Society for the Study of Economic Inequality

Working Paper Series

Robust cross-country analysis of inequality of opportunity

Francesco Andreoli
Alessio Fusco

ECINEQ WP 2019 - 492

ECINEQ 2019 - 492
March 2019

www.ecineq.org

Robust cross-country analysis of inequality of opportunity*

Francesco Andreoli†
University of Verona, Italy and LISER, Luxembourg

Alessio Fusco
Luxembourg Institute of Socio-Economic Research - LISER

Abstract

International rankings of countries based on inequality of opportunity indices may not be robust vis-à-vis the specific metric adopted to measure opportunities. Indices often aggregate relevant information and neglect to control for normatively irrelevant distributional factors. This paper shows that gap curves can be estimated from crosssectional data and adopted to test hypotheses about robust cross-country comparisons of (in)equality of opportunity.

Keywords: Inequality of opportunity, EU-SILC, gap curves.

JEL Classification: D63, J62, C14.

*This note benefited from comments by Tony Atkinson, Luna Bellani, Tim Goedemé, Sigita Grundiza, Eric Marlier, Vincenzo Prete, Philippe Van Kerm, Roy van der Weide and several conferences participants. Financial support from Eurostat (Net-SILC2), the French Agence Nationale de la Recherche (Ordineq grant ANR-16-CE41-0005-01) and the Luxembourg Fonds National de la Recherche (IMCHILD grant INTER/NORFACE/16/11333934, PREFERME CORE grant C17/SC/11715898 and IMeRSe CORE grant C15/SC/10266168) is gratefully acknowledged. Fusco also thanks the World Bank Development Research Group for its hospitality. Replication code is made available on the authors' web-pages. The usual disclaimer applies.

†**Contact details**: F. Andreoli (corresponding author), Department of Economics, University of Verona, Via Cantarane 24, 37129 Verona, IT and LISER; Email: francesco.andreoli@liser.lu. A. Fusco, LISER, 11 Porte des Sciences, L-4366, Esch-sur-Alzette, LU; Email: alessio.fusco@liser.lu.

1. Introduction

Equality of opportunity (EOp) theory distinguishes between illegitimate sources of inequality that deserve compensation (e.g. parental background) and legitimate ones (e.g. effort) – see Roemer and Trannoy (2016). Opportunities are unequally distributed if some individuals enjoy an illegitimate advantage with respect to others, in relation to circumstances beyond their control. Inequality of Opportunity (IOp) indices operationalize this notion (Ramos and Van de gaer, 2016).

Empirical studies focusing on earnings inequality have found that IOp is only a small fraction of total inequality. This might be due to observability constraints (Niehues and Peichl 2014), as well as to heterogeneity in earnings opportunities. Firstly, IOp indices *aggregate* heterogeneity in the distribution of illegitimate advantage across circumstances groups, thus discarding potentially relevant information. Secondly, IOp indices *neglect* the role of covariates which are not illegitimate drivers of inequality (such as age and marital status), but that are correlated with circumstances (older cohorts have on average less educated parents) and explain earnings heterogeneity (older cohorts/married individuals display higher earnings, see Balcázar 2015). Parametric models have been developed (Ferreira and Gignoux, 2011) to account for the role of covariates, at the cost of introducing specification bias.

International rankings based on IOp indices may hence not be robust vis-à-vis the selected IOp metric. We address this issue by introducing *gap curves* in the cross-country analysis of IOp. A gap curve depicts the gap between opportunity profiles attributed to different circumstances in a given country. When there is no gap in opportunity profiles, there is strong evidence of EOp. Otherwise, IOp prevails. Our first contribution is to show that gap curves (i) can be used to tests hypothesis about EOp in each country and (ii) can be contrasted across countries to test for differences in IOp. The normative underpinnings of the ordering induced by non-intersecting gap curves have been detailed in Andreoli *et al* (2019). Our second contribution shows that unconditional gap curves for each country can be flexibly estimated using distribution regression methods, controlling for the effect of irrelevant covariates on opportunity profiles. We tests EOp and IOp for earnings in Europe using the European Union-Survey on Income and Living Conditions (EU-SILC).

2. Concepts

Let earnings $y_s(c,\varepsilon)$ in country s depend on circumstances $c \in \{c_1, \dots, c_N\}$, defining *types*, and on distributional factors ε, characterizing within-type earning heterogeneity. Some factors in ε are rewarded by the market, such as effort or talent, while others are irrelevant from a normative perspective, such as age and marital status. The object of interest is the *opportunity set*, depicting the distribution of potential earnings accruing to type-c individuals. The set coincides with the conditional cdf $F_s(y|c)$ for a group of homogenous individuals. In applied analysis, $F_s(y|c)$ is often (non-parametrically) estimated from data about earnings and circumstances, neglecting the contribution of potential earnings heterogeneity driven by normatively irrelevant covariates.

The gap curve $\Gamma_s(c, c', p)$ depicts the empirical distribution of the unfair gap between opportunity sets of types c and c', and is defined:

$$\Gamma_s(c, c', p) := F_s^{-1}(p|c) - F_s^{-1}(p|c') \ \forall c \neq c' \text{ and } p \in [0,1],$$

where p is a conditional quantile of the *inverse* distributions (Figure 1 displays an example). EOp imposes linear restrictions on the gap curve, leading to testable hypothesis:

$$H_0^{EOp} : \Gamma_s(c, c', p) = 0, \forall c \neq c', \forall p \in [0,1].$$

EOp holds in country s (H_0^{EOp} not rejected) whenever opportunity sets coincide across all pairs $c \neq c'$. Otherwise, a form of IOp prevails. Indices can be used to rank countries by IOp. Many IOp indices are related to gap curves (see Andreoli and Fusco, 2017), including the Gini-opportunity index by Lefranc *et al*. (2008):

$$GO(s) := I\left(\mu_{c_1} \cdot (1 - G_{c_1}), \dots, \mu_{c_N} \cdot (1 - G_{c_N})\right) = \frac{1}{2\mu_s} \sum_i \sum_j w_{c_i} \cdot w_{c_j} \cdot \left| \int_0^1 (1-p) \cdot \Gamma_s(c_i, c_j, p) dp \right|,$$

where w_c is circumstance c weight and μ_s the average earnings in the country. The GO incorporates efficiency (μ_c) and equity (Gini index G_c) concerns about the distribution of opportunities $F_s(y|c)$.

Rankings of countries based on IOp indices, including GO, are not robust to the selected IOp metric. Gap curves allow to compare countries on the basis of the whole distribution of opportunity gaps between any pair $c \neq c'$. Our baseline null hypothesis is that gap curves in countries s and s' coincide:

$$H_0^{IOp}: \Delta\Gamma(c, c', p) = \Gamma_s(c, c', p) - \Gamma_{s'}(c, c', p) = 0, \forall c \neq c' \; \forall p \in [0,1].$$

Two countries are indistinguishable from a IOp perspective whenever H_0^{IOp} cannot be rejected. Rejection implies that fairness gaps between types differ across countries, albeit in an unrestricted way. Gap curves for types $c \neq c'$ may cross, in which case the distribution of unfair advantage depends on relevant distributional factors and countries cannot be robustly ordered. Alternatively, the gap curve for types $c \neq c'$ in country s dominates that of country s', i.e., $\Gamma_s(c, c', .)$ is never below and it is sometimes above $\Gamma_{s'}(c, c', .)$. If there is strong dominance across all pairs $c \neq c'$ for which $\Delta\Gamma(c, c', .) \neq 0$, then there is robust evidence that opportunities are more unequally distributed in country s compared to s' (Andreoli et al., 2019).

3. Estimation of unconditional gap curves

We estimate gap curves at a finite number of intercepts $p \in \{p_1, \dots, p_M\}$ using Recentered Influence Function (RIF) approximations of the quantile function (Firpo et al, 2009). The influence function of a quantile, $IF\left(F^{-1}(p)\right) = \frac{1-p}{f\left(F^{-1}(p)\right)}$, measures the effect (by linearizing the inverse function) of a contamination in the data on that specific quantile p. The RIF estimator yields unbiased estimates of the *unconditional* quantiles of the distribution. We provide RIF estimators for unconditional gap curves. For a given country and circumstance type, we first estimate linear probability regressions of an indicator $1(y_\iota \geq y)$, taking value 1 when observed income y_ι is larger than a predetermined threshold y, on parental circumstances and covariates X_ι (such as age and marital status):

$$\forall s, \forall i = 1, \dots, C: \; 1(y_\iota \geq y) = \alpha_i^s(y) + \sum_{j \neq i} \beta_{ij}^s(y) \cdot 1(\iota \; is \; of \; type \; c_j) + \boldsymbol{X_\iota} \cdot \boldsymbol{\gamma}_i^s(y) + u_\iota(y).$$

Income thresholds y in model i coincide with the observed quantiles p of the conditional distributions $F_s(y|c_i)$ for each type separately.

The coefficients α, β and γ can be estimated using cross-sectional data. For country s and quantile p, we estimate two effects: $\beta_{ij}^{s}(y)$ and $\beta_{ji}^{s}(y)$. The first effect is the difference in probability of achieving larger earnings than $y = F_s^{-1}(p|c_i)$ for type i as opposed to type j. It measures the probability gap $F_s(y|c_j) - F_s(y|c_i)$ at earnings $y = F_s^{-1}(p|c_i)$. Similarly, the second effect measures $F_s(y|c_i) - F_s(y|c_j)$ at earnings $y = F_s^{-1}(p|c_j)$. We apply the IF formula above to obtain estimates of the gap curve coordinates, expressed in the space of earnings. Since the quantiles $F_s^{-1}(p|c_i)$ and $F_s^{-1}(p|c_j)$ do not generally coincide, we use the average effect as a reliable estimate of the unconditional gap curve. This gives:

$$\hat{\Gamma}_s(c_i, c_j, p) = \frac{1}{2}\left[\frac{\beta_{ji}^{s}\left(F_s^{-1}(p|c_j)\right)}{f_s\left(F_s^{-1}(p|c_j)|c_j\right)} - \frac{\beta_{ij}^{s}\left(F_s^{-1}(p|c_i)\right)}{f_s\left(F_s^{-1}(p|c_i)|c_i\right)}\right] \quad \forall \mathrm{c} \neq \mathrm{c}' \text{ and } p \in [0,1],$$

where $f_s(y|c_i)$ is the density of the conditional distribution (non-parametrically identified) of type c_i earnings opportunities.

