

Income Redistribution and Self-Selection of Immigrants: Evidence from Administrative Data

Giacomo Corneo* Guido Neidhöfer†

Preliminary draft! Please, do not cite without permission!

21st June 2018

Abstract: We test the predictions of the Roy-Model about the self-selection of immigrants using an administrative dataset that includes 88 % of all Italians registered in a foreign country. The data at our disposal comprises 13 countries: Argentina, Australia, Belgium, Brazil, Canada, France, Great-Britain, Germany, The Netherlands, New-Zealand, Switzerland, the US, and Venezuela. Our results confirm the predictions of the model: the higher are the returns to skills, the lower is the individual degree of selection of Italian immigrants. Hereby, we measure returns to skills by the level of redistribution – i.e. the (relative) difference of market and after-tax inequality in the host country in the year of arrival. Our findings hold after inclusion of covariates at the individual and country level, controlling for migration costs, and testing for stochastic dominance over the entire skill distributions of migrants and stayer. Finally, we run a discrete choice model that shows a substantial effect of potential net income returns on the choice of the destination country. Our analysis also sheds light on the factors associated with the migration decision and the self-selection of immigrants.

Keywords: Roy-Model, Self-selection, Migration, Administrative Data

*Freie Universität Berlin

†DIW Berlin and Freie Universität Berlin. Contact: guido.neidhoefer@fu-berlin.de

1 Introduction

Theoretical and empirical studies on the selection process of migrants agree on one fundamental aspect: Migrants are not a random draw from the population of their home country neither are they undistinguishable in their observable and unobservable characteristics from the native population of their host country. Besides this basic consensus, different theories and findings on the patterns characterizing the self-selection of immigrants coexist.

For instance, [Borjas \(1987\)](#) applies the Roy-Model of self-selection and argues that the returns to human capital in source and destination country determine whether high or low skilled individuals migrate: the higher are the returns to human capital in a country, the more high skilled individuals will tend to migrate in this particular country, particularly from countries with lower returns to skills. This hypothesis has been empirically confirmed recently by [Moraga \(2010\)](#) and [Parey et al. \(2017\)](#), among others. On the other side, [Chiswick \(1999\)](#) argues in favour of a general positive selectivity of migrants in line with the predictions of standard human capital theory since [Sjaastad \(1962\)](#); another hypothesis that has as well been confirmed in empirical studies like [Liebig and Sousa-Poza \(2004\)](#) and [Chiquiar and Hanson \(2005\)](#).

Most studies that tested the self-selection of immigrants, so far, rely on macro-data containing aggregate information about the characteristics of migrants, census data that allows to analyse flows and stocks of migrants from one particular country to the other (mostly from Mexico to the US), or survey data reporting pre-migration earnings or migration intentions. Further evaluations using novel data sources seem therefore necessary to deepen our understanding about the process of immigrants' self-selection. The strength of our analysis is that we are able to test the selection process on an administrative dataset that contains almost all Italian families living outside of Italy.

The data at our disposal encompasses 88 % of Italians registered abroad. Overall, our sample comprises more than four million Italian citizens living in 13 foreign countries: Argentina, Australia, Belgium, Brazil, Canada, France, Great-Britain, Germany, The Netherlands, New-Zealand, Switzerland, the US, and Venezuela. Of these, about 1.3 million have an own migration experience

(i.e. were born in Italy). To the best of our knowledge, the only other study testing the selection process of migrants with administrative data on almost the entire population of emigrants is [Borjas et al. \(2018\)](#) on Danish migration register data.

Studies analysing the self-selection of migrants with micro-data mostly observed flows from developing countries to developed countries, hence from typically poor and unequal to rich and less unequal countries (as pointed out by [Hatton, 2014](#)), or as mentioned above from Denmark, a country with a relatively flat income distribution, to the rest of the world. Hence, another important contribution of our study is that we observe migrants from a country with medium level of redistribution, Italy, who migrated to more unequal countries like Argentina, Brazil, Venezuela, New Zealand and the US, and to less unequal ones like Belgium, France, Germany and the Netherlands. Furthermore, Italy is the 12th country in the world by its absolute number of emigrants.

With this powerful data source at our disposal, we test whether lower redistribution, our measure for high returns to human capital, attracts positively selected individuals in observable and unobservable characteristics. Hereby, selection on observable characteristics is measured by predicting the labour earnings of immigrants have they stayed in Italy and as the relative educational position of migrants with respect to their reference group, i.e. stayers born in the same year living in the Italian region where the individual resided before moving to the foreign country. The averages for the population of stayers are computed using Italian household survey data. Selection on unobservable characteristics is measured by the probability to be unemployed or to have a high occupational position given the level of education.

All estimations confirm the predictions of the Roy-Model: lower redistribution is associated with higher levels of skill selectivity of Italian immigrants. This relationship holds controlling for demographic characteristics of the migrants, country characteristics like GDP per capita and the unemployment rate, migration costs, approximated by the distance of the country of residence to the Italian border, the existence of migration agreements between the two countries and the share of peers from the same Italian province residing in the same country, and finally even country fixed effects. Furthermore, the skill distribution of migrants in countries with relatively low levels of

redistribution stochastically dominates the distribution of stayer and of migrants in countries with high levels of redistribution.

Last, we estimate a discrete choice model of the decision to migrate as a function of predicted net income gains in the country of destination, as well as other personal characteristics and place characteristics of the country of destination. Also this exercise confirms the predictions of the Roy-Model of income maximization as driver of migration decisions.

2 Returns to Skills and Self-selection of Migrants

In economic models of human migration, the point of departure is usually given by human capital theory. Here, the migration decision of rational agents is a function of the expected lifetime utility gain of the move net to the costs of migration (Sjaastad, 1962). Following this framework the act of moving from one place to another is nothing more and nothing less than an investment in human capital. However, another important aspect is that not all individuals have the capabilities to estimate their costs and benefits of an eventual move, or the right incentives to make the investment. Hence, migrants are a self-selected sample of the populations of host and source countries.

The main dispute in the literature dealing with the self-selection of immigrants is the answer to the question whether immigrants have higher or lower observed and unobserved skills in comparison to non-migrants. The two alternative scenarios are known as positive and negative self-selection, meaning a higher or lower than average degree of skills among the immigrants, respectively. Borjas (1987) argues, applying the Roy-Model (Roy, 1951), that the decision to migrate to one country or another (or stay in the country of origin) depends on the comparative advantages of individuals to obtain the highest possible earnings for their particular level of skills (see also Borjas et al., 1992). Hence, countries with higher returns to skills attract individuals at the top of the skill distribution from countries with lower returns to skills. Vice versa, individuals at the bottom of the skill distribution in their country of origin should have an incentive to move to more egalitarian countries, where the expected earnings for their given skill level are less far away from the earnings of high skilled individuals.

Formally, consider individuals living in country I deciding to migrate either to country G or to country U or staying in their home country. The potential log earnings (w) in country $i \in (I, G, U)$ depend on the individual's skill level (s), as well as the returns to skills (ρ) and the social minimum (ω) in i :

$$w_i = \omega_i + \rho_i s \quad (1)$$

We assume that the degree of returns to skills is highest in U and lowest in G ($\rho_U > \rho_I > \rho_G$), while the reverse applies to the level of social minimum ($\omega_U < \omega_I < \omega_G$). Individuals will migrate if their expected net gains in the foreign country will exceed the net gains in their home country. Hence individuals will migrate from I to U if for their given skill level

$$w_U - c_U > \max(w_I, w_G - c_G), \quad (2)$$

and to country G if

$$w_G - c_G > \max(w_I, w_U - c_U), \quad (3)$$

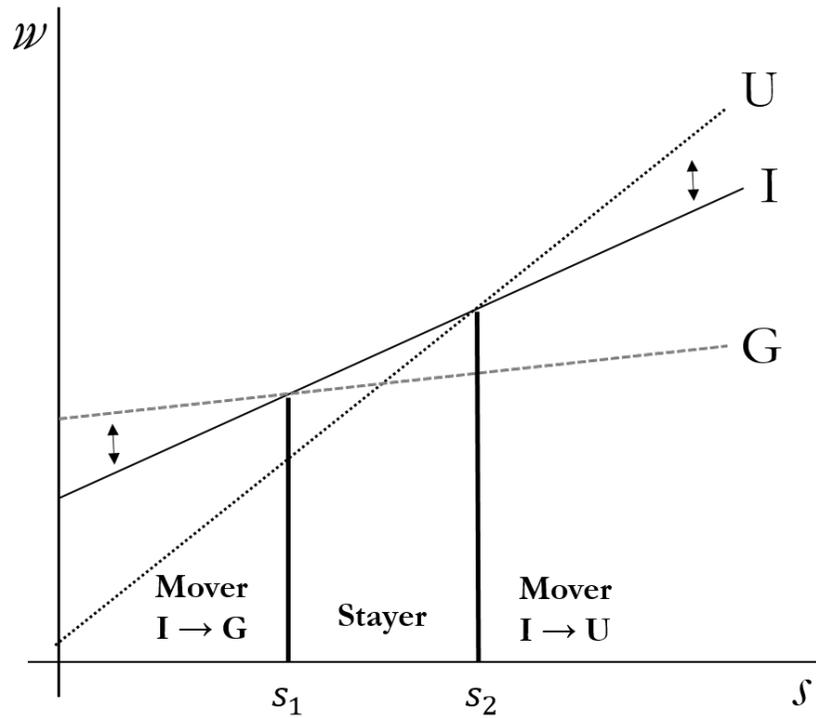
where c_i are the costs of migration to G or U. Consequently, the threshold skill levels that define whether an individual migrates or not and to which country, are

$$s_1 = \frac{(\omega_U - c_U - \omega_I)}{(\rho_I - \rho_U)}, \quad (4)$$

$$s_2 = \frac{(\omega_G - c_G - \omega_I)}{(\rho_I - \rho_G)}. \quad (5)$$

Figure 1 shows this relationship graphically for the case of migration to both foreign countries. Returns to skills are lower (higher) in G (in U) than in I, hence individuals migrate from I to G (to U) if their skill level is below (above) the threshold s_1 (s_2). For individuals with skill levels in between the two thresholds s_1 and s_2 there is no incentive to migrate. Hence, there is a situation of negative self-selection from country I to G and positive self-selection from I to U.

Figure 1: Positive and Negative Self-Selection of Migrants



Furthermore, as it is easy to see, if the relative returns to skills between the source and the host country change, this affects the average degree of migrants' selection. For instance, if the relative returns between U and I become lower – for example as consequence of a marginal rise of ρ_U with constant ρ_I – the threshold s_2 shifts to the left, i.e. the last stayer has a lower skill level than before.¹ This shift has no effect on the general pattern of positive self-selection of immigrants from I to U, but since the average skill level in the source country is endogenously determined by the migration process, it has an effect on the average degree of self-selection. The population of stayer has a lower average degree of skills and, under regularity assumption on the distribution of skills (e.g. log-normal distribution), the average skill level of the immigrants in U rises in comparison to the stayers. The same applies to the average degree of selection of immigrants in G with rising ρ_G , as

¹The effect of highly skilled individuals leaving their country of origin and consequently lowering the stock of skills is also called brain drain in the migration literature. However, other models show that the brain drain might also contribute to human capital formation in the sending countries if the higher returns to education that cause the brain drain also constitute an incentive for people with migration intentions to invest in education (Beine et al., 2008).

long as $C_G < \omega_G - \omega_I$. In contrast, when ρ_G declines, s_1 shifts to the right: the average skill level of the immigrants in G becomes lower in comparison to the stayer.

