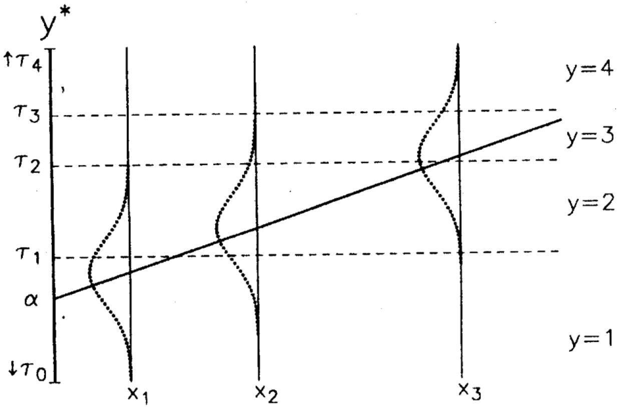


Distribution of y^* Given x for Ordered Regression



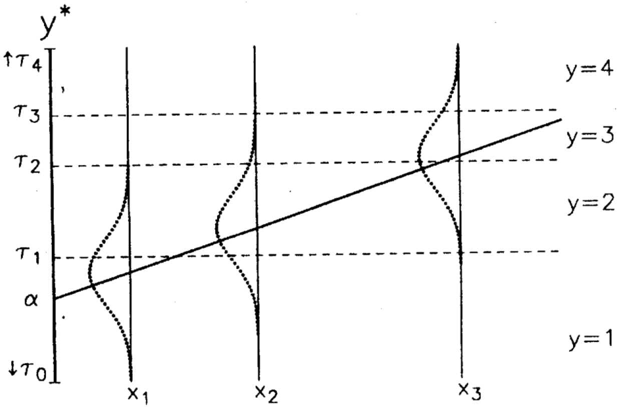
- We imagine a regression line through the underlying latent variable.

Distribution of y^* Given x for Ordered Regression



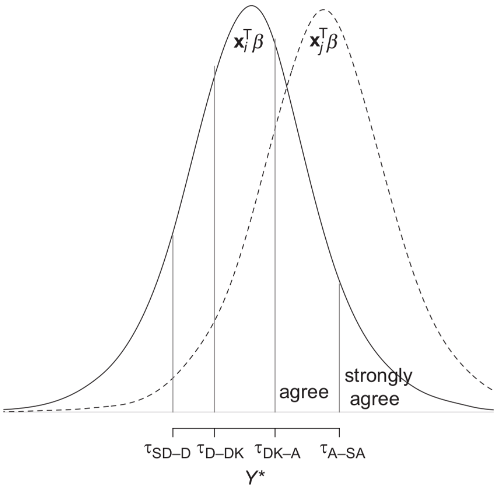
- We imagine a regression line through the underlying latent variable.
- There are logistic or standard normal distributions around the regression line.

Distribution of y^* Given x for Ordered Regression

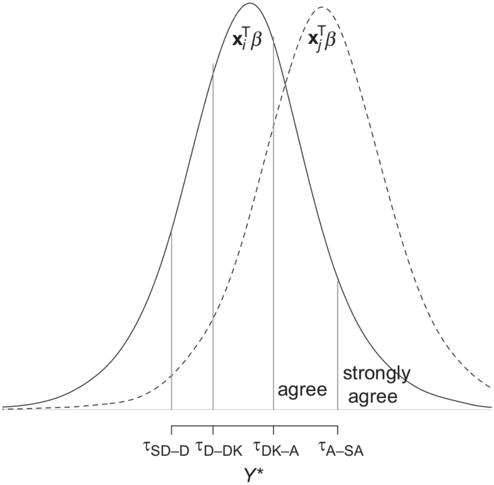


- We imagine a regression line through the underlying latent variable.
- There are logistic or standard normal distributions around the regression line.
- For each x value, we can derive the probability to be in each level from the cumulative distribution function.

Distribution of y^* Given x for Ordered Regression

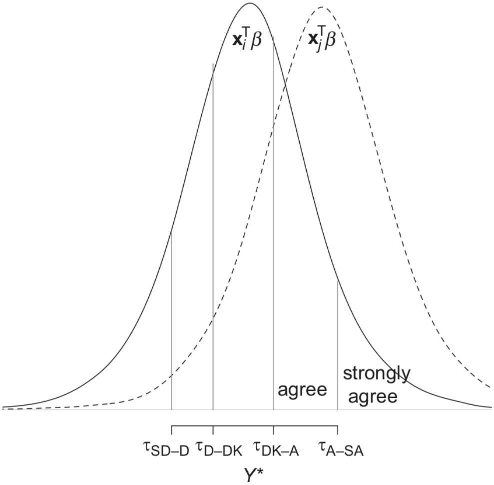


Distribution of y^* Given x for Ordered Regression



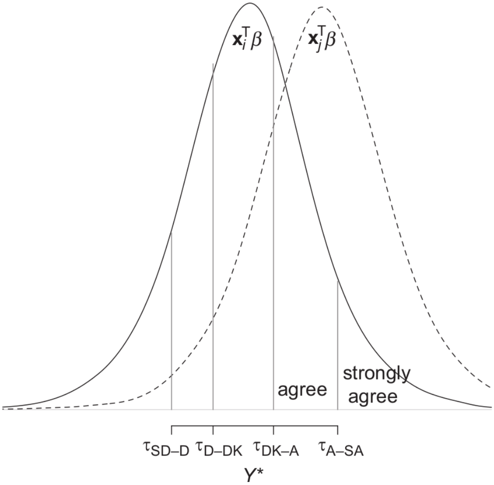
- This is the same thing, illustrated slightly differently.