Gap curves are estimated at earnings deciles ($M = 10$), their variance-covariance matrices are bootstrapped. Assuming normality, H_0^{EOp} and H_0^{IOp} can be tested against an unrestricted alternative using χ_{M-1}^2-distributed joint equality tests for vectors of quantiles estimates. We use t-tests of quantile-specific differences in gap curves to test $\Gamma_s(c, c', p) - \Gamma_{s'}(c, c', p) \geq 0, \forall \mathrm{p} \in [0,1]$ (dominance) for those pairs $c \neq c'$ for which H_0^{IOp} is rejected among countries s and s'. $GO(s)$ is estimated from empirical gap curves via numerical integration methods, thus controlling for normatively irrelevant factors.

4. Empirical illustration

We use the 2011 EU-SILC module on “intergenerational transmission of disadvantage” to test EOp and IOp for earnings acquisition across 16 European countries. Parental education (high, medium, low) defines three types. Our sample includes male full-time employed aged 30-50 (see Andreoli and Fusco 2017 for details). Estimates are always conditional on age and marital status.

Table 1: Tests for EOp and IOp, 16 EU-SILC countries, 2011.

Country	GO	Comparison country															
		NL	FI	DE	SK	NO	SE	IS	AT	BE	PL	UK	EE	LT	HU	IE	LU
NL	0.023	2	0	0	n.o.	0	0	0	0	0	0	1	0	0	0	1	2
FI	0.026	=	2	1	0	0	0	0	1	1	0	0	0	0	1	1	2
DE	0.028	=	=	3	n.o.	0	0	0	0	0	n.o.	1	0	n.o.	1	0	2
SK	0.028	=	=	=	2	0	0	1	2	2	1	2	0	1	1	2	2
NO	0.028	=	=	=	=	2	0	0	0	0	0	0	0	n.o.	0	1	2
SE	0.032	=	=	=	=	=	3	0	0	0	0	0	0	n.o.	0	1	2
IS	0.036	=	=	=	=	=	=	2	0	0	0	0	0	0	0	0	2
AT	0.042	>	=	=	=	=	=	=	3	0	n.o.	0	0	1	n.o.	1	2
BE	0.043	>	=	>	>	=	=	=	=	3	0	0	0	1	0	1	2
PL	0.045	>	>	>	>	>	=	=	=	=	3	1	0	n.o.	0	2	2
UK	0.046	>	>	>	>	>	=	=	=	=	=	3	n.o.	n.o.	n.o.	1	2
EE	0.053	>	>	>	>	>	=	=	=	=	=	=	2	0	0	2	2
LT	0.055	=	=	=	=	=	=	=	=	=	=	=	=	3	1	2	3
HU	0.062	>	>	>	>	>	>	>	>	>	>	=	=	=	3	1	2
IE	0.070	>	>	>	>	>	>	>	>	>	>	=	=	=	=	2	0
LU	0.101	>	>	>	>	>	>	>	>	>	>	>	>	>	>	>	3

Note: Earnings opportunities of three types: low, medium and high parental education.

Countries in Table 1 are arranged by increasing GO. Countries that display similar GO levels are statistically indistinguishable, while the ranking of the other countries stems from marginal differences in IOp (below diagonal, “=” indicates insignificant differences at 5% level). We use gap curves to qualify these results.

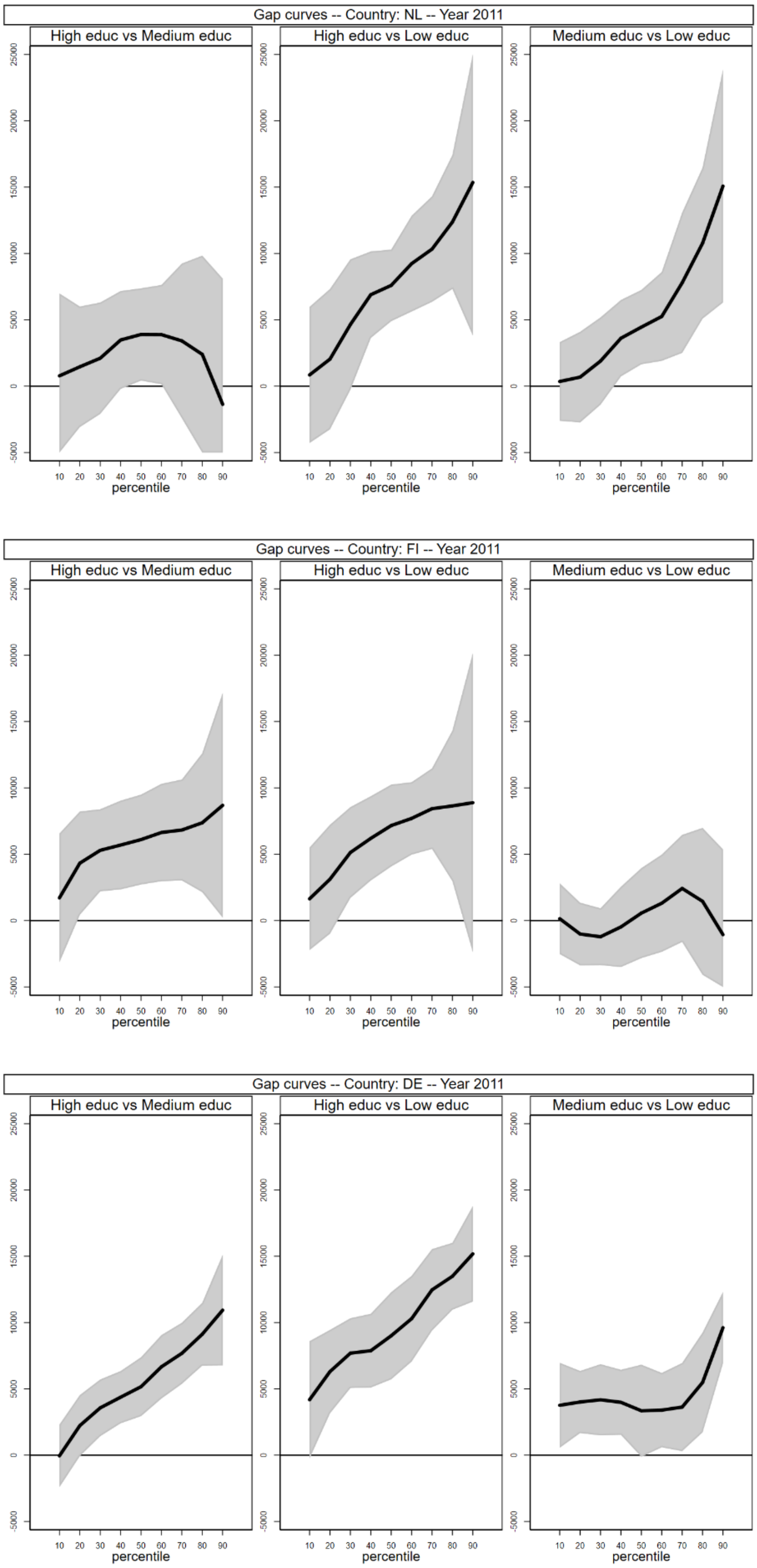
Our first result (diagonal in Table 1) is that in about half of the countries, EOp is rejected across all three comparisons: high vs medium, high vs low and medium vs low parental education. In the remaining countries (including the Nordic), H_0^{EOp} is not rejected for only one pair of types. This is strong evidence against EOp in Europe. We contrast gap curves across countries to test for robust IOp orderings.

Our second result (above diagonal in Table 1) concerns the IOp ranking: H_0^{IOp} is not rejected in 63 over 120 comparisons (in this case, we report “0”). These countries display similar levels of IOp as their gap curves coincide for all pairs $c \neq c'$. The result contrasts the ordering produced by GO (for instance, $GO(UK) > GO(FI)$ although H_0^{IOp}

is not rejected between these two countries), thus unveiling the consequences of aggregating heterogeneity. In each of the remaining cases (57), there exists at least a pair of types for which gap curves do not coincide (H_0^{IOp} rejected). If the gap curves cross, the two countries are not robustly ordered ("n.o.") in terms of IOp. Otherwise, gap curves are clearly ordered, with column-countries in Table 1 robustly display more IOp than row-countries (the table reports the cases in which dominance in gap curves holds). In a large majority of comparisons for which H_0^{IOp} is rejected (45/57), countries can be robustly ordered. In particular, Luxembourg and Ireland are the most opportunity-unequal countries in Europe, while the UK and Belgium are robustly ranked as more opportunity unequal than all the low-IOp countries.

The graphs of the gap curves depict the full heterogeneity in opportunity gaps within and across countries. Figure 1 displays gap curves for the least opportunity-unequal countries (all equal in terms of GO). In the Netherlands, IOp originates from the earnings penalty attributable to low-educated parents. Conversely, unfair advantage in Finland is clustered on children raised by high-educated parents. In both countries, unfair gaps increase with earnings opportunities, suggesting complementarities between parental background and distribution factors. Patterns of disadvantage in Germany resemble that in Finland, with an important difference: children with low-educated parents suffer a significant earnings penalty with respect to children with middle-educated parents, albeit disadvantage is unrelated to distributional factors. Gap curves dominance allows to conclude that Germany displays robustly more IOp than Finland, an evidence not captured by IOp indices. Many other cross-country comparisons display similar patterns.

Figure 1. Gap curves in selected countries (with 95% CI).



