Long Term Trends in IEO in Italy
Birth cohorts 1899-1980

An analysis using uniform association models with heterogeneous linearly constrained scalings

Cinzia Meraviglia
University of Eastern Piedmont, Italy

Harry Ganzeboom
Free University Amsterdam, The Netherlands

International Research Workshop
Inequality of educational opportunity in Italy
Milan (IT), June 19 2009

Why Origin-Education?

- The OE association is so crucial because E can act as a factor for either mobility or immobility
- Empirical results not always confirm the most cherished hypothesis of E as a factor favouring mobility
- At an international level:
  - Shavit and Blossfeld (1993): persistent inequality an most countries (including Italy), except in Sweden and The Netherlands
  - Breen et al. (2005, 2009): various degree of change of IEO; two countries (Italy and Ireland) show no change/unclear pattern

Italian exceptionalism

- Lipset and Bendix (1959): ascription in Italy much higher than in other countries
- …
- Breen and Luijkx (2004): Italy is one of the least open countries in Europe
- Pisati and Schizzerotto (2004): no change in overall inequality (OD) over time

IEO in Italy

- Contrasting evidence:
  - IEO did not change over the last 20-30 years, except for some change concerning the agricultural classes;
  - IEO shows some signs of change over time;
- Evidence is mixed even when using the very same data set!
  - Pisati (2002), Breen et al. (2005, 2009) vs. Shavit and Westerbeek (1998)

Is there a methodological problem?

- Yes, actually more than one:

1. Data used
2. Measurement of crucial variables
3. Models applied
4. Analysis on M and W, or M + W

1. Data

- Two main groups of studies, and some exceptions:
  - Data collected/released before year 2000 (1985 survey): no change of IEO
    - 4500 cases at most
    - Exception: Shavit and Westerbeek 1998
  - Studies conducted after year 2000 (1985 + 1997): either no change (P&S 2004, B&al. 2005) or some degree of change in IEO (B&S 2005+)
    - 12000 cases ca.
    - Exceptions: G&T 1993; B&S 2005 (IALS); Checchi et al. 2007 (SHIW)

2. Measurement: Resp's E

- Degree actually attained
- Generally 4 categories (or years-equivalent)
  - Primary or lower (5) / low secondary (8) / high secondary (13) / tertiary (19)
- Not always clear how categories are built
  - Complete + incomplete primary + illiterate + incomplete low secondary
  - Complete low secondary + incomplete high sec.
  - 2-3 yrs vocational high sec. + 4-5 yrs degree (*maturità*)
  - Lower + higher tertiary

2. Measurement: Origin

- Social origin measured as:
  - Occupation: father / mother / dominance
  - Education: father / mother / dominance
  - Occupation + education
    - Barone 2006, Checchi&al. 2007: parental O or E has the same role in IEO
    - S&W 1998: downward trend in IEO in respect only to parent's E

2. Measurement: Time

- Birth cohorts
  - 10 or 5 years spaced
  - Variety of starting years (1899 ~ 1930) and ending years (1959 ~ 1980)
  - Multiplicative specification (B&S 2006, 2007, 2008)
- Birth years
  - Unit is the year or paired years (S&W 1998)

3. Models

- Log-linear / logit (Mare) models find no change or unclear pattern
  - Schadee&Schizzerotto 1987; Schizzerotto 1988; Cobalti 1990; Cobalti&Schizzerotto 1993, 1994; Pisati 2002; Breen et al. 2005, 2009
- Variants of loglinear models find (some) decrease
  - cumulative logit (B&S 2005+), scaled uniform association (M&G 2007), ordered probit (Checchi&al. 2007)
- OLS regression models find (limited) increase (!) in IEO
  - G&T 1993; Cobalti&Schizzerotto 1993

4. Gender

- Some studies include it as a control variable
  - S&W 1998; Checchi et al. 2007
- Some others do separate analyses
  - Cobalti 1990 (some constant difference between M and W, no difference in IEO trends);
  - G&T 1993 (no change for M, increasing IEO for W); M&G 2007, this study (no change for M, downward trend for W)
- Does the conclusion of "unclear pattern" also depend on pooling W and M together?