The implications of this theoretical framework were firstly confirmed by empirical analyses of international and internal migrants in the US: [Borjas \(1987\)](#) shows that the degree of income inequality in the home country, a measure for the returns to skills, is a predictor for the type of selectivity of migrants, while [Borjas et al. \(1992\)](#) show that interstate variations in the returns to skills affect the skill structure of migration inflows. In contrast, [Chiswick \(1999\)](#) states that immigrants tend always to be favourably self-selected also in presence of higher levels of income inequality. In his view, [Borjas \(1987\)](#)'s empirical results only show that income inequality attenuates the degree of selection, but not the generally positive selection pattern. [Liebig and Sousa-Poza \(2004\)](#) confirm this last hypothesis applying an empirical analysis using cross-country survey data on migration intentions. Other empirical papers testing the selection of migrants from one or more source countries to one or more destination countries, with individual level and aggregated data, obtained contrasting results; recent studies that review the empirical literature and also test this theory are [Parey et al. \(2017\)](#) and [Patt et al. \(2017\)](#).

One common explanation for the differing results so far has been argued to lie in the important role of migration costs for the migration decision and the selection process. For instance, [Chiquiar and Hanson \(2005\)](#) show that if migration costs are negatively correlated with skills, both positive and negative self-selection into countries with higher returns are possible solutions. Apart from transportation costs and the value of friends, family and culture left behind, other factors are determinant for the costs of migration like immigration policies or migration networks ([Hatton, 2014](#)). [McKenzie and Rapoport \(2010\)](#) show indeed that self-selection is more likely to be positive in places with low migration networks and more likely to be negative in places where many migrants from the same origin countries reside. Hence, the literature is still quite open on the topic and it seems of particular importance to add further evidence on the self-selection of immigrants.

3 Data

3.1 Administrative Data on Italians Abroad

The micro-data basis of our empirical tests is the Registry of Italians resident abroad (*Anagrafe degli italiani residenti all'estero*, AIRE), an administrative registry dataset provided us by the Italian Ministry of Foreign Affairs through the Italian Embassy in Germany. Table 1 shows the number of registered people in the AIRE data at our disposal. All Italians who are at least one year abroad or born outside of Italy are required to register to the AIRE by law. Furthermore, some bureaucratic tasks, e.g. renewing an Italian passport or ID card and voting, are only possible being recorded in the AIRE. Furthermore, for Italians residing abroad the registration in the AIRE is a necessary condition to avoid to pay the income tax in Italy. The dataset at our disposal contains individual information on 4,079,646 registered Italian citizens in 13 different foreign countries between 2014 and 2015 as well as information on their spouses and children without Italian citizenship. Our data encompasses approximately 88 % of all Italian citizens worldwide.² The individual information contained in the registry data are: Date of birth, date of arrival in the host country, sex, place of birth, place of residence, education, profession and the last municipality of residence in Italy before migration.

For our analysis, we are only interested in immigrants with own immigration experience. Hence we focus only on individuals that were born in Italy and exclude the foreign-born children and grandchildren of immigrants (so-called second and third generation immigrants) from our final sample. Furthermore, since we want to capture the selection mechanism, we assure that individuals in our sample had already finished their educational career when they moved to the foreign country. We do so excluding all individuals that registered to AIRE when being younger than 20 years old. Finally, to avoid bias deriving from, first, individuals who did not finished their educational career, and, second, differential mortality rates between people with different educational levels, we restrict

²The only country missing in our sample of the ten countries with the highest concentration of Italian immigrants worldwide in 2015 is Spain.

Table 1: Number of registered individuals in the registry of Italians resident abroad (AIRE)

	Registered in AIRE				
	Total	Italian citizenship	Born in Italy	Age 30-64	Arrival 1960-2015
ARG	1,191,059	893,974	119,008	6,765	1,866
AUS	221,292	149,246	53,900	20,572	11,479
BEL	333,235	273,415	95,438	42,453	25,500
BRA	597,232	450,939	36,804	10,132	5,373
CAN	188,289	137,289	72,909	20,884	12,749
CH	695,081	607,084	220,133	119,338	88,199
FRA	214,512	170,023	70,736	32,255	19,712
GBR	308,077	263,916	130,100	73,549	55,138
GER	813,254	694,694	300,863	204,207	134,159
NLD	48,895	41,346	16,767	10,726	4,492
NZL	5,056	4,052	1,497	961	699
USA	334,093	250,176	133,498	72,010	51,083
VEN	218,351	143,492	28,801	6,071	2,461
Total	5,168,426	4,079,646	1,280,454	619,923	412,910

Subsequent columns show the respective subtotal.

our sample to the age range 30 to 75, i.e. to the cohorts 1940 to 1985. Table 1 shows descriptive statistics of individual and country characteristics from the final sample.

3.2 Relative Skills and predicted net returns to skills

We use additional sources of microdata to estimate the relative skills and the net returns to skills for every individual in our administrative dataset. To verify the level of individual selection of movers (people registered to AIRE) relative to the population of the stayers (Italians residing in Italy), we complement this dataset with an established Italian household survey collected by the Bank of Italy: the Survey on Household Income and Wealth (SHIW). The SHIW collects since 1960 information on Italian families including individual characteristics for each single household member. For the present study, we use the comparable survey waves 1977 to 2014 and normalize the sampling weights for each single year. We use this data on Italians living in Italy to measure the relative skills of movers, running an augmented Mincer regression of *log* labour earnings on the survey sample of stayer and using the coefficients of the regression to predict the counterfactual *log*

Table 2: Descriptive Statistics of the Sample – Country Averages

Year of birth	Year of arrival	Share of female	Years of education	Rural origin	Capital	Internal migrant	GDP p.c.	GDP growth	Unemployment rate	Distance (in tsd km)	Bilateral agreement	Policy toward high skilled
ARG	1948.83	0.53	8.39	0.45	0.21	0.11	6.33	4.16	8.41	10.200	1	0
AUS	1958.68	0.40	9.79	0.31	0.00	0.13	24.19	3.70	5.18	12.100	1	0
BEL	1960.24	0.42	10.51	0.20	0.06	0.15	24.13	2.47	7.57	0.530	1	0
BRA	1961.08	0.20	12.68	0.15	0.01	0.21	7.07	3.23	8.21	7.000	1	1
CAN	1952.83	0.43	8.74	0.39	0.02	0.09	14.75	3.88	6.36	5.000	1	1
CH	1961.66	0.39	10.28	0.32	0.01	0.16	49.66	1.71	3.06	0.000	1	1
FRA	1966.65	0.46	13.23	0.17	0.32	0.18	30.62	1.80	9.18	0.000	1	0
GBR	1971.28	0.42	13.17	0.17	0.48	0.17	35.29	2.18	6.68	0.900	1	0
GER	1960.66	0.37	8.62	0.31	0.04	0.11	23.13	2.09	5.85	0.070	1	0
NLD	1973.91	0.37	14.43	0.14	0.28	0.19	46.05	1.10	5.46	0.640	1	0
NZL	1970.54	0.42	13.72	0.13	0.14	0.18	30.39	2.37	5.28	18.000	0	0
USA	1960.30	0.43	11.58	0.20	0.01	0.14	30.13	2.87	6.12	6.100	0	1
VEN	1950.67	0.40	10.36	0.24	0.47	0.16	3.88	3.32	8.45	7.700	0	1
Total	1962.10	0.40	10.55	0.27	0.11	0.14	30.94	2.34	5.81	1.952	1	0

Sources: Individual characteristics from AIRE. Sample is restricted to people born in Italy, 30-64 years old, and who registered in AIRE after the age of 20. Unemployment rate, GDP growth, GDP per capita from World Bank Data. Distance to Italy measured from border to border (Google Maps). Migration agreement is equal to one if there is/have been bilateral migration agreements between Italy and the country of destination; information retrieved from different sources. Policies oriented towards high skilled is equal to one if the policies of the country in the last 50 years have been more oriented at attracting high skilled immigrants or disincentive low skilled immigrants; information retrieved from the DEMIG Policy Database.

labour earnings that mover would have obtained in Italy. The included variables in the regression are years of schooling, age, quadratic age, sex, Italian region of origin, Italian region of birth and an indicator on whether the individual is an internal migrant (i.e. does not live in the same region or did not migrate from the same region where he or she was born). For this exercise, we exclude individuals older than 65.

As further sensitivity analysis, we estimate also the relative educational position with respect to the reference group and use it as an additional measure for the individual degree of self-selection. The relative educational position of individual i born in year b is defined as the relative difference of his or her years of education y with respect to the average of stayers born in the same year residing in his or her Italian region of origin j : $s_{ijb} = \frac{(y_{ijb} - \bar{y}_{jb})}{\bar{y}_{jb}}$. The average of stayers of each birth cohort from 1940 to 1985 in the 20 Italian regions is computed using the Italian household survey SHIW. Results applying this specification of relative skills are included in the Appendix.

We use harmonized microdata from the Luxembourg Income Study (LIS) to estimate the net returns to skills in the country of residence for mover, the counterfactual returns to skills in Italy and every other of the possible destination countries in our dataset of mover, as well as the returns to skills of stayer in Italy and the counterfactual returns to skills of stayer in all possible destinations. To measure this, we use the survey samples of every single destination country around the year 2014. We again estimate an augmented Mincer regression on (log and absolute) household disposable income including the variables sex, age, quadratic age, education, and an indicator on whether at least one child lives in the household. Last, we predict the net incomes of Italian immigrants in all possible destinations and in Italy using the coefficients of this regression. Unfortunately, LIS data is not available for Argentina, New Zealand, and Venezuela. Furthermore, the last available survey for Belgium dates back to the year 2000. Therefore, the parts of the analysis that analyse predicted net income returns are restricted to the remaining nine possible destination countries and Italy.

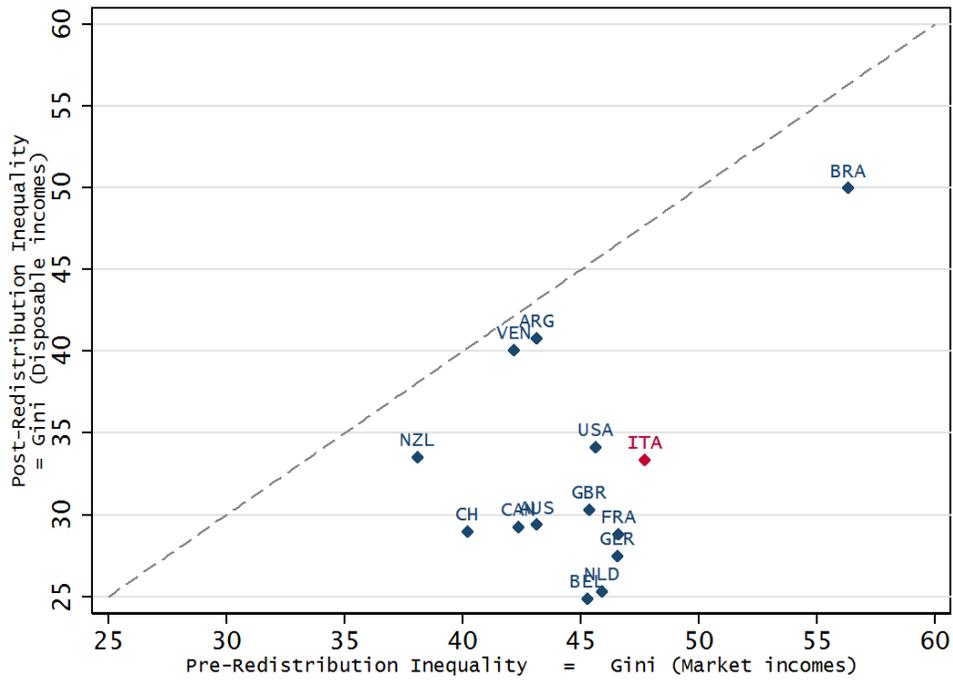
3.3 Country characteristics

Finally, we collect data on country characteristics from different sources: for instance, the Standardized World Income Inequality Database (SWIID), the World Income Inequality Database (WIID), World Bank Macro Data (WB-Data), and the Andrew Young School World Tax Indicators (WTI). From the SWIID we retrieve net and market income inequality indices and compute measures for absolute and relative redistribution, the first measured by the difference between the market and net Gini index, the second by this difference divided by the market Gini (see [Reynolds and Smolensky, 1977](#); [Solt, 2016](#)). Inequality and particularly the level of redistribution are our proxies for the returns to skills and therefore our main variables of interest. [Figure 2](#) shows the average level of pre and post-redistribution inequality for the countries in our sample, and for Italy. The distance from the 45 degree line shows the contribution (in Gini points) of taxes and transfers to the reduction of market income inequality. [Figure 3](#) shows the time trends in the evolution of net income inequality with respect to Italy. As is evident, the countries in our sample are quite heterogeneous in their levels of inequality and redistribution, in particular with respect to Italy, and with an interesting amount of variation over time.