5. Conclusions

Gap curves are useful to identify and test for robust IOp rankings of countries. Using distribution regression methods, we are able to (i) estimate the full distribution of the fairness gaps implied by a gap curve while (ii) controlling for normatively irrelevant covariates, two aspects neglected by IOp indices. Our empirical illustration shows that (i) EOp in Europe is strongly rejected, (ii) in about half of cross-country comparisons, we are able to robustly rank countries by IOp, and (iii) even in least opportunity-unequal countries, gap curves reveal substantial differences in the way high or low educated parental background induces advantages or penalties in earnings, and in the way (dis)advantage correlates with effort\talents.

References

Andreoli, F., Fusco, A. (2017) The evolution of inequality of opportunity across Europe: EU-SILC evidence. Ch. 24 in A.B. Atkinson, A.C. Guio and E. Marlier (Eds.) *Monitoring Social Inclusion in Europe*, Eurostat, pp. 435-448.

Andreoli, F., Havnes, T., Lefranc, A. (2019), Equalization of opportunity: definitions, implementable conditions and application to early-childhood policy evaluation, *The Review of Economics and Statistics*, 101(2):1-15.

Balcázar, C. (2015). Lower bounds on inequality of opportunity and measurement error, *Economics Letters*, 137:102-105.

Ferreira F., Gignoux, J. (2011), "The measurement of inequality of opportunity: theory and an application to Latin America", *The Review of Income and Wealth*, 57(4): 622-657.

Firpo, S., Fortin, N., Lemieux, T. (2009), "Unconditional quantile regressions", *Econometrica*, 77(3) :953-973.

Lefranc, A., Pistolesi, N., Trannoy A. (2008) "Inequality of Opportunities vs. Inequality of Outcomes: Are Western Societies All Alike?" *Review of Income and Wealth*, 54:513–46.

Niehues J. and Peichl A. (2014) "Upper bounds of inequality of opportunity: theory and evidence for Germany and the US", *Social Choice and Welfare* 43(1):73-99.

Ramos, X., Van de gaer, D. (2016). Approaches to inequality of opportunity: Principles, measures and evidence, *Journal of Economic Surveys* 30(5):855-883.

Roemer, J. E., Trannoy, A. (2016). Equality of opportunity: Theory and measurement, *Journal of Economic Literature* 54(4):1288-1332.

Data article

Title: ***Application of the EU-SILC 2011 module "intergenerational transmission of disadvantage" to robust analysis of inequality of opportunity.***

Authors: Francesco Andreoli and Alessio Fusco

Affiliations: University of Verona and LISER; Luxembourg Institute of Socio-Economic Research

Contact email: F. Andreoli (corresponding author) Department of Economics, University of Verona, Via Cantarane 24, 37129 Verona, IT and LISER, 11 Porte des Sciences, L-4366, Esch-sur-Alzette, LU. Email: francesco.andreoli@liser.lu

Abstract

This data article describes the original data, the sample selection process and the variables used in Andreoli and Fusco (2019) to estimate gap curves for a sample of European countries. Raw data are from 2011 roaster of EU-SILC, cross-sectional sample of module "intergenerational transmission of disadvantage". This article reports descriptive statistics of the using sample. It also discusses the algorithm adopted to estimate the main effects and details the content of additional Stata files stored on the online repository. These additional files contain raw estimates from bootstrapped samples, which form the basis for estimating gap curves and their variance-covariance matrices. The data article also reports representations of gap curves for all 16 selected countries.

Specifications Table

Subject area	*Economics*
More specific subject area	*Public economics, welfare economics*
Type of data	*Tables and graphs*
How data was acquired	*Access to EU-SILC 2011 wave granted within the NETSILC2 collaborative network. Data available from Eurostat upon request, see Microdata Access Workflow Tool.*
Data format	*Raw data (not uploaded on the server), anonymized sample used in the analysis (uploaded), bootstrapped estimators (uploaded) are all in Stata format.*
Experimental factors	*NA*
Experimental features	*NA*
Data source location	*NA*
Data accessibility	*Raw data are not available on the public repository. They can be accessed through Eurostat upon request, see Microdata Access Workflow Tool. An anonymized using sample is made available.*

Value of the data

- EU-SILC data represent the baseline survey introduced by the European Commission and managed by Eurostat to monitor and compare standard of living across European countries.
- Data are highly harmonized across countries, and collected by central statistical institutes. This guarantees a high degree of comparability of countries in terms of the main variables we consider to define earnings opportunities and parental circumstances.
- Data are available free of charge in selected institutions in Europe (such as LISER). Users can apply for a visiting scheme which grants resources (material and knowledge-based) to the users of these data.

Data

The raw data are taken from the *European Union - Statistics on Income and Living Conditions* (EU-SILC) 2011 module on intergenerational transmission of disadvantage, where measures of parental background for a sufficiently large number of respondents are available. This module provides repeated cross-sectional information on the socioeconomic background of origin of the individuals interviewed in EU-SILC, along with standard relevant measures of labour market outcomes. In particular, the 2011 module contains retrospective information about the parental background experienced by the respondents when aged between 12 and 16 (see Atkinson *et al*, 1983 for pros and cons of retrospective data). This unique base provides (to a large extent) comparable data allowing similar definitions for variables measuring outcome and circumstances across countries and time. An assessment of the 2011 EU-SILC module can be found here: http://ec.europa.eu/eurostat/web/income-and-living-conditions/data/ad-hoc-modules. Due to use restrictions rules, the raw data cannot be uploaded on the repository. We add workable Stata files reporting the information on data cleaning and sample selection routine.

On the repository, we report an anonymized version of the working sample we use to run our estimates. This sample is taken from EU-SILC 2011 module data (cross-section). We focus on a subset of 16 countries: Austria (AT), Belgium (BE), Germany (DE), Estonia (EE), Finland (FI), Hungary (HU), Ireland (IE), Iceland (IS), Lithuania (LT), Luxembourg (LU), the Netherlands (NL), Norway (NO), Poland (PL), Sweden (SE), Slovakia (SK) and the United Kingdom (UK). Our interest is on individual measures of income opportunities. To estimate opportunity profiles, we restrict attention to males aged between 30 and 50 who worked full time as an employee for at least 7 months in the income reference period. In addition, individuals who declared that they were living in another private household, foster home, collective household or institution were excluded. Following Raitano and Vona (2015), we use the intergenerational module weights.

The 2011 EU-SILC module contain retrospective information about parents' educational attainment, occupational status, labour market activity status, family composition as well as presence of financial difficulties during respondents' teenage years. We focus on the educational attainment of the father as the relevant circumstance. This choice, which is in line with previous literature, is driven by comparability motives and by sample size requirements at the moment of estimating the unfair disadvantage distribution. To construct circumstances, individuals are first partitioned in three types (or groups) according to their father's education. The *high education* type consists of individuals who lived in a household where the father attained the first (e.g. bachelor, master or equivalent) or second (e.g. PhD or equivalent) stage of tertiary education; the *medium education* type consists of individuals who lived in a household where the father attained upper secondary education and post-secondary, non-tertiary education. Finally, the *low education* type consists of individuals who lived in a household where the father at most completed lower secondary education. Table 1 summarizes the rule adopted to generate the circumstance variable.

Table 1: Defining circumstances

Type	**Variable in EU-SILC**: ***pt110***: highest ISCED level of education attained by the father
Low education	- father could neither read nor write in any language - *low level* (pre-primary, primary education or lower secondary education)
Medium education	- *medium level* (upper secondary education and post-secondary non tertiary education)
High education	- *high level* (first stage of tertiary education and second stage of tertiary education)

Our outcome variable of interest is the annual gross employee cash or near cash income. It is defined as the monetary component of the compensation in cash payable by an employer to an employee, and it includes the value of any social contributions and income taxes payable by an employee or by the employer on behalf of the employee to social insurance schemes or tax authorities. This variable reflects the relation between the labour income and individual circumstances before state intervention. Two caveats apply to this particular metric of opportunities. First, this variable is defined at the level of the individual, implying that labour supply decisions are assumed to be made at individual level, thus neglecting household bargaining issues. Second, wages represent yearly evaluations of performances, since we focus on individuals who spent more than six months in the income reference period as full-time workers. The observed earnings were converted in purchasing power standard (PPS) using the conversion rates provided on the CIRCABC user group. For references, see: https://circabc.europa.eu/w/browse/3c60eeec-aca4-4db7-a035-0a6d892e6069.

Additionally, we consider information about marriage status (we use an indicator for married male respondents) and age of respondents.

Our selected running sample is made of 41533 male respondents. The distribution of parental education circumstances, average earnings by parental education, average age and proportion of married individuals are reported in Table 2. Data are collected in the *example_econletters.dta* file in Stata format (optimized for Stata 13).

Table 2: Summary statistics of running sample

Country	N	Types			Earnings				Age	Married
		High	Medium	Low	All	High	Medium	Low		
AT	2887	0.10	0.43	0.48	37,320	49,367	39,829	32,604	40.4	0.69
BE	2446	0.19	0.23	0.57	38,788	54,702	37,742	33,792	40.1	0.65
DE	5345	0.30	0.58	0.11	41,444	44,228	40,642	38,108	41.4	0.75
EE	1777	0.18	0.43	0.40	12,966	17,494	13,398	10,508	40.4	0.64
FI	1949	0.21	0.22	0.56	31,245	41,842	30,229	27,627	40.4	0.61
HU	3825	0.10	0.36	0.54	11,548	19,096	12,506	9,476	39.8	0.69
IE	1122	0.14	0.22	0.65	40,408	52,155	48,067	35,358	40.2	0.74
IS	835	0.14	0.50	0.35	35,873	40,840	37,189	31,950	40.1	0.59
LT	1716	0.11	0.29	0.60	9,546	13,485	10,424	8,426	41.4	0.87
LU	2883	0.13	0.31	0.56	48,562	67,307	57,617	39,039	39.7	0.69
NL	2310	0.21	0.27	0.52	44,900	52,415	48,198	40,212	40.1	0.64
NO	1622	0.28	0.43	0.29	40,774	47,395	39,119	36,872	40.2	0.57
PL	5805	0.06	0.49	0.45	13,641	19,894	14,599	11,726	39.9	0.86
SE	1349	0.16	0.24	0.60	30,673	39,868	32,158	27,583	39.7	0.48
SK	2977	0.10	0.60	0.31	10,809	15,002	10,699	9,702	40.3	0.80
UK	2685	0.17	0.25	0.58	43,383	57,191	46,342	38,034	40.4	0.66
Total	41533	0.16	0.40	0.44	29,447	41,888	29,187	25,230	40.3	0.71

Experimental Design, Materials and Methods

We use Recentered Influence Function methods (Firpo, Fortin and Lemieux, 2009) to recover effects of circumstances on earnings quantiles, while controlling for age and marital status. We estimate standard errors and variance-covariance matrices via bootstrapped resampling procedures on baseline data, where stratification by country, year and region of residence ("psu" variable in *example_econletters.dta*) is accounted for (see Goedemé, 2013).