Contributions present paper

1. DATA
   - Unprecedented database (N=48000 ca.) from all available data (but Istat) covering cohorts 1900-1980
2. MODEL
   - Heterogeneously scaled uniform association models (variant of RC-II models), which allows for detailed data and powerful test of trend
3. QUALITY CHECKS
   - controls for survey characteristics

Data

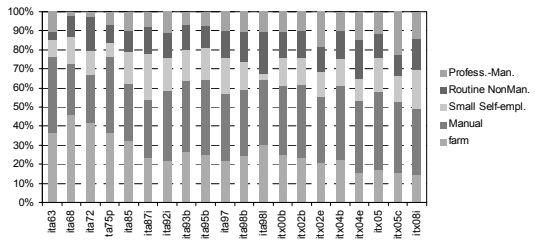
1963	Social mobility survey (Lopreato)
1968	Italian mass election survey (Barnes)
1972	Italian mass election survey (Barnes and Sani)
1975	Political Action project (Sartori, Marradi, Sani)
1985	Social mobility survey
1987 1992	ISSP Social Inequality I, II
1997	ILFI

1993 1995 1998 2000 2002 2004	Survey on Household Income and Wealth (SHIW), Bank of Italy
1998	IALS round II
2003 2006	ESS
2005	Osservatorio Nord Ovest (Ricolfi)
2005	Social desirability survey
2008	ISSP Religiosity III

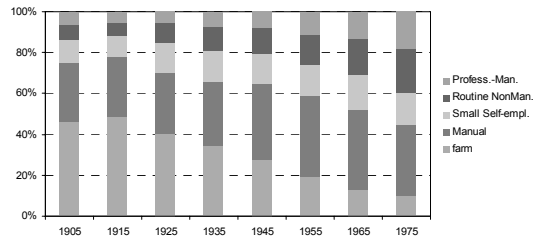
Variables

- Education: four categories
  - Primary (in/complete) + illiterate + incomplete low sec.
  - Lower secondary + incomplete high sec
  - Higher secondary (2-3 yrs voc. + 4-5 yrs)
  - Tertiary (lower + higher + post graduate)
- Parental occupation:
  - Father's occupation (if F's not available, mother's occupation)
  - Five EGP classes from harmonized ISCO68/88 code
  - I+II / IIIa+IIIb / IVa+IVb / V+VI / VIIa / VIIb+IVc
- Birth cohorts: 12 categories, variable width (10-yrs wide 1899-1930, 5 yrs wide 1931-1980)
  - 1899-1910 / 1911-1920 / 1921-1930 / 1931-1935 / 1936-1940/ 1941-1945 / 1946-1950 / 1951-1955 / 1956-1960 / 1961-1965 / 1966-1970 / 1971-1975 / 1976-1980

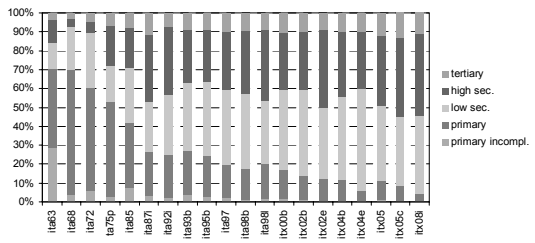
Origin by study, Men & Women



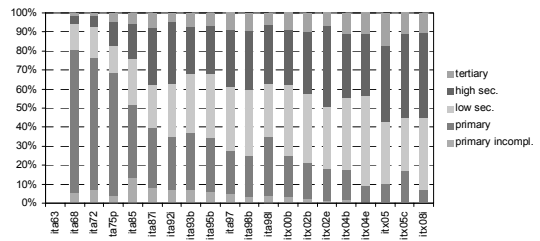
Origin by cohort, Men & Women



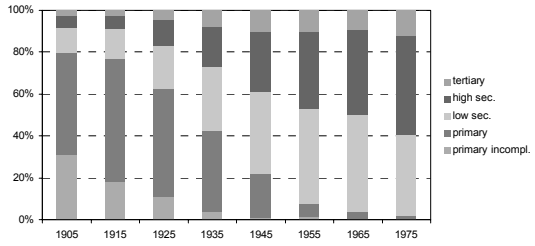
Education by study, Men



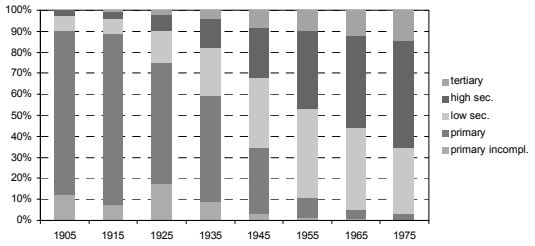
Education by study, Women



Education by cohort, Men



Education by cohort, Women



Scaled association model (1)

- In a given table R × C:

$$\ln \theta_{ij} = \varphi(\mu_{i+1} - \mu_i)(\nu_{j+1} - \nu_j)$$

- φ is the log-odds ratio for R / C categories one unit apart
- It expresses the (uniform) association between R and C, however scaled by the μ_i and ν_j
- We standardize μ_i and ν_j ➔ metric becomes like Pearson's correlation

Scaled association model (2)

- The μ_i are the distances between row (Origin) categories, given φ, i.e. given the pattern of association between columns (Education) categories
- Same for ν_j
- If all adjacent row and column categories have a distance of 1, then the model equals a standard uniform association model

