A new dataset on educational inequality

Meschi, E.; Scervini, F.

Publication date
2010
Document Version
Final published version

GINI

A New Dataset of Educational Inequality

Elena Meschi and Francesco Scervini

GINI Discussion Paper 3
December 2010

GROWING INEQUALITIES' IMPACTS

Acknowledgement

We thank Gabriele Ballarino, Michela Braga, Massimiliano Bratti, Daniele Checchi, Antonio Filippin, Carlo Fiorio, Marco Leonardi for suggestions and useful discussions since the early stages of this work. Precious insights by participants in the gini's Measurements and Methods Workshop, Amsterdam, October 2010, are also gratefully acknowledged.

December 2010

General contact: gini@uva.nl

Correspondending addresses: Elena Meschi & Francesco Scervini. University of Milan – deas, Department of Economics, Business and Statistics – Via Conservatorio, 7 – 20122 – Milan (Italy)

Elena Meschi. Department of Quantitative Social Science – Institute of Education, 20 Bedford Way, London WC1H0AL (UK)

Bibliographic Information

Meschi, E., Scervini, F. (2010). A new dataset on educational inequality. Amsterdam, AIAS, GINI Discussion Paper 3.

This paper can be downloaded from our website www.gini-research.org.

GINI GROWING INEQUALITIES' IMPACTS

A New Dataset on Educational Inequality

Elena Meschi

University of Milan — DEAS

Department of Quantitative Social Science — Institute of Education

elena.meschi@unimi.it

Francesco Scervini

University of Milan — DEAS

francesco.scervini@unimi.it

15 December 2010
DP 3

Table of contents

Abstract

This paper describes a new dataset collecting measures of educational level and inequality for 31 countries over several birth cohorts. Drawing on four representative international datasets (ESS, EIJ-SILC, IALS and ISSP), we collect measures of individual educational attainment and aggregate them to generate synthetic indices of education level and dispersion by countries and birth cohorts. The paper provides a detailed description of the procedures and methodologies adopted to build the new dataset, analyses the validity and consistency of the measures across surveys and discusses the relevance of these data for future research.

The .csv, .dta, .xml dataset is readily available to GINI members (see website data portal); other scholars may put in a request to the authors.

JEL Classification: I21, D30, Y1

Keywords: new dataset, educational attainment, educational inequality, dispersion indices, cross-country, cohorts

1. Introduction

The positive effects of human capital are widely recognised in the economic literature. Accumulation of human capital is proved to be crucial for economic growth (see, among others, Hanushek and Kimko (2000); Krueger and Lindahl (2001); de la Fuente and Domenech (2006)), and beneficial to individuals and societies. From the individual point of view, more educated people are more likely to have better labour market outcomes in terms of employability and wages (see the survey by Harmon et al. (2003)), but they also experience better health, fertility, well-being, less probability of engaging in crime and other non monetary outcomes (see for example Lochner and Moretti (2004) and the survey by Grossman (2006)). Moreover, education has wide spill-over effects, generating not only private, but also public benefits. Increasing education seems in fact to influence positively also social cohesion, citizenship, and political participation (Milligan et al., 2004).

The effects of the distribution of educational attainment have in contrast not been so widely studied by the economic literature. Nevertheless, the way human capital is distributed across the population may also have important economic consequences, affecting for example income distribution (see Checchi (2004); De Gregorio and Lee (2002); Park (1996)), and economic growth (see Lopez et al. (1998)). However, the empirical evidence on this topic is still very scant, possibly due to the lack of crosscountry comparable data on educational inequality.

We have created an original dataset that provides cross-country measures of educational attainment and inequality, drawing on four international representative surveys (ESS[1], EU-SILC[2], IALS[3], ISSP[4]). The dataset covers a sample of 31 countries[5], and provides a wide set of indicators of educational attainment and educational inequality over several birth cohorts. In particular we observe 13 cohorts, aggregated at 5-years intervals, where the first cohort is made of individuals born between 1920 and 1924, and the last one includes people born between 1980 and 1984.

Our dataset presents some remarkable novelties compared to the existing datasets on educational attainment (e.g. Barro and Lee (1996, 2001, 2010); Cohen and Soto (2007)). First of all, our data are organised by birth cohort rather than survey years. This allows to enlarge the period covered, as we have information from the beginning of the 20s, while existing datasets generally start from the 50s or 60s (see, for example, Barro and Lee (2010); Cohen and Soto (2007)). Moreover, this approach allows to observe the evolution of educational attainment for individuals born in different periods, possibly characterized by distinct institutional features of the school systems and this is particularly useful for analyses of the determinants of educational outcomes. Since formal education occurs

1 European Social Survey
2 European Union Statistics on Income and Living Conditions
3 International Adult Literacy Survey
4 International Social Survey Programme
5 The dataset covers 31 countries, 25 EU members states (all EU members except Cyprus and Malta), plus other six sizeable OECD countries (Australia, Canada, Japan, Norway, Switzerland, United States)

mainly in early stage of life and remains invariant over the life cycle, a cohort approach seems more sensible in this case. Second, we use multiple sources to create the aggregate measures and this improves the robustness and reliability of the data. Third, our dataset presents three alternative (but complementary) measures of education. Beside the standard variable on years of education completed, it also contains information on highest qualification achieved and on individual actual competences, that are considered better measures of individual human capital (Hanushek and Kimko, 2000; Hanushek and Woessmann, 2010). Fourth, our measure of years of schooling is directly derived from microdata and therefore based on people's answers and not computed by imputation starting from the information on educational attainment that typically only distinguishes six or less categories (see Barro and Lee (2010)). Finally, our dataset is the first one to provide information not only on educational level, but also on educational distribution, using a wide range of synthetic indices.

The aim of this paper is to explain the procedures implemented to generate the set of indices, describe the data sources, discuss some methodological issues, and check the consistency of the measures across different surveys. In particular, section 2 describes the original measures of education used to create our statistics, and section 3 explains how we aggregate the individual data and mentions some selection rules adopted to increase the homogeneity of the sample. The four data sources are described in section 4 that also discusses methodological issues in each survey. Section 5 presents the entire set of variables included in our datasets. The consistency of our measures across surveys is analysed and commented in section 6.

2. Measures of educational attainment

Measuring education in an international context is not straightforward, due to the difficult comparability of educational data across countries. We shall use three widely used measures of education, each one presenting advantages and drawbacks.

First, we use an indicator of the duration of formal schooling, measured by the number of years of education. The main advantage of this variable is that years of schooling can be easily computed and compared across countries. On the other side, this measure refers only to the "quantity" of education, disregarding its effectiveness, both in terms of results and in terms of different careers and achievements. In other words, focusing on years of education does not allow to discern general/academic from vocational secondary schools, it does not consider the effective attainment of any certificate/diploma/degree and it misleadingly accounts for year repetitions. Therefore, one has to keep in mind that the same years of schooling may entail different amount of learning and heterogeneous quality in different countries. Because of these reasons, such a measure is intuitively rather bad in assessing the qualification achieved, but it is suitable to compute inequality indices. Indeed, even if it is not properly continuous, it spans from 0 to 25-30 years, depending on highest level educational institutions, with an acceptable level of variability.

The second measure is based on attainment levels, capturing the highest qualification achieved. The advantage of this variable with respect to years of education is that it accounts for different duration of analogous school cycles. Moreover the use of a categorical indicator also allows to specify the type of education completed (i.e. academic vs. vocational tracks, etc.). On the other hand, a drawback of this measure is that it is a discrete categorization that disregards by definition intermediate cases, such as dropouts or partial attendance and furthermore, being a categorical variable, it performs poorly in generating aggregation statistics measuring inequality. Another important drawback of this measure in cross-country analyses is that levels, types and duration of specific educational programmes depend on the institutional structure of educational systems and, given the high degree of differentiation of educational systems across countries and over time, it is difficult to construct a classification of educational qualification that is valid and comparable internationally.

A common internationally harmonised measure of educational attainment is the International Standard Classification of Education (ISCED) developed by UNESCO[6] in 1976 and then revised in 1997[7] that aims at ranking and classifying individuals according to the maximum level of educational attained. ISCED should make it possible to

6 United Nations Educational, Scientific and Cultural Organisation
7 See Unesco (2006); Oecd (1999)

readily draw a comparable picture of education distribution across countries, independently of the duration of cycles, the length of compulsory schooling, and the heterogeneous structure of vocational and academic tracks. In particular, the classification distinguishes seven levels of education ranging from pre-primary education to second stage of tertiary education. Mapping tables to link national educational programmes to these categories are available in OECD (1999), and EUROSTAT (2005). However, although the international standards are clearly defined, it is sometimes problematic for countries to integrate these standards with the national systems. In fact, as pointed out by Eurydice studies[8], not all school systems find univocal correspondence to ISCED: among European countries, for instance, many Nordic and Eastern countries9 do not distinguish institutionally between primary and lower secondary levels, so that there is no distinction between ISCED level 1 and ISCED level 2. In other countries it is difficult to differentiate between ISCED 3 and 4 because the certificates awarded are the same, and just the pathway taken differs; the distinction between ISCED 4 and ISCED 5 is also not straightforward in countries with a unitary university structure that does not separate vocational and academic/professional tertiary studies (see Schneider (2007, 2009, 2010) for a detailed discussion of these problems). Moreover, information on ISCED levels are not directly asked to individuals when collecting data surveys, but they are derived from the answers they provide to country-specific questions. It may happen, therefore, that national questionnaires include less than 7 categories for educational attainment (this is the case, for instance, in most of the waves of ISSP), or that they do not separate vocational from academic schools. Furthermore, even if national questionnaires are detailed enough, it is still possible to observe mistakes in the assignment of national educational programmes to ISCED categories. In fact, Schneider (2007, 2010) studied the implementation of ISCED in ESS and showed that, in many cases, the national teams did not follow the definitions established by UNESCO and OECD and different classification decisions were taken with respect to similar educational programmes in different countries

In order to reduce measurement error due do imprecise coding, we create a simplified version of ISCED, that eliminates some controversial distinctions and includes four category only (ISCED4 hereafter). Table 1 illustrates the conversion of the original ISCED categories into our variable ISCED4.

8 www.eurydice.org

9 Bulgaria, Czech Republic, Denmark, Estonia, Latvia, Hungary, Slovenia, Slovak Republic, Finland, Sweden, Iceland, Norway, Turkey

Table 1: Official isced classification and correspondence to authors' simplified version

ISCED classification		Simplified ISCED4 classification	
Level	Stage of education	Level	Stage of education
0	Pre-primary	1	Primary or below
1	Primary or first stage of basic education		
2	Lower secondary or second stage of basic education	2	Lower secondary
3	(Upper) secondary	3	(Upper) secondary
4	Post-secondary non-tertiary		
5	First stage of tertiary	4	Any tertiary
6	Second stage of tertiary		

The third measure of education consists of achievement data measuring people's actual competences. We gather this information from IALS that provides detailed assessment of the literacy skills of respondents (see section 4.3 for a description of the tests). The main advantage of achievement data is that by testing what people actually know, they measure effective skills and are thus related to both the quantity and quality of schooling. Another advantage of this measure is that these tests were specifically designed for cross-country comparison and are therefore readily comparable across country. Moreover, test scores are continuous-like variables and are therefore suitable to create inequality indices. On the other hand, skills are not stable over the life cycle and are likely to be affected by many factors other than school.

Unfortunately, not all the measures are available and reliable in all the data sources used in the paper. Section 4 explains and discusses which variables were taken from the different surveys.

3. Data aggregation and selection rules

After having selected variables on individual education from the four surveys (see section 4 for details), we aggregate the data by countries and birth cohorts and generate the synthetic indices of educational attainment and inequality described in the next section. We construct a pseudo panel, exploiting cohorts of birth as time dimension. Since education occurs mainly in early stage of life, formal education is unchanged over the life cycle of every individual out of the schooling system. The more immediate consequence of this observation is that yearly variation in statistics on education are mainly due to the difference between the very low fraction of individuals who exit the education system in the meanwhile and the even lower fraction of individuals who eventually died. Therefore observing the education pattern over cohorts is much more variable and informative. Another advantage of focusing on cohorts rather than survey years is that we can pull the different surveys' waves together, thus increasing the sample size for each country. In particular we create 13 cohorts at 5-years intervals. The first cohort is made of individuals born between 1920 and 1924, the second between 1925 and 1929 and so on until the last cohort made of people born between 1980 and 1984 (see table 4).

In order to generate a meaningful and comparable set of indicators, we put some effort in selecting the most suitable observations, by trimming the sample over two dimensions in all the considered datasets: first of all, we exclude all individuals aged less than 25. The rationale behind this choice is to exclude individuals who did not complete their study at the time of the survey. This is a source of great disturbances, since demographic trends could heavily affect the results as far as the share of young individuals still enrolled in formal schooling lowers the level of education and increases its dispersion. We choose 25 years as a threshold as this is the standard definition of the beginning of adult age according to most of the literature in the field (see for instance Barro and Lee (2010)). Second, we drop individuals reporting an amount of education over 30 years, assuming that it is the highest reasonable duration of a complete educational career, starting at 5-6 years of age and ending with a Ph.D., and that additional years of education do not reflect real improvements but either misreported data or repeated years. In principle, there is a window of individuals older than 25 and with less than 30 years of education who are in the sample even if still enrolled in tertiary education. Such individuals should in principle be excluded, but since it is not possible to discern whether they are still in education, we are forced to include them in our sample. However, the share of such individuals is so little that it is not expected to significantly affect the results.

