



UNIVERSIDADE CATÓLICA PORTUGUESA

# Regional ex-ante Inequality of Opportunity in Portugal

Inês Barros de Oliveira

Católica Porto Business School  
Abril de 2020



UNIVERSIDADE CATÓLICA PORTUGUESA

# Regional ex-ante Inequality of Opportunity in Portugal

Trabalho Final na modalidade de Dissertação  
apresentado à Universidade Católica Portuguesa  
para obtenção do grau de mestre em Gestão

por

Inês Barros de Oliveira

sob orientação de  
José Gaspar e Ricardo Ribeiro

Católica Porto Business School  
Abril de 2020

# Acknowledgements

I would like to express my gratitude to numerous people whose support was crucial for the conclusion of this dissertation.

Firstly, I would like to thank my supervisors Professor Ricardo Ribeiro and Professor José Gaspar for guiding and helping me throughout the development of my thesis and for always answering my questions promptly.

Secondly, I extend my gratitude to Direção Geral de Estatísticas da Educação e Ciência and to Instituto Nacional de Estatística for providing the data required for the development of this study.

Lastly, and most of all, a special thanks to my family, for giving me strength, support and for always caring so much about my personal and professional future.

# Abstract

Inequality of opportunity is a concept related to the part of inequality that exists due to circumstances beyond the control of the individual and should be considered as ethically offensive (non-acceptable). Factors such as gender, race and family background, that people cannot change through effort, should not affect individual outcome. In this sense, public policies should be designed and implemented in order to reduce the outcome inequalities that arise from unequal opportunities.

This dissertation focuses on studying the impact on net income in Portugal caused by inequality of opportunity and identifying regional differences based on different degrees of urbanization. This research could be an important contribution to this field of study because although the concept of IOp has already been discussed by different authors, it hasn't been deeply analyzed in Portugal, especially at a regional level.

With the results achieved, it will be possible to identify what could be done in terms of public action to mitigate the impact of these uneven circumstances, since the conception of equality of opportunity is related to social justice.

Keywords: Inequality of opportunity, social justice, circumstances, measure inequality, inequality regional differences

# Resumo

O conceito de desigualdade de oportunidade diz respeito à componente da desigualdade que existe devido a circunstâncias não controláveis pelo indivíduo e deve ser considerada eticamente ofensiva (não aceitável). Variáveis como o género, raça e antecedentes familiares, que as pessoas não podem alterar através do seu esforço, não devem afetar o rendimento individual. Nesse sentido, as políticas governamentais devem ser concebidas e implementadas com o propósito de reduzir as desigualdades resultantes de oportunidades desiguais.

Esta dissertação centra-se no estudo do impacto no rendimento líquido em Portugal causado pela desigualdade de oportunidade e na identificação de diferenças regionais com base em diferentes graus de urbanização. Esta pesquisa pretende ser uma contribuição importante para este campo de estudo porque, apesar do conceito de IOp ter já sido discutido por diferentes autores, não foi ainda aprofundadamente analisado em Portugal, particularmente ao nível regional.

A partir dos resultados obtidos, será possível identificar o que pode ser feito em termos de ação pública para mitigar o impacto dessas circunstâncias desiguais, uma vez que o conceito de igualdade de oportunidade está relacionado com justiça social.

Palavras-chave: Desigualdade de oportunidade, justiça social, circunstâncias, medição da desigualdade, desigualdades regionais

# Contents

Acknowledgements.....	3
Abstract.....	4
Resumo.....	5
Contents.....	6
List of Tables and Figures.....	7
1. Introduction.....	9
2. Literature Review.....	13
2.1. Inequality.....	13
2.2. Inequality of opportunity.....	18
2.2.1 Approaches to inequality of opportunity.....	22
2.2.1.1 The ex-post approach.....	23
2.2.1.2 The ex-ante approach.....	23
3. Study Development.....	26
4. Empirical Analysis.....	30
4.1 Data Description.....	30
4.2 Estimation of Inequality of Opportunity.....	39
4.2.1 Analysis by Degree of Urbanization.....	42
4.3 Shapley Decomposition.....	47
4.3.1 Analysis by Degree of Urbanization.....	48
5. Conclusion.....	51
Bibliography.....	54
Appendix A – Variable Description.....	57
Appendix B – Net Income Descriptive Statistics.....	59
Appendix C – Predicted income statistics.....	60
Appendix D – Detailed IOp results.....	61