Scaled association model (3)

- μ_i and ν_j parameters are estimated on two constraints:
  - on average they sum to zero
  - the standard deviation is 1 weighted by the data
- The μ_i and ν_j can be linearly constrained over tables (cohorts):

$$\mu_{ik} = \mu_{i0}(1 + \beta Y)$$
$$\nu_{jk} = \nu_{j0}(1 + \beta Y)$$

Some lexical notes…

- Homogeneous
  - one scaling parameter per R (C) per table
- Heterogeneous
  - as many scaling parameters as R (C) categories per table
- Heterogeneous (but) linearly constrained
  - as many scaling parameters as R (C) categories per table, however they have to develop linearly across tables (i.e. lay on a regression line over time)
- Heterogeneous linearly constrained and parallel
  - as in the previous case, plus (all/some) regression lines must be parallel between them

Models

- Reference models:
  - UniDiff, also with linear constraints
  - Constant Social Fluidity
- Scaled association models:
  - **Fully heterogeneous models**: both O and E scaling parameters free to vary across cohorts, with/out constraints on overall trend across cohorts
  - **Heterogeneous models** with linear constraints on O scalings and/or E scalings
  - **Heterogeneous models** with linear and parallel constraints on O scalings and/or E scalings

Model fit: Models without scalings

				Men		Women		Men & Women	
		DF Spss	DF Corr	L2	BIC	L2	BIC	L2	BIC
A1	Uniform Differences		121	252.6	-970.0	215.4	-1001.9	318.8	-985.1
A2	Uniform Differences * linear trend		131	280.9	-1042.8	233.5	**-1084.5**	346.0	-1065.6
A3	Constant Social Fluidity (LEM)		132	281.0	**-1052.8**	254.3	-1073.7	354.3	**-1068.1**
A3	Constant Social Fluidity (SPSS)	132	132	282.8	-1051.0	260.8	-1067.2	355.2	-1067.2

Model fit: Fully heterogeneous models

				Men		Women		Men & Women	
		DF Spss	DF Corr	L2	BIC	L2	BIC	L2	BIC
B1	OC, EC, O_cE_cC		72	**165.3**	**-562.2**	**156.6**	**-567.8**	**232.2**	**-543.7**
B1	Idem, SPSS	132	72	**206.7**	**-529.2**	**213.0**	**-518.9**	**231.8**	**-552.0**
B2	OC, EC, O_cE_cC_lin	142	82	**237.2**	**-600.8**	**225.8**	**-606.9**	**272.1**	**-619.5**
B3	OC, EC, O_cE_c	143	83	**238.1**	**-610.6**	**272.2**	**-572.9**	**289.3**	**-615.6**

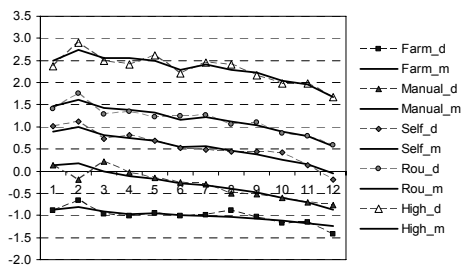
Model fit: Linearly constrained O scalings

				Men		Women		Men & Women	
		DF Spss	DF Corr	L2	BIC	L2	BIC	L2	BIC
C1	**OC, EC, O_LE_cC_lin**	142	113	268.7	-873.7	245.3	-890.0	323.3	-893.7
C2	**C1 + O4 homogeneous**	142	114	268.7	-883.8	245.3	-900.0	323.3	-904.5
C3	**C1 + O3 homogeneous**	142	115	269.1	-893.7	245.2	-910.1	323.7	-915.0
C4	**C1 + O5 homogeneous**	142	116	269.6	-902.9	244.3	-922.2	323.5	-926.2
C5	**C1 + O2 homogeneous**	142	117	269.9	-912.5	243.3	-933.2	323.1	-937.8
C6	**C1 + O1 homogeneous**	142	118	274.7	-918.1	266.0	-920.7	351.0	-921.6

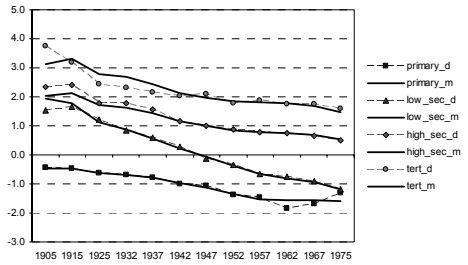
Model fit: Linearly constrained O+E scalings