Finally, we keep only the cells (country-cohort) with more than 50 observations. Calculating summary statistics based on small sample sizes may in fact lead to imprecise measures. Therefore dropping the cells with less that 50 observations should reduce the measurement error. Moreover, we report the number of observations for each country and cohort, in order to allow the weighting of the aggregate measures according to the sample size in each cell.

4. Data sources

4.1. European Social Survey (ESS)

The European Social Survey (ESS) is biennial survey that covers over 30 mostly European countries and provides detailed information on individuals attitudes, beliefs, and behaviour collected from nationally representative population samples.10 It consists of repeated cross-sections initiated in 2002/2003 (first round) and then carried out in 2004/2005 (2nd round), 2006/2007 (3rd round) and 2008/2009 (4th round). The survey mainly focuses on people's attitudes and values, but it also contains several variables capturing the social background of respondents as well as their partners and parents. As far as education is concerned, the ESS includes three measures of educational attainment: actual (full-time equivalent) years of education, highest level of education in a country-specific format, and highest level of education in the internationally harmonised format (ISCED) derived from the national questions. From ESS, we will thus take the two variables on years of schooling and ISCED levels. Interestingly, since ESS provides educational attainment country-specific format, we were able to correct for misclassifications of country-specific variables into ISCED, using the guidelines provided in Eurydice, OECD and following suggestions in Schneider (2007, 2009, 2010).

4.2. European Union Statistics on Income and Living Conditions (EU-SILC)

The European Union Statistics on Income and Living Conditions (EU-SILC)[11] is a collection of timely and comparable multidimensional microdata covering EU countries plus Iceland and Norway.

EU-SILC is a project developed by EUROSTAT, run yearly since 2004 and including both cross-section and longitudinal surveys. Since we use birth cohorts as time dimension and since education is time-invariant, we are interested only in the cross-sectional element. Unfortunately, opposite to other surveys, for EU-SILC we cannot pull all the waves together, because even cross-sectional data contain a longitudinal dimension, and it is not possible to detect repeated observations in different cross-section files. This means that merging different waves would result in including some individuals up to four times. Therefore we had to select only one wave and, among the five available cross-sections, we chose the 2008 wave, be-

10 ESS is a project started in 2001, directed by a consortium of seven academic institutions and funded by The European Commission, The European Science Foundation and National academic funding bodies (see http://www.europeansocialsurvey.org).

11 http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/eu_silc

cause it includes the largest number of countries and observations, and it is also the most recent available.12 The main advantages of EU-SILC consist in the wide country coverage (all eu-27 countries (plus Iceland, Norway and Switzerland) and the large sample size within each country. The number of individuals in each country is in fact substantially higher than that in IALS and ESS and similar to that in ISSP, which is collected using repeated waves (see section 4.4). Therefore in EU-SILC, no cohorts are excluded either because of the low number of observations or because of consistency issues. However, there is also a major shortcoming in the data: there is no information on years of schooling, so that we can only compute statistics on the highest level of education attained.

4.3. International Adult Literacy Survey (IALS)

The International Adult Literacy Survey (IALS) is a survey collecting information on adult literacy in various countries, with the aim of comparing literacy levels across countries. The survey was coordinated by Statistics Canada and it was implemented in different years - 1994, 1996, 1998 - for different countries using a common questionnaire. In 1994, nine countries (Canada,[13] France,[14] Germany, Ireland, Netherlands, Poland, Sweden, Switzerland - German and French-speaking regions - and United States) were surveyed. Five additional countries (Australia,[15] Flemish Belgium, Great Britain, New Zealand and Northern Ireland) were included in 1996; and other countries (Chile, Czech Republic, Denmark, Finland, Hungary, Italy, Norway, Slovenia and the Italian-speaking region of Switzerland) participated in the third round of data collection in 1998. Our final IALS sample includes 16 countries, namely Flemish Belgium, Czech Republic, Denmark, Finland, Germany, Hungary, Ireland, Italy, Netherlands, Norway, Poland, Slovenia, Sweden, Switzerland, United Kingdom and United States. Since data were collected in late 90s, the youngest cohort included in our dataset is the 1970-1974 cohort, available for ten countries only. Moreover, because of the small sample size for most countries, the first two cohorts (1920-1924 and 1925-1929) include only Netherlands and Sweden.

The central element of the survey is the direct assessment of the literacy skills of respondents, but the background questionnaire also includes several information on individual socio-demographic characteristics. Regarding education, the survey contains the standard question on the number of years of formal education completed and the highest level attained. We can therefore use these questions to create all the indices described above.

12 As in November 2010, data relative to 2008 are available for all countries apart from France, for which we used data in EU-SILC 2007.
13 Canada is not included in our dataset since there are no informations available on the year of birth, which makes it impossible to generate cohorts.
14 France decided to withdraw from the study in November 1995, citing concerns over comparability and it is therefore not included in the dataset.
15 Data for Australia are not included in the main IALS dataset, and are available only through the Australian Bureau of Statistics, for confidentiality reasons.

As measure of individual competences we use the results obtained in different tests. In particular, IALS contains three tests, each one measuring a particular dimension of literacy: prose, document, and quantitative. Prose literacy is defined as the knowledge and skills needed to understand and use information from texts including editorials, news stories, poems, and fiction. Document literacy measures the knowledge and skills required to locate and use information contained in various formats, including job applications, payroll forms, transportation schedules, maps, tables, and graphics. Quantitative literacy is defined as the knowledge and skills required to apply arithmetic operations (see IALS User guide for further details). The results of the tests are scaled in the range of 0-500 and our measure of individual "skills" is calculated by averaging the scores obtained in the three tests.

4.4. International Social Survey Programme (ISSP)

ISSP[16] is a continuing annual programme of cross-national collaboration on surveys covering a wide range of topics. Data collection is annual, ranging from 1985 to 2008 (last wave available in late 2010), covering 26 countries, most of which are surveyed only in a subset of years (only United Kingdom and United States are available for the whole period), for a total of 388 datasets.

Even if ISSP does not focus specifically on education and the main topic changes every year, its characteristics make it very attractive for analysing in details the evolution of average years of schooling: data on education are obtained by the personal characteristics section, which is included in every survey and contains information on two of our three dimensions, namely years of education and attainment. However, only the former is exploitable, since it is not subject to recoding or rescaling and is therefore absolutely comparable both cross-country and over time. Opposite, attainment is virtually impossible to be analysed, since the coding changed several times both across countries and over time, making the harmonisation very challenging, since it would require comparing answers from almost 400 different datasets. Since ISSP is not coordinated by European Union agencies, the survey also includes non-European countries, such as Japan, Australia, Canada.

A careful analysis of the relevant variable (Years of education) across surveys let emerge many issues that have been faced in order to generate a reliable and consistent set of indicators. First of all, it frequently happened that two next surveys were submitted together to the same sample of individuals. In this case, in order not to replicate observations, it was necessary to drop one of the two surveys. For instance, the 2005 survey "Work Orientation III" was submitted in Bulgaria together with the 2004 survey "Citizenship" to the same sample of 1,121 individuals. In this case, therefore, one of the two samples were excluded, in particular the one relative to 2004, since data have

16 http://www.issp.org/index.php

been collected in July 2005 (by the way, since we focus on the cohort dimension, the exact time of data collection is not relevant). This procedure led to the exclusion of 31 datasets.[17]

Second, datasets were compared across waves, in order to detect possible data inconsistencies. Over 357 datasets remaining, we detected evident data inconsistency only for France 1999, in which the average level of years of education was 6 years higher than both previous and subsequent waves, due to some kind of measurement error (years of education takes an unrealistic minimum value of 10). Moreover, we also exclude datasets for Bulgaria, Canada, Germany and Spain in 1999, since data on years of education were missing for all individuals. In addition, we also exclude all observations for United Kingdom and United States because data were unfortunately bottom- and top-coded (8-20 years for UK, 0-20 years for US), making the comparison to other countries impossible.

Aggregation of all surveys generates a single dataset including more than 300,000 observations, that are in turn partitioned over countries and cohorts, according to the general criteria. The only cells with less than 50 individuals are Finland in the first cohort, Italy in the two youngest cohorts and Canada and Hungary in the last one.

4.5. Summary

Table 2 summarises the measures of education chosen from the four surveys. While the variable on years of education is available and reliable in ESS, IALS and ISSP, the information on the highest qualification achieved is only available in ESS, IALS and EU-SILC. Finally, competences are only included in IALS, which was specially designed to test adults' skills.

Table 2: Summary of measures of education in the four surveys

	ESS	EU-SILC	IALS	ISSP
Years of education	x		x	x.
Qualifications	x	x	x	
Competences			x	

Overall, we have information on 31 countries, most of which are European and covered in all the surveys, as reported in Table 3. Extra-European countries are available in IALS (e.g United States) or ISSP (e.g. Japan and Australia only).[18] However, the number of countries covered in each survey is not constant across cohorts. For some countries, the number of observations for some cohorts (typically the first or the last ones) is lower than 50 and we decided to exclude these cells in order to avoid to calculate imprecise statistics based on small sample sizes. Table 4 reports the number of countries covered in each survey for the different cohorts.

17 Australia 1993; Austria 1987, 2001, 2003, 2007, Bulgaria 1996, 1998, 2004; Canada 1998, 2000, 2003, 2005, Ireland 1990, 1995, 2004, 2005, 2007; Italy 1990; Latvia 2002, 2006; Netherland 2003; Poland 2003; Portugal 2003, 2005; Slovak Republic 2007; Slovenia 1993, 1999, 2001; Switzerland 1996, 2004; United States 2005.

18 Given the relative homogeneity of education data across surveys, we could have potentially merged observations from different surveys and calculated a unique summary statistics. However, since the four surveys contain different individual weights, it would have been difficult to compute aggregate measures representative of the real population and we therefore decided to keep the datasets separate.

Table 3: Countries covered by different surveys

Countries	ESS	EU-SILC	IALS	ISSP	Countries	ESS	EU-SILC	IALS	ISSP
Australia				x	Japan				x
Austria	x	x		x	Latvia	x	x		x
Belgium	x	x			Lithuania		x		
Belgium (Flanders)			x	x	Luxembourg	x	x		
Bulgaria	x	x		x	Netherlands	x	x	x	x
Canada				x	Norway	x	x	x	x
Czech Republic	x	x	x	x	Poland	x	x	x	x
Denmark	x	x	x	x	Portugal	x	x		x
Estonia	x	x			Romania	x	x		
Finland	x	x	x	x	Slovak Republic	x	x		x
France	x	x		x	Slovenia	x	x	x	x
Germany	x	x	x	x	Spain	x	x		x
Greece	x	x			Sweden	x	x	x	x
Hungary	x	x	x	x	Switzerland	x		x	x
Ireland	x	x	x	x	United Kingdom	x	x	x	
Italy	x	x	x	x	United States			x	

Table 4: Number of countries covered by surveys and cohorts

Cohorts	ESS	EU-SILC	IALS	ISSP
20-24	22		2	23
25-29	26		2	24
30-34	26	25	16	24
35-39	26	25	16	24
40-44	26	25	16	24
45-49	26	25	16	24
50-54	26	25	16	24
55-59	26	25	16	24
60-64	26	25	16	24
65-69	26	25	16	24
70-74	26	25	10	24
75-79	26	25		23
80-84	24	25		21

5. Variables

For each country and cohort, we calculate different measures of educational level and several indices of dispersion. All the statistics presented in the dataset have been computed using survey weights, which allow to make inference on the whole population. Educational levels are measured using the following statistics:

- Weighted average: $\mu_x = \frac{\sum_i x_i f_i}{\sum_i f_i}$, where x is either years of education or skills, f is the weight and i denotes the N individuals in the population;
- Percentages P of individuals who completed at least each ISCED level: $\%ISCED_k = \frac{\sum_{i=1}^{n}(1|ISCED_i \geq k)}{n}, k = 1, 2, 3, 4.$

Educational inequality is measured computing the following set of dispersion indices based on years of education and competences. These measures have been frequently used for income and consumption, but very few studies have calculated them on education (see for example Checchi (2004) and Thomas et al. (2001)). One point that is worth mentioning is that the cardinality of the two measures (years of education and competences) is not theoretically doubtless (is a child obtaining a score 400 twice as competent as a child obtaining 200? Does a year of university give the same education as a year of primary school?). However, we disregard this issue here, since there are no generally agreed means to overcome this problem.

