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

Outline

Process and Outcome

Empirical example

The seqlogit package

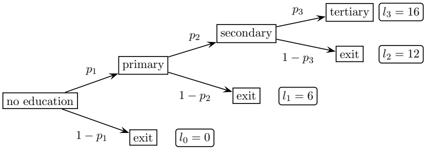
Sequential logit model

- This model is know 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 blossfeld93) (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



sequential logit to end result

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

$$E(ed) = (1-\hat{p}_{1i})l_0 + \hat{p}_{1i}(1-\hat{p}_{2i})l_1 + \hat{p}_{1i}\hat{p}_{2i}(1-\hat{p}_{3i})l_2 + \hat{p}_{1i}\hat{p}_{2i}\hat{p}_{3i}l_3$$

Effect of explanatory variable on outcome

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:

Effect of explanatory variable on outcome

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:

$$
\begin{aligned}
&\frac{\partial E(ed)}{\partial SES} = \\
&\{1 \times \hat{p}_{1i}(1-\hat{p}_{1i}) \times [(1-\hat{p}_2)l_1 + \hat{p}_2(1-\hat{p}_3)l_2 + \hat{p}_2\hat{p}_3 l_3 - l_0]\}\lambda_1 + \\
&\{\hat{p}_{1i} \times \hat{p}_{2i}(1-\hat{p}_{2i}) \times [(1-\hat{p}_3)l_2 + \hat{p}_3 l_3 - l_1]\}\lambda_2 + \\
&\{\hat{p}_{1i}\hat{p}_{2i} \times \hat{p}_{3i}(1-\hat{p}_{3i}) \times [l_3 - l_2]\}\lambda_3
\end{aligned}
$$

Effect of explanatory variable on outcome

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

Effect of explanatory variable on outcome

$$\frac{\partial E(ed)}{\partial SES} =$$
$$\{1 \times \hat{p}_{1i}(1-\hat{p}_{1i}) \times [(1-\hat{p}_2)I_1 + \hat{p}_2(1-\hat{p}_3)I_2 + \hat{p}_2\hat{p}_3 I_3 - I_0]\}\lambda_1 +$$
$$\{\hat{p}_{1i} \times \hat{p}_{2i}(1-\hat{p}_{2i}) \times [(1-\hat{p}_3)I_2 + \hat{p}_3 I_3 - I_1]\}\lambda_2 +$$
$$\{\hat{p}_{1i}\hat{p}_{2i} \times \hat{p}_{3i}(1-\hat{p}_{3i}) \times [I_3 - I_2]\}\lambda_3$$

Effect of explanatory variable on outcome

proportion at risk

$$\frac{\partial E(ed)}{\partial SES} =$$
$$\{1 \times \hat{p}_{1i}(1-\hat{p}_{1i}) \times [(1-\hat{p}_2)I_1 + \hat{p}_2(1-\hat{p}_3)I_2 + \hat{p}_2\hat{p}_3I_3 - I_0]\}\lambda_1 +$$
$$\{\hat{p}_{1i} \times \hat{p}_{2i}(1-\hat{p}_{2i}) \times [(1-\hat{p}_3)I_2 + \hat{p}_3I_3 - I_1]\}\lambda_2 +$$
$$\{\hat{p}_{1i}\hat{p}_{2i} \times \hat{p}_{3i}(1-\hat{p}_{3i}) \times [I_3 - I_2]\}\lambda_3$$

Effect of explanatory variable on outcome

variance of the variable indicating whether one passes or not

$$\frac{\partial E(ed)}{\partial SES} =$$
$$\{1 \times \hat{p}_{1i}(1-\hat{p}_{1i}) \times [(1-\hat{p}_2)I_1 + \hat{p}_2(1-\hat{p}_3)I_2 + \hat{p}_2\hat{p}_3 I_3 - I_0]\}\lambda_1 +$$
$$\{\hat{p}_{1i} \times \hat{p}_{2i}(1-\hat{p}_{2i}) \times [(1-\hat{p}_3)I_2 + \hat{p}_3 I_3 - I_1]\}\lambda_2 +$$
$$\{\hat{p}_{1i}\hat{p}_{2i} \times \hat{p}_{3i}(1-\hat{p}_{3i}) \times [I_3 - I_2]\}\lambda_3$$

