Deceleration of the Trend in Inequality of Educational Opportunity in the Netherlands

Maarten L. Buis

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

- Introduction
- Main results
- Inequality of Educational Opportunity
  - method:
    - Multiple indicators for socioeconomic status
    - Using lowess smooth to estimate trend and change in trend
    - Using bootstrap to estimate confidence envelopes
- results
- conclusions
Previous research primarily found a negative trend in IEO.
Previous research

- Previous research primarily found a negative trend in IEO.
- This has been confirmed with:
  - linear regression
  - ordered logits
  - loglinear models (uniform association, scaled association: RC-2)
  - sequential logits (Mare model)
Many models have been used, but is there any reason to prefer one over the other?
The problems

Many models have been used, but is there any reason to prefer one over the other?

Even the most recent accounts (Ganzeboom & Luijkx, 2004a, 2005b) find a linear trend.
The problems

Many models have been used, but is there any reason to prefer one over the other?

Even the most recent accounts (Ganzeboom & Luijkx, 2004a, 2005b) find a linear trend.

However, there is reason to believe that the trend cannot continue.
The problems

Many models have been used, but is there any reason to prefer one over the other?

Even the most recent accounts (Ganzeboom & Luijkx, 2004a, 2005b) find a linear trend.

However, there is reason to believe that the trend cannot continue.

When do we begin to observe a deceleration of the trend?
Main results: measuring IEO

A linear regression of SES on highest level of education is a good measure of the IEO for the entire educational system, because:
Main results: measuring IEO

A linear regression of SES on highest level of education is a good measure of the IEO for the entire educational system, because:

It is a good approximation of a weighted sum of Mare coefficients that give more weight to transitions that:
Main results: measuring IEO

- A linear regression of SES on highest level of education is a good measure of the IEO for the entire educational system, because:

- It is a good approximation of a weighted sum of Mare coefficients that give more weight to transitions that:
  - affect more people, and
Main results: measuring IEO

- A linear regression of SES on highest level of education is a good measure of the IEO for the entire educational system, because:
  - It is a good approximation of a weighted sum of Mare coefficients that give more weight to transitions that:
    - affect more people, and
    - give access to more higher levels of education.
Main results: measuring IEO

- A linear regression of SES on highest level of education is a good measure of the IEO for the entire educational system, because:

- It is a good approximation of a weighted sum of Mare coefficients that give more weight to transitions that:
  - affect more people, and
  - give access to more higher levels of education.

- Ordered logistic regression and log linear models don’t have this relationship with the Mare model.
Main results: non-linear trend
Inequality of Educational Opportunity

Education consists of a sequence of transitions.
Inequality of Educational Opportunity

- Education consists of a sequence of transitions.
- Mare model measures effects of SES on each individual transition.
Inequality of Educational Opportunity

- Education consists of a sequence of transitions.
- Mare model measures effects of SES on each individual transition.
- Total IEO would be a mean of transition specific IEOs, where transitions receive more weight if:
  - more people are at risk,
  - close to 50% pass the transition,
  - transition gives more access to higher subsequent levels of education.
Inequality of Educational Opportunity

- Education consists of a sequence of transitions.
- Mare model measures effects of SES on each individual transition.
- Total IEO would be a mean of transition specific IEOs, where transitions receive more weight if:
  - more people are at risk,
Inequality of Educational Opportunity

Education consists of a sequence of transitions.

Mare model measures effects of SES on each individual transition.

Total IEO would be a mean of transition specific IEOs, where transitions receive more weight if:
- more people are at risk,
- close to 50% pass the transition, and
Inequality of Educational Opportunity

- Education consists of a sequence of transitions.
- Mare model measures effects of SES on each individual transition.
- Total IEO would be a mean of transition specific IEOs, where transitions receive more weight if:
  - more people are at risk,
  - close to 50% pass the transition, and
  - transition gives more access to higher subsequent levels of education.
Inequality of Educational Opportunity

Education consists of a sequence of transitions.

Mare model measures effects of SES on each individual transition.

Total IEO would be a mean of transition specific IEOs, where transitions receive more weight if:

- more people are at risk,
- close to 50% pass the transition, and
- transition gives more access to higher subsequent levels of education.