In particular, we derived the following set of inequality measures comparable across countries and over time (see Cowell (2009) for an exhaustive treatment of inequality indices and their properties):

- Standard deviation $\sigma_x = \sqrt{\frac{\sum_i (x_i f_i - \mu_x)^2}{N-1}}$, a "standardised" measure of the variance of a variable;
- Coefficient of variation: $cv_x = \frac{\sigma_x}{\mu_x}$, the standard deviation normalised by the mean;
- Gini index: $G_x = \frac{1}{2\mu} \sum_i \sum_j \frac{|x_i - x_j|}{N^2}, \forall i, j \in N$, the most used inequality index in the literature
- Generalized entropy family: $GE(\alpha) = \frac{1}{\alpha^2 - \alpha} \left[\frac{1}{N} \sum_i \left(\frac{x_i}{\mu} \right)^{\alpha} - 1 \right]$, a set of indices separable and bounded between zero and one. Opposite to the more widely used Gini index, these indices allow to perfectly decompose total inequality in "within" and "between" group inequality. Among all the possible values of, we choose the more frequently used:
  - Theil index, with $\alpha = 1$: $T_x = \frac{1}{N} \sum_i \left(\frac{x_i}{\mu} \ln \frac{x_i}{\mu} \right)$, introduced by Theil (1967) and extensively described among others by Conceicao and Ferreira (2000),
  - Mean logarithmic deviation (MLD), with $\alpha = 0$: $T_x = \frac{1}{N} \sum_i \left(\ln \frac{\mu}{x_i} \right)$, analogous to the Theil index, but characterized by a lower sensitivity index;

- Atkinson indices: $A_x = 1 - \frac{1}{\mu}\left(\frac{1}{N}\sum_i x_i^{1-\epsilon}\right)^{\frac{1}{1-\epsilon}}$ $(\epsilon = 0.5, 1, 2)$, a measure that allows different sensitivities to transfers to the low tail of the distribution. The formula above simplifies as follows:

$$A_x = 1 - \frac{1}{\mu}\left(\frac{1}{N}\sum_i x_i^{1-\epsilon}\right)^{\frac{1}{1-\epsilon}} = \begin{cases} 1 - \frac{1}{\mu}\left(\frac{1}{N}\sum_i \sqrt{x_i}\right)^2 & \text{if } \epsilon = 0.5 \\ 1 - \prod_i \frac{x_i}{\mu} & \text{if } \epsilon = 1 \\ 1 - \frac{\mu_h}{\mu} & \text{if } \epsilon = 2 \end{cases}$$

  where μ_h is the harmonic mean.

- Deciles (pc10, pc20 to pc90) and quartiles (pc25, pc50, pc75), in order to give an intuition of the shape of underlying distribution and to allow the readers to build interdecile ratios.

6. Consistency of measures across surveys

Finally, we check the consistency of the aggregate measures of educational attainment and educational inequality across the surveys.

Graphs in Appendices plot for each variable the measure obtained using the different data sources. Each point in the graph corresponds to a specific cell (country and cohort), calculated using the survey indicated in the axes and the 45 degree line represents the perfect equality of the computed measures.

If we look at figure 1 that analyses the consistency of the average years of schooling, we notice that we managed to achieve a good consistency across surveys, as all the points are substantially aligned around the 45 degree line. The same consideration applies to all the statistics based on years of schooling (see figures 2 to 20) which seem to be consistent across different data sources. An exception is the standard deviation of years of education which turns out to be fairly different in the three surveys, but this is not surprising since this variable is not normalised by the mean and it is likely to be affected by the different sample size. Therefore we can conclude that the measures of dispersion calculated on years of education are consistent and robust to the use of different sources of microdata, which proves their good reliability.

As far as attainment levels are concerned, the aggregate measures seem to be more dependent to the surveys used to generate the statistics. Figures 21 to 23 in fact show that the proportion of individuals that completed respectively lower secondary, upper secondary and tertiary education somehow differ when computed in ESS, EU-SILC or IALS.

This evidence confirms the difficult cross-country comparability of educational attainment, whose categorisation depends on the specific institutional features of national school systems. Even if an internationally harmonised measure has been used, the classification of educational levels may still be subject to some measurement error. On the contrary, years of schooling can be easily computed and compared internationally and indeed the indicators of educational levels and inequality based on this simple measure appear to be fairly consistent across surveys.

References

Barro, Robert J, and Jong-Wha Lee (1996) 'International measures of schooling years and schooling quality.' *American Economic Review* 86(2), 218-23

Barro, Robert J., and Jong-Wha Lee (2001) 'International data on educational attainment. updates and implications.' *Oxford Economic Papers* 3, 541-563

_ (2010) 'A new data set of educational attainment in the world, 1950-2010.' NBER Working Papers 15902, National Bureau of Economic Research, Inc. http://www. nber.org/papers/w15902.pdf

Checchi, Daniele (2004) 'Does educational achievement help to explain income inequality?' In *Inequality, Growth and Poverty in an Era of Liberalization and Globalization*, ed. Andrea Cornia (Oxford University Press) chapter 4

Cohen, Daniel, and Marcelo Soto (2007) 'Growth and human capital: good data, good results.' *Journal of Economic Growth* 12(1), 51-76

Conceicao, Pedro, and Pedro M. Ferreira (2000) 'The young person's guide to the theil index: Suggesting intuitive interpretations and exploring analytical applications.' *UTIP Working Paper Series*. http://ssrn.com/paper=228703

Cowell, Frank (2009) 'Measuring inequality.' http://darp.lse.ac.uk/MI3

De Gregorio, Jose, and Jong-Wha Lee (2002) 'Education and income inequality: New evidence from cross-country data.' *Review of Income and Wealth* 48(3), 395-416

de la Fuente, Angel, and Rafael Domenech (2006) 'Human capital in growth regressions: How much difference does data quality make?' *Journal of the European Economic Association* 4(1), 1-36

Eurostat (2005) *UOE-UNESCO-OECD-Eurostat data collection on education: Mapping of national education programmes to* isced-97 UOE-UNESCO-OECD-Eurostat

Grossman, Michael (2006) 'Education and nonmarket outcomes.' In *Handbook ofthe Economics of Education*, ed. Erik Hanushek and F. Welch (Elsevier) chapter 10, pp. 577633

Hanushek, Eric A., and Dennis D. Kimko (2000) 'Schooling, labor-force quality, and the growth of nations.' *American Economic Review* 90(5), 1184-1208

Hanushek, Eric A., and Ludger Woessmann (2010) 'The economics of international differences in educational achievement.' Technical Report

Harmon, Colm, Hessel Oosterbeek, and Ian Walker (2003) 'The returns to education: Microeconomics.' *Journal of Economic Surveys* 17(2), 115-156

Krueger, Alan B., and Mikael Lindahl (2001) 'Education for growth: Why and for whom?' *Journal ofEconomic Literature* 39(4), 1101-1136

Lochner, Lance, and Enrico Moretti (2004) 'The effect of education on crime: Evidence from prison inmates, arrests, and self-reports.' *American Economic Review* 94(1), 155189

Lopez, Ramon, Vinod Thomas, and Yan Wang (1998) 'Addressing the education puzzle : the distribution of education and economic reform.' Policy Research Working Paper Series 2031, The World Bank, December. http://ideas.repec.org/p/wbk/wbrwps/ 2031.html

Milligan, Kevin, Enrico Moretti, and Philip Oreopoulos (2004) 'Does education improve citizenship? Evidence from the United States and the United Kingdom.' *Journal of Public Economics* 88(9-10), 1667-1695

OECD (1999) Classifying educational programmes. *Manual for ISCED-97 implementation in OECD countries* OECD (Paris)

Park, Kang H. (1996) 'Educational expansion and educational inequality on income distribution.' *Economics of Education Review* 15(1), 51-58

Schneider, Silke L. (2007) 'Measuring educational attainment in cross-national surveys: The case of the European Social Survey.' Paper presented at the educ research group workshop of the equalsoc network

_ (2009) 'Confusing credentials: The cross-nationally comparable measurement of educational attainment.' PhD dissertation, Nuffield College, Oxford: University of Oxford

_ (2010) 'Nominal comparability is not enough: (in-)equivalence of construct validity of cross-national measures of educational attainment in the european social survey.' *Research in Social Stratification and Mobility* 28(3), 343 - 357

Theil, Henri (1967) *Economics and Information Theory* (Amsterdam: North Holland)

Thomas, Vinod, Yan Wang, and Xibo Fan (2001) 'Measuring education inequality -Gini coefficients of education.' Policy Research Working Paper Series 2525, The World Bank, January

UNESCO (2006) *Global education digest 2006. Comparing education statistics across the world* Institute for Statistics (Montreal)