Effect of explanatory variable on outcome

expected increase in the level of education after passing

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

Effect of explanatory variable on outcome

expected level of education for those that pass

$$\frac{\partial E(ed)}{\partial SES} =$$
$$\{1 \times \hat{p}_{1i}(1 - \hat{p}_{1i}) \times [(1 - \hat{p}_2)I_1 + \hat{p}_2(1 - \hat{p}_3)I_2 + \hat{p}_2\hat{p}_3I_3 - I_0]\}\lambda_1 +$$
$$\{\hat{p}_{1i} \times \hat{p}_{2i}(1 - \hat{p}_{2i}) \times [(1 - \hat{p}_3)I_2 + \hat{p}_3I_3 - I_1]\}\lambda_2 +$$
$$\{\hat{p}_{1i}\hat{p}_{2i} \times \hat{p}_{3i}(1 - \hat{p}_{3i}) \times [I_3 - I_2]\}\lambda_3$$

Effect of explanatory variable on outcome

minus the expected level of education for those that fail

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

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.

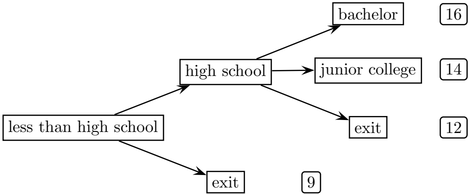
Variables

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

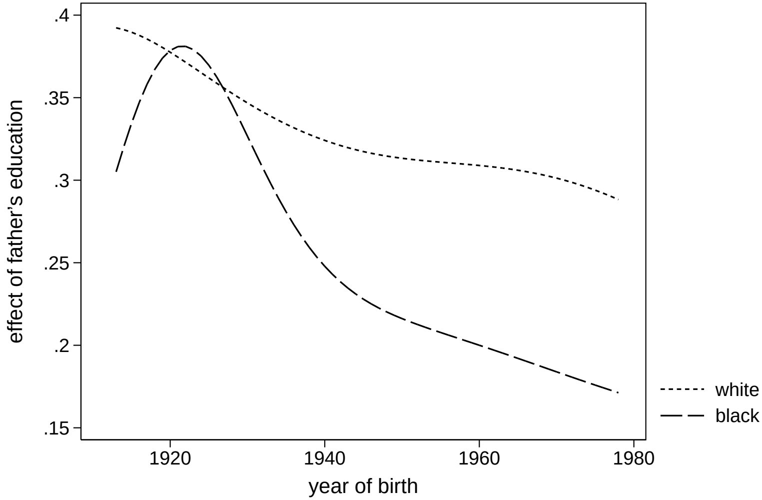
Variables

- 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



Change in effect on outcome over cohorts



Decomposition of effect on outcome

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

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:
  IEOut = w_1 λ_1 + w_2 λ_2 + w_3 λ_3
- The contribution of the first transition is: w_1 λ_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

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

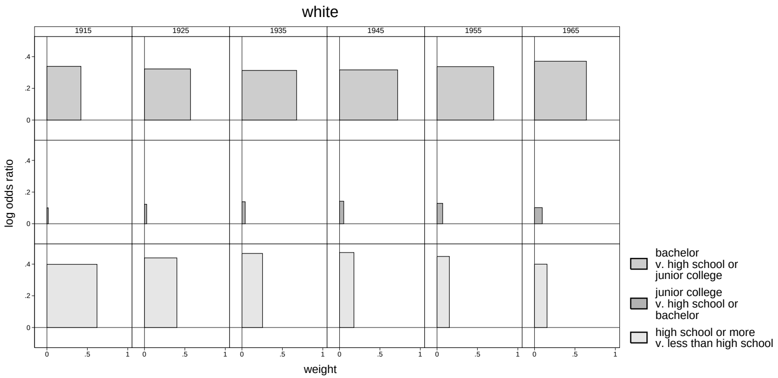
Decomposition of effect on outcome

- The effect on outcome is a weighted sum of effects on transitions:
  IEOut = w_1 λ_1 + w_2 λ_2 + w_3 λ_3
- The contribution of the first transition is: w_1 λ_1
- This can be visualized as the area of a rectangle with width w_1 and height λ_1.