Effect of a unit change in $SES$ on the highest achieved level of education is consistent with these criteria.
Example:

- Four level educational system, so three transitions
- One explanatory variable: $SES$
- Probability that individual $i$ passes transition $k$ given that it has passed all previous transitions:

$$p_{ki} = \frac{\exp(\alpha_k + \gamma_k SES_i)}{1 + \exp(\alpha_k + \gamma_k SES_i)}$$

- $\gamma_k$ is the transition specific inequality for transition $k$
- The four levels of education are given values $l_0$ to $l_3$. 
Example (continued)

\[ E(ed) = (1 - p_{1i})l_0 + p_{1i}(1 - p_{2i})l_1 + p_{1i}p_{2i}(1 - p_{3i})l_2 + p_{1i}p_{2i}p_{3i}l_3 \]
Example (continued)

\[ E(ed) = (1 - p_{1i})l_0 + p_{1i}(1 - p_{2i})l_1 + p_{1i}p_{2i}(1 - p_{3i})l_2 + p_{1i}p_{2i}p_{3i}l_3 \]

\[ \frac{\partial E(ed)}{\partial SES} = \]
\[ \{ 1 \times p_{1i}(1 - p_{1i}) \times [(l_1 - l_0) + p_{2i}(l_2 - l_1) + p_{2i}p_{3i}(l_3 - l_2)] \}\gamma_1 + \]
\[ \{ p_{1i} \times p_{2i}(1 - p_{2i}) \times [(l_2 - l_1) + p_{3i}(l_3 - l_1)] \}\gamma_2 + \]
\[ \{ p_{1i}p_{2i} \times p_{3i}(1 - p_{3i}) \times [(l_3 - l_2)] \}\gamma_3 \]
Mare and linear regression

The linear regression coefficient tells by how much the highest level of education changes for a unit change in $SES$. 
Mare and linear regression

- The linear regression coefficient tells by how much the highest level of education changes for a unit change in $SES$.

- Conceptually this is exactly the right summary measure of IEO.
Mare and linear regression

- The linear regression coefficient tells by how much the highest level of education changes for a unit change in \( SES \).

- Conceptually this is exactly the right summary measure of IEO.

- However, this is based on a linear approximation of the relationship between highest education and SES.
Mare and linear regression

- The linear regression coefficient tells by how much the highest level of education changes for a unit change in SES.
- Conceptually this is exactly the right summary measure of IEO.
- However, this is based on a linear approximation of the relationship between highest education and SES.
- In order to check whether this approximation is acceptable both a Mare model and a linear regression were estimated on Dutch data.
Data

- International Stratification and Mobility File (ISMF) on the Netherlands.


- 40,000 respondents aged between 24 and 65 have complete information on child's, father's and mother's education and father's occupation.

- Number of cases are unequally distributed over cohorts.
Expected levels of education

Comparing linear regression and Mare model

Deceleration of the Trend in Inequality of Educational Opportunity in the Netherlands – p. 12/29
Ordered logistic regression

Proportional odds assumption

\[
\frac{\partial}{\partial SES} \ln \left( \frac{\Pr(ed \leq 0)}{\Pr(ed > 0)} \right) = \frac{\partial}{\partial SES} \ln \left( \frac{\Pr(ed \leq 1)}{\Pr(ed > 1)} \right) = \frac{\partial}{\partial SES} \ln \left( \frac{\Pr(ed \leq 2)}{\Pr(ed > 2)} \right)
\]
Ordered logistic regression

Proportional odds assumption

\[
\frac{\partial}{\partial SES} \ln \left( \frac{\Pr(\text{ed} \leq 0)}{\Pr(\text{ed} > 0)} \right) = \frac{\partial}{\partial SES} \ln \left( \frac{\Pr(\text{ed} \leq 1)}{\Pr(\text{ed} > 1)} \right) = \frac{\partial}{\partial SES} \ln \left( \frac{\Pr(\text{ed} \leq 2)}{\Pr(\text{ed} > 2)} \right)
\]

Proportional odds assumption with Mare model

\[
\frac{\partial}{\partial SES} \ln \left( \frac{1 - p_{1i}}{p_{1i}} \right) = \frac{\partial}{\partial SES} \ln \left( \frac{1 - p_{1i}p_{2i}}{p_{1i}p_{2i}} \right) = \frac{\partial}{\partial SES} \ln \left( \frac{1 - p_{1i}p_{2i}p_{3i}}{p_{1i}p_{2i}p_{3i}} \right)
\]
Ordered logistic regression

Proportional odds assumption

\[
\frac{\partial}{\partial SES} \ln \left( \frac{\Pr(ed \leq 0)}{\Pr(ed > 0)} \right) = \frac{\partial}{\partial SES} \ln \left( \frac{\Pr(ed \leq 1)}{\Pr(ed > 1)} \right) = \frac{\partial}{\partial SES} \ln \left( \frac{\Pr(ed \leq 2)}{\Pr(ed > 2)} \right)
\]