[Figure 4](#) ranks the countries by their average level of redistribution over all observed years. On the x-axis redistribution is defined in absolute terms while on the y-axis in relative terms as defined above. The country ranking is rather similar applying both measures and depicts basically three main groups of countries with high, medium and low levels of redistribution. The first group comprises Belgium, the Netherlands, Germany and France. The second, besides Italy, also Great-Britain, Australia, Switzerland, Canada and the US. The third, New-Zealand, Brazil, Venezuela and Argentina.

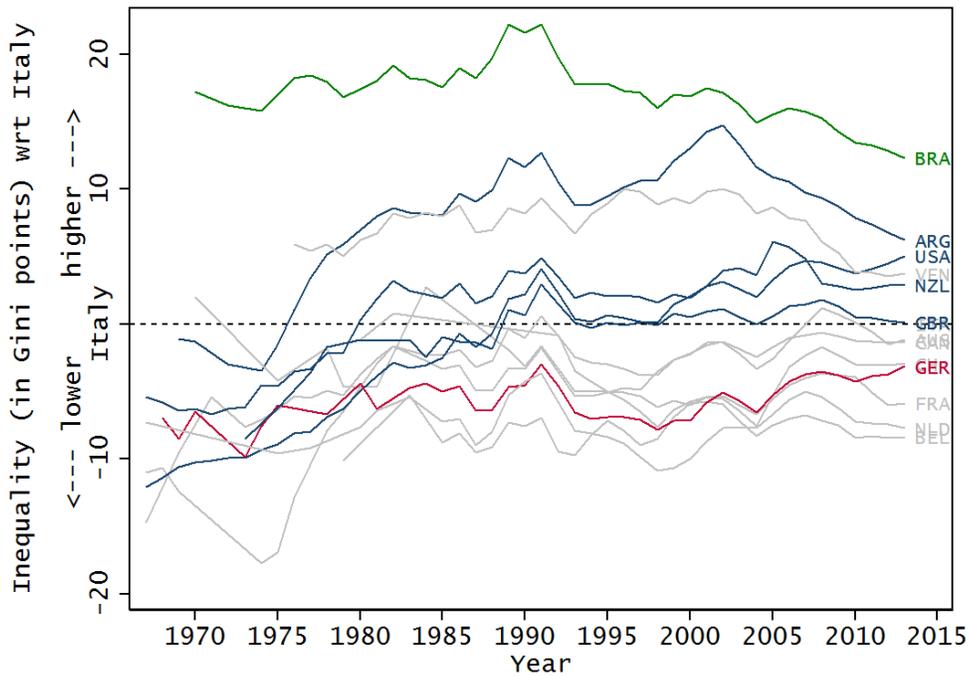
To cross-check the results obtained using SWIID we perform the same analysis using WIID and obtain no significant differences in results. The latter is a dataset with sensibly less observations for each country than the former, but has been argued in past to rely on a more consistent methodology and data basis than SWIID (see [Jenkins, 2015](#)). Anyway, for the countries in our sample, including Italy, we find a very high degree of congruence between SWIID and WIID data with a correlation of

Figure 2: Levels of Inequality Pre- and Post-Redistribution



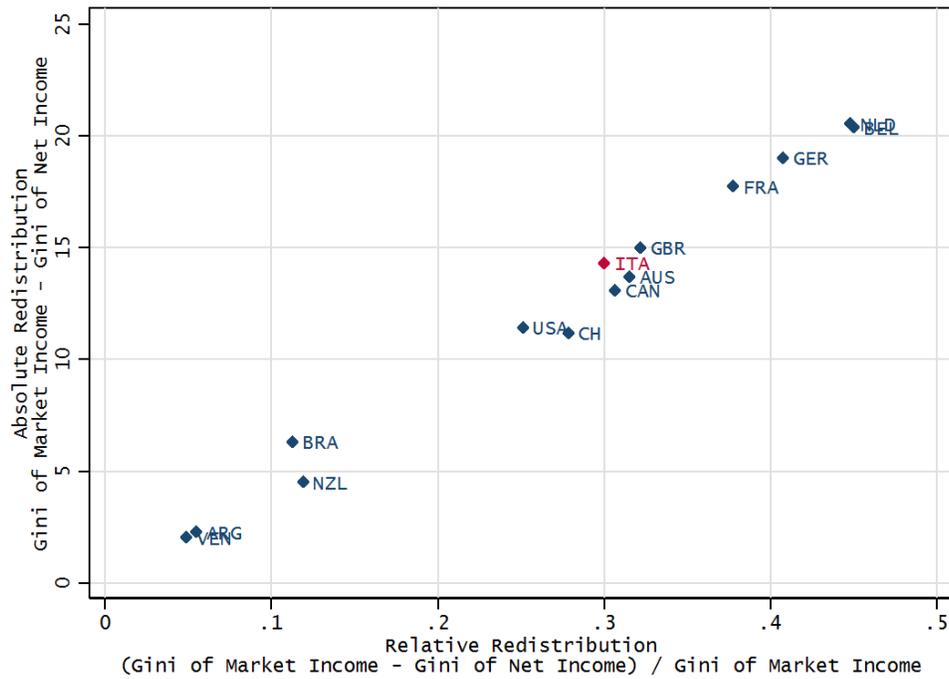
Source: SWIID, own calculations. Average over all available years.

Figure 3: Disposable income inequality trends with respect to Italy



Source: SWIID, own calculations.

Figure 4: Absolute and relative redistribution



Source: SWIID, own calculations. Average over all available years.

about 0.95. From different sources we retrieve other country characteristics which could also affect the type and level of selection of immigrants and hence act as control variables in the regressions; see the notes of Table 2. Last, we perform the analysis also with the Marginal Tax Rate Progression retrieved from WTI data as further proxy for returns to skills.

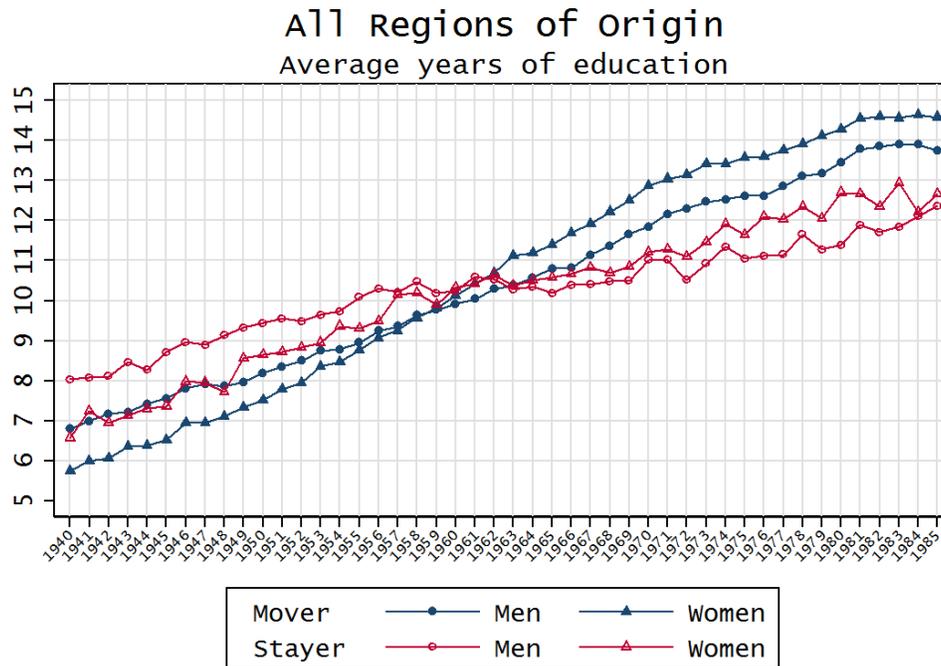
4 Descriptive Evidence

4.1 Selection on Educational Attainment

The comparison of the education of Italian immigrants in different countries already gives some important hints about differential self-selection patterns. Figure 5 shows the average years of education of individuals in our administrative data on Italians abroad (Mover), compared to the average of Italians living in Italy estimated from Italian household survey data (Stayer).³ On average male

³Years of schooling are coded following this scheme: No school degree, 0 years. Uncompleted compulsory schooling, 5 years. Compulsory schooling, 8 years. Beyond compulsory education, 13 years. Tertiary degree, 16 years.

Figure 5: Average Years of Education of Italian Mover and Stayer by Year of Birth. Italians of all regional origins and subdivided by gender.

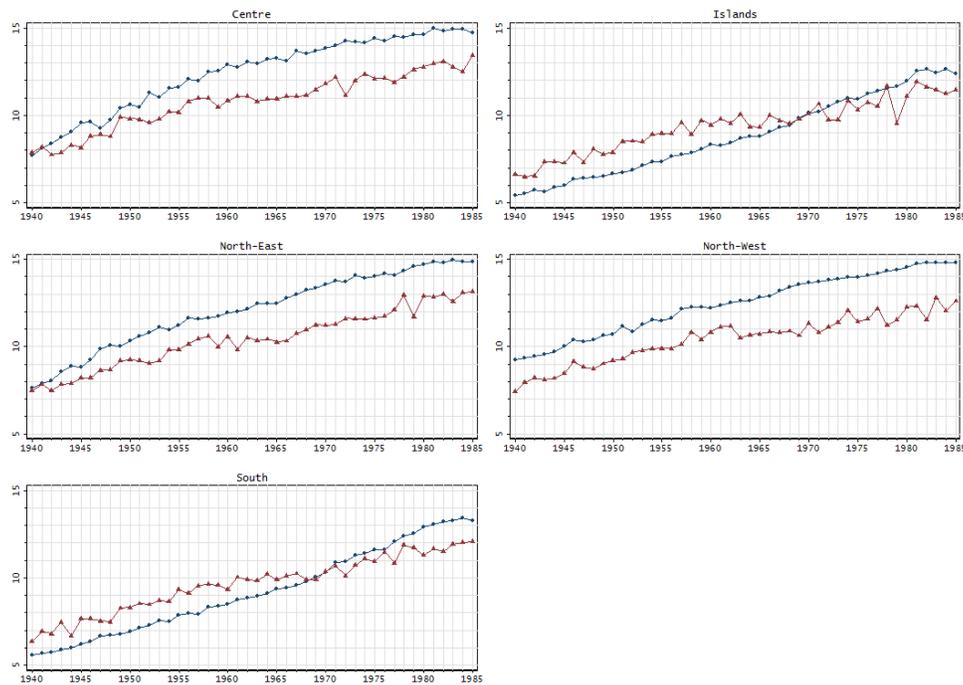


Source: Averages for Mover are own calculations using AIRE, averages for Stayer are own calculations using SHIW.

and female Italians born from 1940 to 1961 and living abroad in 2015 have lower educational attainment than the average of the Italian population of people born in the same years. After this year, subsequent cohorts of migrants have a relatively higher average education by about one year of education. It shows up that especially the share of individuals with a completed tertiary education degree experienced a dramatic rise. Furthermore, we observe a similar development in gender differences among movers and stayers: older men have higher education than women, while younger women are more educated than men.

Figure 6 shows these differences in average educational attainment by Italian region of origin. Mover from the Centre and North of Italy have constantly a higher education than stayer, while mover from South Italy and the Islands have lower or rather similar average education than stayer. This finding highlights the crucial importance to evaluate the degree of selection of people with respect to their region of origin rather than with respect to the national average.