# List of Tables and Figures

Figure 1 - The Lorenz curve framework (De Maio, 2007, p. 850).....	15
Table 1 - Original number of observations.....	32
Table 2 - Final number of observations .....	33
Table 3 - Number of observations per type of income .....	34
Table 4 – Employed Net Income descriptive statistics .....	35
Table 5 - Detail and evolution of individual’s characteristics and family background.....	36
Table 6 - Detail and evolution of individual’s characteristics and family background for large urban areas .....	38
Table 7 - Detail and evolution of individual’s characteristics and family background for small urban areas .....	38
Table 8 - Detail and evolution of individual’s characteristics and family background for rural areas.....	389
Table 9 - Regression coefficients for employed income at a national level ....	40
Table 10 - Estimated income statistics at a national level .....	41
Table 11 - Regression coefficients for employed income for large urban areas .....	43
Table 12 - Estimated income statistics for large urban areas.....	44
Table 13 - Regression coefficients for employed income for small urban areas .....	44
Table 14 - Estimated income statistics for small urban areas .....	45
Table 15 - Regression coefficients for employed income for rural areas .....	45
Table 16 - Estimated income statistics for rural areas.....	46
Table 17 – Shapley decomposition results in 2005 and 2011 at the national level for employed income.....	48

Table 18 - Shapley decomposition results in 2005 and 2011 for employed income for large urban areas .....	49
Table 19 - Shapley decomposition results in 2005 and 2011 for employed income for small urban areas .....	49
Table 20 - Shapley decomposition results in 2005 and 2011 for employed income for rural areas.....	50
Table 1.A – Description of individual and family characteristics in 2005 .....	57
Table 2.A - Description of individual and family characteristics in 2011 .....	58
Table 1.B – Net Income Descriptive Statistics.....	59
Table 1.C - Income statistics for total and self-employed income .....	60
Table 1.D – Shapley decomposition results in 2005 and 2011 at the national level for total income.....	61
Table 2.D – Shapley decomposition results in 2005 and 2011 for Large urban areas .....	61
Table 3.D - Shapley decomposition results in 2005 and 2011 for Small urban areas for total income.....	62
Table 4.D - Shapley decomposition results in 2005 and 2011 for Rural areas for total income .....	62
Table 5.D - Shapley decomposition results in 2005 and 2011 at the national level for self-employed income .....	63
Table 6.D - Shapley decomposition results in 2005 and 2011 for Large urban areas for self-employed income.....	63
Table 7.D - Shapley decomposition results in 2005 and 2011 for Small urban areas for self-employed income.....	64
Table 8.D – Shapley decomposition results in 2005 and 2011 for Rural areas for self-employed income .....	64

# Chapter 1

## Introduction

With the growth of society and its increasing level of maturity, several social dilemmas arise that require a deep understanding and attention, inequality being one of them. Measuring inequality is crucial to answer a great variety of questions regarding the income distribution over time, disparities between countries, the effect of taxes in income distribution and many other important topics (Atkinson, 1970).

Inequality of opportunity (IOp) is a concept related to the part of inequality that arises from circumstances that cannot be controlled by individuals and, thus, considered as ethically non-acceptable. Gender, race and family background are factors that people cannot change through effort and hence should not affect individual outcome.

The egalitarian philosophers such as Rawls (1971), Dworkin (1981), Arneson (1989) and Cohen (1989) are (among others) the ones that initiated the discussion of the principle of inequality of opportunity. Also, Roemer (1993, 1998, and 2002) has an important contribution to the measurement of inequalities of opportunity in income.

This research is a contribution for the study of the relation between inequality of opportunities and regional dimension aiming to answer the following research questions: What percentage of the regional income differences in Portugal can be exclusively attributed to inequality of opportunity and how is this percentage evolving from 2005 to 2011? What circumstances are responsible for creating inequality of opportunity across regions with different degrees of urbanization in this country? More than being able to quantify the percentage of inequality of opportunity, it is crucial to understand the regional disparities because the

effectiveness of the public policies pursued to mitigate the level of inequality is directly related to the knowledge of the reality of each region. Other than that, regional differences could become an obstacle to economic growth if they generate a lack of social cohesion.

The main contributions of this study to the existing literature on the measurement of IOp are the study of this measure for the particular of Portugal, the analysis of its evolution in two years (2005 and 2011) and the identification of the differences observed by degree of urbanization. The investigation of the main determinants of inequality of opportunity is also an important contribution.