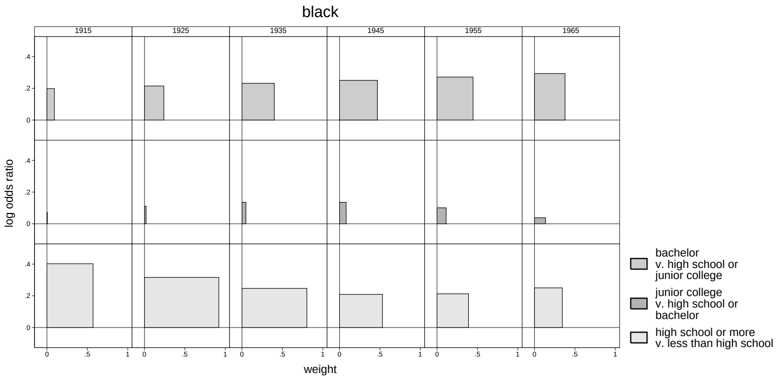
Decomposition of effect on outcome

- 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 λ_1.
- The effect on the outcome is the sum of the areas of these rectangles

Decomposition of effect on outcome for white men



Decomposition of effect on outcome for black men



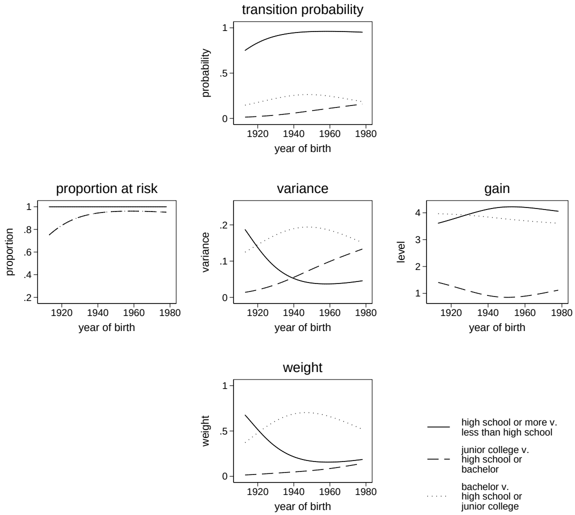
Decomposition of weights

- The weights are:
  at risk $\times$ variance $\times$ gain

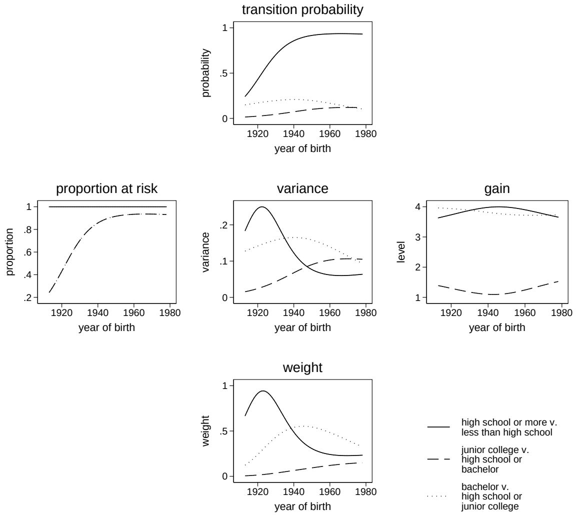
Decomposition of weights

- 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



Decomposition of the weights for black men



Outline

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.

seqlogit

- The dependent variable The highest achieved level

seqlogit

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

seqlogit

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

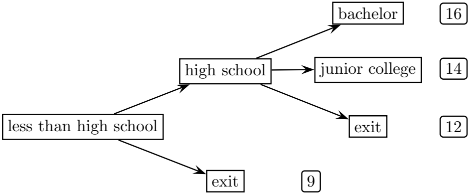
seqlogit

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

example:

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

Simplified model of the US educational system



`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

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

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`

- 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.

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

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.
- 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.

seqlogitdecomp

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

`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

- `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:
```
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)
```

Conclusion

- 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:

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.