Distribution of y^* Given x for Ordered Regression



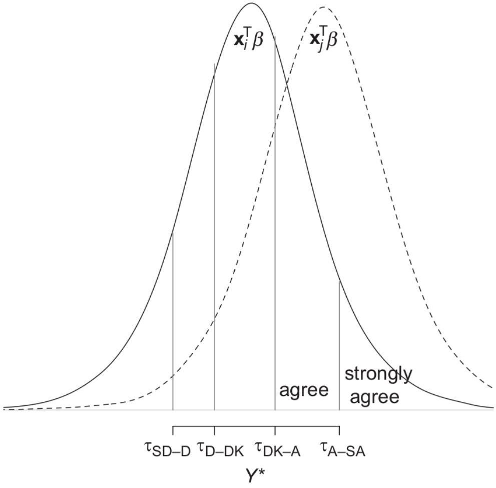
- This is the same thing, illustrated slightly differently.
- For each predicted $\mathbf{x}_i^\top \boldsymbol{\beta}$ observation, we draw the logistic or standard normal distribution around the point.

Distribution of y^* Given x for Ordered Regression



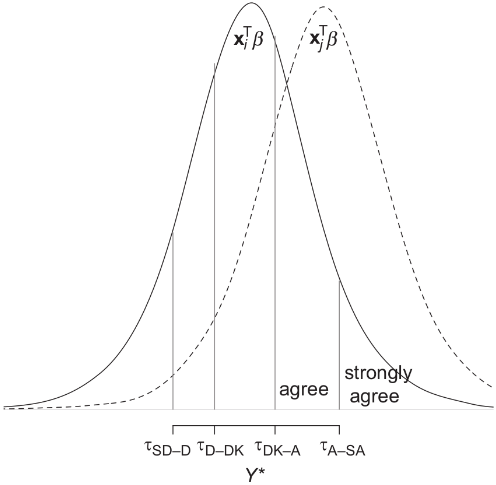
- This is the same thing, illustrated slightly differently.
- For each predicted $\mathbf{x}_i^\top \boldsymbol{\beta}$ observation, we draw the logistic or standard normal distribution around the point.
- We can use this to compute the probability for each x value to be in each interval.

Distribution of y^* Given x for Ordered Regression



- This is the same thing, illustrated slightly differently.
- For each predicted $\mathbf{x}_i^\top\boldsymbol{\beta}$ observation, we draw the logistic or standard normal distribution around the point.
- We can use this to compute the probability for each x value to be in each interval.
- This is how the model maps the linear component onto the discrete observations.

Distribution of y^* Given x for Ordered Regression



- This is the same thing, illustrated slightly differently.
- For each predicted $\mathbf{x}_i^\top \boldsymbol{\beta}$ observation, we draw the logistic or standard normal distribution around the point.
- We can use this to compute the probability for each x value to be in each interval.
- This is how the model maps the linear component onto the discrete observations.

We use the *cdf* to compute these probabilities!

Example: Calculating Interval Probabilities

Let's say the cut points are $\tau_1 = -3$; $\tau_2 = 1.4$; $\tau_3 = 6.3$.

Example: Calculating Interval Probabilities

Let's say the cut points are $\tau_1 = -3$; $\tau_2 = 1.4$; $\tau_3 = 6.3$.

Let's say $r = \mathbf{x}_i^\top \boldsymbol{\beta} = 1.75$.

Example: Calculating Interval Probabilities

Let's say the cut points are $\tau_1 = -3$; $\tau_2 = 1.4$; $\tau_3 = 6.3$.

Let's say $r = \mathbf{x}_i^\top \boldsymbol{\beta} = 1.75$.

Then the probability that the value belongs to the first interval is:

$$P(r \leq \tau_1) = \int_{-\infty}^{\tau_1} f(r)dr = F(\tau_1)$$

where $f(\cdot)$ is the *pdf* and $F(\cdot)$ the *cdf* of the logistic or standard normal distribution.

Example: Calculating Interval Probabilities

Let's say the cut points are $\tau_1 = -3$; $\tau_2 = 1.4$; $\tau_3 = 6.3$.

Let's say $r = \mathbf{x}_i^\top \boldsymbol{\beta} = 1.75$.

Then the probability that the value belongs to the first interval is:

$$P(r \leq \tau_1) = \int_{-\infty}^{\tau_1} f(r)dr = F(\tau_1)$$

where $f(\cdot)$ is the *pdf* and $F(\cdot)$ the *cdf* of the logistic or standard normal distribution.

We can compute this in `R`:

```
pnorm(-3, mean = 1.75)
## [1] 1.017083e-06
```

Example: Calculating Interval Probabilities

Let's say the cut points are $\tau_1 = -3$; $\tau_2 = 1.4$; $\tau_3 = 6.3$.

Let's say $r = \mathbf{x}_i^\top \boldsymbol{\beta} = 1.75$.

Then the probability that the value belongs to the first interval is:

$$P(r \leq \tau_1) = \int_{-\infty}^{\tau_1} f(r)dr = F(\tau_1)$$

where $f(\cdot)$ is the *pdf* and $F(\cdot)$ the *cdf* of the logistic or standard normal distribution.

We can compute this in R:

```
pnorm(-3, mean = 1.75)
## [1] 1.017083e-06
```

That's a very low probability.

Example: Calculating Interval Probabilities

Probability that the value is in the second interval:

Example: Calculating Interval Probabilities

Probability that the value is in the second interval:

$$P(\tau_1 < r \leq \tau_2) = \int_{-\infty}^{\tau_2} f(r)dr - \int_{-\infty}^{\tau_1} f(r)dr = F(\tau_2) - F(\tau_1)$$

Example: Calculating Interval Probabilities

Probability that the value is in the second interval:

$$P(\tau_1 < r \leq \tau_2) = \int_{-\infty}^{\tau_2} f(r)dr - \int_{-\infty}^{\tau_1} f(r)dr = F(\tau_2) - F(\tau_1)$$

```
pnorm(1.4, mean = 1.75) - pnorm(-3, mean = 1.75)
## [1] 0.3631683
```

Example: Calculating Interval Probabilities

Probability that the value is in the second interval:

$$P(\tau_1 < r \leq \tau_2) = \int_{-\infty}^{\tau_2} f(r)dr - \int_{-\infty}^{\tau_1} f(r)dr = F(\tau_2) - F(\tau_1)$$

```
pnorm(1.4, mean = 1.75) - pnorm(-3, mean = 1.75)
## [1] 0.3631683
```

