Linking process to outcome

The seqlogit package

Maarten L. Buis

Department of Social Research Methodology
Vrije Universiteit Amsterdam
http://home.fsw.vu.nl/m.buis/
Introduction

- **Aim:** describe the effect of explanatory variables on a process and its outcome.
Introduction

- **Aim**: describe the effect of explanatory variables on a process and its outcome.
- **Process**: a sequence of ‘choices’
Introduction

- **Aim**: describe the effect of explanatory variables on a process and its outcome.
- **Process**: a sequence of ‘choices’, e.g. moving from one type of education to the next.
Introduction

- **Aim:** describe the effect of explanatory variables on a process and its outcome.
- **Process:** a sequence of ‘choices’, e.g. moving from one type of education to the next.
- **Outcome:** The final outcome of the process
Introduction

▸ **Aim:** describe the effect of explanatory variables on a process and its outcome.

▸ **Process:** a sequence of ‘choices’, e.g. moving from one type of education to the next.

▸ **Outcome:** The final outcome of the process, e.g. highest achieved level of education.
Outline

Process and Outcome

Empirical example

The seqlogit package
Outline

Process and Outcome

Empirical example

The seqlogit package
Sequential logit model

- This model is known under a variety of other names:
  - sequential response model (Maddala 1983),
  - continuation ratio logit (Agresti 2002),
  - model for nested dichotomies (Fox 1997), and
  - the Mare model (Shavit and Blossfeld 1993) (after Mare 1981)
Sequential logit model

- Model each choice separately using a (m)logit on the sub-sample that is ‘at risk’

**Figure:** Hypothetical educational system

```
no education
  ↘ p1
   ↘ 1 - p1
      ↘ exit 1 - p1
          ↘ exit 1 - p1
              ↘ exit 0

primary
  ↘ p2
   ↘ 1 - p2
      ↘ exit 1 - p2
          ↘ exit 1 - p2
              ↘ exit 0

secondary
  ↘ p3
   ↘ 1 - p3
      ↘ exit 1 - p3
          ↘ exit 1 - p3
              ↘ exit 0

tertiary
  ↘ l3 = 16
   ↘ l2 = 12
      ↘ l1 = 6
```
The `seqlogit` package translates sequential logit to end result

$$\hat{p}_{ki} = \frac{\exp(\alpha_k + \lambda_k SES_i)}{1 + \exp(\alpha_k + \lambda_k SES_i)} \quad \text{if} \quad y_{k-1i} = 1$$

$$E(ed) = (1 - \hat{p}_{1i})l_0 + \hat{p}_{1i}(1 - \hat{p}_{2i})l_1 + \hat{p}_{1i}\hat{p}_{2i}(1 - \hat{p}_{3i})l_2 + \hat{p}_{1i}\hat{p}_{2i}\hat{p}_{3i}l_3$$
The effect of the explanatory variable SES on the outcome is the increase in expected highest achieved level of education for a unit increase in SES, i.e. a first derivative:
The effect of the explanatory variable SES on the outcome is the increase in expected highest achieved level of education for a unit increase in SES, i.e. a first derivative:

\[
\frac{\partial E(ed)}{\partial SES} = \\
\{1 \times \hat{p}_{1i}(1 - \hat{p}_{1i}) \times [(1 - \hat{p}_{2})l_1 + \hat{p}_2(1 - \hat{p}_3)l_2 + \hat{p}_2\hat{p}_3 l_3 - l_0]\} \lambda_1 + \\
\{\hat{p}_{1i} \times \hat{p}_{2i}(1 - \hat{p}_{2i}) \times [(1 - \hat{p}_3)l_2 + \hat{p}_3 l_3 - l_1]\} \lambda_2 + \\
\{\hat{p}_{1i}\hat{p}_{2i} \times \hat{p}_{3i}(1 - \hat{p}_{3i}) \times [l_3 - l_2]\} \lambda_3
\]
Effect of explanatory variable on outcome

\[
\frac{\partial E(\text{ed})}{\partial \text{SES}} = \\
\{1 \times \hat{p}_1(1 - \hat{p}_1) \times [(1 - \hat{p}_2)l_1 + \hat{p}_2(1 - \hat{p}_3)l_2 + \hat{p}_2\hat{p}_3l_3 - l_0]\} \lambda_1 + \\
\{\hat{p}_1 \times \hat{p}_2(1 - \hat{p}_2) \times [(1 - \hat{p}_3)l_2 + \hat{p}_3l_3 - l_1]\} \lambda_2 + \\
\{\hat{p}_1\hat{p}_2 \times \hat{p}_3(1 - \hat{p}_3) \times [l_3 - l_2]\} \lambda_3
\]
Effect of explanatory variable on outcome