Figure 6: Average Years of Education of Italian Mover and Stayer by Year of Birth and Geo-graphic Macro-Region of Origin.



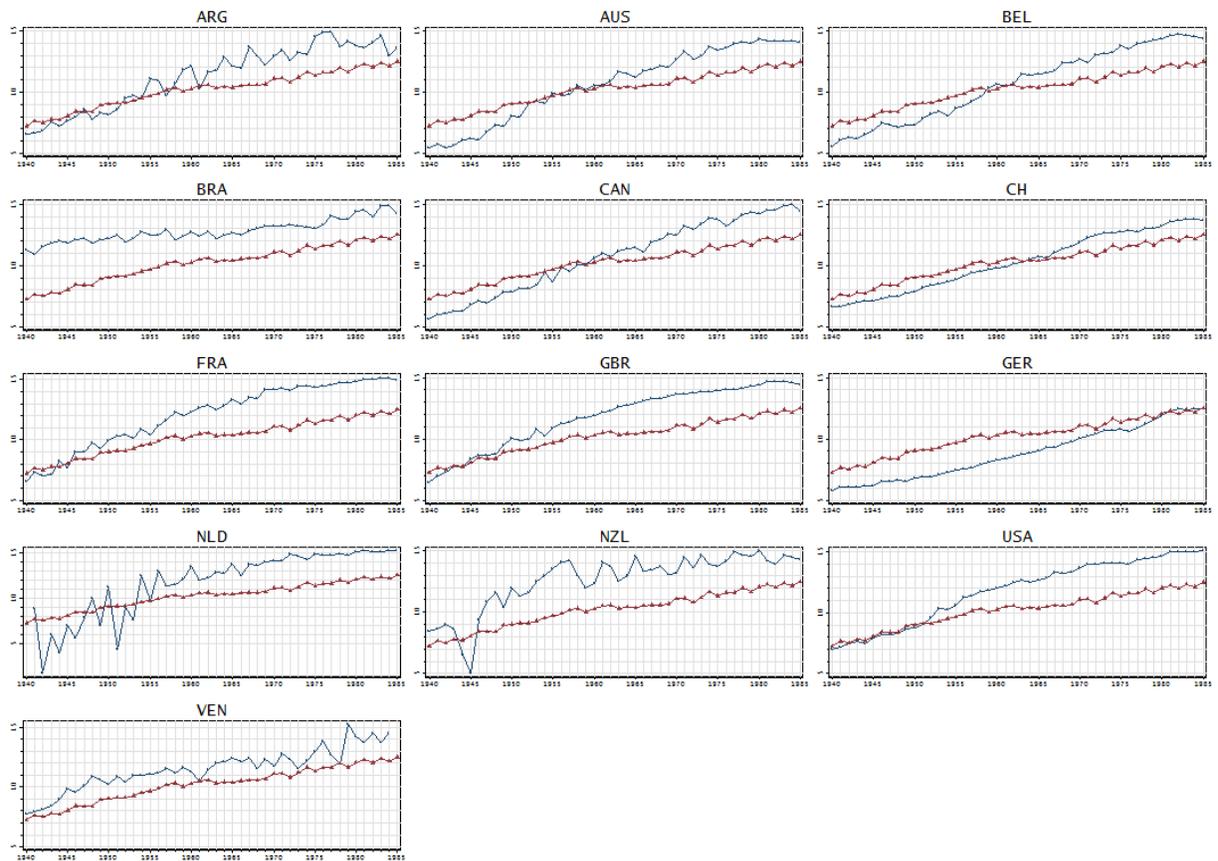
Source: Averages for Mover are own calculations using AIRE, averages for Stayer are own calculations using SHIW.

Figure 7 shows different patterns by country of destination, highlighting some interesting contrasts. Italians living in countries like Brazil, Great-Britain, New-Zealand, and the US are almost constantly more educated than their peers living in Italy, while the average education of Italian immigrants of all ages in Germany is lower than the average education of stayer.

4.2 Income Redistribution and Self-Selection of Immigrants

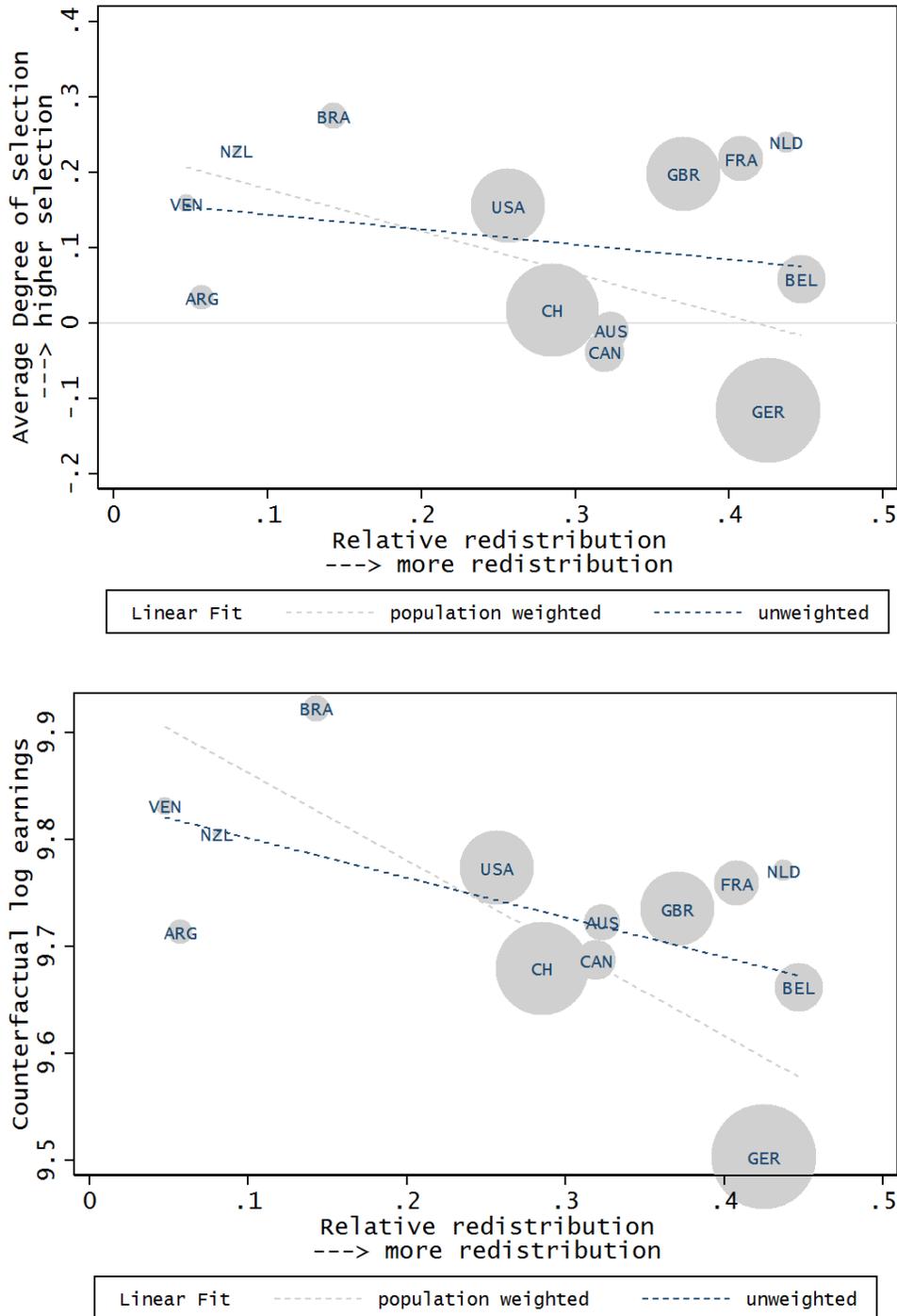
Figure 8 show the result of a first, stylized cross-country analysis on the relationship between redistribution and the degree of self-selection of immigrants. The average amount of relative redistribution over all available years in SWIID is plotted in relation to the average degree of selection of Italian immigrants residing in that particular country. In the upper part of the figure, selection is measured by the relative educational position, in the lower part by the predicted counterfactual log earnings.

Figure 7: Average Years of Education of Italian Mover and Stayer by Year of Birth and Country of Residence. Italians of all regional origins.

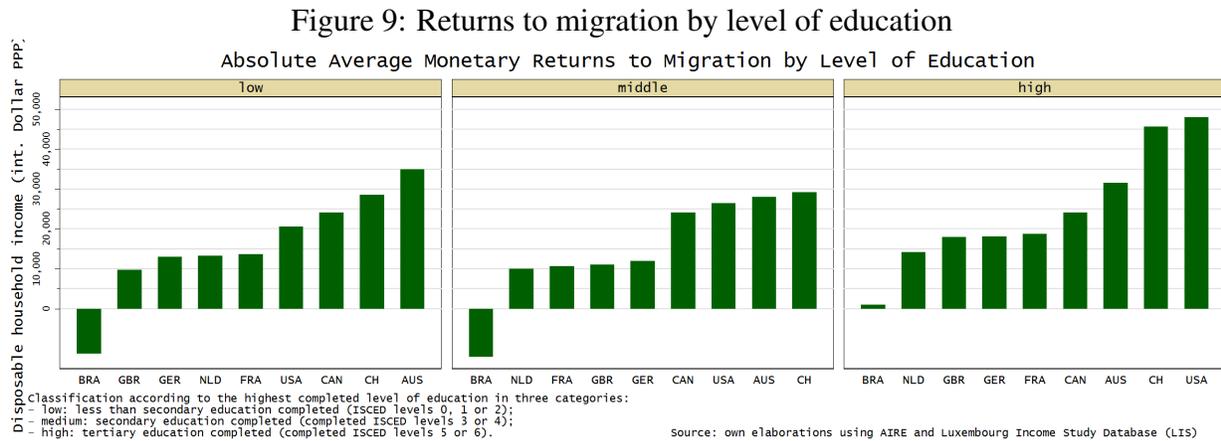


Source: Averages for Mover are own calculations using AIRE, averages for Stayer are own calculations using SHIW.

Figure 8: Returns to skills and self-selection of immigrants



Notes: All variables are averages over the complete observation period. Sources: AIRE, SWIID, own calculations.



Both measures show the same pattern of association between returns to skills, approximated by the level of relative redistribution, and the degree of selection of immigrants. The population weighted correlations are -0.38 and -0.65 respectively. Hence, this stylized analysis shows the first evidence in favour of the Roy-Model: Countries with higher returns to skills seem to attract immigrants from the upper part of the skill distribution of their home country.

4.3 Net Returns to Migration

Figure 9 shows the average yearly net returns to migration estimated with LIS data for the Italian migrants in our sample. Net returns are hereby measured in international US-Dollars using purchasing power parity at 2011 prices. The analysis is restricted to the nine countries for which recent harmonized survey data is available. Furthermore, we assume that the disposable incomes of Italian immigrants in the destination countries are not substantially different from the incomes of natives and other immigrants.

We observe that the predicted monetary returns for highly educated migrants are around 4,000 Dollars per month in the US and Switzerland, and substantially lower in the other countries. In contrast, low educated Italian migrants in Australia have about 2,900 Dollars more at disposition than they would have in their home country. In the US, highly educated Italian migrants earn more than twice as much as low educated migrants. Net returns of Italian migrants in the four EU

countries in our sample is rather similar between 800 and 1,600 Dollars per month. Migration to Brazil has on average only positive returns for highly educated migrants.