Third and fourth interval:

Example: Calculating Interval Probabilities

Probability that the value is in the second interval:

$$P(\tau_1 < r \leq \tau_2) = \int_{-\infty}^{\tau_2} f(r)dr - \int_{-\infty}^{\tau_1} f(r)dr = F(\tau_2) - F(\tau_1)$$

```
pnorm(1.4, mean = 1.75) - pnorm(-3, mean = 1.75)
## [1] 0.3631683
```

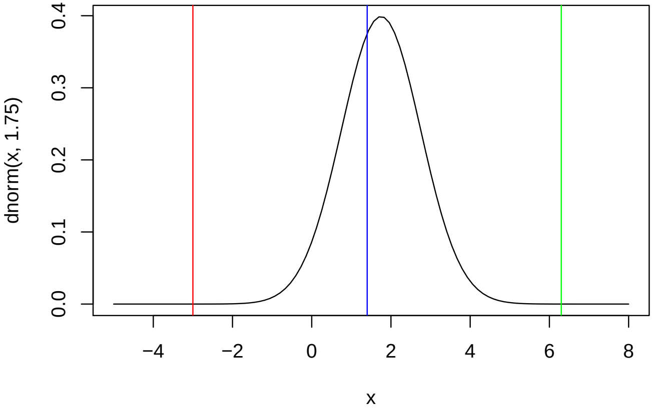
Third and fourth interval:

```
pnorm(6.3, mean = 1.75) - pnorm(1.4, mean = 1.75)
## [1] 0.636828

1 - pnorm(6.3, mean = 1.75)
## [1] 2.682296e-06
```

Visual Representation of these Results

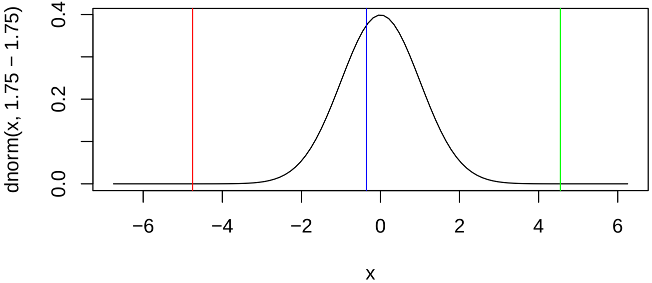
```
x <- seq(-5, 8, length.out = 100)
plot(x, dnorm(x, 1.75), type = "l")
abline(v = -3, col = "red")
abline(v = 1.4, col = "blue")
abline(v = 6.3, col = "green")
```



Since the Φ and Λ distributions are centred around zero, we can obtain the same result by subtracting the predicted value each time and thus shifting the distribution:

$$P(y_i = 1|\mathbf{x}_i) = P(\tau_0 \le \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i < \tau_1|\mathbf{x}_i)$$
$$\Leftrightarrow P(y_i = 1|\mathbf{x}_i) = P(\tau_0 - \mathbf{x}_i^\top \boldsymbol{\beta} \le \epsilon_i < \tau_1 - \mathbf{x}_i^\top \boldsymbol{\beta}|\mathbf{x}_i)$$

```
x <- seq(-5 - 1.75, 8 - 1.75, length.out = 100)
plot(x, dnorm(x, 1.75 - 1.75), type = "l")
abline(v = -3 - 1.75, col = "red")
abline(v = 1.4 - 1.75, col = "blue")
abline(v = 6.3 - 1.75, col = "green")
```



Ordered Probit: Individual Probabilities

Probability of observation i to belong into interval m:

$$P(y_i = m) = \begin{cases} \Phi(\tau_1 - \mathbf{x}_i^\top \boldsymbol{\beta}) & \text{for } m: \text{first level} \\ \Phi(\tau_2 - \mathbf{x}_i^\top \boldsymbol{\beta}) - \Phi(\tau_1 - \mathbf{x}_i^\top \boldsymbol{\beta}) & \text{for } m: \text{second level} \\ \dots & \text{for } m: \text{third etc level} \\ 1 - \Phi(\tau_M - \mathbf{x}_i^\top \boldsymbol{\beta}) & \text{for } m: \text{last level} \end{cases}$$

where Φ denotes the cumulative distribution function of the standard normal distribution:

$$\Phi(\epsilon) = \int_{-\infty}^{\epsilon} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{r^2}{2}\right\} dr$$

Ordered Logit: Individual Probabilities

Probability of observation i to belong into interval m:

$$P(y_i = m) = \begin{cases} \Lambda(\tau_1 - \mathbf{x}_i^\top \boldsymbol{\beta}) & \text{for } m : \text{first level} \\ \Lambda(\tau_2 - \mathbf{x}_i^\top \boldsymbol{\beta}) - \Lambda(\tau_1 - \mathbf{x}_i^\top \boldsymbol{\beta}) & \text{for } m : \text{second level} \\ \dots & \text{for } m : \text{third etc level} \\ 1 - \Lambda(\tau_M - \mathbf{x}_i^\top \boldsymbol{\beta}) & \text{for } m : \text{last level} \end{cases}$$

where Λ denotes the cumulative distribution function of the logistic distribution:

$$\Lambda(\epsilon) = \frac{\exp\{\epsilon\}}{1 + \exp\{\epsilon\}}$$

Likelihood Function

Likelihood Function

So the individual probabilities per observation are:

$$P(y_i = m | \mathbf{x}_i, \boldsymbol{\beta}, \boldsymbol{\tau}) = F(\tau_m - \mathbf{x}_i^\top \boldsymbol{\beta}) - F(\tau_{m-1} - \mathbf{x}_i^\top \boldsymbol{\beta})$$