A Consistency graphics — Inequality indices

Figure 1

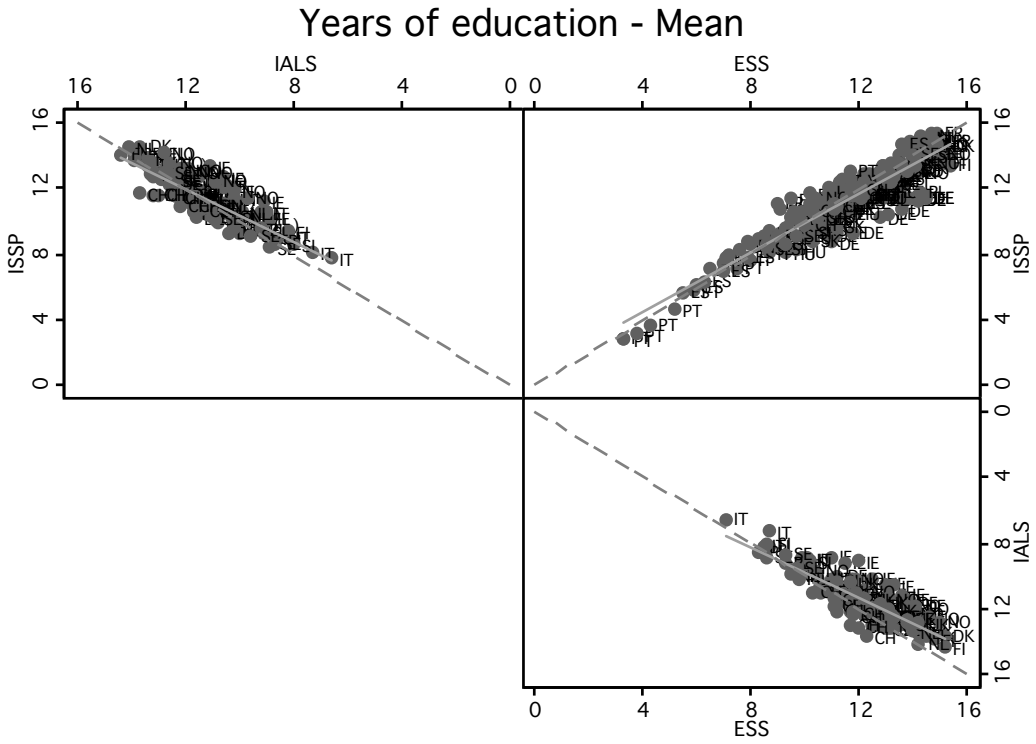


Figure 2

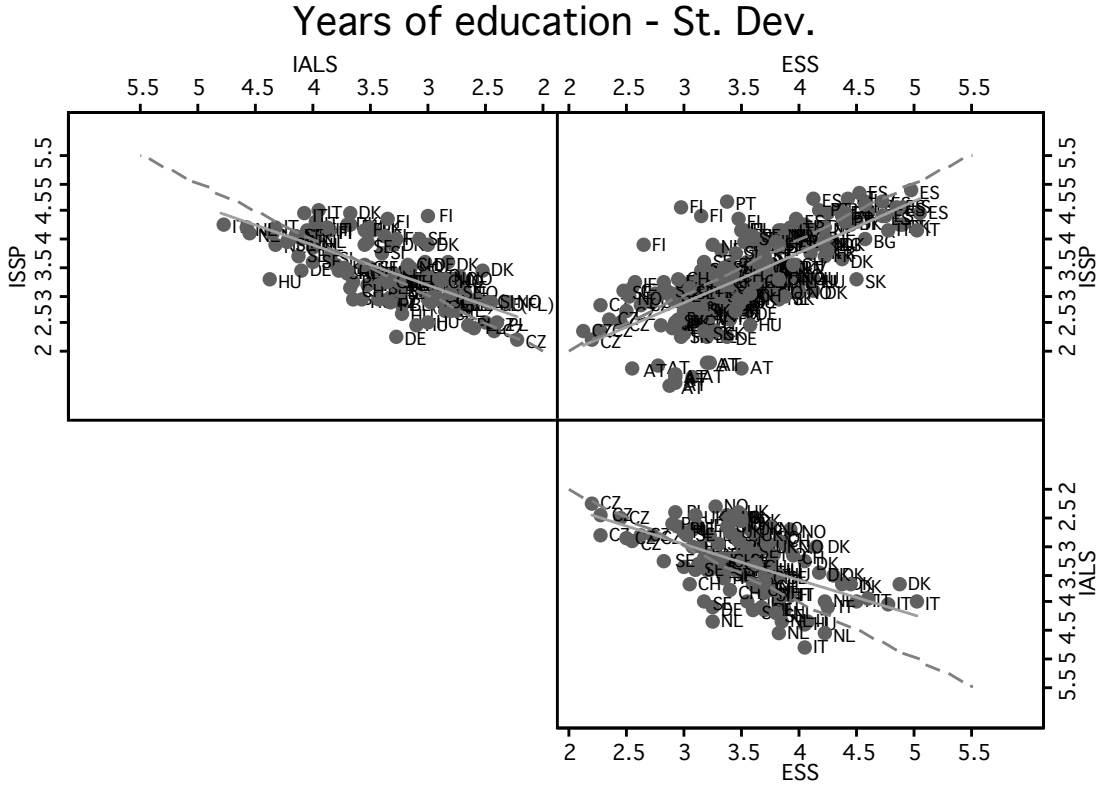


Figure 3

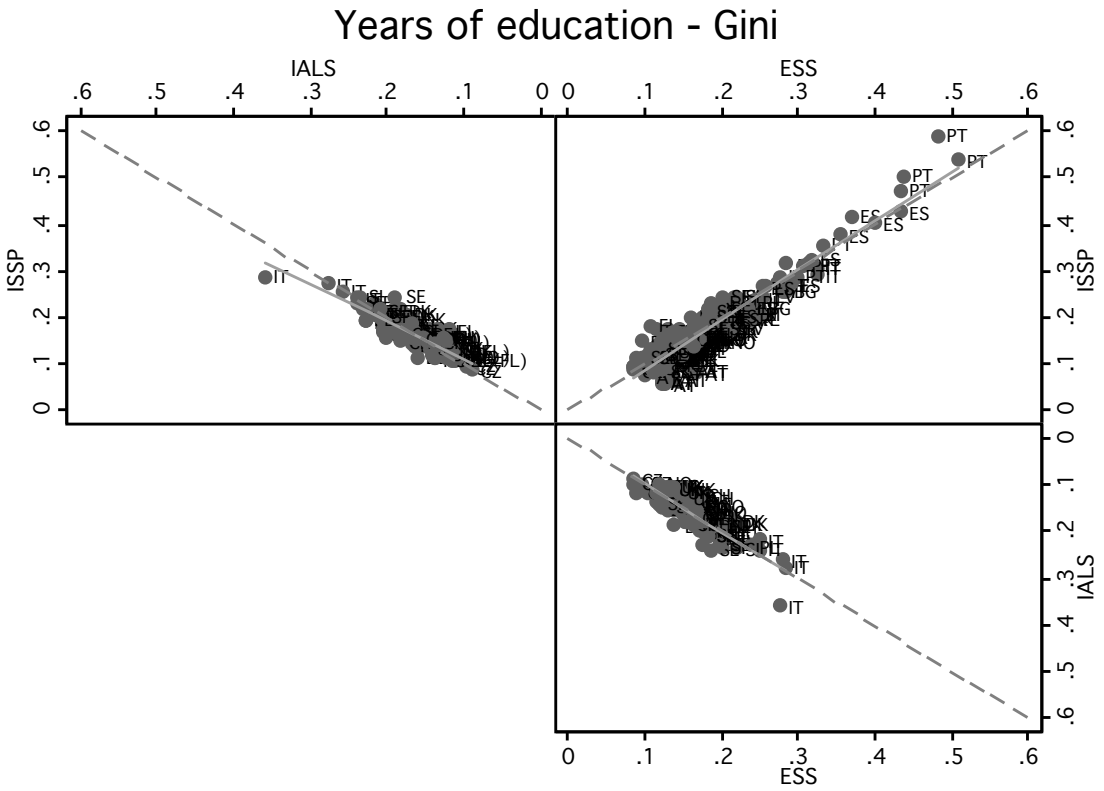


Figure 4

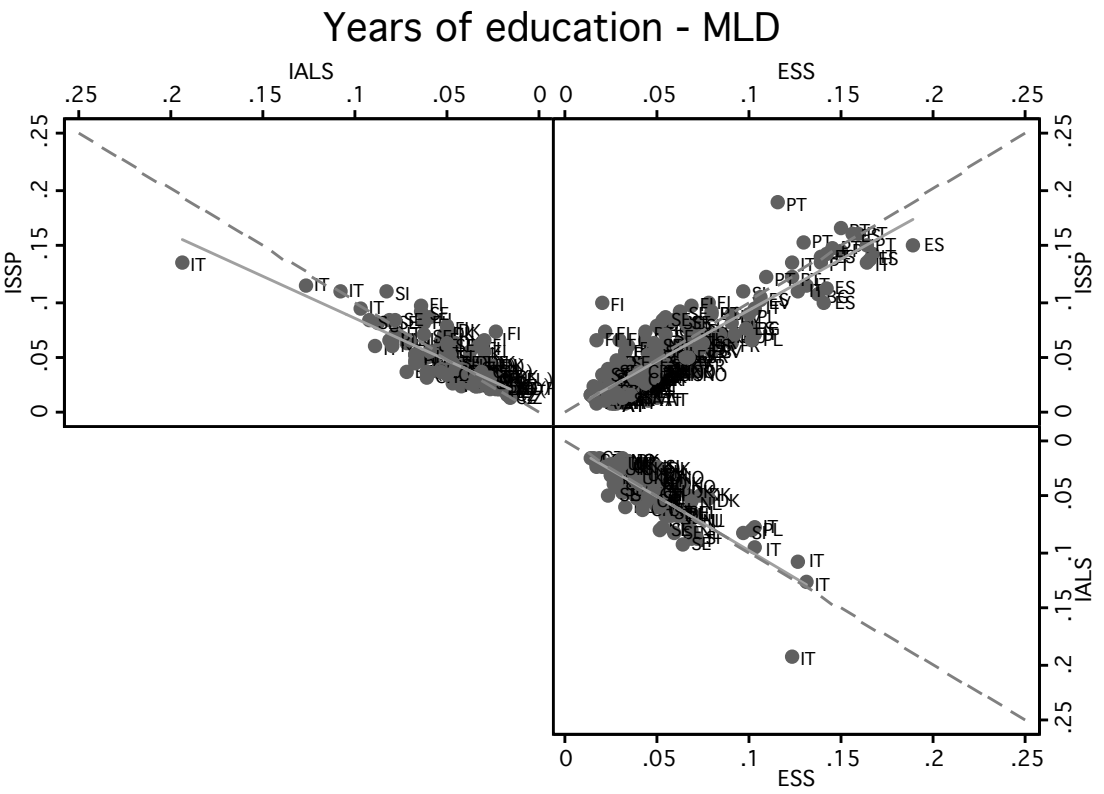