In the next exercise, we bring the Roy Model of self-selection to our data in a graphical representation. Figure 10 plots the predicted net income in the host country on the y-axis, and the predicted counterfactual log labour earnings in Italy on the x-axis. The slope of the single curves is indicative for the degree of returns to skills of each country. As expected, the curves are rather steep for the US and Brazil, which are countries with high returns to skills and relatively low levels of redistribution. The curves of countries like Germany, Canada, and the Netherlands show a lower gradient, comparable to the slope of Italy. It is conspicuous that for all countries but Brazil individuals with the lowest level of skills have lower net incomes in Italy than in any other destination country.

5 Empirical Set-Up and Results

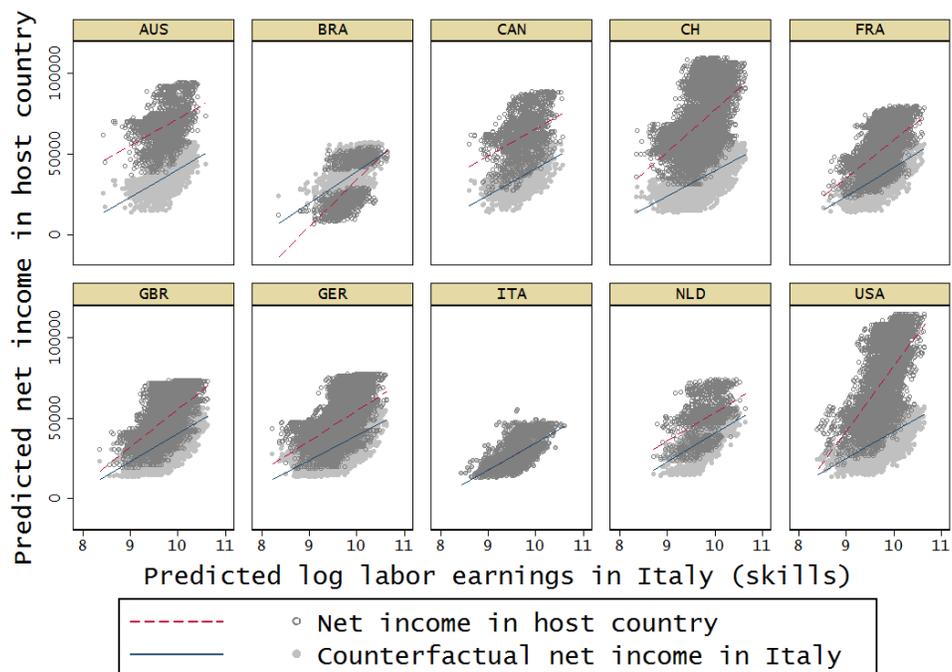
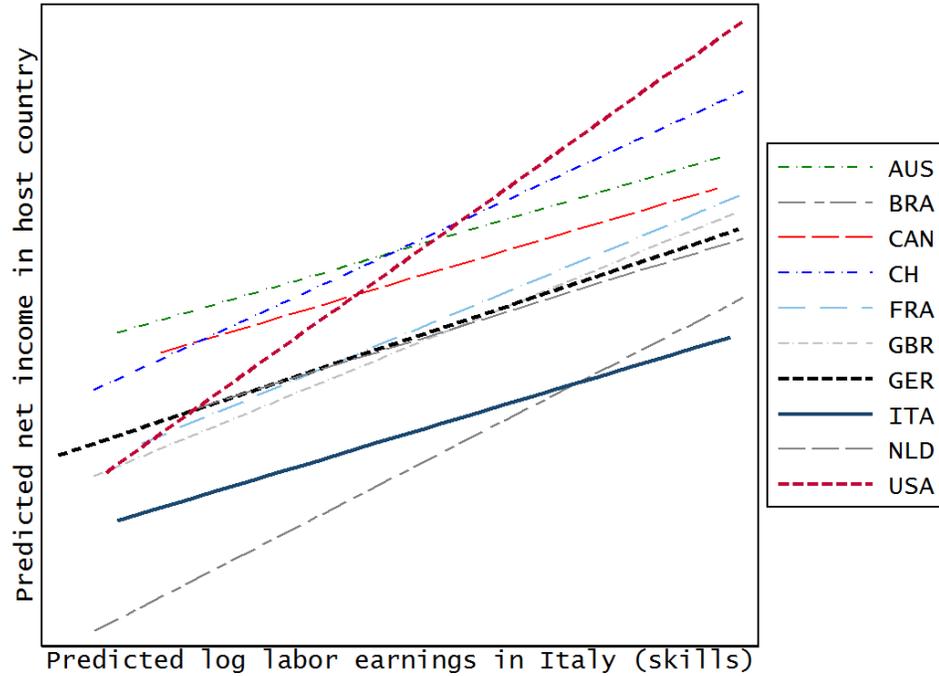
5.1 Returns to skills and degree of selection

To measure the association between the level of redistribution as a proxy measure for the returns to skills and the self-selection of immigrants, we run the following linear regression:

$$s_{ijt}^* = \alpha + \beta T_{tc} + \delta' Z_{tc} + \gamma' X_{ijt} + \zeta' M_c + \lambda_j + \tau_t + \varphi_c + \varepsilon_{ijt} \quad (6)$$

s^* are the relative skills measured by the predicted counterfactual log labour earnings that individual i from the Italian region of origin j who registered in year t to the registry of Italians living abroad in country c would have obtained had he stayed in Italy. T are the returns to skills in country c in year t . The coefficient β shows whether returns to skills are associated with the relative skill level of immigrants'. In our main specification, we measure returns to skills by the level of relative redistribution. To smooth the time variation and avoid bias resulting from temporary shocks and measurement error in the inequality measures, the value associated to each year t is the average from t to the last year with available information (mostly 2015). The assumption behind this pro-

Figure 10: Self selection of Italian immigrants



cedure is furthermore that rational agents take future developments into account when choosing to stay in the foreign country of residence.⁴

The same procedure is applied to the other macroeconomic variables in Z . Z is a vector of controls for country characteristics that vary by t : i) unemployment rate, ii) GDP-growth, iii) GDP per capita. These variables are indicators that may shape the expectations of individuals willing to migrate about the overall conditions in a particular country. X is a vector of controls for individual characteristics that have not been used for the prediction of the counterfactual income: year of birth (polynomial of second degree), month of birth, year of arrival of the individual and the first immigrated member of the same household, a dummy indicating whether the individual was the first household member to arrive in the country of destination, rural or urban location of origin (definition: < 150 inhabitants/km²), and a dummy variable indicating if the individual lives in the capital of the host country. M is a vector of controls for non-varying country characteristics that act as proxies for the costs of migrating to a particular country: i) distance from the Italian border, ii) a dummy signaling whether this country and Italy signed an immigration agreement and iii) the share of peers from the same Italian province residing in that particular country. λ_j , τ_t and φ_c are fixed effects for the Italian region of origin, the year of arrival and country of destination. We restrict the coefficients of the control variables to be zero in some of the estimations. Standard errors are clustered at the country-year level.

Table 3 shows the results of estimating equation 6 on the sample of Italian immigrants worldwide. The dependent variable are the predicted counterfactual log labour earnings in Italy as measure of skills. The coefficient of the variable that indicates the relative level of redistribution in the country of destination is negative and highly significant in all specification of the model. For instance, our most conservative estimates obtained including country fixed effects in the regressions show that an increase of relative redistribution by 10 percentage points is associated to a 5 percent decrease in the degree of self-selection in education of Italian immigrants. Hence, returns to skills, measured here by the amount of redistribution, seems to be associated with the degree of immig-

⁴Results associating the macroeconomic variables only in the year of arrival can be found in the Appendix.

Table 3: Returns to skills and degree of selection

	(1)	(2)	(3)	(4)
Relative Redistribution	-0.861*** (0.0866)	-0.580*** (0.0657)	-0.682*** (0.0716)	-0.549* (0.316)
Share of people from the same province in country of residence		-0.339*** (0.0412)	-0.328*** (0.0407)	-0.245*** (0.0459)
Other family members migrated earlier (0/1)		-0.187*** (0.00791)	-0.185*** (0.00776)	-0.181*** (0.00774)
Rural place of origin (0/1)		-0.128*** (0.00506)	-0.127*** (0.00496)	-0.122*** (0.00481)
Resident in the capital (0/1)		0.0938*** (0.00552)	0.0907*** (0.00559)	0.0741*** (0.00489)
Distance of country of residence from Italian border (in 1000 km)		0.00148* (0.000796)	-0.000795 (0.00101)	
Migration agreement between Italy and country of residence (0/1)		0.00198 (0.00598)	0.0118* (0.00655)	
Migration policies oriented towards high skilled (0/1)		0.0158* (0.00857)	0.0174* (0.00999)	
Unemployment rate			0.0123*** (0.00215)	-0.00349 (0.00279)
GDP growth			0.0181*** (0.00630)	0.0177*** (0.00449)
GDP per capita			0.000623*** (0.000209)	0.00457*** (0.000597)
Constant	9.962*** (0.0288)	-2059.8*** (86.78)	-2072.5*** (87.87)	-2046.4*** (86.38)
Demographic controls	No	Yes	Yes	Yes
Country F.E.	No	No	No	Yes
Observations	218236	217571	217571	218236
R ²	0.038	0.250	0.252	0.259

Dependent variable is the predicted counterfactual log labour earnings in Italy as measure of skills.

rants' selectivity, as well as their likelihood to be positively self-selected from their population of origin.

Two variables capturing so called network effects show the expected negative relationship with the selectivity of immigrants. Past research argued that networks lower the cost of migration making it more attractive for low skilled individuals to migrate, and hence lowering, on average, the pattern of positive self-selection of particular immigrant groups (e.g. [Mckenzie and Rapoport, 2007](#); [McKenzie and Rapoport, 2010](#)). Our results confirm this findings. The presence of people from the same province of origin is negatively and significantly associated with the skill level of migrants.

Moreover, the presence of a family member in the country of residence is associated with around 20 percent lower degree of skill selection. We observe furthermore that Italian immigrants originating from rural places have a lower degree of selection, while those residing in the capitals of their country of destination are more likely to be positively selected.

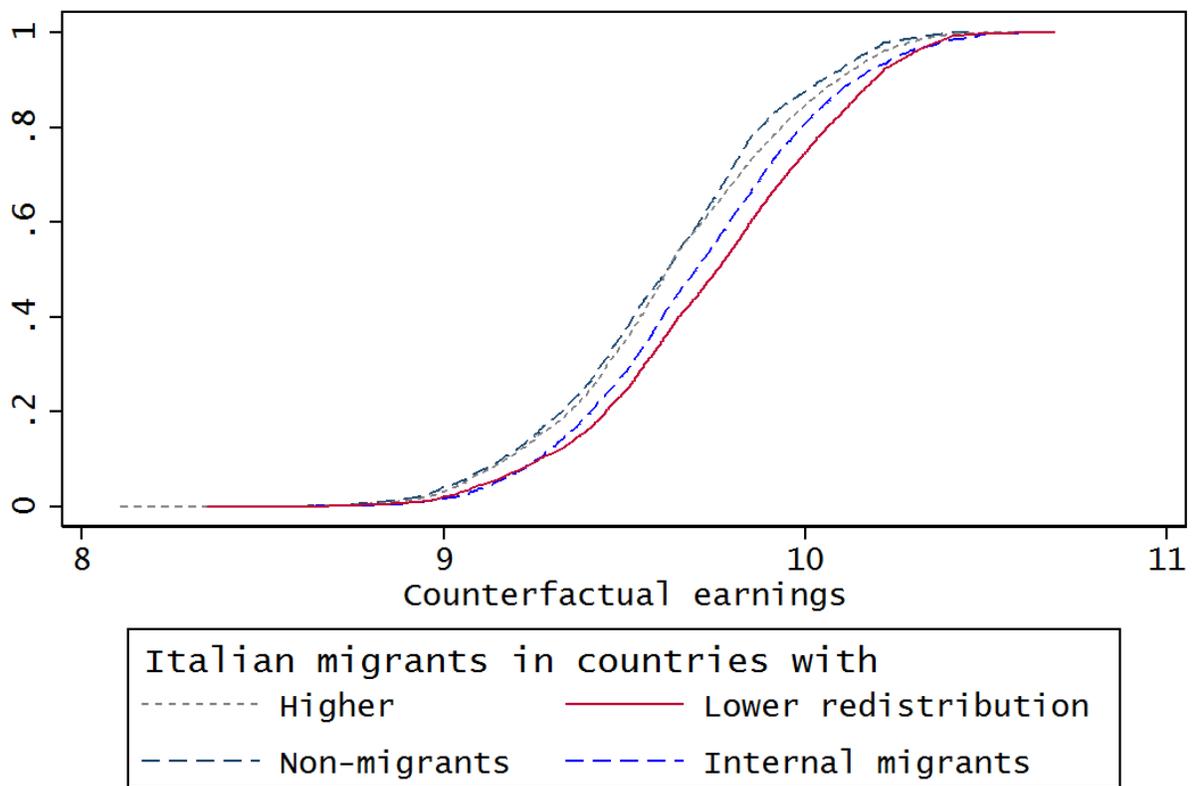
Among country characteristics, the distance of the country of destination from Italy is very weakly associated with selectivity. The same applies to the presence of bilateral migration agreements between Italy and the country of destination. Last, unemployment, growth and GDP per capita in the country of destination are positively associated with the degree of self-selection of Italian immigrants.