The estimation algorithm proceed s as follows:

1) draw a bootstrapped sample from the using sample;
2) estimate RIF regression parameters, income levels and pdf at given preselected deciles for each bootstrapped sample;

3) calculate gap curves for each country, differences in gap curves across countries for each pair of types and aggregated inequality of opportunity indices for each country and their variations across countries;
4) reiterate the bootstrap procedure 250 times;
5) compute averages and standard error of gap curves, differences in gap curves, IOp indices and store results;
6) produce graphs of gap curves and of their 95% confidence interval based on bootstrapped standard errors at specific earnings deciles identified in point 2);
7) estimate variance-covariance matrices from bootstrapped data and use them to test relevant hypothesis, then test these hypothesis and count cases (passed on pairwise comparisons of types) for which an hypothesis is accepted or rejected.
8) Report results in the form of tables.

The whole procedure requires to generate output datasets which we store in the folder "\output" available in the repository. Notably, this folder contains the following datasets, all created from the resampling procedure:

- *bs_frale.dta*: reports estimates of regression coefficients estimates for RIF regressions, by country (country), income decile (percentile) and bootstrapped replica (rep).
- *bs2_frale.dta*: reports estimates of income deciles (pdf_pcty_X) and the corresponding type-specific pdf level (pdf_pcty_X) for each type X=1,2,3 by country (country), income decile (percentile) and bootstrapped replica (rep).
- *meanGap0.dta*, reports average estimates of gap curves based on the using sample.
- *meanGap.dta*, reports average estimates of gap curves based on bootstrapped samples.
- *Chi2_data.dta*, collects data about gap curves by country (deciles estimates).
- *eop.dta*, reports values of test statistics for H_0^{EOp}, see Andreoli and Fusco (2019).
- *gapcountry.dta*, reshaped database, reports gap curves estimates by country (columns).
- *dataiop.dta*, reports the differences in gap curves of type X versus type Y across row country and column country Z, giving G_X_Y_cZ by country (country), income decile (percentile) and bootstrapped replica (rep).
- *iop.dta*, for each pair of countries (country country2), produce t-tests for differences in average gaps across types X and Y (test_G_X_Y_c) alongside the number of cases where equality in average gaps is accepted or rejected. Moreover, the file reports test statistics for equality in gap curves (Chi2G_X_Y), ascertain if H_0^{IOp} is rejected or not for each comparison (accept_X) and then reports number of cases where H_0^{IOp} is rejected or accepted.
- *GO_bs.dta*, reports estimates of GO index by country and of differences in GO index across countries. SE (bootstrapped) reported for levels and differences in GO index.

Table 1 in Andreoli and Fusco (2019) is based on these estimates. Tests for H_0^{EOp} and H_0^{IOp} against unrestricted alternatives require to impose equality constraints on vectors of parameter estimates that are jointly normally distributed (by assumption). Tests putting failure of gap curves dominance at the null against strong dominance at the alternative (a test Andreoli and Fusco 2019 use to verify gap curve dominance in those cross-countries comparisons where H_0^{IOp} is rejected) can be estimated from t-tests for differences in gap curves at specific quantiles (see Andreoli 2018 for a discussion about these tests).

Figure 1 in Andreoli and Fusco (2019) is obtained by stacking graphs of gap curves of selected countries. We report below all gap curves (and their 95% confidence intervals) estimated from the running sample. The figures are obtained from data in *gapcountry.dta* are collected in the folder \output\graphs in the repository.