Figure 5

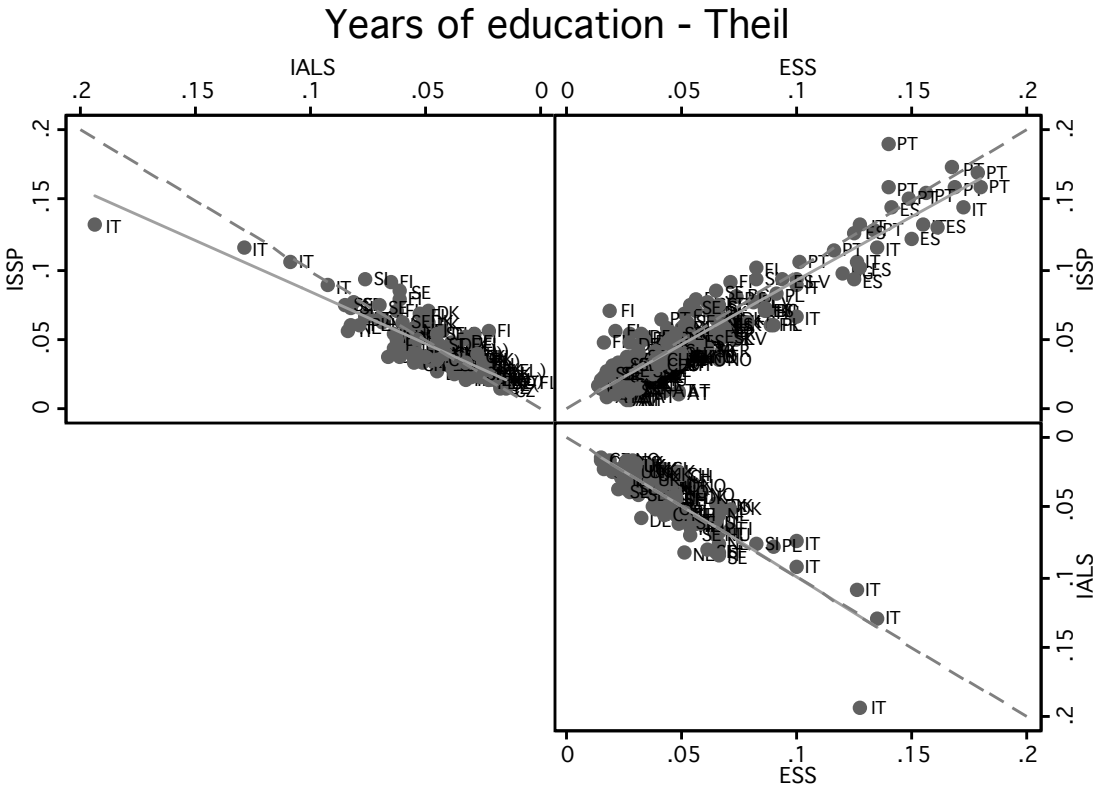


Figure 6

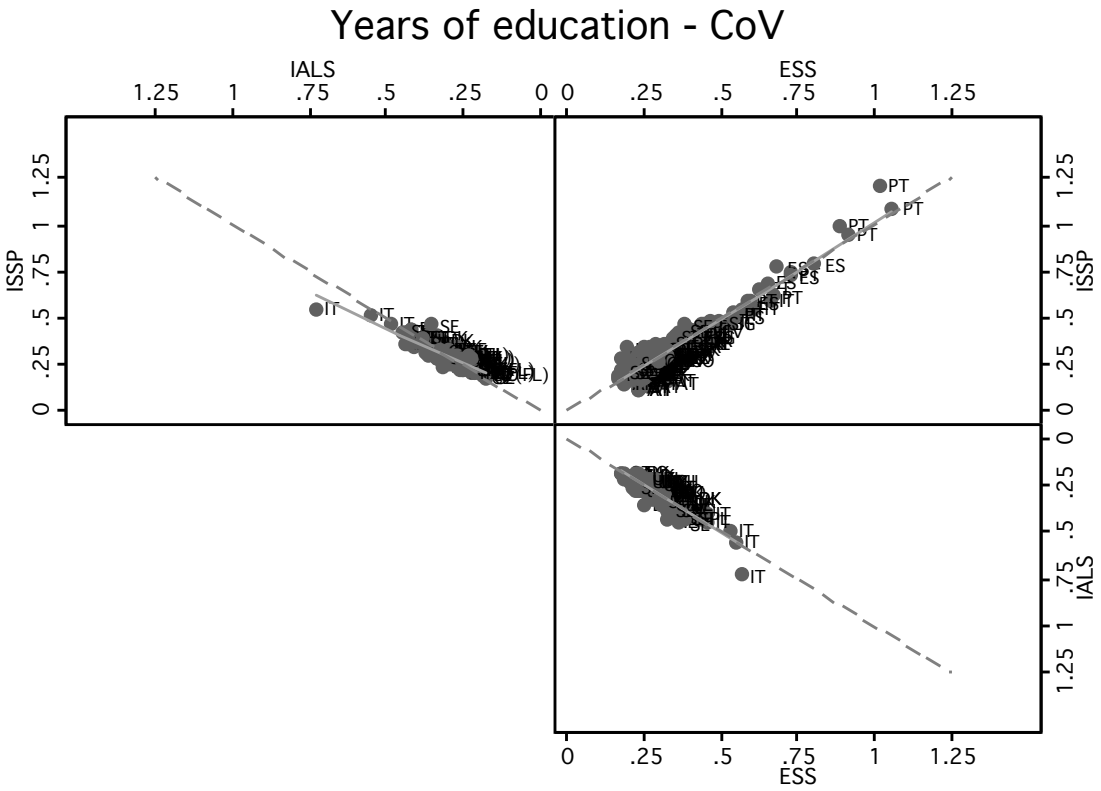


Figure 7

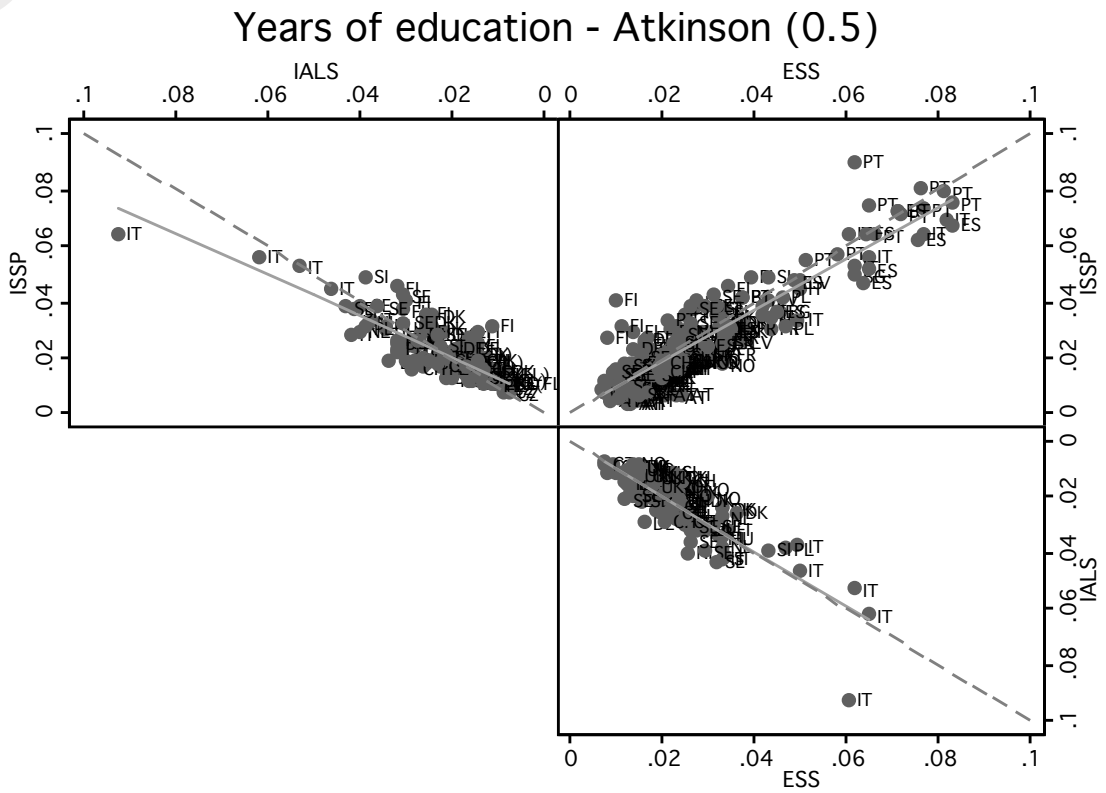


Figure 8

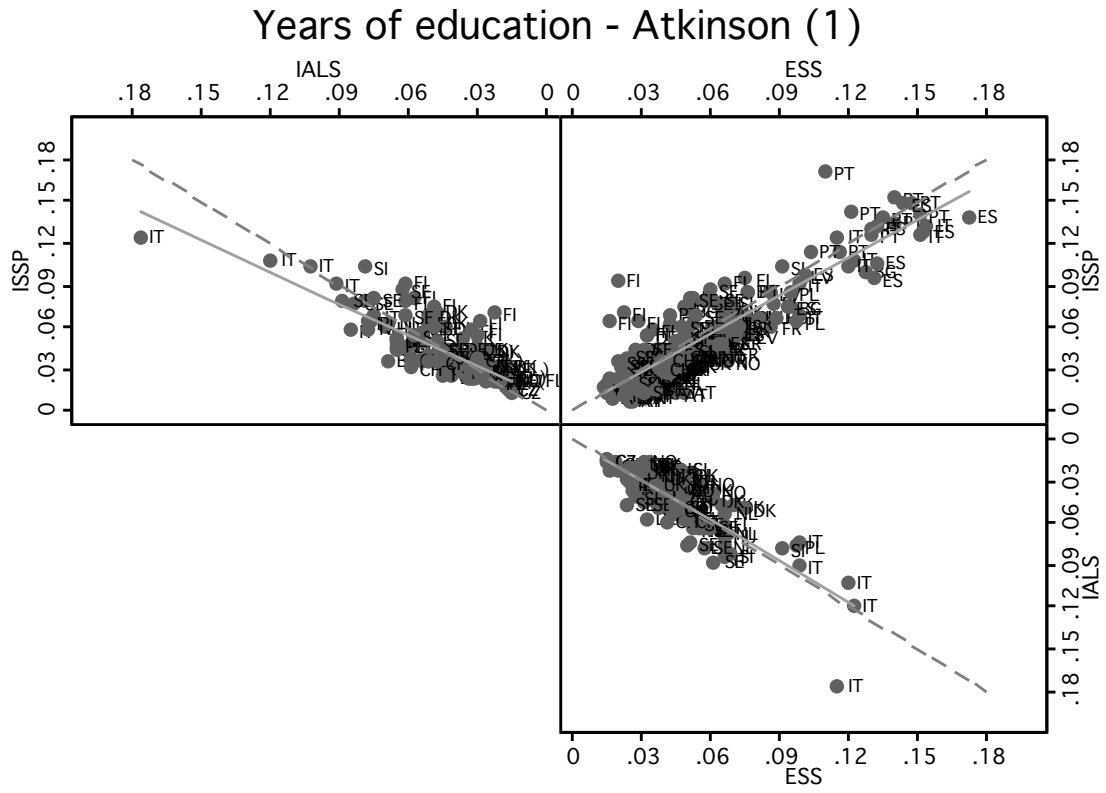
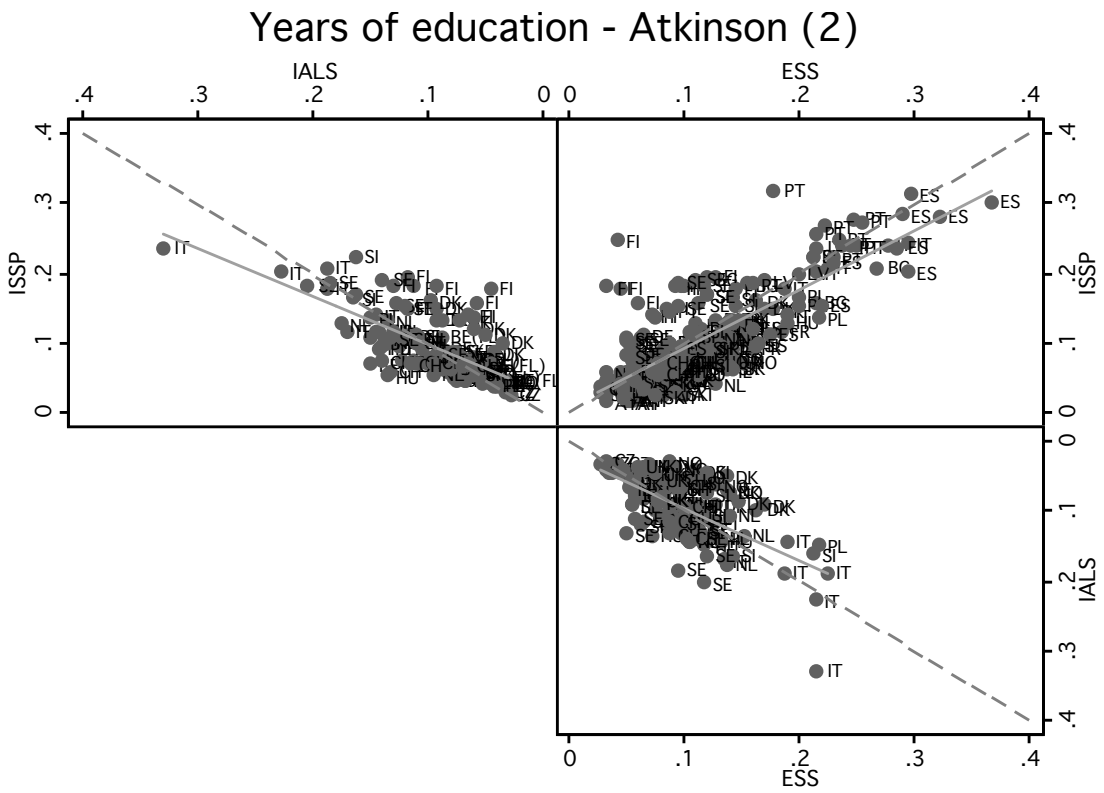