Likelihood Function

So the individual probabilities per observation are:

$$P(y_i = m|\mathbf{x}_i, \boldsymbol{\beta}, \boldsymbol{\tau}) = F(\tau_m - \mathbf{x}_i^\top \boldsymbol{\beta}) - F(\tau_{m-1} - \mathbf{x}_i^\top \boldsymbol{\beta})$$

To construct the likelihood function, we compute the joint probability for all observations $i \dots n$ and all intervals $m \dots M$:

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\tau}|\mathbf{y}, \mathbf{X}) = \prod_{m=1}^{M} \prod_{y_i=m} \left[F(\tau_m - \mathbf{x}_i^\top \boldsymbol{\beta}) - F(\tau_{m-1} - \mathbf{x}_i^\top \boldsymbol{\beta})\right]$$

Likelihood Function

So the individual probabilities per observation are:

$$P(y_i = m|\mathbf{x}_i, \boldsymbol{\beta}, \boldsymbol{\tau}) = F(\tau_m - \mathbf{x}_i^\top \boldsymbol{\beta}) - F(\tau_{m-1} - \mathbf{x}_i^\top \boldsymbol{\beta})$$

To construct the likelihood function, we compute the joint probability for all observations $i \dots n$ and all intervals $m \dots M$:

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\tau}|\mathbf{y}, \mathbf{X}) = \prod_{m=1}^{M} \prod_{y_i=m} \left[F(\tau_m - \mathbf{x}_i^\top \boldsymbol{\beta}) - F(\tau_{m-1} - \mathbf{x}_i^\top \boldsymbol{\beta})\right]$$

We take the product for all observations and all intervals, but it's multiplied only for those cases where y_i is observed to equal m.

Likelihood Function

So the individual probabilities per observation are:

$$P(y_i = m|\mathbf{x}_i, \boldsymbol{\beta}, \boldsymbol{\tau}) = F(\tau_m - \mathbf{x}_i^\top \boldsymbol{\beta}) - F(\tau_{m-1} - \mathbf{x}_i^\top \boldsymbol{\beta})$$

To construct the likelihood function, we compute the joint probability for all observations $i \ldots n$ and all intervals $m \ldots M$:

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\tau}|\mathbf{y}, \mathbf{X}) = \prod_{m=1}^{M} \prod_{y_i=m} \left[F(\tau_m - \mathbf{x}_i^\top \boldsymbol{\beta}) - F(\tau_{m-1} - \mathbf{x}_i^\top \boldsymbol{\beta})\right]$$

We take the product for all observations and all intervals, but it's multiplied only for those cases where y_i is observed to equal m.

Corresponding log likelihood:

Likelihood Function

So the individual probabilities per observation are:

$$P(y_i = m|\mathbf{x}_i, \boldsymbol{\beta}, \boldsymbol{\tau}) = F(\tau_m - \mathbf{x}_i^\top \boldsymbol{\beta}) - F(\tau_{m-1} - \mathbf{x}_i^\top \boldsymbol{\beta})$$

To construct the likelihood function, we compute the joint probability for all observations $i \dots n$ and all intervals $m \dots M$:

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\tau}|\mathbf{y}, \mathbf{X}) = \prod_{m=1}^{M} \prod_{y_i=m} \left[F(\tau_m - \mathbf{x}_i^\top \boldsymbol{\beta}) - F(\tau_{m-1} - \mathbf{x}_i^\top \boldsymbol{\beta})\right]$$

We take the product for all observations and all intervals, but it's multiplied only for those cases where y_i is observed to equal m.

Corresponding log likelihood:

$$\log \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\tau}|\mathbf{y}, \mathbf{X}) = \sum_{m=1}^{M} \sum_{y_i=m} \log \left[F(\tau_m - \mathbf{x}_i^\top \boldsymbol{\beta}) - F(\tau_{m-1} - \mathbf{x}_i^\top \boldsymbol{\beta})\right]$$

Estimation and Identification

- Maximising this log likelihood gives us estimates both for the β and τ parameters.

Estimation and Identification

- Maximising this log likelihood gives us estimates both for the β and τ parameters.
- Since we get τ intercepts for each interval, the intercept that we normally estimate for all observations is superfluous.

Estimation and Identification

- Maximising this log likelihood gives us estimates both for the β and τ parameters.
- Since we get τ intercepts for each interval, the intercept that we normally estimate for all observations is superfluous.
- In fact, it would be collinear with the τ parameters if we included it, and the model could not be estimated.

Estimation and Identification

- Maximising this log likelihood gives us estimates both for the β and τ parameters.
- Since we get τ intercepts for each interval, the intercept that we normally estimate for all observations is superfluous.
- In fact, it would be collinear with the τ parameters if we included it, and the model could not be estimated.
- Therefore we just drop the intercept and interpret the cut point τ estimates as our intercepts.

Estimation and Identification

- Maximising this log likelihood gives us estimates both for the β and τ parameters.
- Since we get τ intercepts for each interval, the intercept that we normally estimate for all observations is superfluous.
- In fact, it would be collinear with the τ parameters if we included it, and the model could not be estimated.
- Therefore we just drop the intercept and interpret the cut point τ estimates as our intercepts.
- Alternatively, we could drop the first τ estimate and keep the intercept. It does not make a difference for the β estimates.

Estimation and Identification

- Maximising this log likelihood gives us estimates both for the β and τ parameters.
- Since we get τ intercepts for each interval, the intercept that we normally estimate for all observations is superfluous.
- In fact, it would be collinear with the τ parameters if we included it, and the model could not be estimated.
- Therefore we just drop the intercept and interpret the cut point τ estimates as our intercepts.
- Alternatively, we could drop the first τ estimate and keep the intercept. It does not make a difference for the β estimates.
- Estimation as usual computationally via IWLS, BFGS etc, for example via the `optim` function in `R`.