\[ \frac{\partial E(\text{ed})}{\partial \text{SES}} = \{1 \times \hat{p}_{1i}(1 - \hat{p}_{1i}) \times [(1 - \hat{p}_2)l_1 + \hat{p}_2(1 - \hat{p}_3)l_2 + \hat{p}_2\hat{p}_3l_3 - l_0]\} \lambda_1 + \{\hat{p}_{1i} \times \hat{p}_{2i}(1 - \hat{p}_{2i}) \times [(1 - \hat{p}_3)l_2 + \hat{p}_3l_3 - l_1]\} \lambda_2 + \{\hat{p}_{1i}\hat{p}_{2i} \times \hat{p}_{3i}(1 - \hat{p}_{3i}) \times [l_3 - l_2]\} \lambda_3 \]
Effect of explanatory variable on outcome

Proportion at risk

\[
\frac{\partial E(ed)}{\partial SES} = \\
\{1 \times \hat{p}_{1i}(1 - \hat{p}_{1i}) \times [(1 - \hat{p}_{2})l_1 + \hat{p}_{2}(1 - \hat{p}_{3})l_2 + \hat{p}_{2}\hat{p}_{3}l_3 - l_0]\} \lambda_1 + \\
\{\hat{p}_{1i}\hat{p}_{2i}(1 - \hat{p}_{2i}) \times [(1 - \hat{p}_{3})l_2 + \hat{p}_{3}l_3 - l_1]\} \lambda_2 + \\
\{\hat{p}_{1i}\hat{p}_{2i}\hat{p}_{3i}(1 - \hat{p}_{3i}) \times [l_3 - l_2]\} \lambda_3
\]
Effect of explanatory variable on outcome

variance of the variable indicating whether one passes or not

\[
\frac{\partial E(ed)}{\partial SES} = \\
\{1 \times \hat{p}_{1i}(1 - \hat{p}_{1i})\times[(1 - \hat{p}_2)l_1 + \hat{p}_2(1 - \hat{p}_3)l_2 + \hat{p}_2\hat{p}_3l_3 - l_0]\}\lambda_1 + \\
\{\hat{p}_{1i}\times\hat{p}_{2i}(1 - \hat{p}_{2i})\times[(1 - \hat{p}_3)l_2 + \hat{p}_3l_3 - l_1]\}\lambda_2 + \\
\{\hat{p}_{1i}\hat{p}_{2i}\times\hat{p}_{3i}(1 - \hat{p}_{3i})\times[l_3 - l_2]\}\lambda_3
\]
Effect of explanatory variable on outcome

expected increase in the level of education after passing

\[
\frac{\partial E(\text{ed})}{\partial \text{SES}} = \\
\{1 \times \hat{p}_{1i}(1 - \hat{p}_{1i}) \times [(1 - \hat{p}_{2})l_1 + \hat{p}_{2}(1 - \hat{p}_{3})l_2 + \hat{p}_{2}\hat{p}_{3}l_3 - l_0] \} \lambda_1 + \\
\{\hat{p}_{1i} \times \hat{p}_{2i}(1 - \hat{p}_{2i}) \times [(1 - \hat{p}_{3})l_2 + \hat{p}_{3}l_3 - l_1] \} \lambda_2 + \\
\{\hat{p}_{1i}\hat{p}_{2i} \times \hat{p}_{3i}(1 - \hat{p}_{3i}) \times [l_3 - l_2] \} \lambda_3
\]
Effect of explanatory variable on outcome

expected level of education for those that pass

\[
\frac{\partial E(\text{ed})}{\partial \text{SES}} = \\
\{1 \times \hat{p}_{1i}(1 - \hat{p}_{1i}) \times [(1 - \hat{p}_2)l_1 + \hat{p}_2(1 - \hat{p}_3)l_2 + \hat{p}_2\hat{p}_3l_3 - l_0]\} \lambda_1 + \\
\{\hat{p}_{1i} \times \hat{p}_{2i}(1 - \hat{p}_{2i}) \times [(1 - \hat{p}_3)l_2 + \hat{p}_3l_3 - l_1]\} \lambda_2 + \\
\{\hat{p}_{1i}\hat{p}_{2i} \times \hat{p}_{3i}(1 - \hat{p}_{3i}) \times [l_3 - l_2]\} \lambda_3
\]
Effect of explanatory variable on outcome