				Men		Women		Men & Women	
		DF Spss	DF Corr	L2	BIC	L2	BIC	L2	BIC
D1	**OC, EC, O_LE_cC_lin**	142	117	269.9	-912.5	243.3	-933.2	323.1	-937.8
D2	**OC, EC, O_LE_LC_lin**	142	133	294.7	-1059.6	248.9	-1099.0	356.5	-1087.5
D3	**D2 + E4 homogeneous**	142	134	293.0	-1071.3	251.8	-1106.2	359.0	-1095.8
D4	**D3 + E3 homogeneous**	142	135	295.3	-1079.1	255.2	-1112.8	365.9	-1099.6
D5	**D4 + E2 homogeneous**	142	136	382.3	-1002.2	325.4	-1052.6	528.4	-947.9
D6	***D4 +C_lin (no trend)***	*142*	*137*	*295.8*	*-1088.7*	*298.9*	*-1079.1*	*379.9*	*-1096.4*

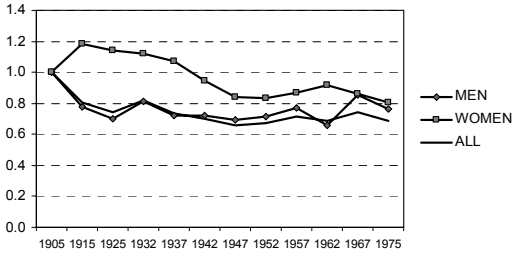
Scaling parameters: Origin × Cohort



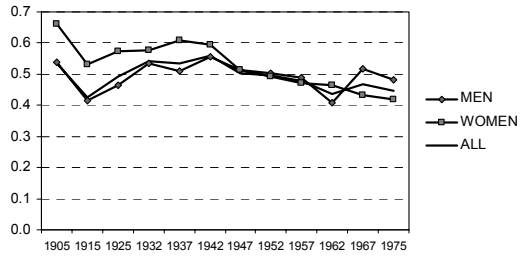
Scaling parameters: Education × Cohort



OE association: UniDiff



OE association: Model D4



Sensitivity checks

- See paper / handouts, Table 10
- We estimated model D4 on each of our 20 data sets separately
  - (Model D4: O + E scalings linearly constrained; E scalings for primary, high secondary snd tertiary are linear and parallel to each other)
- Men: the trend over cohorts is highly variable across studies
- Women: the trend is still there in all studies

Trends with quality controls

Table 11: Scaled uniform association model D4 with data quality controls

	Controls	MEN		WOMEN		MEN & WOMEN	
		B0	B1	B0	B1	B0	B1
1	+ COHNR*TABLE	0.545	-0.061	0.761	-0.372	0.579	-0.131
			-1.6		-7.6		-4.4
2	+ FOCC*TABLE	0.570	-0.090	0.795	-0.409	0.610	-0.166
			-2.2		-8.3		-5.4
3	+ EDCAT*TABLE	0.571	-0.086	0.795	-0.408	0.609	-0.162
			-2.1		-8.2		-5.3
3	+ OE*TABLE	*	-0.057	*	-0.320	*	-0.124
			-1.0		-5.2		-3.0

Conclusions (1)

- Scaled association models represent the data well: BIC's are lower than for UniDiff
- Scaling parameters for education confirm Mare pattern
  - Large distances at lower levels of education
  - Small distances at higher levels of education
  - The heterogeneous model shows a clear change in relative position in levels (lower secondary closer to primary)
- Scaling parameters for classes show:
  - Large distances at the top of the hierarchy
  - Large distance at bottom of the hierarchy (farmers)
  - No trend over time, except for farm

Conclusions (2)

- Association parameters:
  - Minor but significant downward linear trend for men
  - Major linear downward linear trend for women
- Quality controls do not make much of a difference: when you use 20 surveys, the differences balance out.

Models

1. OC, EC, β_COE	UniDiff
2. OC, EC, $O_C E_C$C	• OE varies across cohorts • heterogeneous scalings for O + E
2a. OC, EC, OE_CC	• OE varies across cohorts • homogeneous scalings for O • heterogeneous scalings for E
2b. OC, EC, OEC	• OE varies across cohorts • homogeneous scalings for O + E
2c. OC, EC, OE_{C_lin}C	• OE varies across cohorts • homogeneous scalings for O • heterogeneous linearly constrained scalings for E
2d. OC, EC, OE_{C_lin}C_lin	• OE linearly constrained across cohorts • homogeneous scalings for O • heterogeneous linearly constrained scalings for E
2e. OC, EC, OE_{C_lin}	• OE constant across cohorts • homogeneous scalings for O • heterogeneous linearly constrained scalings for E
3. OC, EC, OEC_lin	• OE linearly constrained across cohorts • homogeneous scalings for O + E
4. OC, EC, OE	• OE constant across cohorts • homogeneous scalings for O + E