Proportional odds assumption with Mare model

\[
\frac{\partial}{\partial SES} \ln \left( \frac{1 - p_{1i}}{p_{1i}} \right) = \frac{\partial}{\partial SES} \ln \left( \frac{1 - p_{1i}p_{2i}}{p_{1i}p_{2i}} \right) = \frac{\partial}{\partial SES} \ln \left( \frac{1 - p_{1i}p_{2i}p_{3i}}{p_{1i}p_{2i}p_{3i}} \right)
\]

\[ -\gamma_1 = -\frac{\gamma_2}{p_{1i}} = -\frac{\gamma_3}{p_{1i}p_{2i}} \]
Stereotyped Ordered regression

\[
\ln \left( \frac{\Pr(y = q)}{\Pr(y = r)} \right) = (\alpha_q - \alpha_r)\beta_0 + (\phi_q - \phi_r)(\beta_{SES})
\]
Stereotyped Ordered regression

\[ \ln \left( \frac{\Pr(y = q)}{\Pr(y = r)} \right) = (\alpha_q - \alpha_r)\beta_0 + (\phi_q - \phi_r)(\beta SES) \]

Identifying constraints:

- The \( \alpha \) and \( \phi \) associated with the lowest level are constrained to be zero, and
- the \( \alpha \) and \( \phi \) associated with the highest level are constrained to be one
Stereotyped Ordered regression

\[
\ln \left( \frac{\Pr(y = q)}{\Pr(y = r)} \right) = (\alpha_q - \alpha_r)\beta_0 + (\phi_q - \phi_r)(\beta SES)
\]

Identifying constraints:

- The \(\alpha\) and \(\phi\) associated with the lowest level are constrained to be zero, and
- the \(\alpha\) and \(\phi\) associated with the highest level are constrained to be one

\[
\frac{\partial}{\partial SES} \ln \left( \frac{\Pr(y = 3)}{\Pr(y = 0)} \right)
\]
Stereotyped Ordered regression

\[
\ln \left( \frac{\Pr(y = q)}{\Pr(y = r)} \right) = (\alpha_q - \alpha_r)\beta_0 + (\phi_q - \phi_r)(\beta SES)
\]

Identifying constraints:

- The \(\alpha\) and \(\phi\) associated with the lowest level are constrained to be zero, and
- the \(\alpha\) and \(\phi\) associated with the highest level are constrained to be one

\[
\frac{\partial}{\partial SES} \ln \left( \frac{\Pr(y = 3)}{\Pr(y = 0)} \right) = \frac{\partial}{\partial SES} \ln \left( \frac{p_1 p_2 p_3}{1 - p_1} \right)
\]
Stereotyped Ordered regression

\[ \ln \left( \frac{\Pr(y = q)}{\Pr(y = r)} \right) = (\alpha_q - \alpha_r)\beta_0 + (\phi_q - \phi_r)(\beta SES) \]

Identifying constraints:

- The \( \alpha \) and \( \phi \) associated with the lowest level are constrained to be zero, and
- the \( \alpha \) and \( \phi \) associated with the highest level are constrained to be one

\[ \frac{\partial}{\partial SES} \ln \left( \frac{\Pr(y = 3)}{\Pr(y = 0)} \right) = \frac{\partial}{\partial SES} \ln \left( \frac{p_1 p_2 p_3}{1 - p_1} \right) = \gamma_1 + (1 - p_2)\gamma_2 + (1 - p_3)\gamma_3 \]
Outline

Introduction

Main results

Inequality of Educational Opportunity

Method:
- Multiple indicators for socioeconomic status
- Using lowess smooth to estimate trend and change in trend
- Using bootstrap to estimate confidence envelopes

Results

Conclusions
Socioeconomic Status

Multiple indicators are used in order to reduce measurement error.
Socioeconomic Status

- Multiple indicators are used in order to reduce measurement error.
- Father’s occupational status and father’s and mother’s highest achieved level of education.
Socioeconomic Status

- Multiple indicators are used in order to reduce measurement error.
  
- Father’s occupational status and father’s and mother’s highest achieved level of education.

\[ ed_i = \alpha_{tg} + \beta_{tg} (\lambda_1 fisei_i + \lambda_2 fed_i + \lambda_3 med_i) \]

\[ SES \]
Socioeconomic Status

- Multiple indicators are used in order to reduce measurement error.

- Father’s occupational status and father’s and mother’s highest achieved level of education.

\[ ed_i = \alpha_{tg} + \beta_{tg} \left( \lambda_1 fisei_i + \lambda_2 fed_i + \lambda_3 med_i \right) \]

- Standard deviation of latent variable is constrained to 1 in 1970.
Multiple indicators are used in order to reduce measurement error.