minus the expected level of education for those that fail

\[
\frac{\partial E(\text{ed})}{\partial \text{SES}} = \\
\{1 \times \hat{p}_{1i}(1 - \hat{p}_{1i}) \times [(1 - \hat{p}_2)l_1 + \hat{p}_2(1 - \hat{p}_3)l_2 + \hat{p}_2\hat{p}_3l_3 - l_0]\} \lambda_1 + \\
\{\hat{p}_{1i} \times \hat{p}_{2i}(1 - \hat{p}_{2i}) \times [(1 - \hat{p}_3)l_2 + \hat{p}_3l_3 - l_1]\} \lambda_2 + \\
\{\hat{p}_{1i}\hat{p}_{2i} \times \hat{p}_{3i}(1 - \hat{p}_{3i}) \times [l_3 - l_2]\} \lambda_3
\]
In words:

- effect on outcome = weighted sum of effects on transition probabilities
In words:

- effect on outcome = weighted sum of effects on transition probabilities
- weights = at risk $\times$ variance $\times$ gain
Outline

Process and Outcome

Empirical example

The seqlogit package
Data

- General Social Survey (GSS).
- 20 surveys held between 1977 and 2004 with information on cohorts 1913-1978.
- 13,400 men aged between 27 and 65 with complete information.
Variables

- Father’s highest achieved level of education measured in (pseudo) years.
Variables

- Father’s highest achieved level of education measured in (pseudo) years.
- Respondent’s highest achieved Level of education in (pseudo) years
Variables

- Father’s highest achieved level of education measured in (pseudo) years.
- Respondent’s highest achieved Level of education in (pseudo) years
- Time measured as a restricted cubic spline with one knot in 1946.
Simplified model of the US educational system

- Less than high school
- High school
- Junior college
- Bachelor
- Exit

Node labels and transitions:
- Less than high school to high school: 12
- High school to junior college: 14
- High school to exit: 9
- Junior college to bachelor: 16
- Exit to bachelor: 16
- Exit to less than high school: 12

Maarten L. Buis
Linking process to outcome
Change in effect on outcome over cohorts

The `seqlogit` package

Maarten L. Buis

Linking process to outcome
Decomposition of effect on outcome

- The effect on outcome is a weighted sum of effects on transitions:

\[ IEOut = w_1 \lambda_1 + w_2 \lambda_2 + w_3 \lambda_3 \]

- The contribution of the first transition is: \( w_1 \lambda_1 \)
- This can be visualized as the area of a rectangle with width \( w_1 \) and height \( \lambda_1 \).
- The effect on the outcome is the sum of the areas of these rectangles.
Decomposition of effect on outcome

The effect on outcome is a weighted sum of effects on transitions:
IEOut = w_1 \lambda_1 + w_2 \lambda_2 + w_3 \lambda_3
Decomposition of effect on outcome

- The effect on outcome is a weighted sum of effects on transitions:
  \[ \text{IEOut} = w_1 \lambda_1 + w_2 \lambda_2 + w_3 \lambda_3 \]
- The contribution of the first transition is: \( w_1 \lambda_1 \)
Decomposition of effect on outcome

- The effect on outcome is a weighted sum of effects on transitions:
  \[ \text{IEOut} = w_1 \lambda_1 + w_2 \lambda_2 + w_3 \lambda_3 \]
- The contribution of the first transition is: \( w_1 \lambda_1 \)
- This can be visualized as the area of a rectangle
Decomposition of effect on outcome

- The effect on outcome is a weighted sum of effects on transitions:
  \[ \text{IEOut} = w_1 \lambda_1 + w_2 \lambda_2 + w_3 \lambda_3 \]
- The contribution of the first transition is: \( w_1 \lambda_1 \)
- This can be visualized as the area of a rectangle with width \( w_1 \)
Decomposition of effect on outcome

- The effect on outcome is a weighted sum of effects on transitions:
  \[ \text{IEOut} = w_1 \lambda_1 + w_2 \lambda_2 + w_3 \lambda_3 \]
- The contribution of the first transition is: \( w_1 \lambda_1 \)
- This can be visualized as the area of a rectangle with width \( w_1 \) and height \( \lambda_1 \).
Decomposition of effect on outcome

- The effect on outcome is a weighted sum of effects on transitions:
  \[ IEOut = w_1 \lambda_1 + w_2 \lambda_2 + w_3 \lambda_3 \]
- The contribution of the first transition is: \( w_1 \lambda_1 \)
- This can be visualized as the area of a rectangle with width \( w_1 \) and height \( \lambda_1 \).
- The effect on the outcome is the sum of the areas of these rectangles
Decomposition of effect on outcome for white men