Estimation and Identification

- Maximising this log likelihood gives us estimates both for the β and τ parameters.
- Since we get τ intercepts for each interval, the intercept that we normally estimate for all observations is superfluous.
- In fact, it would be collinear with the τ parameters if we included it, and the model could not be estimated.
- Therefore we just drop the intercept and interpret the cut point τ estimates as our intercepts.
- Alternatively, we could drop the first τ estimate and keep the intercept. It does not make a difference for the β estimates.
- Estimation as usual computationally via IWLS, BFGS etc, for example via the `optim` function in `R`.
- There are several ready-to-use implementations in `R`, for example the `polr` function in the `MASS` package.

3. Ordered Logit and Probit in `R`

World Values Survey: Poverty

```
library("MASS")
model1 <- polr(poverty ~ religion + degree + country +
    age + gender, data = WVS, method = "logistic",
    Hess = TRUE)
model2 <- polr(poverty ~ religion + degree + country +
    age + gender, data = WVS, method = "probit",
    Hess = TRUE)

class(model1)
## [1] "polr"

WVS$poverty_ols <- as.numeric(WVS$poverty)
head(WVS$poverty_ols)
## [1] 1 2 1 3 1 2

model3 <- glm(poverty_ols ~ religion + degree +
    country + age + gender, data = WVS)
```

Model Output

```
summary(model1)
## Call:
## polr(formula = poverty ~ religion + degree + country + age +
##     gender, data = WVS, Hess = TRUE, method = "logistic")
##
## Coefficients:
##                  Value Std. Error t value
## religionyes    0.17973   0.077346   2.324
## degreeyes      0.14092   0.066193   2.129
## countryNorway -0.32235   0.073766  -4.370
## countrySweden -0.60330   0.079494  -7.589
## countryUSA     0.61777   0.070665   8.742
## age            0.01114   0.001561   7.139
## gendermale     0.17637   0.052972   3.329
##
## Intercepts:
##                         Value   Std. Error t value
## Too Little|About Right  0.7298  0.1041      7.0128
## About Right|Too Much    2.5325  0.1103     22.9496
##
## Residual Deviance: 10402.59
## AIC: 10420.59
```

Comparison of the Models

	Ordered Logit	Ordered Probit	Linear Model
religionyes	$0.18\ (0.08)^{*}$	$0.11\ (0.05)^{*}$	$0.08\ (0.03)^{**}$
degreeyes	$0.14\ (0.07)^{*}$	$0.08\ (0.04)^{*}$	$0.05\ (0.02)$
countryNorway	$-0.32\ (0.07)^{***}$	$-0.25\ (0.05)^{***}$	$-0.16\ (0.03)^{***}$
countrySweden	$-0.60\ (0.08)^{***}$	$-0.41\ (0.05)^{***}$	$-0.25\ (0.03)^{***}$
countryUSA	$0.62\ (0.07)^{***}$	$0.37\ (0.04)^{***}$	$0.25\ (0.03)^{***}$
age	$0.01\ (0.00)^{***}$	$0.01\ (0.00)^{***}$	$0.00\ (0.00)^{***}$
gendermale	$0.18\ (0.05)^{***}$	$0.10\ (0.03)^{**}$	$0.06\ (0.02)^{**}$
(Intercept)			$1.37\ (0.04)^{***}$
AIC	10420.59	10370.25	11444.39
BIC	10479.91	10429.57	11503.71
Log Likelihood	-5201.30	-5176.13	-5713.20
Deviance	10402.59	10352.25	2633.90
Num. obs.	5381	5381	5381

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Comparison of the Models

	Ordered Logit	Ordered Probit	Linear Model
religionyes	0.18 (0.08)*	0.11 (0.05)*	0.08 (0.03)**
degreeyes	0.14 (0.07)*	0.08 (0.04)*	0.05 (0.02)
countryNorway	−0.32 (0.07)***	−0.25 (0.05)***	−0.16 (0.03)***
countrySweden	−0.60 (0.08)***	−0.41 (0.05)***	−0.25 (0.03)***
countryUSA	0.62 (0.07)***	0.37 (0.04)***	0.25 (0.03)***
age	0.01 (0.00)***	0.01 (0.00)***	0.00 (0.00)***
gendermale	0.18 (0.05)***	0.10 (0.03)**	0.06 (0.02)**
(Intercept)			1.37 (0.04)***
AIC	10420.59	10370.25	11444.39
BIC	10479.91	10429.57	11503.71
Log Likelihood	−5201.30	−5176.13	−5713.20
Deviance	10402.59	10352.25	2633.90
Num. obs.	5381	5381	5381

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

The linear model is biased, fits worse, and yields different results.

Interpretation

Interpretation

- Ordered regression models are notoriously hard to interpret.

Interpretation

- Ordered regression models are notoriously hard to interpret.
- What does a coefficient mean substantively, let's say for age?

Interpretation

- Ordered regression models are notoriously hard to interpret.
- What does a coefficient mean substantively, let's say for age?
- With each additional year of age, the odds of getting from "too little" to "about right" or from "about right" to "too much" increase by `(exp(0.01114) - 1) * 100 =` 1.1202281 per cent.

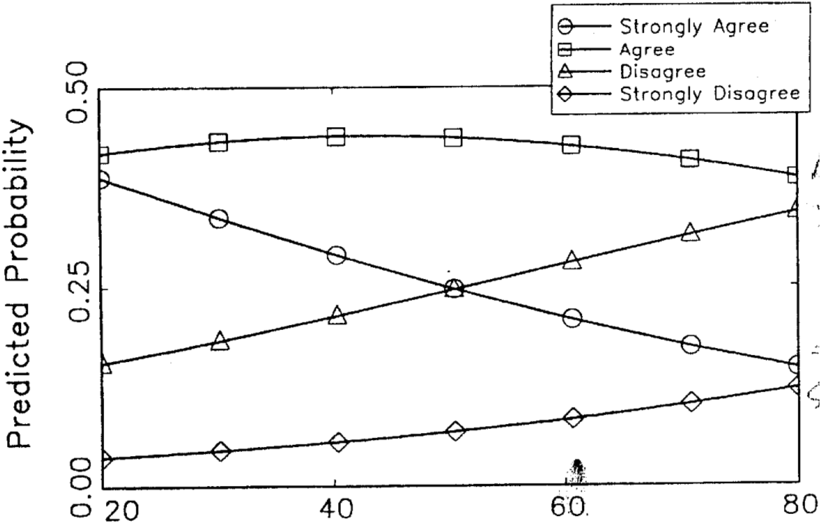
Interpretation

- Ordered regression models are notoriously hard to interpret.
- What does a coefficient mean substantively, let's say for age?
- With each additional year of age, the odds of getting from "too little" to "about right" or from "about right" to "too much" increase by `(exp(0.01114) - 1) * 100 =` 1.1202281 per cent.
- As this is conditional on the state of all variables, it makes more sense to interpret models by employing prediction and based on scenarios – just like with other GLMs.