5.2 Stochastic Dominance

The regression coefficients of relative redistribution shown in Table 3 measure the average association between returns to skills and the self-selection of immigrants. However, as shown by [Borjas et al. \(2018\)](#) the Roy-Model of self selection also implies a first order stochastic dominance relationship between the distributions of mover and stayer. We test these predictions using our administrative data on Italian immigrants and the survey data on Italians in Italy. As argued for instance by [Chiquiar and Hanson \(2005\)](#), the complete group of stayer might not be a suitable comparison group for mover because of the selection on unobserved characteristics that shapes the income distributions as well. Hence, we subdivide the group of stayers in two separate groups: i) internal migrants, i.e. individuals that migrated within Italy, and ii) non-migrants, i.e. individuals that reside in their region of birth. Mover, i.e. individuals in our administrative data, are also divided in two groups according to the amount of redistribution they experienced in the country of destination in comparison to the redistribution in Italy: i) Italian migrants in countries with higher redistribution, and ii) Italian migrants in countries with lower redistribution. Figure 11 plots the cumulative distributions of the predicted counterfactual earnings in Italy of these four groups.

Indeed, this test confirms again the hypotheses of the Roy-Model of self-selection. The distribution of skills of migrants in countries with lower redistribution dominates the distribution of

Figure 11: Cumulative Distributions



stayer and migrants in countries with higher redistribution. At the same time, the distribution of immigrants in countries with higher redistribution is rather similar to the distribution of non-migrants, but stochastically dominated by the distribution of internal migrants. A Kolmogorov-Smirnov test of the equality of distributions shows that all the differences between the distributions of the four groups visualized in Figure 11 are statistically significant.

5.3 Selection on unobservable skills

To test the relationship between returns to skills and the selection on unobservable skills of migrants (e.g. abilities or motivation) we analyse if the likelihood to attain certain occupations changes with the level of redistribution holding observable skills constant. We apply a Probit model on a binary variable indicating the occupation status h of individual i who registered in year t to the registry of Italians living abroad in country c . We adopt two different specifications for h : i) one if the individual is unemployed or inactive and zero if in an employment situation, ii) one if the individual is an executive or manager and zero if unemployed or in another type of occupation. Furthermore, educational attainment y^* are included in the equation as binary variable that is one if the individual attained beyond compulsory education and zero otherwise.

The following model is estimated where the level of redistribution is interacted with educational attainment:

$$Prob(h_{ijtc} = 1) = \Phi(\vartheta T_{tc} \cdot y_{ijtc}^* + \beta T_{tc} + \theta y_{ijtc}^* + \gamma' X_{ijtc} + \delta' Z_{tc}) \quad (7)$$

Individual level covariates are included in X , country level covariates in Z . The marginal effect of the interaction term ϑ computed at different values of T shows how the likelihood to be unemployed or to have a high occupational status (executive or manager) varies with returns to skills for immigrants with high and low education, respectively.

While it is safe to assume that educational achievements is a (quasi) time-invariant characteristic in adulthood, this does not apply to occupation. Hence, we have to restrict our sample further to people available to the labour market. For instance, as Figure 12 highlights, there are substan-

tial cross-country difference in female labour participation across countries. To avoid that these differences affect our estimates we restrict this part of the analysis on male immigrants.

Table 4 shows the estimated coefficients of the Probit models. In the first and third columns of both specifications of the dependent variable the coefficient of the interaction term is restricted to zero. The marginal effects of the interaction terms for different levels of relative redistribution are plotted in Figure 4. Also in this application, our findings confirm the hypothesis of the Roy-Modell, pointing at a positive relationship between returns to skills and the self-selection of immigrants on unobservable characteristics. First, controlling for education, redistribution is associated with a higher likelihood of being unemployed or inactive and a lower likelihood to be an executive or manager. However, in the case of the latter dependent variable, including country fixed effects the coefficient is not statistically significant from zero in both applications. The reason for this could be that the largest variation in the level of redistribution takes place between countries, while within countries this variable changes only marginally.

The evidence in favour of the Roy-Modell is reinforced looking at the likelihoods to attain certain occupation types for individuals with different educational attainments. As is evident from the marginal effects shown in in Figure 4, the likelihood to be unemployed is significantly lower for individuals with higher educational attainments. Interestingly, this likelihood rises for people regardless of their educational level with rising redistribution and the difference in the likelihood between educational levels is higher the higher is the level of redistribution. The same pattern is observed for the likelihood to be executive or manager. Especially among individuals with higher educational levels the likelihood to be executive or manager is substantially higher when the level of redistribution is rather low.

All evidence so far confirms the hypothesis of the Roy-Modell: higher returns to skills attract immigrants with higher observable and unobservable skills.

Figure 12: Occupation and Education of Italian immigrants by country

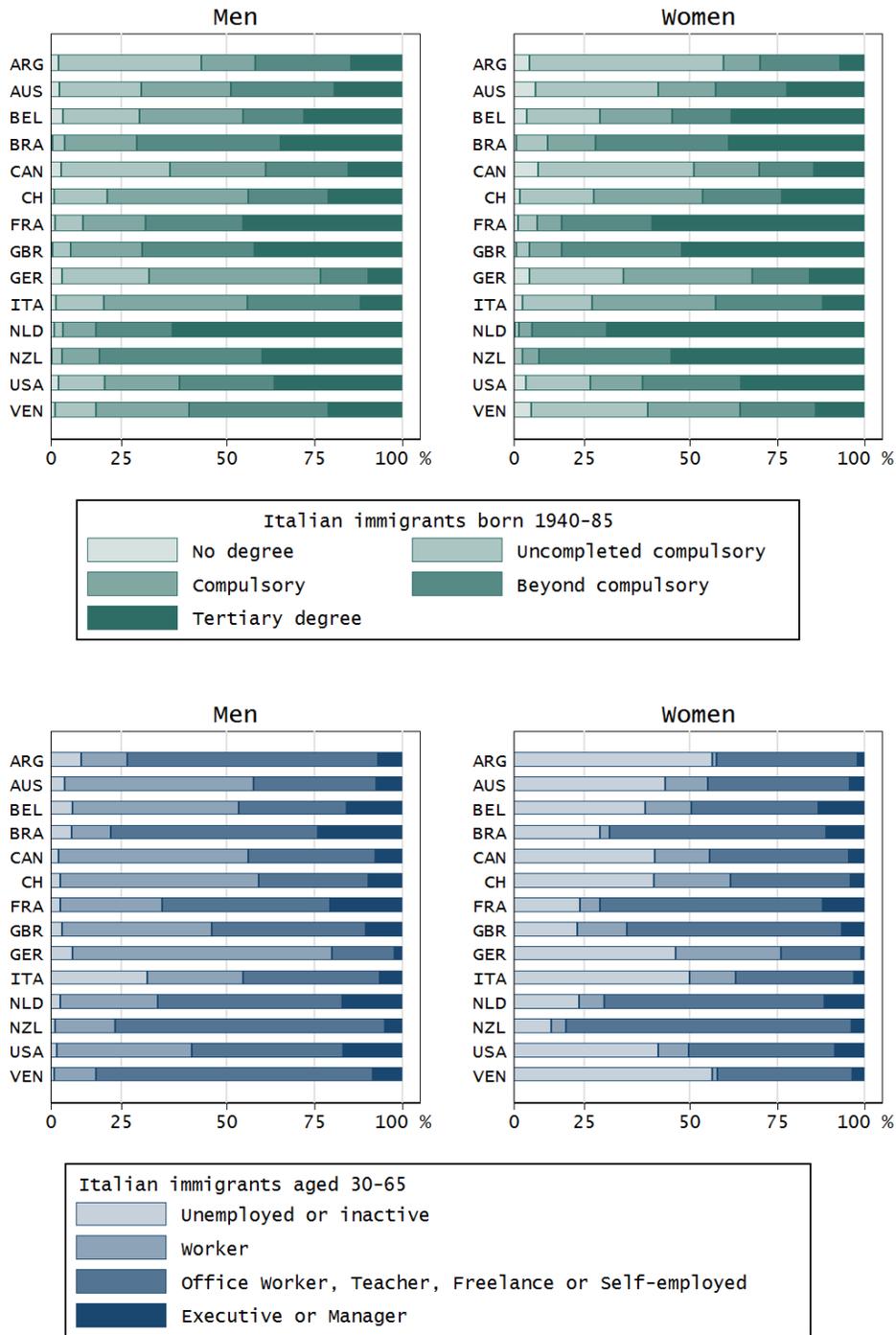
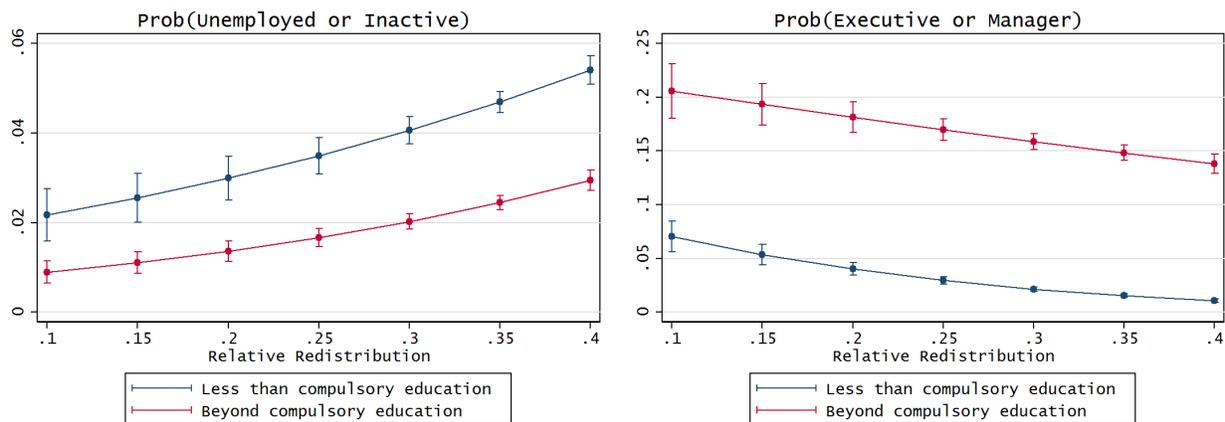