![Graph showing the decomposition of effect on outcome for white men. The graph displays the log odds ratio for different weight categories and educational levels over time (1915 to 1965).]
Decomposition of effect on outcome for black men

Maarten L. Buis
Decomposition of weights

- The weights are: at risk × variance × gain
Decomposition of weights

- The weights are: at risk $\times$ variance $\times$ gain
- These three elements are all a function of the proportions that pass the transitions
Decomposition of the weights for white men

Maarten L. Buis
Decomposition of the weights for black men

- Transition probability
- Proportion at risk
- Variance
- Gain
- Weight

Maarten L. Buis  
Linking process to outcome
The `seqlogit` package

- `seqlogit` will estimate a sequential logit model
The `seqlogit` package

- `seqlogit` will estimate a sequential logit model
- It can predict the weights and its components
The `seqlogit` package

- `seqlogit` will estimate a sequential logit model
- It can predict the weights and its components
- `seqlogitdecomp` shows the decomposition of the effect on the outcome into effects on the transitions and their weights.
The seqlogit package

seqlogit

- The dependent variable: The highest achieved level

Example:
```
seqlogit degree south padeg coh padegXcoh, tree(0:1 2 3 , 1:2:3 )
```
The `seqlogit` package

- The dependent variable: The highest achieved level
- The explanatory variables

Example:
```
seqlogit degree south padeg coh padegXcoh, tree(0:1 2 3 , 1:2:3 )
```
seqlogit

- The dependent variable: The highest achieved level
- The explanatory variables
- The tree: The way one reaches a level of education
Process and Outcome
Empirical example
The \texttt{seqlogit} package

\texttt{seqlogit}

- **The dependent variable** The highest achieved level
- **The explanatory variables**
- **The tree** The way one reaches a level of education

example:
\begin{verbatim}
seqlogit degree south padeg coh padegXcoh, /*
*/ tree(0:1 2 3 , 1:2:3 )
\end{verbatim}
Simplified model of the US educational system

- less than high school
- high school
  - junior college 14
  - exit 12
- exit 9
- bachelor 16
**predict**

- **trpr** probability of passing transition
predict

- `trpr` probability of passing transition
- `tratrisk` proportion of respondents at risk of passing transition
- `trvar` variance of the indicator variable indicating whether or not the respondent passed the transition
- `trgain` difference in expected highest achieved level between those that pass the transition and those that do not
- `trweight` weight assigned to transition
- `pr` probability that an outcome is the highest achieved outcome.
- `y` expected highest achieved level.
predict

- **trpr** probability of passing transition
- **tratrisk** proportion of respondents at risk of passing transition
- **trvar** variance of the indicator variable indicating whether or not the respondent passed the transition
- **trgain** difference in expected highest achieved level between those that pass the transition and those that do not
- **trweight** weight assigned to transition
The `seqlogit` package

**predict**

- **trpr** probability of passing transition
- **tratrisk** proportion of respondents at risk of passing transition
- **trvar** variance of the indicator variable indicating whether or not the respondent passed the transition
- **trgain** difference in expected highest achieved level between those that pass the transition and those that do not
- **trweight** weight assigned to transition
- **p** probability that an outcome is the highest achieved outcome.
- **y** expected highest achieved level
Those statistics marked with a † need a scaling of the end result (e.g. pseudo years of education)
predict

- Those statistics marked with a † need a scaling of the end result (e.g. pseudo years of education)
- The numerical values of `depvar` are used by default.
Those statistics marked with a † need a scaling of the end result (e.g. pseudo years of education).

The numerical values of depvar are used by default.

They can also be specified using the levels() option.
Those statistics marked with a † need a scaling of the end result (e.g. pseudo years of education).

The numerical values of `depvar` are used by default.

They can also be specified using the `levels()` option.

Example:
```
predict weib*, trweight /*
*/ levels(0=9, 1=12, 2=14, 3=16)
```
seqlogitdecomp

- seqlogitdecomp is used to compare the decomposition across groups, e.g. cohorts.
The `seqlogit` package

`seqlogitdecomp` is used to compare the decomposition across groups, e.g. cohorts.

Differences in the effect on the outcome may be due to:

- Differences in weights, or
- Differences in effects on transitions (log odds ratios)

Both need to be specified.

The weights can be specified by fixing all values of all explanatory variables.