Interpretation

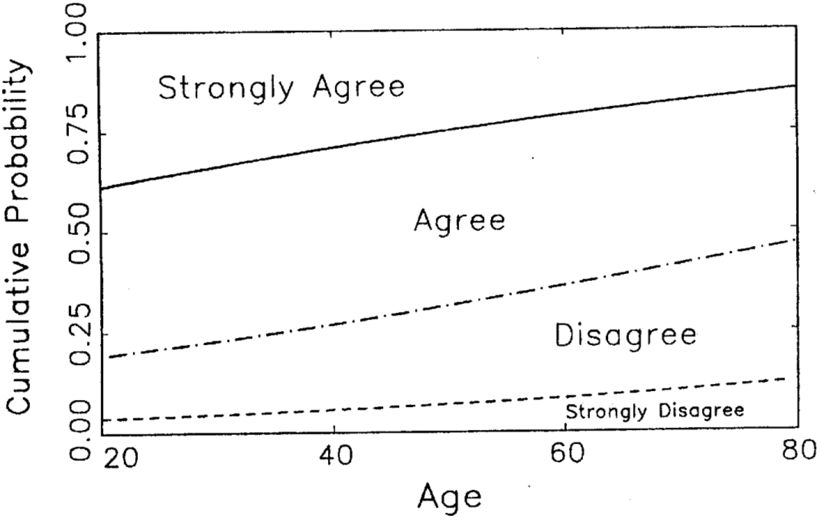
- Ordered regression models are notoriously hard to interpret.
- What does a coefficient mean substantively, let's say for age?
- With each additional year of age, the odds of getting from "too little" to "about right" or from "about right" to "too much" increase by `(exp(0.01114) - 1) * 100 =` 1.1202281 per cent.
- As this is conditional on the state of all variables, it makes more sense to interpret models by employing prediction and based on scenarios – just like with other GLMs.
- You can also look at marginal effects and other quantities, as detailed in Long (1997), but this is less common in applied work in polisci.

Interpretation: Predicted Probabilities



Age on the x axis. Probabilities can go up and down again.
Otherwise same procedure as with logit and probit models.

Interpretation: Cumulative Probability



Same as before, but stacked up to display cumulative probabilities.

4. The Parallel Regression Assumption

The Parallel Regression Assumption

The Parallel Regression Assumption

- Also known as the *proportional odds assumption* (logit case).

The Parallel Regression Assumption

- Also known as the *proportional odds assumption* (logit case).
- We estimate only one slope per independent variable.

The Parallel Regression Assumption

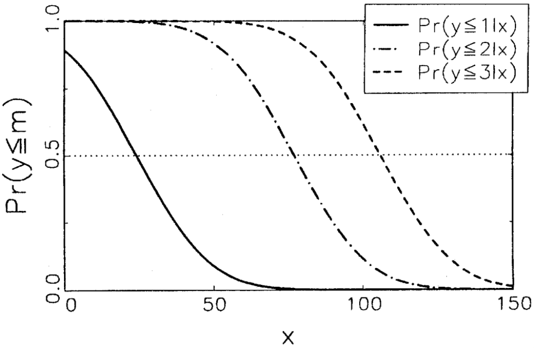
- Also known as the *proportional odds assumption* (logit case).
- We estimate only one slope per independent variable.
- This obviously implies that the slope must be identical across all cutpoints/intervals.

The Parallel Regression Assumption

- Also known as the *proportional odds assumption* (logit case).
- We estimate only one slope per independent variable.
- This obviously implies that the slope must be identical across all cutpoints/intervals.
- If it were not, what would the coefficients tell us at all?

The Parallel Regression Assumption

- Also known as the *proportional odds assumption* (logit case).
- We estimate only one slope per independent variable.
- This obviously implies that the slope must be identical across all cutpoints/intervals.
- If it were not, what would the coefficients tell us at all?



Testing the Parallel Regression Assumption

- There are several ways to test the assumption.

Testing the Parallel Regression Assumption

- There are several ways to test the assumption.
- You can fit separate binary logit or probit models for the different cut-off levels and see if the slopes are similar.

Testing the Parallel Regression Assumption

- There are several ways to test the assumption.
- You can fit separate binary logit or probit models for the different cut-off levels and see if the slopes are similar.
- The Brant test: a Wald-type test statistic that tests for the whole model and for each model term separately if the assumption holds.

Testing the Parallel Regression Assumption

- There are several ways to test the assumption.
- You can fit separate binary logit or probit models for the different cut-off levels and see if the slopes are similar.
- The Brant test: a Wald-type test statistic that tests for the whole model and for each model term separately if the assumption holds.
- The assumption is frequently violated in empirical applications. If it is:

Testing the Parallel Regression Assumption

- There are several ways to test the assumption.
- You can fit separate binary logit or probit models for the different cut-off levels and see if the slopes are similar.
- The Brant test: a Wald-type test statistic that tests for the whole model and for each model term separately if the assumption holds.
- The assumption is frequently violated in empirical applications. If it is:
  - Use separate binary models to fit the different DGPs.

Testing the Parallel Regression Assumption

- There are several ways to test the assumption.
- You can fit separate binary logit or probit models for the different cut-off levels and see if the slopes are similar.
- The Brant test: a Wald-type test statistic that tests for the whole model and for each model term separately if the assumption holds.
- The assumption is frequently violated in empirical applications. If it is:
  - Use separate binary models to fit the different DGPs.
  - Use a multinomial logit or probit model, which do not assume an ordinal measurement level.