Figure 1: gap curves for Austria

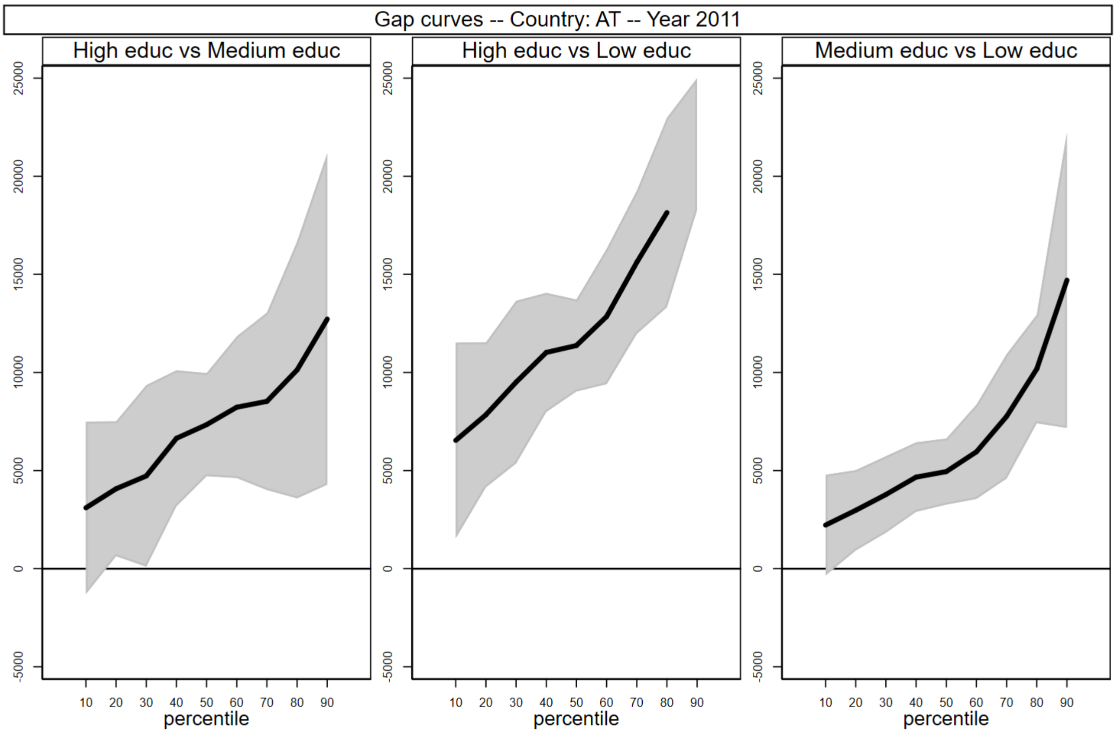


Figure 2: gap curves for Belgium

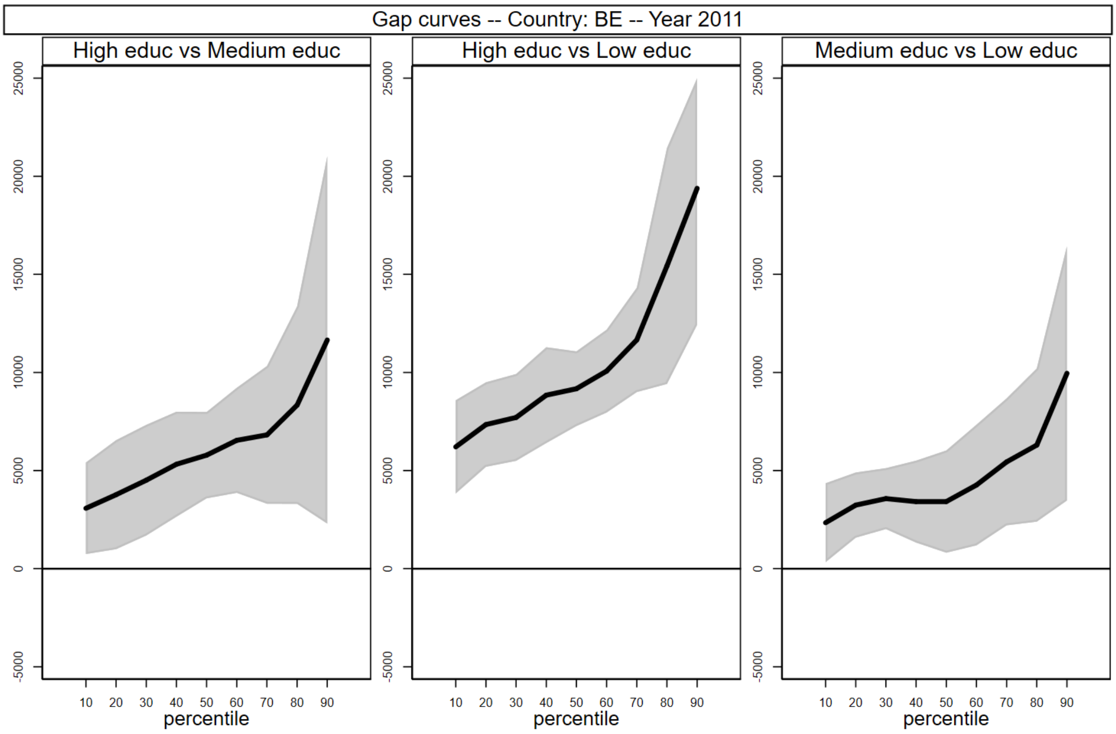


Figure 3: gap curves for Germany

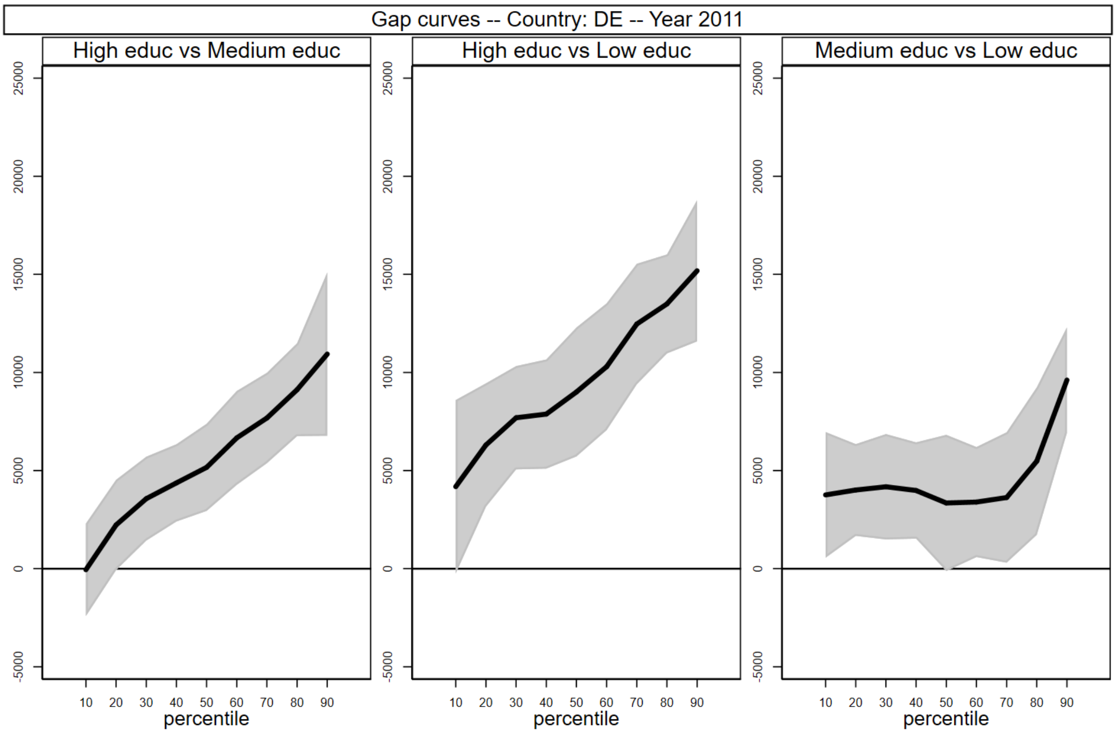


Figure 4: gap curves for Estonia

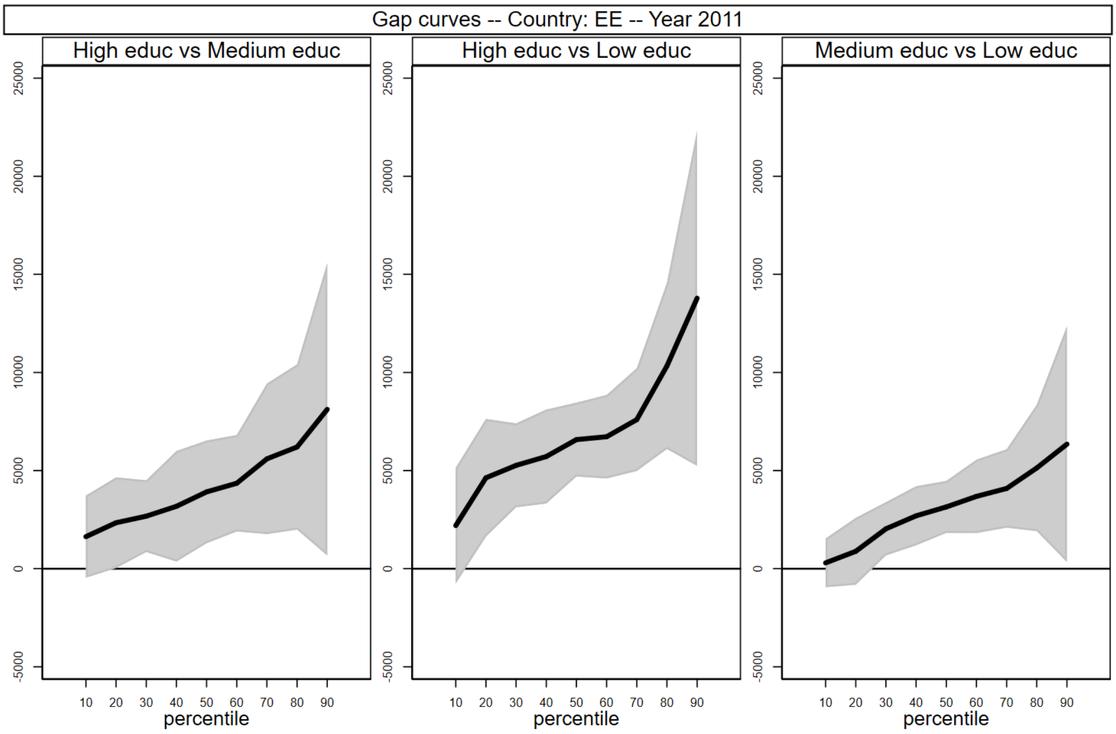


Figure 5: gap curves for Finland

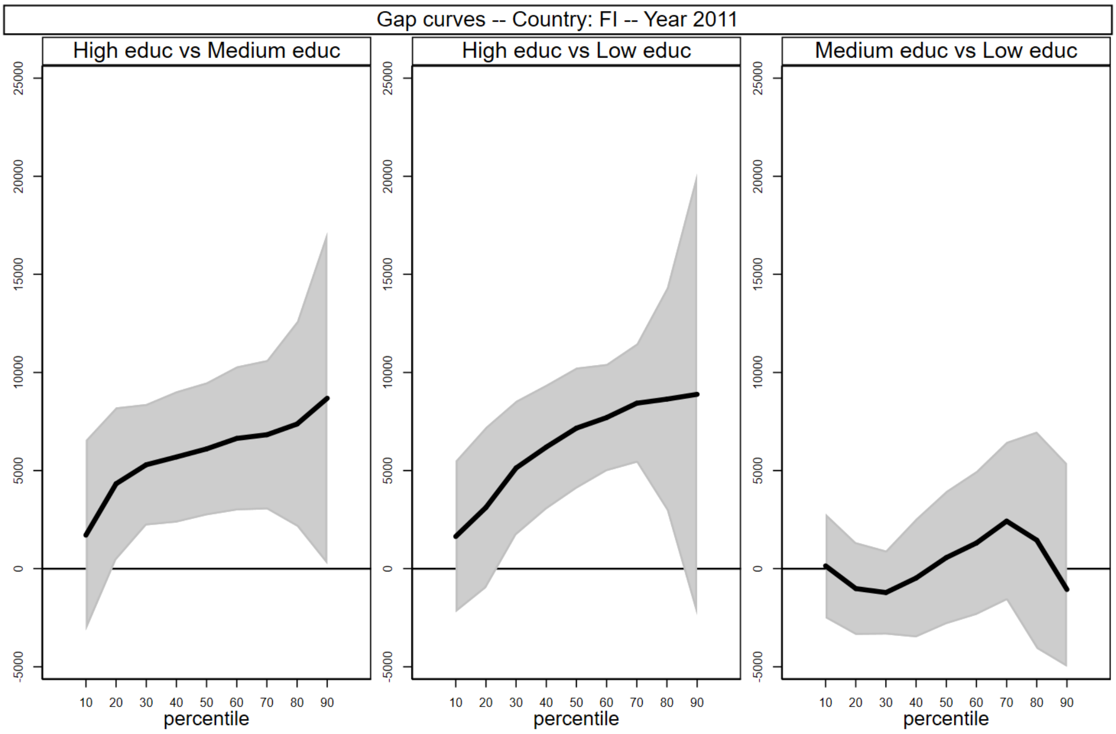


Figure 6: gap curves for Hungary

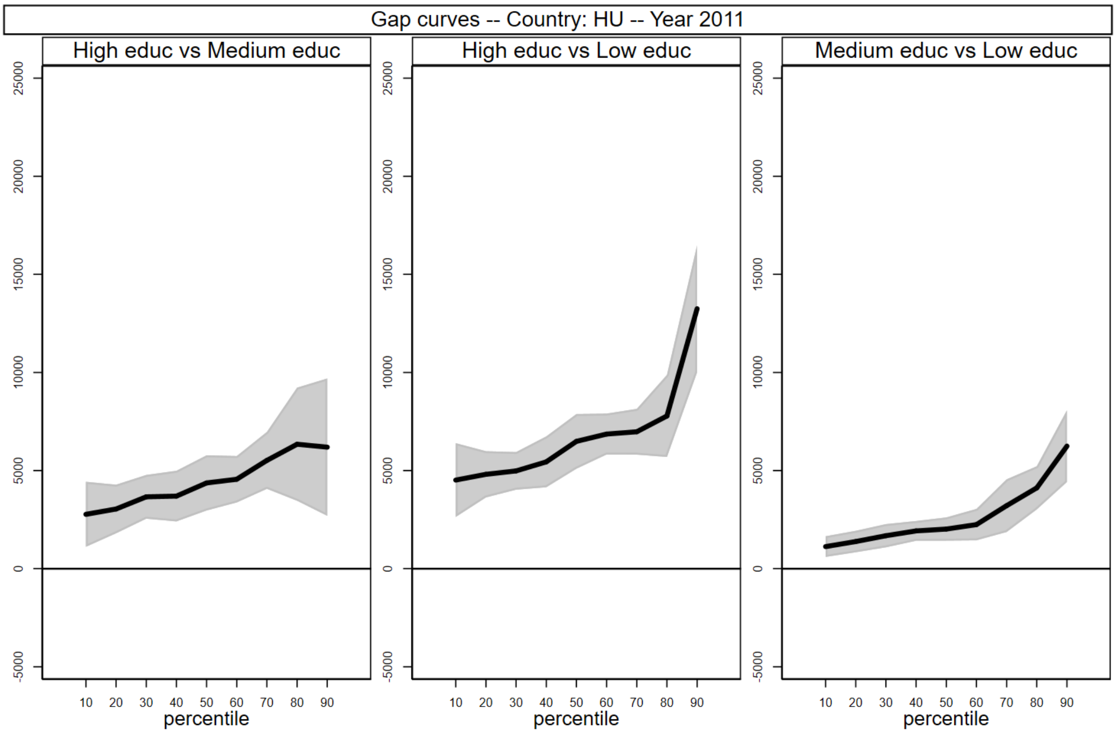


Figure 7: gap curves for Ireland

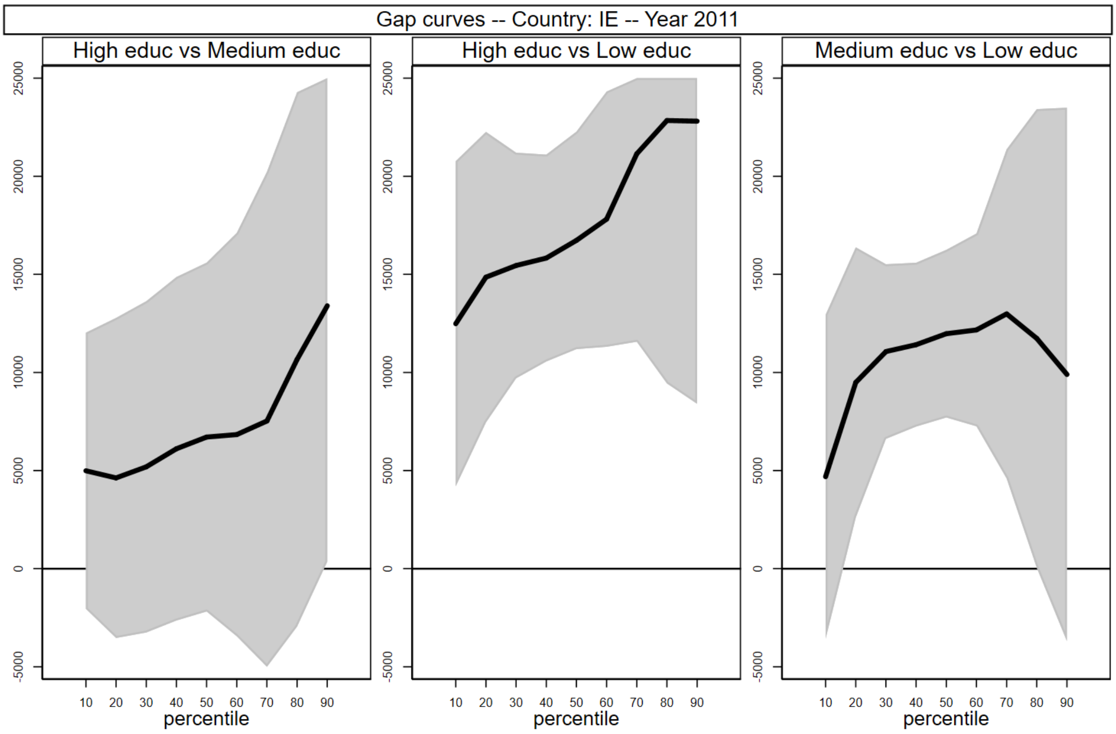


Figure 8: gap curves for Iceland

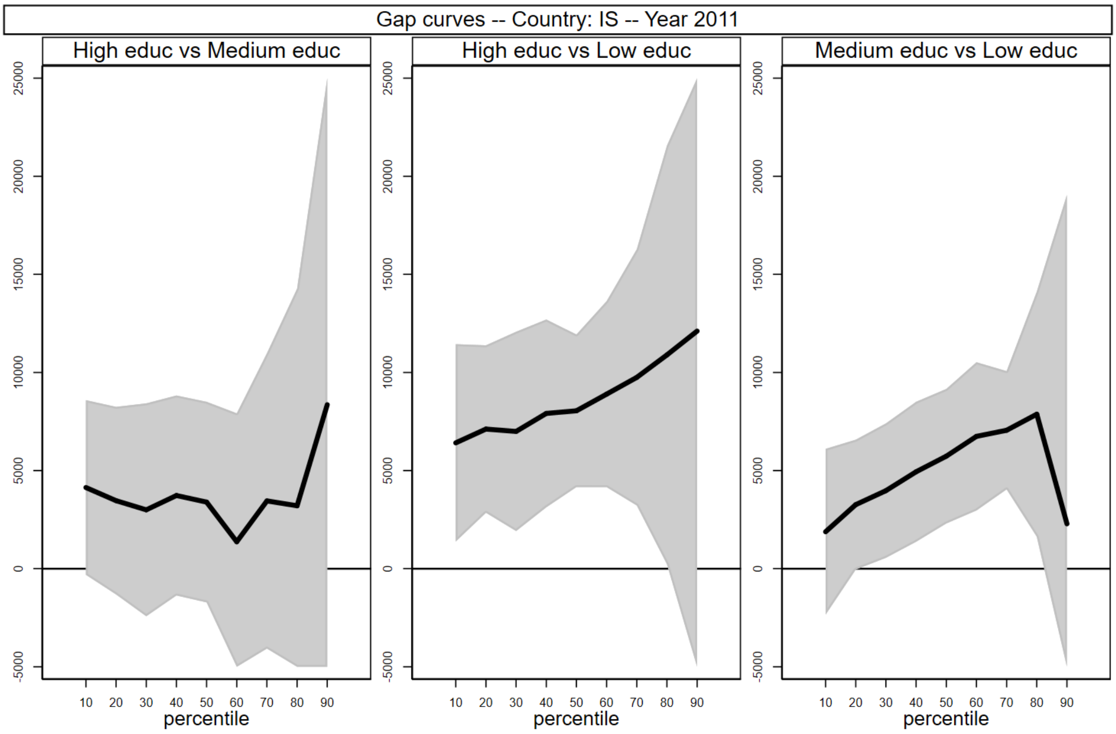


Figure 9: gap curves for Lithuania

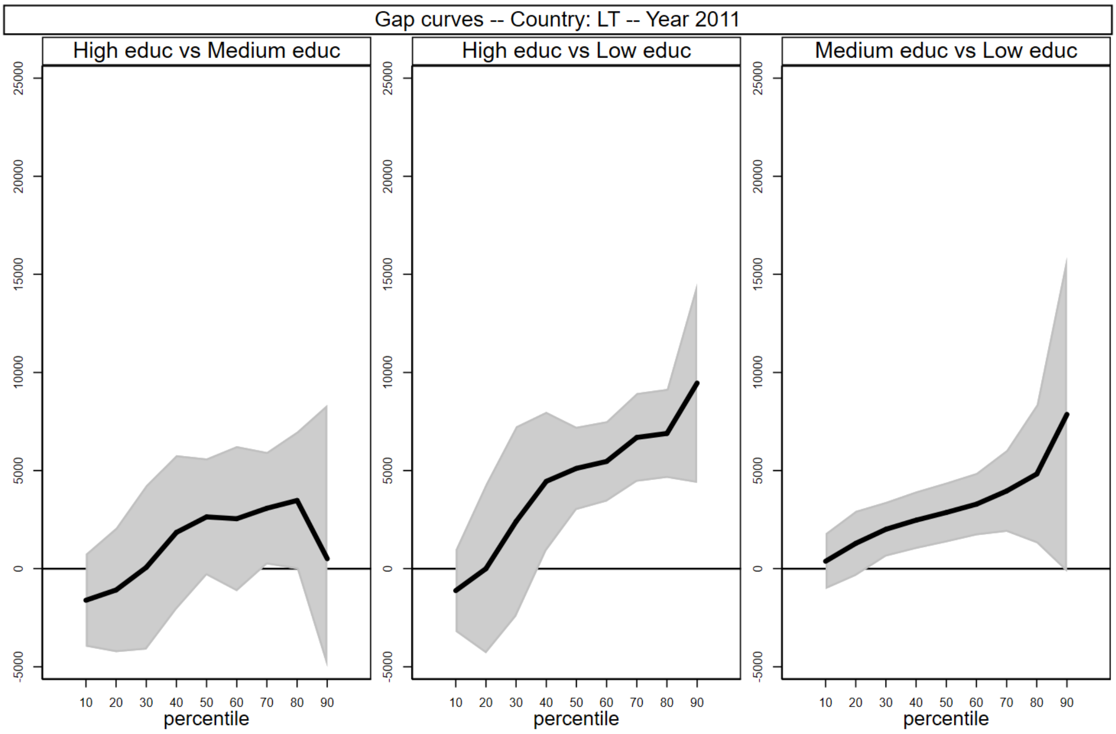


Figure 10: gap curves for Luxembourg