Table 4: Selection on unobservable skills

	Prob(Unemployed)				Prob(Executive or Manager)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Probit estimates								
Beyond compulsory education (0/1)	-0.299*** (0.0193)	-0.380*** (0.0867)	-0.288*** (0.0192)	-0.415*** (0.0870)	1.184*** (0.0237)	0.507*** (0.0805)	1.163*** (0.0233)	0.657*** (0.0847)
Relative Redistribution	1.522*** (0.183)	1.416*** (0.221)	4.651*** (1.548)	4.513*** (1.553)	-1.284*** (0.196)	-3.015*** (0.239)	1.944 (1.324)	0.857 (1.365)
Beyond compulsory · Relative Redistribution		0.223 (0.225)		0.352 (0.229)		2.041*** (0.246)		1.545*** (0.270)
Share of people from the same province in country of residence	0.233*** (0.0573)	0.242*** (0.0586)	0.134** (0.0676)	0.139** (0.0676)	-0.771*** (0.0620)	-0.740*** (0.0612)	-0.469*** (0.0635)	-0.462*** (0.0638)
Other family members migrated earlier (0/1)	0.0589* (0.0318)	0.0585* (0.0318)	0.0590* (0.0318)	0.0588* (0.0319)	-0.0661 (0.0456)	-0.0689 (0.0456)	-0.101** (0.0452)	-0.104** (0.0453)
Rural place of origin (0/1)	-0.0752*** (0.0165)	-0.0754*** (0.0165)	-0.0815*** (0.0166)	-0.0817*** (0.0166)	-0.192*** (0.0179)	-0.194*** (0.0180)	-0.185*** (0.0181)	-0.186*** (0.0181)
Internal migration experience before emigration (0/1)	0.0178 (0.0232)	0.0180 (0.0232)	0.0151 (0.0235)	0.0152 (0.0235)	0.0727*** (0.0158)	0.0739*** (0.0158)	0.0761*** (0.0159)	0.0769*** (0.0160)
Resident in the capital (0/1)	-0.0450 (0.0363)	-0.0483 (0.0367)	-0.0166 (0.0384)	-0.0192 (0.0385)	0.0389 (0.0247)	0.0253 (0.0247)	0.118*** (0.0241)	0.111*** (0.0241)
Unemployment rate	-0.0340*** (0.0115)	-0.0332*** (0.0115)	0.00630 (0.0160)	0.00720 (0.0161)	0.0669*** (0.0122)	0.0682*** (0.0122)	-0.00495 (0.0130)	-0.00185 (0.0131)
GDP growth	-0.0137 (0.0210)	-0.0133 (0.0212)	-0.0183 (0.0238)	-0.0185 (0.0241)	-0.109*** (0.0294)	-0.106*** (0.0296)	0.00377 (0.0250)	0.00236 (0.0251)
GDP per capita	-0.0113*** (0.00112)	-0.0114*** (0.00118)	-0.0125*** (0.00328)	-0.0117*** (0.00334)	0.00573*** (0.00154)	0.00571*** (0.00154)	-0.00490 (0.00336)	-0.00325 (0.00340)
Constant	33.09*** (2.768)	33.17*** (2.783)	32.93*** (2.762)	33.06*** (2.777)	60.01*** (2.717)	61.01*** (2.797)	58.67*** (2.783)	59.45*** (2.848)
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country F.E.	No	No	Yes	Yes	No	No	Yes	Yes
Observations	106214	106214	106214	106214	106137	106137	106137	106137
Pseudo R ²	0.056	0.057	0.060	0.060	0.234	0.235	0.250	0.251

Figure 13: Selection on unobservable skills



Dots show the predicted probabilities for different levels of relative redistribution. Left figure shows the marginal effects of the interaction term in model (2) on Table 4, right figure shows the marginal effects of the interaction term in model (6).

5.4 Monetary Returns and the Choice of Migration

In this step of the analysis we estimate a discrete choice model of the decision to migrate as a function of personal characteristics and place characteristics of the country of destination. These sort of models are usually applied to estimate the determinants of the migration decision (e.g. [Davies et al., 2001](#); [O’Keefe, 2004](#); [Vigdor, 2002](#)). To test the Roy-Model of income maximization, our main interest lies in estimating if net income returns determine the choice of migration. A similar set up has been adopted, for instance, by [Grogger and Hanson \(2011\)](#) on aggregate data to test the selection and sorting of international migrants.

We pool our administrative data on Italian migrants with the Italian survey data on stayer and run an alternative specific conditional logit model ([McFadden, 1974](#)). The explanatory variables are either alternative specific or case specific. The former vary among countries and individuals, while the latter only among individuals. The model is motivated by a random utility framework, which models the potential utility of migrant i in country $c = 1, \dots, C$ as a function of the obtainable net income returns Y^{net} in this country, which vary for each individual depending on his or her level of education and other individual characteristics, and some other country specific characteristics that vary for each individual depending on his or her year of migration or province of origin. The model can be expressed as

$$U_{ic} = \gamma'Z_{ic} + \alpha'_c A_{ic} + \varepsilon_{ic} \quad (8)$$

where the vector U_i quantifies the utility associated to the potential choice of each alternative country of destination, including the possible decision to stay in Italy. The country actually chosen by i is the one that maximizes his or her utility. The vector Z includes Y^{net} as well as the other alternative specific characteristics. A is a vector that contains dummy variables for each country and individual specific characteristics that do not change across alternatives. These must be interacted with each potential choice, yielding coefficients α_c for each potential country of destination. Hereby, we must set one of the countries as the baseline alternative, setting $\alpha_k = 0$ for this baseline

country k . We set Italy – i.e. the choice to stay – as baseline when the full decision set is evaluated, and Switzerland when the estimation are run just on the sample of mover. ε_{ic} is the random component, which is assumed to be independently and identically distributed, with an extreme-value distribution. Under this assumption the probability that i chooses destination country c is

$$Prob(D_{ic} = 1) = \frac{e^{\gamma'Z_{ic} + \alpha'_c A_{ic}}}{\sum_{j=1}^C e^{\gamma'Z_{ij} + \alpha'_c A_{ij}}}. \quad (9)$$

Where D_{ic} is an indicator of i 's decision regarding the country of destination. Each individual chooses among the 10 countries in the choice set, including Italy. Hence, the dataset is expanded to encompass 10 observations for each individual where D is equal to one if c is the actual country of residence and zero otherwise. Z captures the circumstances the individual faces in the actual country of residence and that he or she would face in the other potential destinations, including disposable household income. Individual level control variables included in A are indicators for age, sex and the Italian geographic region of origin. The model is estimated on the whole sample, as well as separately for each education group.

Table 5 shows the estimated coefficients of the conditional logit model including only predicted disposable income as alternative specific variable. Table 6 and Table 7 show the associated marginal effects of net income returns on the choice of the country of destination; the first including Italy in the choice set, the second excluding it. We observe that the coefficient of predicted net income is positive and highly significant. This pattern holds excluding Italy among the possible destinations and hence focusing on the population of migrants. The marginal effects show that a yearly net income raise by 10,000 international Dollars PPP in Italy rises the probability of low and high educated individuals to stay in Italy by around 2 and 6 percent, respectively. To give another example, the average yearly returns to migration of high educated Italian immigrants in Switzerland, 45,000 USD, are associated to a higher likelihood by 5.4 percent to leave Italy and move to Switzerland. These probabilities are substantially lower for low educated individuals.

Table 8 shows the estimated coefficients of the conditional logit model including the full set of alternative specific control variables and excluding Italy as possible destination. The coefficient of

Table 5: Conditional Logit Estimates I

Education Level	w/ Italy				w/o Italy			
	All	Low	Middle	High	All	Low	Middle	High
Choice: Migration and Destination Country								
Predicted Net Income (absolute)	1.316*** (0.0361)	5.041*** (0.270)	4.426*** (0.207)	3.977*** (0.286)	0.336*** (0.00697)	0.242*** (0.0387)	0.383*** (0.0278)	0.246*** (0.0248)
Log-lik.	-1303369.9	-420752.0	-369768.2	-396590.7	-332271.9	-94933.0	-98755.8	-123248.2
Observations	1519960	558070	437330	524560	1326195	481896	378963	465336
Cases	151996	55807	43733	52456	147355	53544	42107	51704
Alternatives	10	10	10	10	9	9	9	9

Probability to stay in Italy or chose one of the 9 destination countries as a function of predicted net income returns to migration.

net income is positive and significant. Furthermore, the probability to chose a country over the other is positively associated with language relatedness, GDP per capita, and relative redistribution, and negatively with the distance of the country from the Italian border. The effect of the unemployment rate is not uniform across education groups, being positive for low and high educated individuals and negative for individuals with compulsory education. The positive association between the share of migrants from the same province of origin and the likelihood to reside in a particular country is merely mechanical and serves here just as control variable.

Overall, the results of this exercise again confirm the predictions of the Roy-Model: income maximization is a substantial determinant of the migration decision.

6 Conclusions

In this study we tested the predictions of the Roy-Model about the self-selection of immigrants using an administrative dataset including about 90 % of Italians living abroad. Our results confirm the predictions of the model: high returns to human capital, which we measure by the degree of income redistribution, are significantly associated with a positive skill selection of Italian immigrants. On the contrary, countries with stronger redistribution seem to attract immigrants from the lower part of the skill distribution. This patterns are confirmed by several distinct exercises and test procedures.

Currently, the coordination of migration policies is a widely debated topic. However, the discussion is mostly focused on migration quotas taking less into account that particular country char-

Table 6: Marginal effects of predicted net income returns on the probability of migration (including Italy in the choice set) – Yearly disposable household income

Education: Low											
	ITA	AUS	BRA	CAN	CH	FRA	GBR	GER	NLD	USA	.
ITA	0.0191	-0.0001	-0.0002	-0.0000	-0.0049	-0.0005	-0.0026	-0.0104	-0.0000	-0.0003	
AUS	-0.0001	0.0001	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	
BRA	-0.0002	-0.0000	0.0002	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	
CAN	-0.0000	-0.0000	-0.0000	0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	
CH	-0.0049	-0.0000	-0.0000	-0.0000	0.0049	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	
FRA	-0.0005	-0.0000	-0.0000	-0.0000	-0.0000	0.0005	-0.0000	-0.0000	-0.0000	-0.0000	
GBR	-0.0026	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	0.0026	-0.0000	-0.0000	-0.0000	
GER	-0.0104	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	0.0104	-0.0000	-0.0000	
NLD	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	0.0000	-0.0000	
USA	-0.0003	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	0.0003	
mean											0.0038
Education: Middle											
	ITA	AUS	BRA	CAN	CH	FRA	GBR	GER	NLD	USA	.
ITA	0.0248	-0.0005	-0.0006	-0.0001	-0.0065	-0.0019	-0.0079	-0.0052	-0.0002	-0.0018	
AUS	-0.0005	0.0005	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	
BRA	-0.0006	-0.0000	0.0006	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	
CAN	-0.0001	-0.0000	-0.0000	0.0001	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	
CH	-0.0065	-0.0000	-0.0000	-0.0000	0.0065	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	
FRA	-0.0019	-0.0000	-0.0000	-0.0000	-0.0000	0.0019	-0.0000	-0.0000	-0.0000	-0.0000	
GBR	-0.0079	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	0.0079	-0.0000	-0.0000	-0.0000	
GER	-0.0052	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	0.0053	-0.0000	-0.0000	
NLD	-0.0002	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	0.0002	-0.0000	
USA	-0.0018	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	0.0019	
mean											0.0050
Education: High											
	ITA	AUS	BRA	CAN	CH	FRA	GBR	GER	NLD	USA	.
ITA	0.0584	-0.0009	-0.0011	-0.0002	-0.0119	-0.0078	-0.0203	-0.0089	-0.0012	-0.0061	
AUS	-0.0009	0.0009	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	
BRA	-0.0011	-0.0000	0.0011	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	
CAN	-0.0002	-0.0000	-0.0000	0.0002	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	
CH	-0.0119	-0.0000	-0.0000	-0.0000	0.0120	-0.0000	-0.0001	-0.0000	-0.0000	-0.0000	
FRA	-0.0078	-0.0000	-0.0000	-0.0000	-0.0000	0.0079	-0.0000	-0.0000	-0.0000	-0.0000	
GBR	-0.0203	-0.0000	-0.0000	-0.0000	-0.0001	-0.0000	0.0205	-0.0000	-0.0000	-0.0000	
GER	-0.0089	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	0.0091	-0.0000	-0.0000	
NLD	-0.0012	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	0.0012	-0.0000	
USA	-0.0061	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	0.0062	
mean											0.0117

Marginal effect of income rise in country A (rows) on Probability to migrate to country B (columns). Diagonal shows the marginal effects of a net income rise in the country on the probability to stay/migrate in/to this country. Panel A shows the estimates of absolute income, one unit is 10,000 international USD PPP. Panel B shows the estimates of log income.