Separate Logit Models

```
WVS_a <- WVS
WVS_a$poverty_ols[WVS_a$poverty_ols == 1] <- 0
WVS_a$poverty_ols[WVS_a$poverty_ols %in% 2:3] <- 1

WVS_b <- WVS
WVS_b$poverty_ols[WVS_b$poverty_ols %in% 1:2] <- 0
WVS_b$poverty_ols[WVS_b$poverty_ols == 3] <- 1

a <- glm(poverty_ols ~ religion + degree + country +
    age + gender, data = WVS_a,
    family = binomial(link = "logit"))

b <- glm(poverty_ols ~ religion + degree + country +
    age + gender, data = WVS_b,
    family = binomial(link = "logit"))
```

Separate Logit Models

```
screenreg(list(a, b), single.row = TRUE)
##
## =========================================================
##                    Model 1                 Model 2
## ---------------------------------------------------------
## (Intercept)           -0.70 (0.11) ***     -2.59 (0.16) ***
## religionyes            0.11 (0.08)          0.37 (0.11) ***
## degreeyes              0.18 (0.07) *        0.02 (0.11)
## countryNorway         -0.13 (0.08)         -1.78 (0.18) ***
## countrySweden         -0.44 (0.08) ***     -2.07 (0.21) ***
## countryUSA             0.36 (0.07) ***      0.90 (0.09) ***
## age                    0.01 (0.00) ***      0.01 (0.00) ***
## gendermale             0.20 (0.06) ***      0.09 (0.08)
## ---------------------------------------------------------
## AIC                 7315.16              3911.40
## BIC                 7367.89              3964.12
## Log Likelihood     -3649.58             -1947.70
## Deviance            7299.16              3895.40
## Num. obs.           5381                 5381
## =========================================================
## *** p < 0.001, ** p < 0.01, * p < 0.05
```

```r
library("brant")
brant(model1)
## --------------------------------------------
## Test for X2 df probability
## --------------------------------------------
## Omnibus 261.22 7 0
## religionyes 6.07 1 0.01
## degreeyes 2.08 1 0.15
## countryNorway 83.64 1 0
## countrySweden 59.58 1 0
## countryUSA 43.85 1 0
## age 0.7 1 0.4
## gendermale 1.69 1 0.19
## --------------------------------------------
##
## H0: Parallel Regression Assumption holds
##                       X2 df  probability
## Omnibus       261.2204841  7 1.130805e-52
## religionyes     6.0674567  1 1.376951e-02
## degreeyes       2.0848479  1 1.487670e-01
## countryNorway  83.6373448  1 5.943954e-20
## countrySweden  59.5786391  1 1.175037e-14
## countryUSA     43.8454840  1 3.553501e-11
## age             0.6951767  1 4.044091e-01
## gendermale      1.6894831  1 1.936691e-01
```

Parallel Regression Assumption: Results

Parallel Regression Assumption: Results

- Age, gender, and possibly degree do not violate the assumption.

Parallel Regression Assumption: Results

- Age, gender, and possibly degree do not violate the assumption.
- Religion and the country dummies are offenders.

Parallel Regression Assumption: Results

- Age, gender, and possibly degree do not violate the assumption.
- Religion and the country dummies are offenders.
- Overall, the model violates the assumption.

Parallel Regression Assumption: Results

- Age, gender, and possibly degree do not violate the assumption.
- Religion and the country dummies are offenders.
- Overall, the model violates the assumption.
- Both methods show these differences, but the Brant test may be more conclusive.

Parallel Regression Assumption: Results

- Age, gender, and possibly degree do not violate the assumption.
- Religion and the country dummies are offenders.
- Overall, the model violates the assumption.
- Both methods show these differences, but the Brant test may be more conclusive.
- We could now do one of four things:

Parallel Regression Assumption: Results

- Age, gender, and possibly degree do not violate the assumption.
- Religion and the country dummies are offenders.
- Overall, the model violates the assumption.
- Both methods show these differences, but the Brant test may be more conclusive.
- We could now do one of four things:
    1. Choose to ignore the evidence and keep the model anyway. This is often done but is bad practice because the model does not really tell us anything.

Parallel Regression Assumption: Results

- Age, gender, and possibly degree do not violate the assumption.
- Religion and the country dummies are offenders.
- Overall, the model violates the assumption.
- Both methods show these differences, but the Brant test may be more conclusive.
- We could now do one of four things:
    1. Choose to ignore the evidence and keep the model anyway. This is often done but is bad practice because the model does not really tell us anything.
    2. Keep the separate binary models and interpret them because the DGP is indeed different.

Parallel Regression Assumption: Results

- Age, gender, and possibly degree do not violate the assumption.
- Religion and the country dummies are offenders.
- Overall, the model violates the assumption.
- Both methods show these differences, but the Brant test may be more conclusive.
- We could now do one of four things:
    1. Choose to ignore the evidence and keep the model anyway. This is often done but is bad practice because the model does not really tell us anything.
    2. Keep the separate binary models and interpret them because the DGP is indeed different.
    3. Treat the three choices as unordered/categorical and use an appropriate model, e. g., multinomial choice models.

Parallel Regression Assumption: Results

- Age, gender, and possibly degree do not violate the assumption.
- Religion and the country dummies are offenders.
- Overall, the model violates the assumption.
- Both methods show these differences, but the Brant test may be more conclusive.
- We could now do one of four things:
  1. Choose to ignore the evidence and keep the model anyway. This is often done but is bad practice because the model does not really tell us anything.
  2. Keep the separate binary models and interpret them because the DGP is indeed different.
  3. Treat the three choices as unordered/categorical and use an appropriate model, e. g., multinomial choice models.
  4. Use a cumulative link model (CLM)...