Father’s occupational status and father’s and mother’s highest achieved level of education.

\[ ed_i = \alpha_{tg} + \beta_{tg} \left( \lambda_1 fisei_i + \lambda_2 fed_i + \lambda_3 med_i \right) \]

Standard deviation of latent variable is constrained to 1 in 1970.

\( \beta \) is the effect of a standard deviation change in family SES on the child’s highest achieved level of education.
Constrained and unconstrained model

- Male
  - Father's occupational status
  - Father's education
  - Mother's education

- Female
  - Father's occupational status
  - Father's education
  - Mother's education

○ unconstrained  ● constrained
Annual estimates of IEO

Deceleration of the Trend in Inequality of Educational Opportunity in the Netherlands – p. 18/29
We have a dataset consisting of estimates of IEO for each annual cohort, which used only information from that cohort.
We have a dataset consisting of estimates of IEO for each annual cohort, which used only information from that cohort.

If we think that IEO develops like a smooth curve over time, than nearby estimates also contain relevant information.
Lowess

- We have a dataset consisting of estimates of IEO for each annual cohort, which used only information from that cohort.
- If we think that IEO develops like a smooth curve over time, than nearby estimates also contain relevant information.
- The lowess curve creates an improved estimate of the IEO for each cohort using information from nearby cohorts.
Lowess curve in 1943

- Point on lowess curve in 1943
- Select closest 60% of the points.
- Give larger weights to nearby points.
- Adjust weights for precision of estimated IEO.
- WLS regression of IEO on time, time squared and time cubed on weighted points.
- Predicted value in 1943, is smoothed value of 1943.
- First derivative in 1943 is trend in 1943.
- Second derivative in 1943 is change in trend in 1943.
Lowess curve in 1943

(a) Observations Within the Window
span = 0.6

(b) Tricube Weights

(c) Tricube (+), Precision (x),
and Joint (o) Weights

(d) Weighted Third Degree Polynomial
(size of circle proportional to weight)
Confidence interval gives the range of results that could plausibly occur just through sampling error.
bootstrap confidence intervals

Confidence interval gives the range of results that could plausibly occur just through sampling error.

Make many ‘datasets’ that could have occurred just by sampling error.
bootstrap confidence intervals

- Confidence interval gives the range of results that could plausibly occur just through sampling error.
- Make many ‘datasets’ that could have occurred just by sampling error.
- Fit lowess curves through each ‘dataset’.
Confidence interval gives the range of results that could plausibly occur just through sampling error.

Make many ‘datasets’ that could have occurred just by sampling error.

Fit lowess curves through each ‘dataset’.

The area containing 90% of the curves is the 90% confidence interval.
The estimates of IEO are regression coefficients with standard errors.
The estimates of IEO are regression coefficients with standard errors.

The standard errors and covariances give information about what values of IEO could plausibly occur in a ‘new’ dataset.
bootstrap confidence intervals

- The estimates of IEO are regression coefficients with standard errors.
- The standard errors and covariances give information about what values of IEO could plausibly occur in a ‘new’ dataset.
- ‘New’ dataset is a random draw from a multivariate normal distribution with mean vector at the estimated IEOs and the estimated variance covariance matrix.
First 25 bootstrapped curves

(a) Lowess Smooths in the First 25 Bootstrap Samples

(b) Trend in IEO in the First 25 Bootstrap Samples

(c) Change in Trend in IEO in the First 25 Bootstrap Samples
Outline

- Introduction
- Main results
- Inequality of Educational Opportunity
  - method:
    - Multiple indicators for socioeconomic status
    - Using lowess smooth to estimate trend and change in trend
    - Using bootstrap to estimate confidence envelopes
- results
- conclusions
Male trend

(b) Lowess Smooth and 90% Confidence Envelope

(c) Trend in IEO and 90% Confidence Envelope

(d) Change in Trend in IEO and 90% Confidence Envelope
Female trend

(b) Lowess Smooth and 90% Confidence Envelope

(c) Trend in IEO and 90% Confidence Envelope

(d) Change in Trend in IEO and 90% Confidence Envelope
Summary of results

Deceleration of the Trend in Inequality of Educational Opportunity in the Netherlands – p. 28/29
Conclusions

- IEO is well measured by linear regression if one is interested in comparing the total IEO of an educational system across time and/or across space.
- The trend in IEO was primarily negative.
- But, the trend in IEO in the Netherlands has slowed down since the 1970s and has become non-significant.