Figure 9



B Consistency graphics — Percentiles

Figure 10

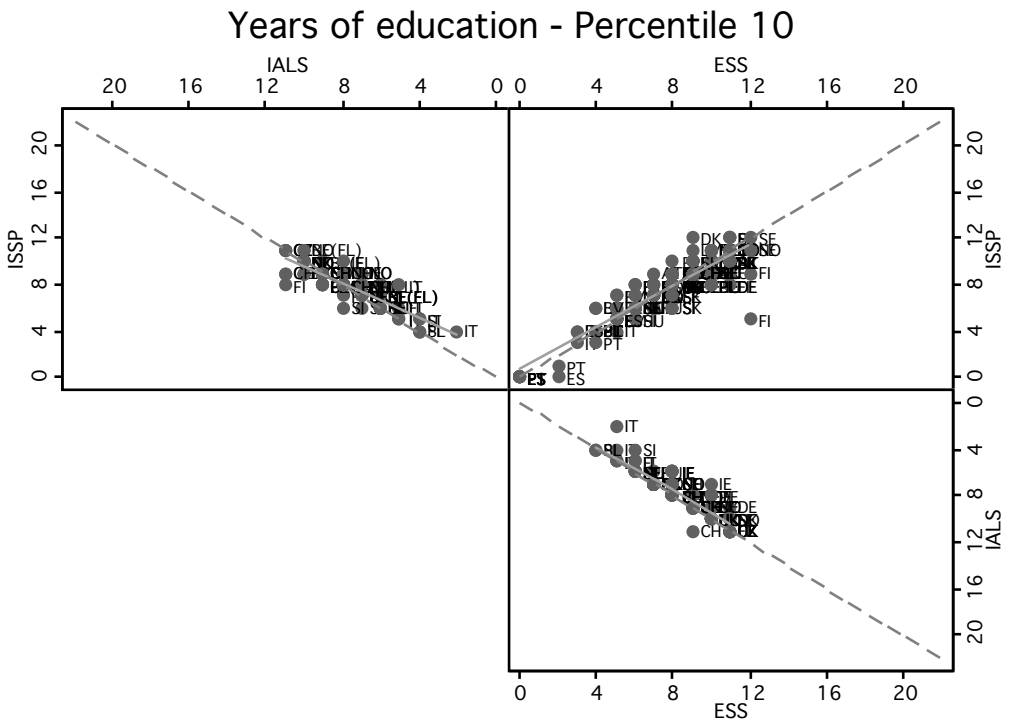


Figure 11

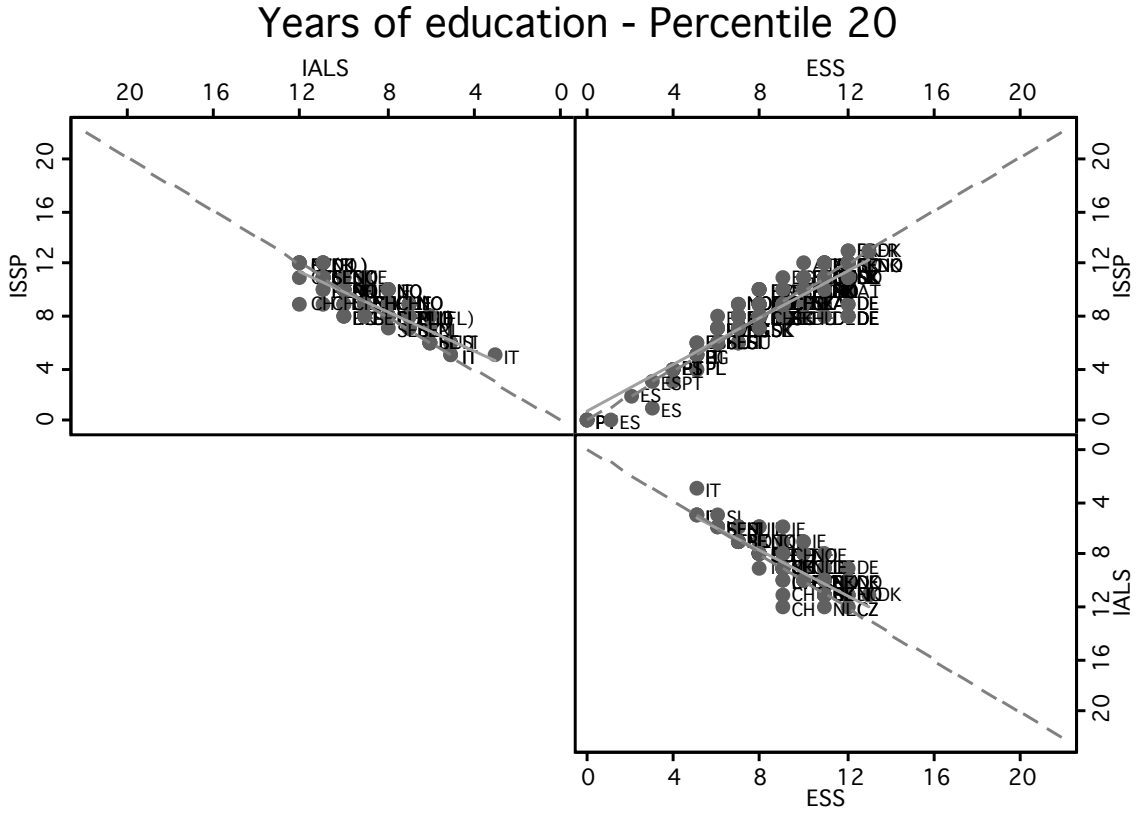


Figure 12

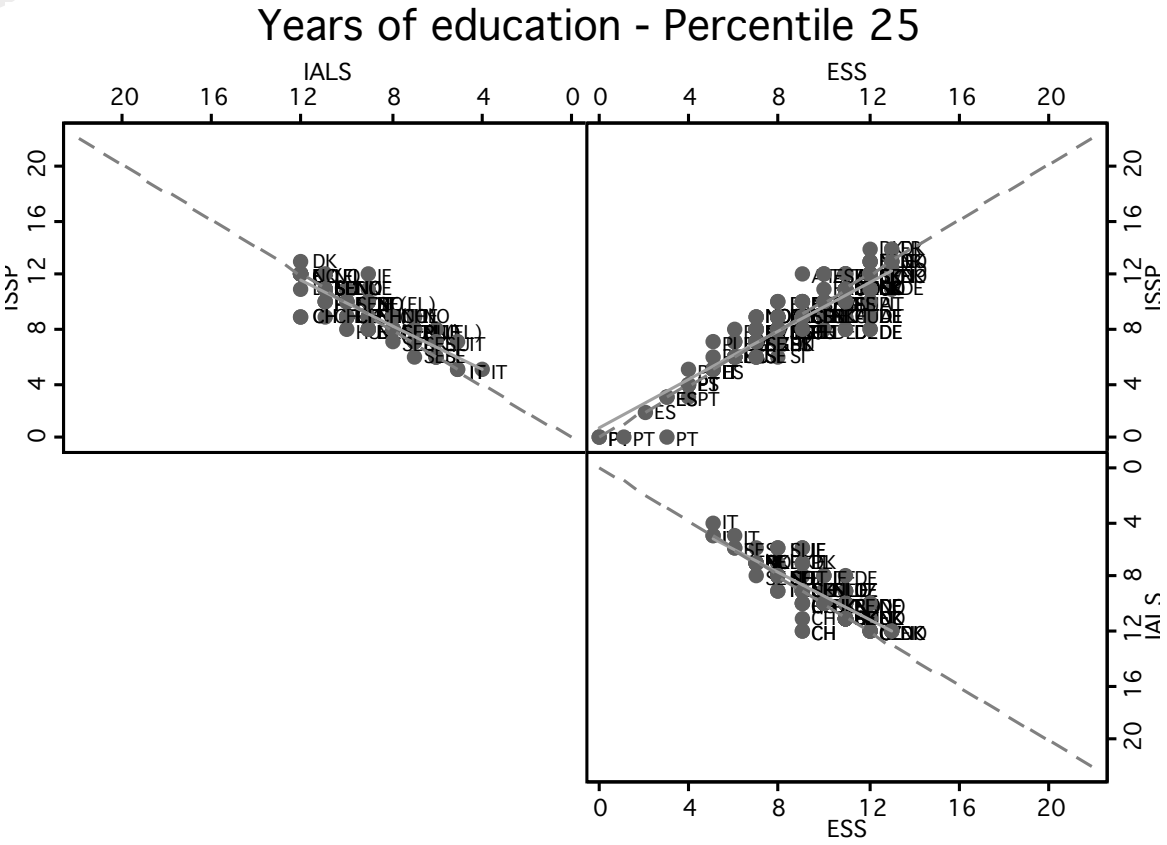


Figure 13

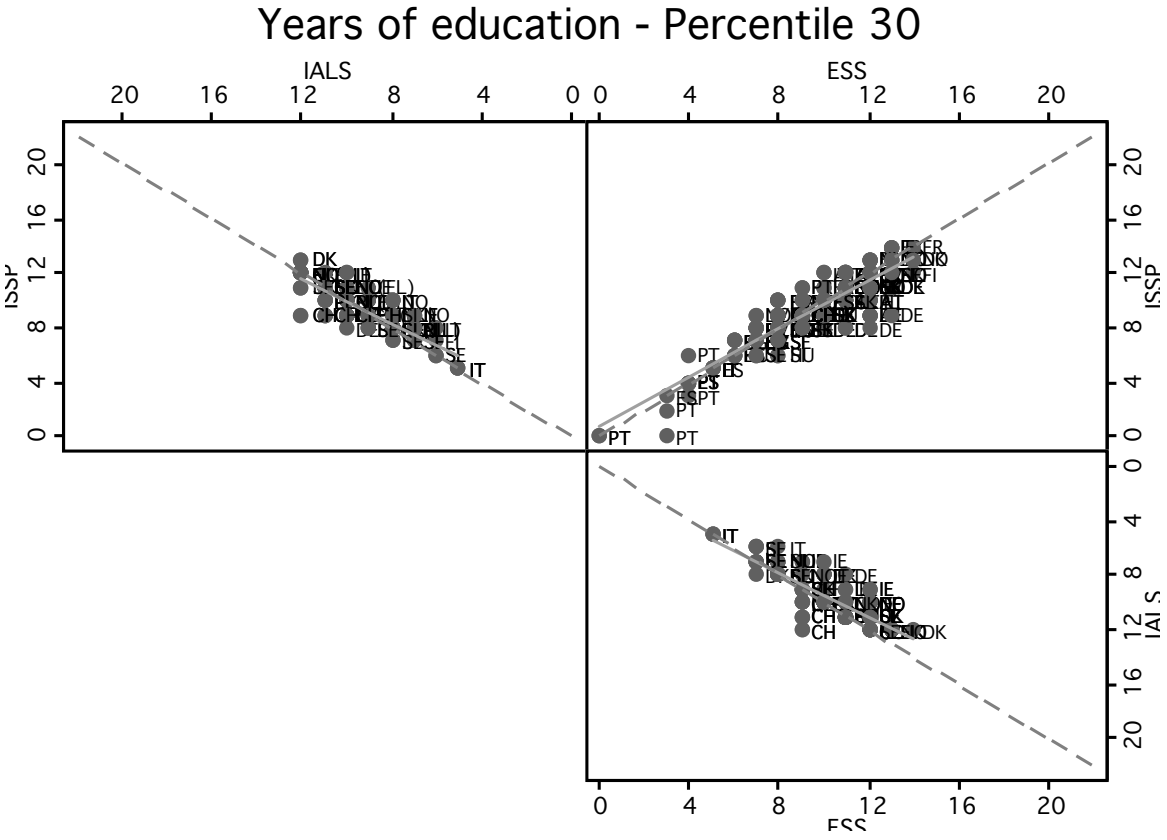


Figure 14

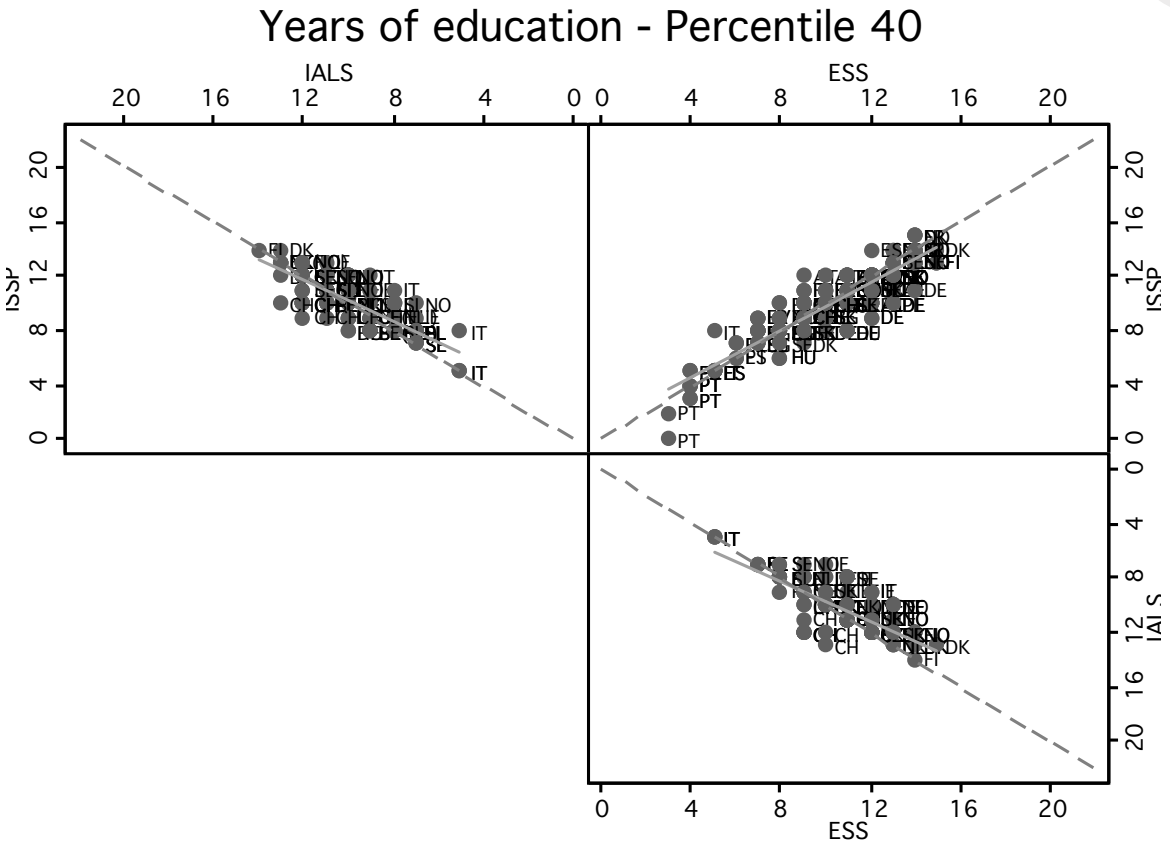