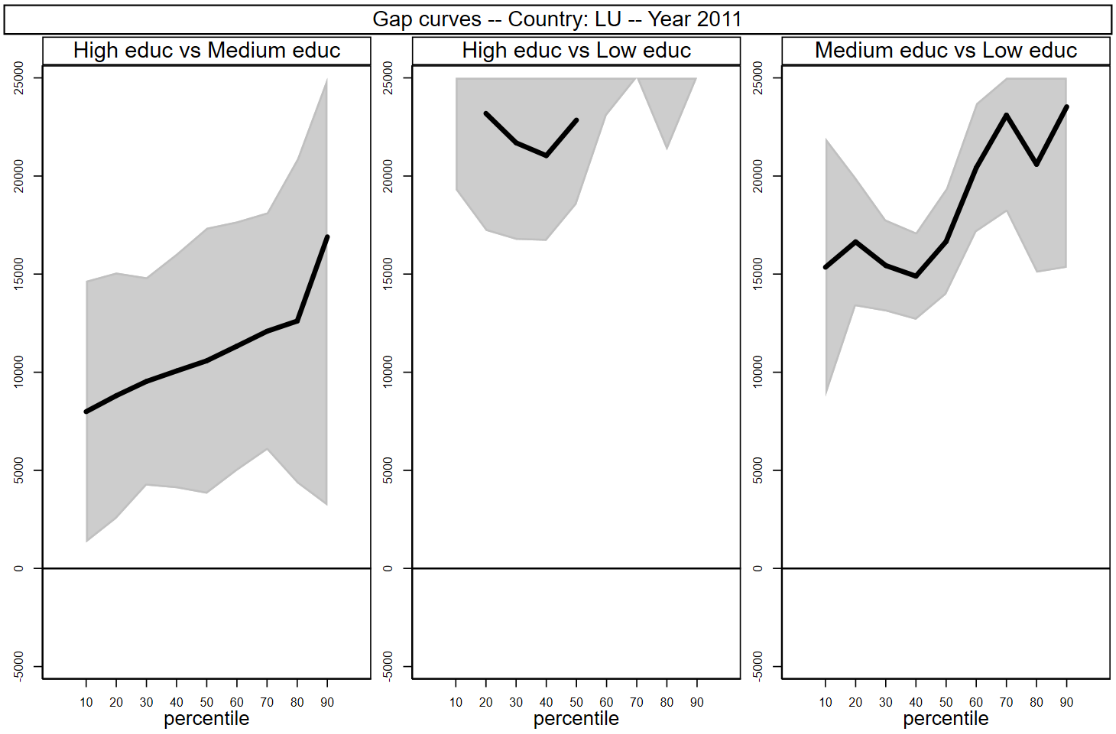


Figure 11: gap curves for the Netherland

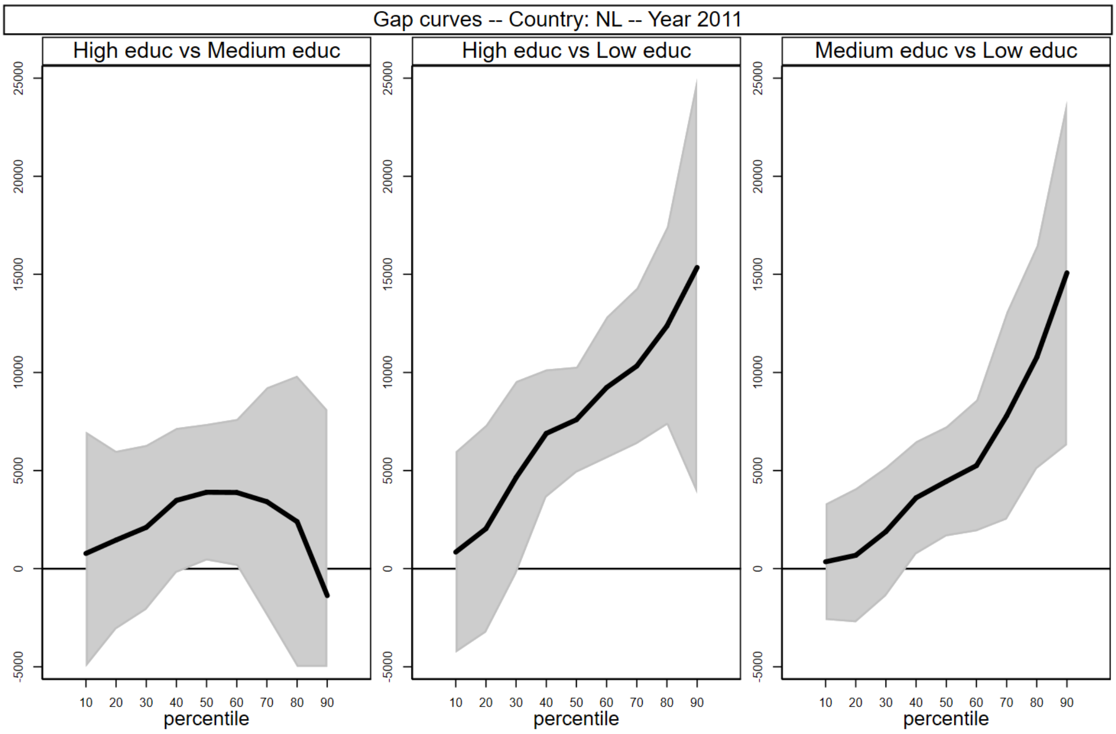


Figure 12: gap curves for Norway

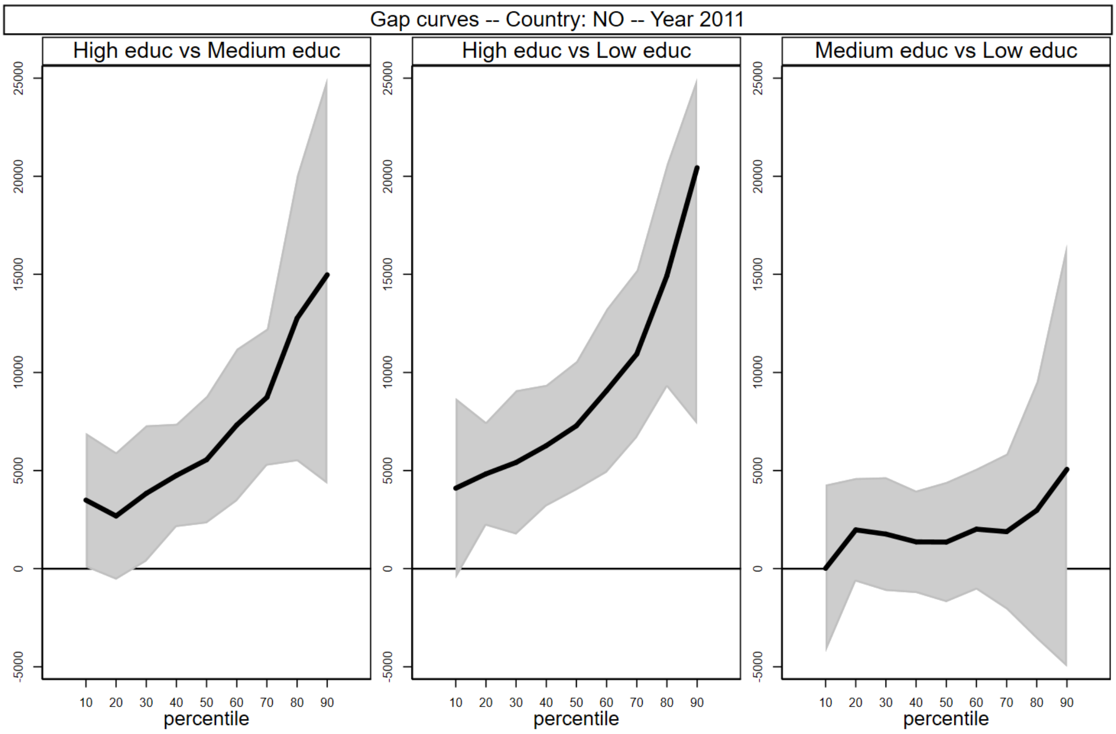


Figure 13: gap curves for Poland

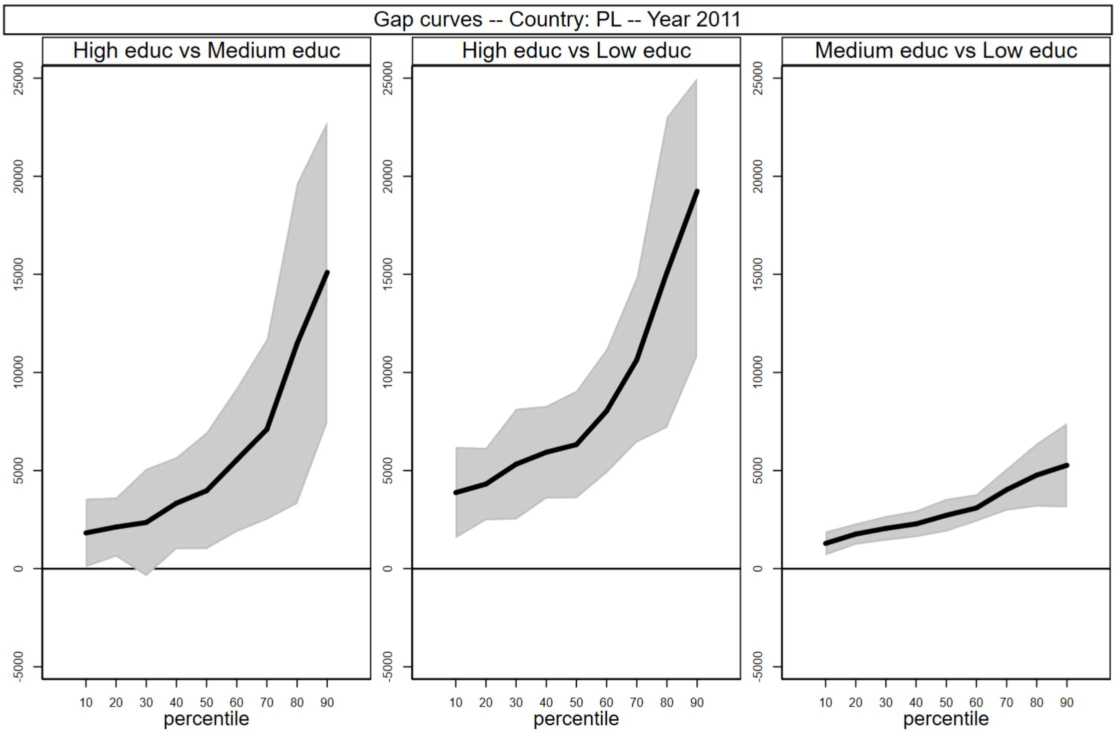


Figure 14: gap curves for Sweden

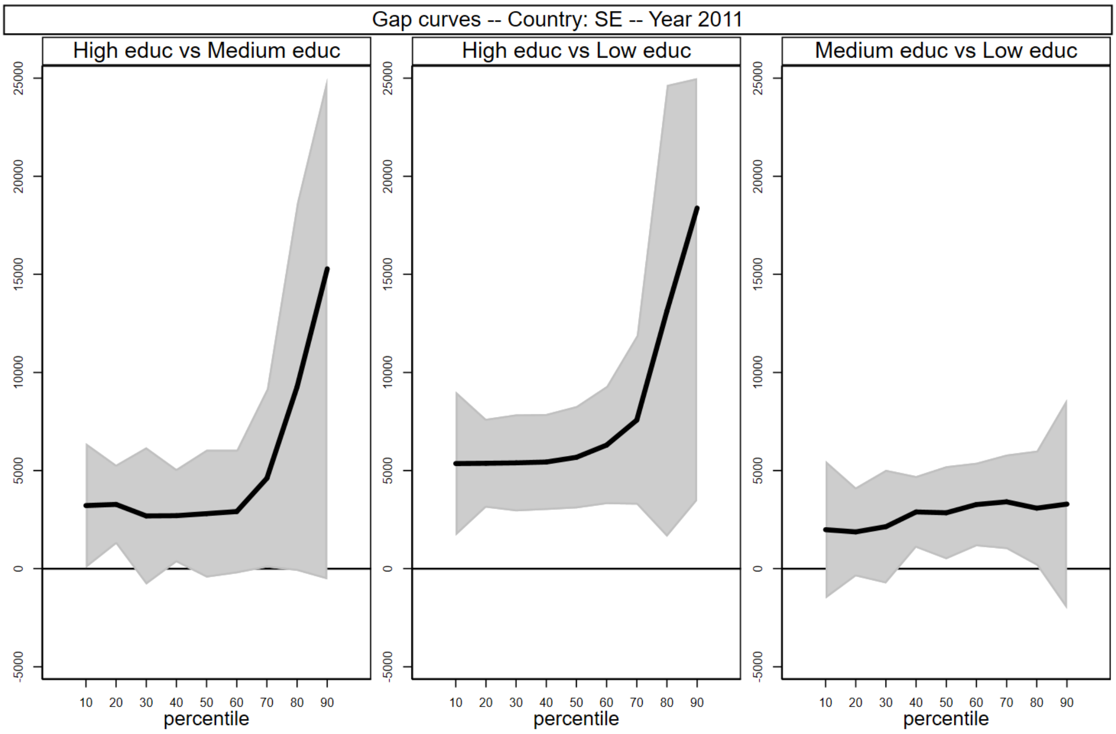


Figure 15: gap curves for Slovakia

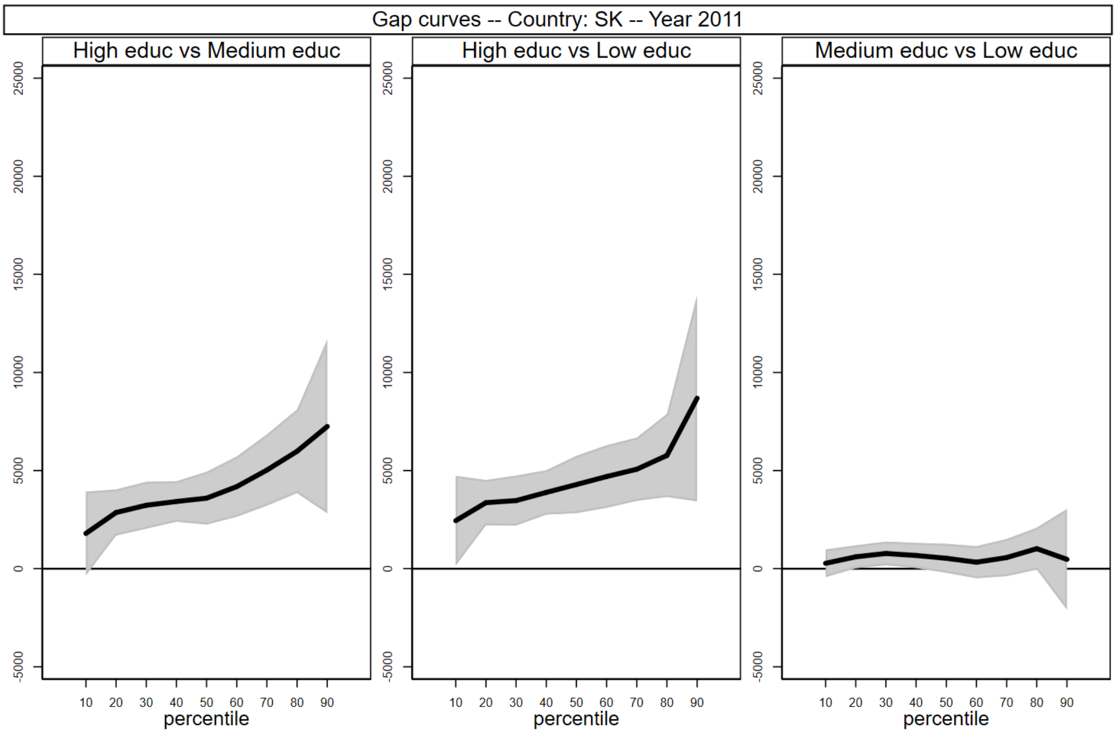
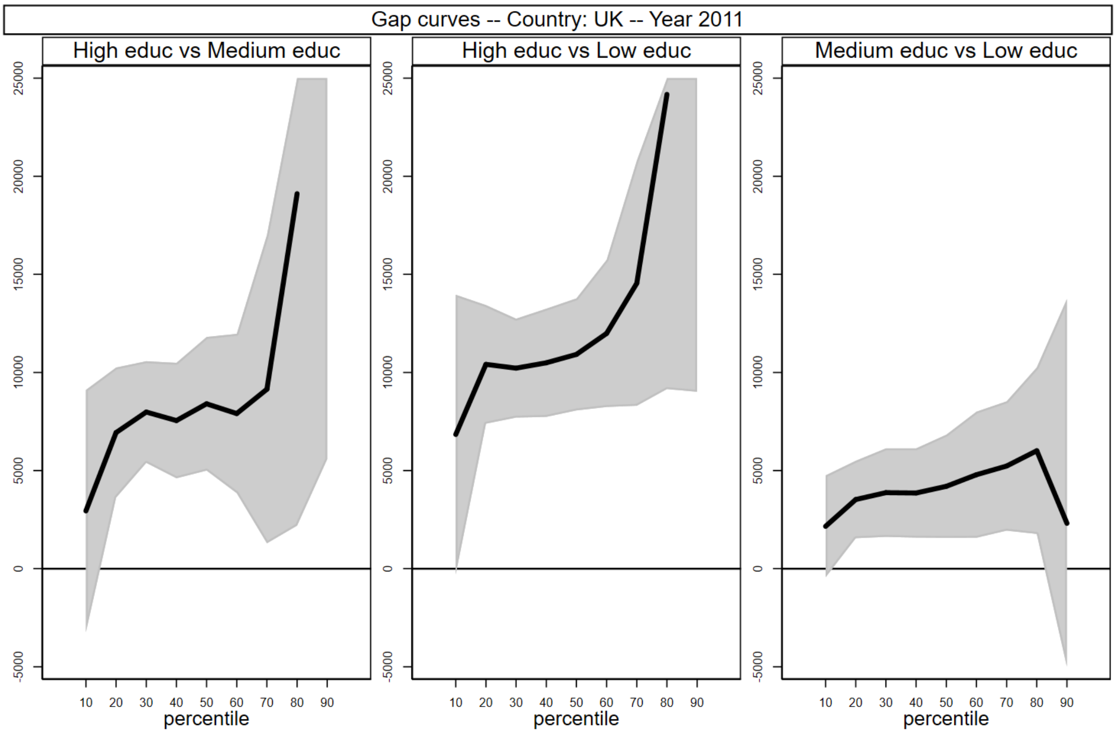


Figure 16: gap curves for the UK



Acknowledgements

Financial support from Eurostat (Net-SILC2), the French Agence Nationale de la Recherche (Ordineq grant ANR-16-CE41-0005-01) and the Luxembourg Fonds National de la Recherche (IMCHILD