The `ordinal` Package

- The `ordinal` package provides two main functions:

The `ordinal` Package

- The `ordinal` package provides two main functions:
  1. `clm` for cumulative link models (including ordered logit and probit).

The `ordinal` Package

- The `ordinal` package provides two main functions:
  1. `clm` for cumulative link models (including ordered logit and probit).
  2. `clmm` for mixed CLMs – same thing but with random slopes and intercepts.

The `ordinal` Package

- The `ordinal` package provides two main functions:
  1. `clm` for cumulative link models (including ordered logit and probit).
  2. `clmm` for mixed CLMs – same thing but with random slopes and intercepts.
- CLMs are more flexible than ordered logit and probit because they allow you to specify some effects as nominal.

The `ordinal` Package

- The `ordinal` package provides two main functions:
  1. `clm` for cumulative link models (including ordered logit and probit).
  2. `clmm` for mixed CLMs – same thing but with random slopes and intercepts.
- CLMs are more flexible than ordered logit and probit because they allow you to specify some effects as nominal.
- This means selected slopes are allowed to vary across levels.

The `ordinal` Package

- The `ordinal` package provides two main functions:
  1. `clm` for cumulative link models (including ordered logit and probit).
  2. `clmm` for mixed CLMs – same thing but with random slopes and intercepts.
- CLMs are more flexible than ordered logit and probit because they allow you to specify some effects as nominal.
- This means selected slopes are allowed to vary across levels.
- This solves the problem with the parallel regression assumption.

Cumulative Link Models (CLM) in R

Let's first re-estimate Model 1, the ordered logit model:

Cumulative Link Models (CLM) in R

Let's first re-estimate Model 1, the ordered logit model:

```
library("ordinal")
model4 <- clm(poverty ~ religion + degree + country +
    age + gender, data = WVS, link = "logit")
```

(Output not shown here because identical to Model 1).

Cumulative Link Models (CLM) in R

Let's first re-estimate Model 1, the ordered logit model:

```
library("ordinal")
model4 <- clm(poverty ~ religion + degree + country +
    age + gender, data = WVS, link = "logit")
```

(Output not shown here because identical to Model 1).

`country` was one of the offending variables. We can now re-specify `country` as a variable that has a nominal effect:

Cumulative Link Models (CLM) in R

Let's first re-estimate Model 1, the ordered logit model:

```
library("ordinal")
model4 <- clm(poverty ~ religion + degree + country +
    age + gender, data = WVS, link = "logit")
```

(Output not shown here because identical to Model 1).

`country` was one of the offending variables. We can now re-specify `country` as a variable that has a nominal effect:

```
model5 <- clm(poverty ~ religion + degree + age +
    gender, nominal = ~ country, data = WVS,
    link = "logit")
summary(model5) # results on next slide
```

```
## formula: poverty ~ religion + degree + age + gender
## nominal: ~country
## data:    WVS
##
##  link  threshold nobs logLik   AIC      niter max.grad cond.H
##  logit flexible  5381 -5020.12 10064.25 7(0)  9.47e-13 1.7e+05
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## religionyes 0.149106   0.076176   1.957  0.05030 .
## degreeyes   0.141428   0.066552   2.125  0.03358 *
## age         0.010605   0.001556   6.816 9.39e-12 ***
## gendermale  0.173844   0.052915   3.285  0.00102 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
##                                      Estimate Std. Error z value
## Too Little|About Right.(Intercept)    0.71784    0.10396   6.905
## About Right|Too Much.(Intercept)      2.36086    0.11461  20.599
## Too Little|About Right.countryNorway  0.12271    0.07779   1.577
## About Right|Too Much.countryNorway    1.78119    0.18170   9.803
## Too Little|About Right.countrySweden  0.44490    0.08240   5.399
## About Right|Too Much.countrySweden    2.06867    0.21338   9.695
## Too Little|About Right.countryUSA    -0.36255    0.07340  -4.939
## About Right|Too Much.countryUSA      -0.87275    0.08666 -10.071
```

Assessing the Parallel Regression Assumption with CLMs

We can now do a likelihood ratio test to see if the parallel regression assumption was actually violated:

Assessing the Parallel Regression Assumption with CLMs

We can now do a likelihood ratio test to see if the parallel regression assumption was actually violated:

```
anova(model4, model5)
## Likelihood ratio tests of cumulative link models:
##
##        formula:
## model4 poverty ~ religion + degree + country + age + gender
## model5 poverty ~ religion + degree + age + gender
##        nominal: link: threshold:
## model4 ~1       logit flexible
## model5 ~country logit flexible
##
##        no.par   AIC  logLik LR.stat df Pr(>Chisq)
## model4      9 10421 -5201.3
## model5     12 10064 -5020.1  362.35  3  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Assessing the Parallel Regression Assumption with CLMs

We can now do a likelihood ratio test to see if the parallel regression assumption was actually violated:

```
anova(model4, model5)
## Likelihood ratio tests of cumulative link models:
##
##        formula:
## model4 poverty ~ religion + degree + country + age + gender
## model5 poverty ~ religion + degree + age + gender
##        nominal: link: threshold:
## model4 ~1       logit flexible
## model5 ~country logit flexible
##
##        no.par   AIC  logLik LR.stat df Pr(>Chisq)
## model4      9 10421 -5201.3
## model5     12 10064 -5020.1  362.35  3  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This is the best way to test the assumption.

Assessing the Parallel Regression Assumption with CLMs

We can now do a likelihood ratio test to see if the parallel regression assumption was actually violated:

```
anova(model4, model5)
## Likelihood ratio tests of cumulative link models:
##
##        formula:
## model4 poverty ~ religion + degree + country + age + gender
## model5 poverty ~ religion + degree + age + gender
##        nominal: link: threshold:
## model4 ~1       logit flexible
## model5 ~country logit flexible
##
##        no.par   AIC  logLik LR.stat df Pr(>Chisq)
## model4      9 10421 -5201.3
## model5     12 10064 -5020.1  362.35  3  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This is the best way to test the assumption.

And it offers a flexible remedy.