Figure 15

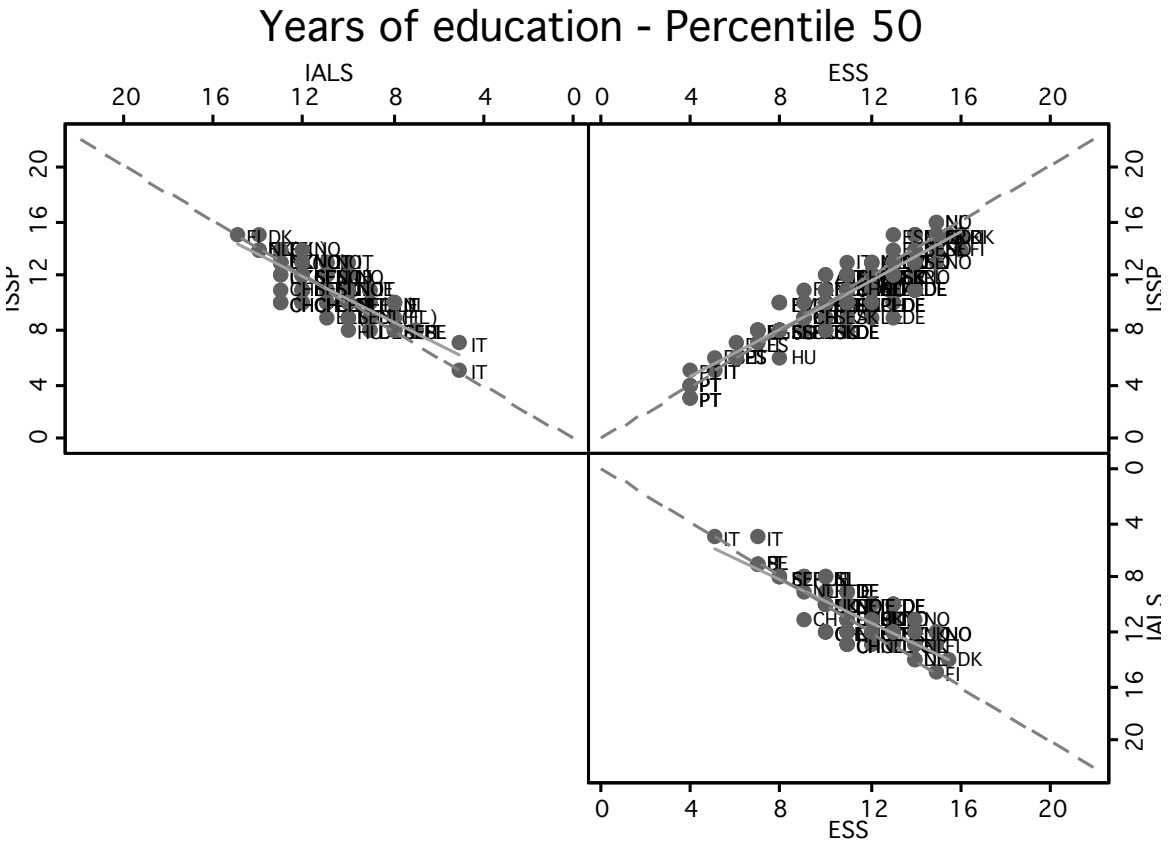


Figure 16

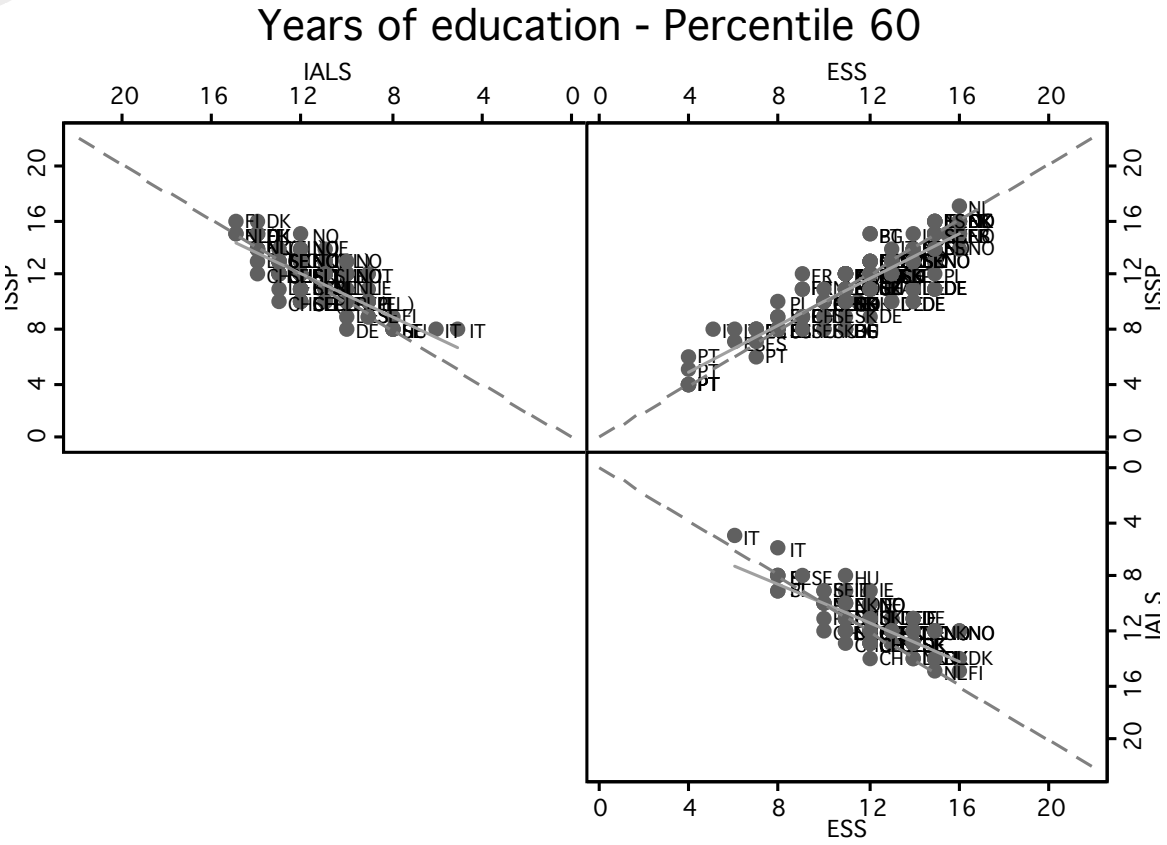


Figure 17

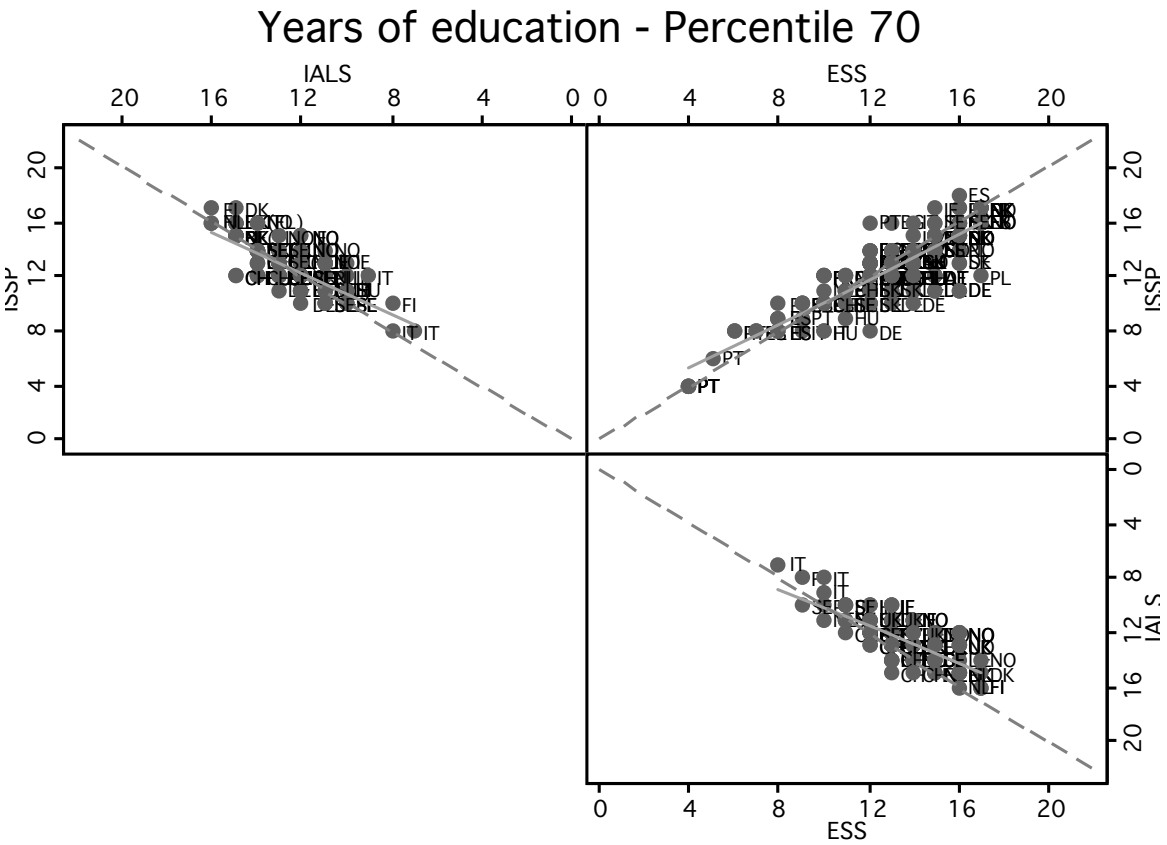


Figure 18

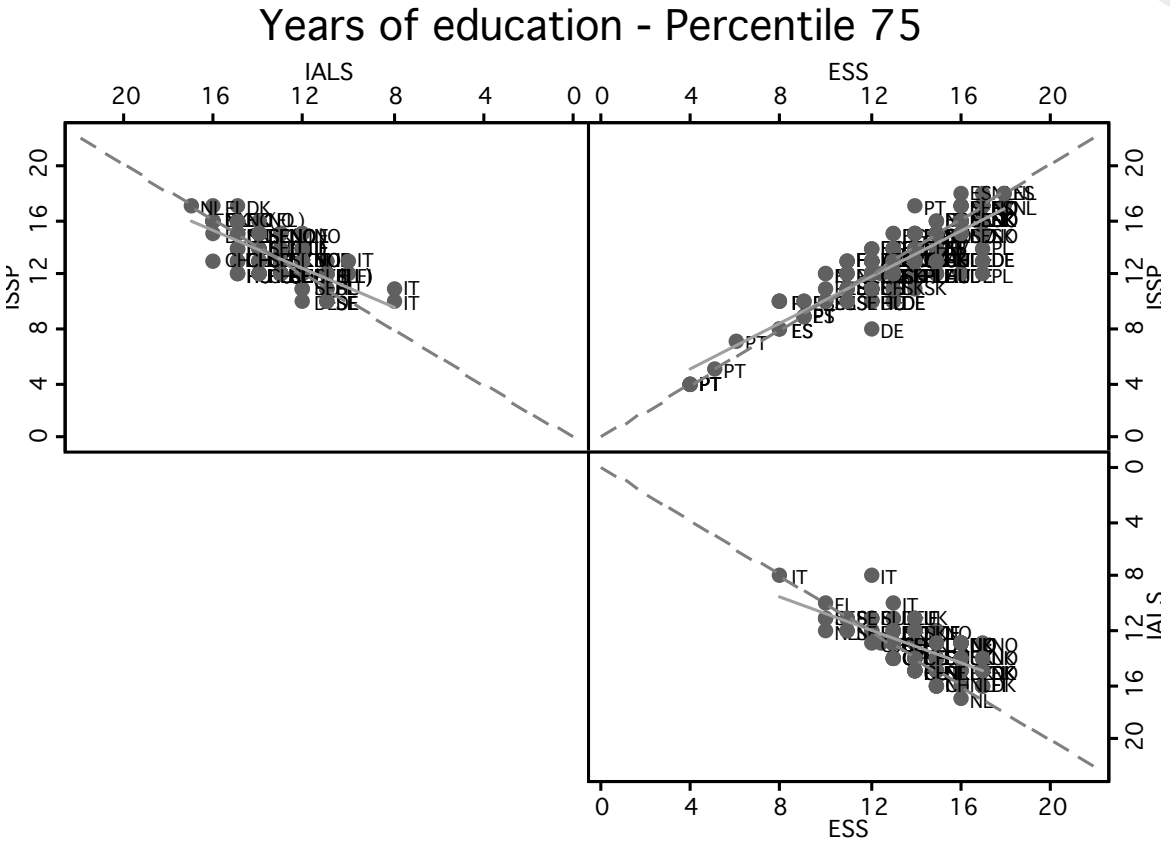


Figure 19

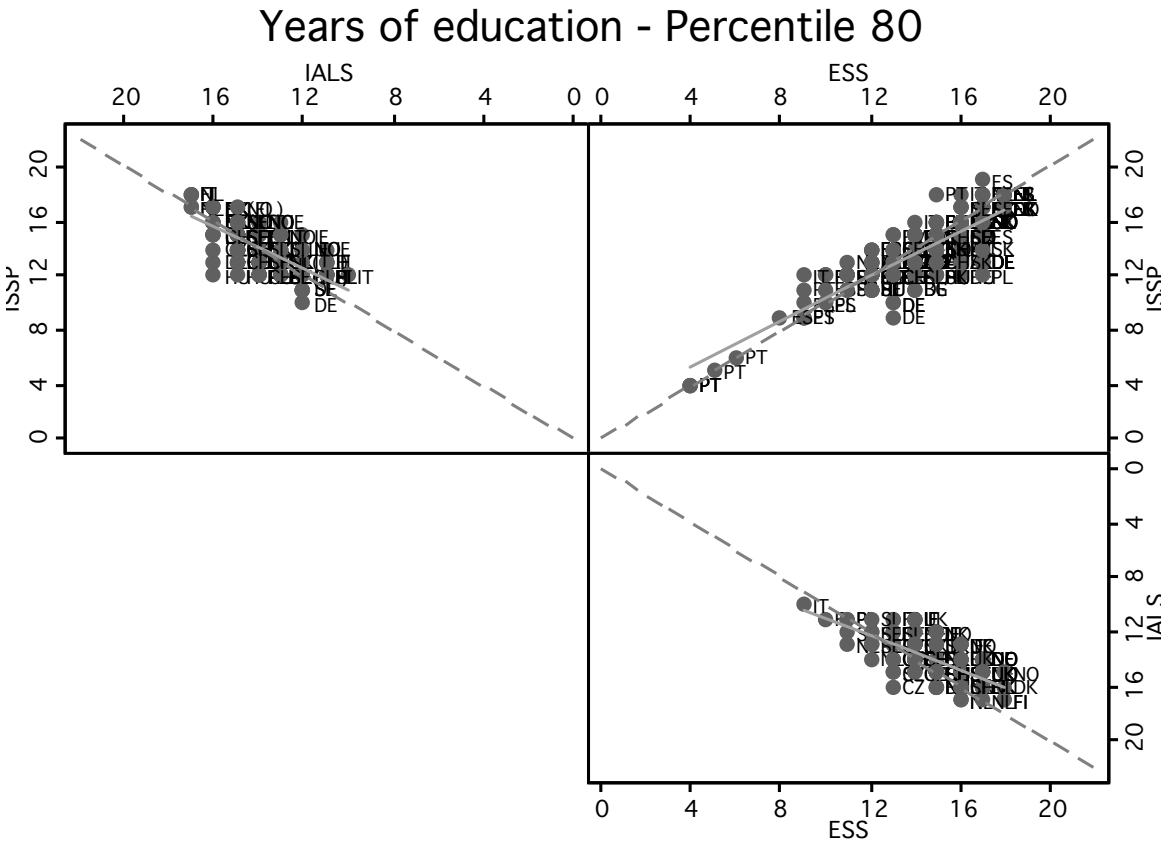
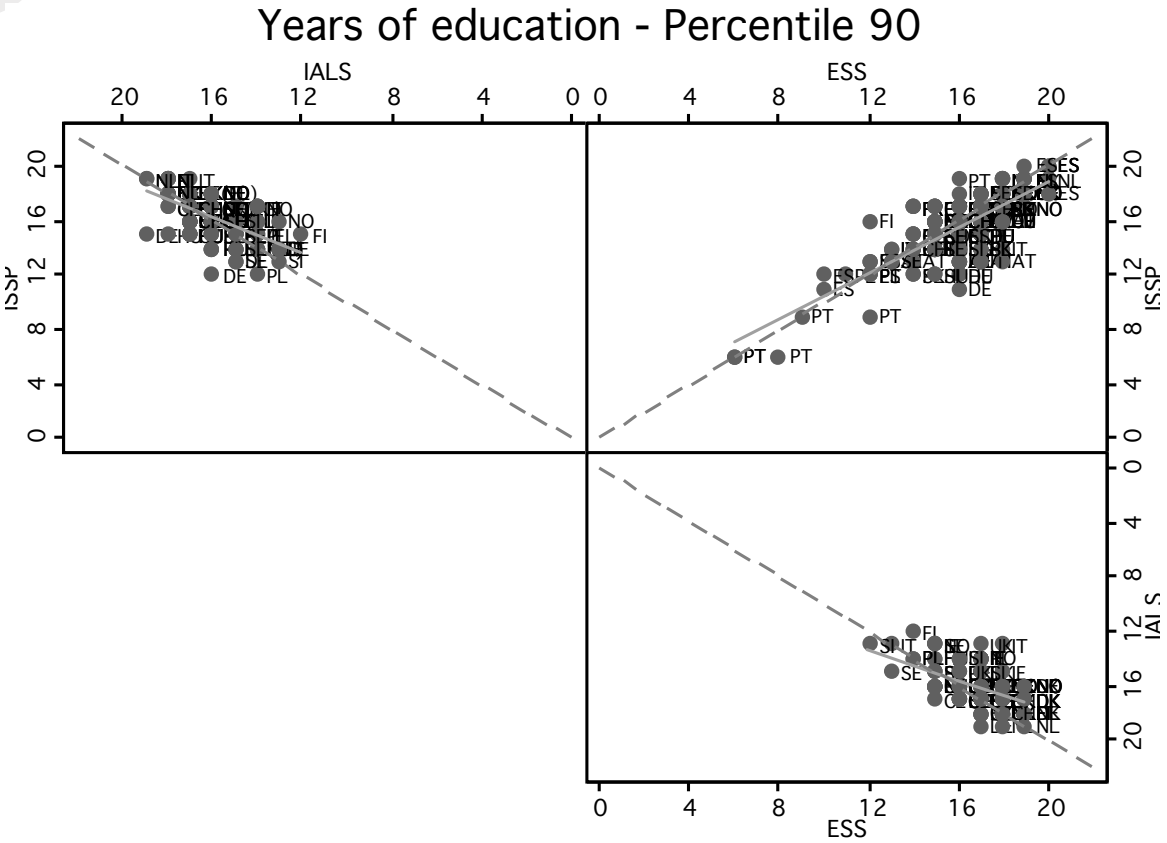