grant INTER/NORFACE/16/11333934, PREFERME CORE grant C17/SC/11715898 and IMeRSe CORE grant C15/SC/10266168) is gratefully acknowledged. Fusco also thanks the World Bank Development Research Group for its hospitality.

References

Andreoli, F. (2018). Robust inference for inverse stochastic dominance, *Journal of Business & Economic Statistics* 36(1):146-159.
Andreoli, F., Fusco, A. (2017) The evolution of inequality of opportunity across Europe: EU-SILC evidence. Ch. 24 in A.B. Atkinson, A.C. Guio and E. Marlier (Eds.) *Monitoring Social Inclusion in Europe*, Eurostat, pp. 435-448.
Andreoli, F., Fusco, A. (2019) Robust cross-country analysis of inequality of opportunity, *mimeo.*
Atkinson, T., Maynard, A.K. and Trinder, C. in conjunction with Corlyon, J., Jenkins, S.P. and Sutherland, H. (1983), *Parents and children. Incomes in Two Generations, Studies in Deprivation and Disadvantage*, Heinemann educational Books, London.

Firpo, S., Fortin, N., Lemieux, T. (2009), "Unconditional quantile regressions", *Econometrica,* 77(3) :953-973.
Goedemé, T. (2013), "How much confidence can we have in EU-SILC? Complex sample designs and the standard error of the Europe 2020 poverty indicators", *Social Indicators Research* 110(1):89-110.
Raitano, M. and Vona, F. (2015), "Measuring the link between intergenerational occupational mobility and earnings: evidence from eight European countries", *Journal of Economic Inequality*. 13(1):83-102.