Table 7: Marginal effects of predicted net income returns on the probability of migration (excluding Italy from the choice set) – Yearly disposable household income

Education: Low										
	AUS	BRA	CAN	CH	FRA	GBR	GER	NLD	USA	.
AUS	0.0039	-0.0000	-0.0001	-0.0010	-0.0001	-0.0003	-0.0021	-0.0000	-0.0003	
BRA	-0.0000	0.0013	-0.0000	-0.0003	-0.0000	-0.0001	-0.0007	-0.0000	-0.0001	
CAN	-0.0001	-0.0000	0.0040	-0.0011	-0.0001	-0.0003	-0.0022	-0.0000	-0.0003	
CH	-0.0010	-0.0003	-0.0011	0.0463	-0.0014	-0.0048	-0.0331	-0.0002	-0.0044	
FRA	-0.0001	-0.0000	-0.0001	-0.0014	0.0054	-0.0004	-0.0029	-0.0000	-0.0004	
GBR	-0.0003	-0.0001	-0.0003	-0.0048	-0.0004	0.0172	-0.0099	-0.0001	-0.0013	
GER	-0.0021	-0.0007	-0.0022	-0.0331	-0.0029	-0.0099	0.0602	-0.0004	-0.0090	
NLD	-0.0000	-0.0000	-0.0000	-0.0002	-0.0000	-0.0001	-0.0004	0.0007	-0.0001	
USA	-0.0003	-0.0001	-0.0003	-0.0044	-0.0004	-0.0013	-0.0090	-0.0001	0.0158	
mean										0.0172
Education: Middle										
	AUS	BRA	CAN	CH	FRA	GBR	GER	NLD	USA	.
AUS	0.0123	-0.0003	-0.0003	-0.0030	-0.0008	-0.0031	-0.0025	-0.0001	-0.0021	
BRA	-0.0003	0.0076	-0.0002	-0.0018	-0.0005	-0.0019	-0.0015	-0.0001	-0.0013	
CAN	-0.0003	-0.0002	0.0100	-0.0024	-0.0007	-0.0025	-0.0020	-0.0001	-0.0017	
CH	-0.0030	-0.0018	-0.0024	0.0692	-0.0060	-0.0225	-0.0176	-0.0011	-0.0148	
FRA	-0.0008	-0.0005	-0.0007	-0.0060	0.0238	-0.0063	-0.0049	-0.0003	-0.0042	
GBR	-0.0031	-0.0019	-0.0025	-0.0225	-0.0063	0.0714	-0.0184	-0.0011	-0.0155	
GER	-0.0025	-0.0015	-0.0020	-0.0176	-0.0049	-0.0184	0.0598	-0.0009	-0.0121	
NLD	-0.0001	-0.0001	-0.0001	-0.0011	-0.0003	-0.0011	-0.0009	0.0044	-0.0007	
USA	-0.0021	-0.0013	-0.0017	-0.0148	-0.0042	-0.0155	-0.0121	-0.0007	0.0523	
mean										0.0345
Education: High										
	AUS	BRA	CAN	CH	FRA	GBR	GER	NLD	USA	.
AUS	0.0057	-0.0001	-0.0001	-0.0011	-0.0006	-0.0016	-0.0008	-0.0001	-0.0011	
BRA	-0.0001	0.0038	-0.0001	-0.0007	-0.0004	-0.0011	-0.0005	-0.0001	-0.0008	
CAN	-0.0001	-0.0001	0.0053	-0.0010	-0.0006	-0.0015	-0.0008	-0.0001	-0.0011	
CH	-0.0011	-0.0007	-0.0010	0.0380	-0.0051	-0.0130	-0.0066	-0.0012	-0.0092	
FRA	-0.0006	-0.0004	-0.0006	-0.0051	0.0238	-0.0074	-0.0038	-0.0007	-0.0052	
GBR	-0.0016	-0.0011	-0.0015	-0.0130	-0.0074	0.0491	-0.0096	-0.0017	-0.0133	
GER	-0.0008	-0.0005	-0.0008	-0.0066	-0.0038	-0.0096	0.0298	-0.0009	-0.0068	
NLD	-0.0001	-0.0001	-0.0001	-0.0012	-0.0007	-0.0017	-0.0009	0.0060	-0.0012	
USA	-0.0011	-0.0008	-0.0011	-0.0092	-0.0052	-0.0133	-0.0068	-0.0012	0.0388	
mean										0.0222

Marginal effect of income rise in country A (rows) on Probability to migrate to country B (columns). Diagonal shows the marginal effects of a net income rise in the country on the probability to stay/migrate in/to this country. Panel A shows the estimates of absolute income, one unit is 10,000 international USD PPP. Panel B shows the estimates of log income.

Table 8: Conditional Logit Estimates II

Education Level	w/o Country FE				w/ Country FE			
	All	Low	Middle	High	All	Low	Middle	High
Choice: Destination Country								
Predicted Net Income (absolute)	0.247*** (0.00705)	0.260*** (0.0387)	0.542*** (0.0281)	0.244*** (0.0239)	0.240*** (0.00730)	0.140*** (0.0397)	0.372*** (0.0286)	0.260*** (0.0255)
Share of migrants from the same province of origin	4.504*** (0.0278)	5.221*** (0.0470)	3.908*** (0.0516)	2.299*** (0.0594)	4.574*** (0.0281)	5.237*** (0.0472)	4.005*** (0.0523)	2.348*** (0.0595)
Distance of country of residence from Italian border (in 1000 km)	-0.000232*** (0.0000122)	-0.000409*** (0.0000274)	-0.000315*** (0.0000205)	-0.000329*** (0.0000204)				
Language relatedness	0.0301*** (0.00150)	0.0437*** (0.00306)	0.0443*** (0.00262)	0.0455*** (0.00273)				
Unemployment rate	0.121*** (0.00494)	0.0440*** (0.0114)	0.0158* (0.00906)	0.0813*** (0.00814)	0.0797*** (0.00597)	0.0308** (0.0142)	-0.0185* (0.0108)	0.0525*** (0.00938)
GDP per capita	0.0204*** (0.000969)	0.0271*** (0.00211)	0.0459*** (0.00176)	0.0472*** (0.00175)	0.0358*** (0.00137)	0.0340*** (0.00271)	0.0661*** (0.00254)	0.0607*** (0.00254)
Relative Redistribution	4.108*** (0.513)	0.959 (1.152)	2.225*** (0.838)	7.525*** (0.848)	1.060 (0.727)	2.288 (1.632)	3.490*** (1.282)	2.954** (1.171)
Log-lik.	-297802.2	-82072.5	-89848.1	-113327.1	-294794.9	-81821.3	-88868.2	-112757.5
Observations	1259214	472396	356697	430121	1259214	472396	356697	430121
Cases	141198	52624	40053	48521	141198	52624	40053	48521
Alternatives	9	9	9	9	9	9	9	9

Probability to chose one of the 9 destination countries as a function of predicted net income returns to migration and other country characteristics.

acteristics might attract certain types of immigrants. Our empirical results again show that migrants are likely to move to countries in order to maximize their income, and that the selection of immigrants substantially depends on the relative returns to skills. Since this patterns of selection seems to apply not only to observable skills, like education, but also to unobservable skills, policy strategies that establish migration quotas based on qualifications are not sufficient for policy makers aimed to attract immigrants from the top of the skill distribution.

One of the few ways in which the skill composition of migrants can be influenced by public policy, is through the tax and transfer system as a tool to change the returns to skills. As a consequence, countries have an incentive to reduce their redistributive effort to attract high skilled immigrants and discourage the immigration of low-skilled immigrants. This situation might lead to a race to the bottom where eventually most countries have lower than optimal levels of taxation and redistribution. An international coordination of income redistribution seems therefore necessary to face the free movement of individuals.

References

BEINE, M., F. DOCQUIER AND H. RAPOPORT, "Brain Drain and Human Capital Formation in Developing Countries: Winners and Losers," *The Economic Journal* 118 (2008), 631–652.

- BORJAS, G. J., "Self-Selection and the Earnings of Immigrants," *The American Economic Review* 77 (1987), 531–553.
- BORJAS, G. J., S. G. BRONARS AND S. J. TREJO, "Self-selection and internal migration in the United States," *Journal of Urban Economics* 32 (sep 1992), 159–185.
- BORJAS, G. J., I. KAUPPINEN AND P. POUTVAARA, "Self-selection of Emigrants: Theory and Evidence on Stochastic Dominance in Observable and Unobservable Characteristics," *The Economic Journal* (may 2018), 1–73.
- CHIQUIAR, D. AND G. HANSON, "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States," *Journal of Political Economy* 113 (2005), 239–281.
- CHISWICK, B. R., "Are Immigrants Favorably Self-Selected?," *American Economic Review Papers and Proceedings* 89 (1999), 181–185.
- DAVIES, P. S., M. J. GREENWOOD AND H. LI, "A Conditional Logit Approach to U.S. State-to-State Migration," *Journal of Regional Science* 41 (2001), 337–360.
- GROGGER, J. AND G. H. HANSON, "Income maximization and the selection and sorting of international migrants," *Journal of Development Economics* 95 (2011), 42–57.
- HATTON, T. J., "The economics of international migration: A short history of the debate," *Labour Economics* 30 (oct 2014), 43–50.
- JENKINS, S. P., "World income inequality databases: an assessment of WIID and SWIID," *Journal of Economic Inequality* 13 (2015), 629–671.
- LIEBIG, T. AND A. SOUSA-POZA, "Migration, self-selection and income inequality: An international analysis," *Kyklos* 57 (2004), 125–146.
- MCFADDEN, D., "Conditional logit analysis of qualitative choice behavior," *Frontiers in Econometrics* 1 (1974), 105–142.

- MCKENZIE, D. AND H. RAPOPORT, "Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico," *Journal of Development Economics* 84 (2007), 1–24.
- MCKENZIE, D. AND H. RAPOPORT, "Self-Selection Patterns in Mexico-U.S. Migration: The Role of Migration Networks," *Review of Economics and Statistics* 92 (2010), 811–821.
- MORAGA, J. F.-H., "New Evidence on Emigrant Selection," *Review of Economics and Statistics* 93 (2010), 72–96.
- O'KEEFE, S., "Locational choice of AFDC recipients within California: A conditional logit analysis," *Journal of Public Economics* 88 (2004), 1521–1542.
- PARÉY, M., J. RUHOSE, F. WALDINGER AND N. NETZ, "The Selection of High-Skilled Emigrants," *The Review of Economics and Statistics* 99 (dec 2017), 776–792.
- PATT, A., J. RUHOSE, S. WIEDERHOLD AND M. FLORES, "International Emigrant Selection on Occupational Skills," (2017).
- REYNOLDS, M. O. AND E. SMOLENSKY, *Public expenditures, taxes, and the distribution of income: the United States, 1950, 1961, 1970* (Academic Press, 1977).
- ROY, A. D., "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers* 3 (1951), 135–146.
- SJAASTAD, L. A., "The Costs and Returns of Human Migration," *Journal of Political Economy* 70 (1962), 80–93.
- SOLT, F., "The Standardized World Income Inequality Database*," *Social Science Quarterly* 97 (2016), 1267–1281.
- VIGDOR, J. L., "The pursuit of opportunity: Explaining selective black migration," *Journal of Urban Economics* 51 (2002), 391–417.