Figure 20


Years of education - Percentile 90
IALS
ESS
ISSP
ESS
ISSP
IALS

C Consistency graphics — Attainmentlevels

Figure 21

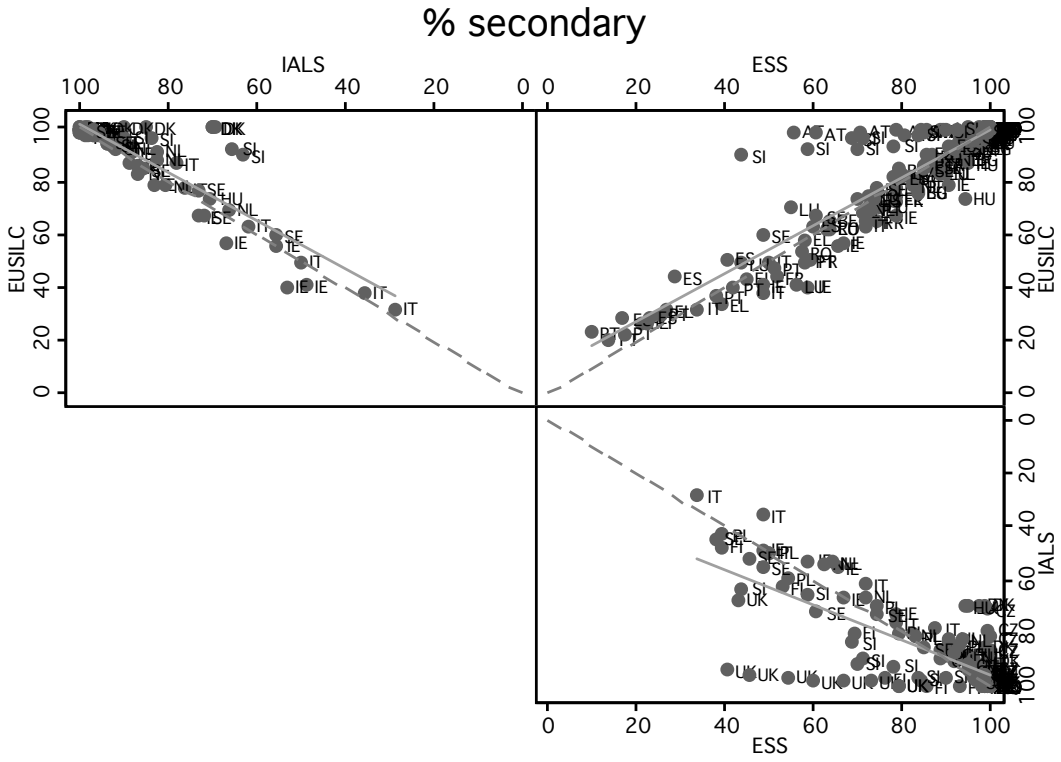


Figure 22

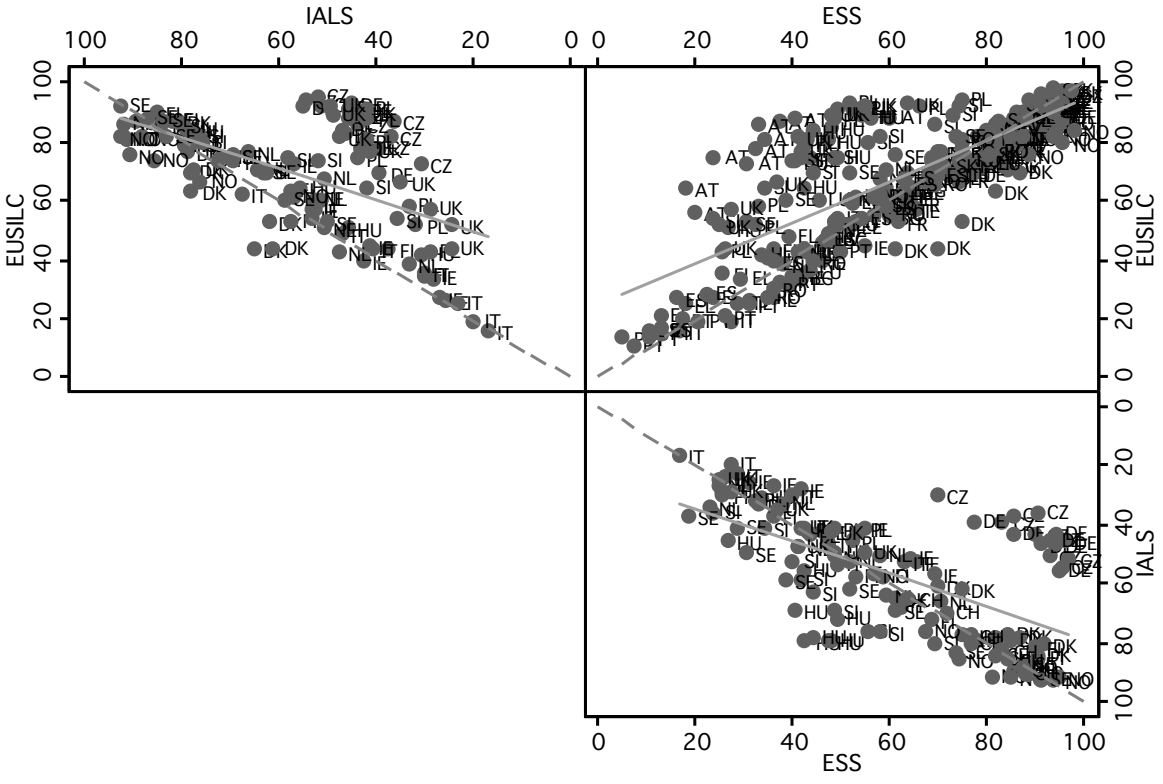
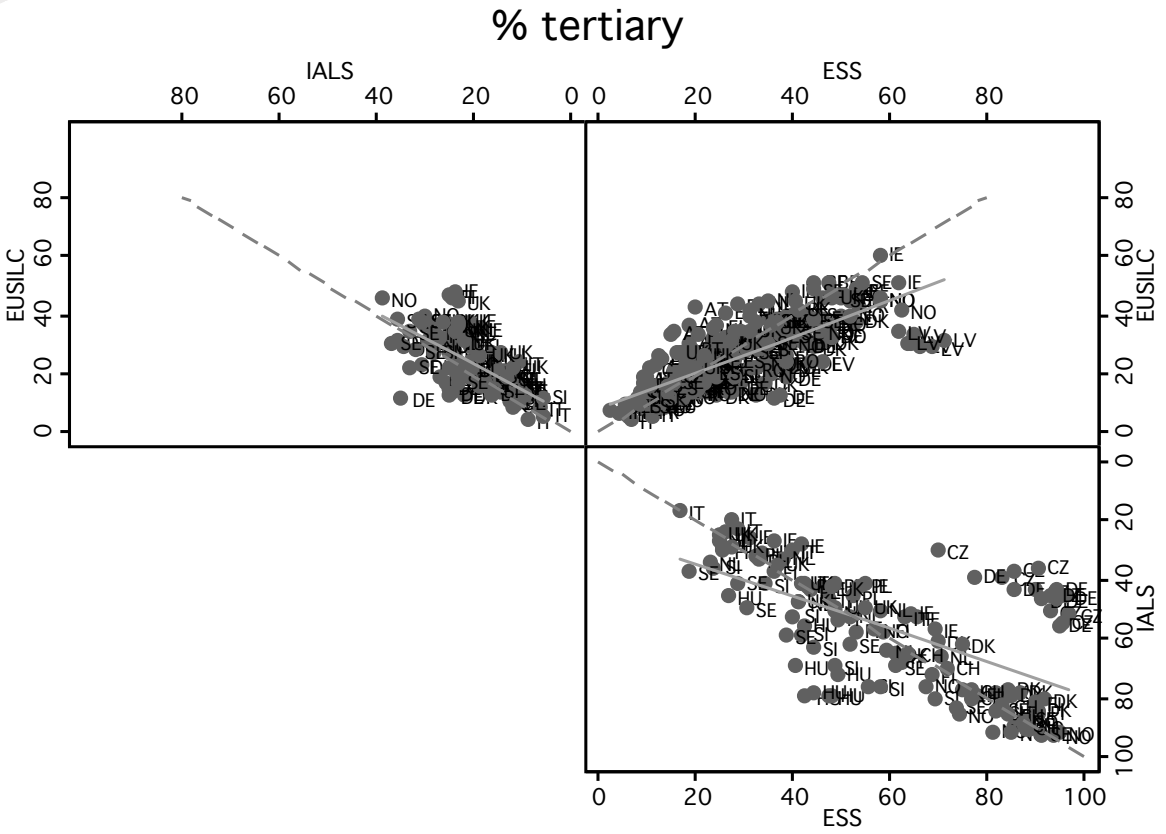


Figure 23



GINI Discussion Papers

Recent publications of GINI. They can be downloaded from the website www.gini-research.org under the subject Papers.

DP 5 **Forthcoming: Household Joblessness and Its Impact on Poverty and Deprivation in Europe**
Marloes de Graaf-Zijl
January 2011

DP 4 **Forthcoming: Inequality Decompositions – A Reconciliation**
Frank A. Cowell and Carlo V. Fiorio
December 2010

DP 3 **A New Dataset of Educational Inequality**
Elena Meschi and Francesco Scervini
December 2010

DP 2 **Coverage and adequacy of Minimum Income schemes in the European Union**
Francesco Figari, Tina Haux, Manos Matsaganis and Holly Sutherland
November 2010

DP 1 **Distributional Consequences of Labor Demand Adjustments to a Downturn. A Model-based Approach with Application to Germany 2008-09**
Olivier Bargain, Herwig Immervoll, Andreas Peichl and Sebastian Siegloch
September 2010

Information on the GINI project

Aims

The core objective of GINI is to deliver important new answers to questions of great interest to European societies: What are the social, cultural and political impacts that increasing inequalities in income, wealth and education may have? For the answers, GINI combines an interdisciplinary analysis that draws on economics, sociology, political science and health studies, with improved methodologies, uniform measurement, wide country coverage, a clear policy dimension and broad dissemination.

Methodologically, GINI aims to:

- exploit differences between and within 29 countries in inequality levels and trends for understanding the impacts and teasing out implications for policy and institutions,
- elaborate on the effects of both individual distributional positions and aggregate inequalities, and
- allow for feedback from impacts to inequality in a two-way causality approach.

The project operates in a framework of policy-oriented debate and international comparisons across all EU countries (except Cyprus and Malta), the USA, Japan, Canada and Australia.

Inequality Impacts and Analysis

Social impacts of inequality include educational access and achievement, individual employment opportunities and labour market behaviour, household joblessness, living standards and deprivation, family and household formation/breakdown, housing and intergenerational social mobility, individual health and life expectancy, and social cohesion versus polarisation. Underlying long-term trends, the economic cycle and the current financial and economic crisis will be incorporated. Politico-cultural impacts investigated are: Do increasing income/educational inequalities widen cultural and political 'distances', alienating people from politics, globalisation and European integration? Do they affect individuals' participation and general social trust? Is acceptance of inequality and policies of redistribution affected by inequality itself? What effects do political systems (coalitions/winner-takes-all) have? Finally, it focuses on costs and benefits of policies limiting income inequality and its effi ciency for mitigating other inequalities (health, housing, education and opportunity), and addresses the question what contributions policy making itself may have made to the growth of inequalities.

Support and Activities

The project receives EU research support to the amount of Euro 2.7 million. The work will result in four main reports and a final report, some 70 discussion papers and 29 country reports. The start of the project is 1 February 2010 for a three-year period. Detailed information can be found on the website.

www.gini-research.org

GINI GROWING INEQUALITIES' IMPACTS

GINI GROWING INEQUALITIES' IMPACTS

Amsterdam Institute for Advanced labour Studies

University of Amsterdam

Plantage Muidergracht 12 ●1018 TV Amsterdam ●The Netherlands

Tel +31 20 525 4199 ●Fax +31 20 525 4301

gini@uva.nl ●www.gini-research.org

SEVENTH FRAMEWORK PROGRAMME

Project funded under the Socio-Economic sciences and Humanities theme.