The effects on transitions can be specified directly.
seqlogitdecomp

- `seqlogitdecomp` is used to compare the decomposition across groups, e.g. cohorts.
- Differences in the effect on the outcome may be due to:
  - differences in weights, or
seqlogitdecomp

- seqlogitdecomp is used to compare the decomposition across groups, e.g. cohorts.
- Differences in the effect on the outcome may be due to:
  - differences in weights, or
  - differences in effects on transitions (log odds ratios)
seqlogitdecomp

- **seqlogitdecomp** is used to compare the decomposition across groups, e.g. cohorts.
- Differences in the effect on the outcome may be due to:
  - differences in weights, or
  - differences in effects on transitions (log odds ratios)
- Both need to be specified
seqlogitdecomp

- seqlogitdecomp is used to compare the decomposition across groups, e.g. cohorts.
- Differences in the effect on the outcome may be due to:
  - differences in weights, or
  - differences in effects on transitions (log odds ratios)
- Both need to be specified
  - The weights can be specified by fixing all values of all explanatory variables
`seqlogitdecomp` is used to compare the decomposition across groups, e.g. cohorts.

Differences in the effect on the outcome may be due to:

- differences in weights, or
- differences in effects on transitions (log odds ratios)

Both need to be specified

- The weights can be specified by fixing all values of all explanatory variables
- The effects on transitions can be specified directly
Specify the weights

▶ Model:

    seqlogit degree south padeg coh padegXcoh, /*
    */ tree(0:1 2 3 , 1:2:3 )

▶ We want to compare cohorts 1920 1940 1960

    seqlogitdecomp, 
    overat( coh 1920 padegXcoh 'mean20' , 
    coh 1940 padegXcoh 'mean40' , 
    coh 1960 padegXcoh 'mean60' )

    overlodds( _b[padeg] + 1920*_b[padegXcoh] , 
    _b[padeg] + 1940*_b[padegXcoh] , 
    _b[padeg] + 1960*_b[padegXcoh] )

    at(south 0)

Locals ‘mean20’, ‘mean40’, and ‘mean60’ contain the mean of padeg times 1920, 1940, 1960 respectively.
Specify the odds ratios

- Model:
  
  \[
  \text{seqlogit} \ degree \ south \ padeg \ coh \ padegXcoh, */
  \]
  
  \[
  */ \ \text{tree}(0:1 \ 2 \ 3 \ , \ 1:2:3 )
  \]

- We want to compare cohorts 1920 1940 1960

  \[
  \text{seqlogitdecomp,}
  \]
  
  overat( coh 1920 padegXcoh ‘mean20’ ,
  coh 1940 padegXcoh ‘mean40’ ,
  coh 1960 padegXcoh ‘mean60’ )

  \[
  \text{overlodds(} \ _b[\text{padeg}] + 1920*\_b[\text{padegXcoh}] ,
  \_b[\text{padeg}] + 1940*\_b[\text{padegXcoh}] ,
  \_b[\text{padeg}] + 1960*\_b[\text{padegXcoh}] )
  \]

  \[
  \text{at(south \ 0)}
  \]
The effect on the outcome depends in an understandable way on the effects on the process.
Conclusion

- The effect on the outcome depends in an understandable way on the effects on the process.
- The effect on the outcome is a weighted sum of the effects on the transition probabilities.
Conclusion

- The effect on the outcome depends in an understandable way on the effects on the process.
- The effect on the outcome is a weighted sum of the effects on the transition probabilities, and the weights increase if:
  - the proportion at risk increases,
  - the proportion that passes is closer to .50,
  - the expected increase in the outcome increases.
Conclusion

- The effect on the outcome depends in an understandable way on the effects on the process.
- The effect on the outcome is a weighted sum of the effects on the transition probabilities, and the weights increase if:
  - the proportion at risk increases,
Conclusion

- The effect on the outcome depends in an understandable way on the effects on the process.
- The effect on the outcome is a weighted sum of the effects on the transition probabilities, and the weights increase if:
  - the proportion at risk increases,
  - the proportion that passes is closer to .50,
Conclusion

- The effect on the outcome depends in an understandable way on the effects on the process.
- The effect on the outcome is a weighted sum of the effects on the transition probabilities, and the weights increase if:
  - the proportion at risk increases,
  - the proportion that passes is closer to .50,
  - the expected increase in the outcome increases
Conclusion

- This relationship can be used to:
Conclusion

- This relationship can be used to:
  - to relate the process to the outcome.
Conclusion

▶ This relationship can be used to:
  ▶ to relate the process to the outcome.
  ▶ identify important and less important transitions,
Conclusion

- This relationship can be used to:
  - to relate the process to the outcome.
  - identify important and less important transitions,
  - to explain differences in effect on outcome with well documented phenomena like educational expansion or racial differences in educational attainment.