Ideally, economic inequalities that arise from factors not controlled by individuals should be compensated by society, since the conception of equality of opportunity is related to social justice. Therefore, the results achieved with this research could be useful to support any public redistributive policy that aims to mitigate the impact of these uneven circumstances.

To quantify the percentage of IOp in Portugal, the dependent variable selected was the net monetary income of an individual, which is a continuous variable with an inherent scale and the estimation of the expected conditional outcome was made following the method used by Ferreira and Gignoux (2011). The data regarding the individual's income as well as their sociodemographic information was collected from the survey called *Inquérito às Condições de Vida e Rendimento* (ICOR) conducted by INE (*Instituto Nacional de Estatística*) to analyze the distribution of income, living conditions and social exclusion in Portugal for the years 2005 and 2011. In addition to the information collected from ICOR, specific ad-hoc modules with data related to family background were also used. The selected circumstances variables were the gender, parent's education, parent's occupational status and parent's job. For the regional dimension, the population was grouped according to three different types: large urban areas, small urban areas and rural areas.

To correlate the circumstances variables with the dependent variable selected it was estimated a linear regression that relates the output variable with a set of circumstances variables. Once this regression was obtained, the estimates of the coefficients for the different circumstance variables were used to calculate an estimate of the net monetary income based only on the circumstance variables, which was in turn use to compute a measure of inequality of opportunity. The inequality of opportunity ratio was obtained dividing the above measure of inequality of opportunity by the total inequality in income.

The results show that, in average, the net monetary income increased 17% from 2005 to 2011 at a national level, with a higher growth in small urban areas and rural areas (25% and 21%, respectively) than in large urban areas (12%). For this same period, there is a decrease on the total inequality level of 18% at a national level with the greater impact on large urban areas (22%) compared with small urban and rural areas (15% and 10% respectively). The inequality of opportunity ratio at a national level represents 19.50% of the total income inequality in 2005, but there's a positive evolution for 2011, where IOp decreases to a level of 11.02%. By degree of urbanization, the results obtained show that, in 2005, inequality of opportunity is higher in rural areas compared to large urban areas. However, and despite the fact that IOp values decreased for all areas between 2005 and 2011, in 2011 there are no significant differences in IOp between degree of urbanization, although the tendency is that IOp starts to be slightly higher in urban areas than in rural ones. There is a similar study made for other countries in Europe, developed by Melo et al. (2019) and therefore it was possible to establish a comparison between countries and try to understand why some countries are better than others at fighting inequality of opportunity, which will be addressed at the end.

After this introduction, Chapter 2 presents a review of the most relevant literature on this topic where: i) a deeper insight into the concepts of total

inequality and inequality of opportunity are provided; and ii) a detailed description of the approaches commonly used to measure inequality is presented. Chapter 3 describes the method used to measure inequality of opportunity and Chapter 4 provides the empirical analysis with: i) the description of data collected; ii) the estimation process of IOp and iii) the Shapley decomposition, also presenting and discussing the results achieved. Finally, Chapter 5 concludes this study with an overview of the key findings and limitations and giving suggestions for future research.

# Chapter 2

## Literature Review

The present chapter starts with the analysis of the global concept of inequality's relevance with a brief review of some studies about this subject. Inequality of opportunity is then introduced as the main topic of this research, and the different approaches for measuring it are accurately described.

### 2.1. Inequality

With the need to maintain social justice comes the need to measure inequality. It is then necessary to understand this concept. As stated by Rohwerder (2016), inequality is often used in the context of social and economic problems and refers to disparities and discrepancies in areas such as education, health, income, opportunities and others.

According to Haughton and Khandker (2009), inequality measures can be used to illustrate differences between groups or within groups and the outcome of the studies can have different policy implications. From a mathematical (statistical) point of view, inequality exists when we compare two or more given quantities and we realize there are differences between them.

There are several methods that can be used to measure the total inequality that basically can be classified according to two different categories. The first category refers to summary measures, including the Gini index and the generalized entropy measures.

The Gini coefficient (or Gini index) is the most widely used statistical measure to analyze and compare inequalities, particularly regarding income. Since this

index sums up the absolute differences between each pair-wise comparison of the individual's income, it results in a statistical measure of income inequality ranging from zero (perfect equality) to one (perfect inequality, one individual has everything). It is derived from the Lorenz curve, which sorts the population from poorest to richest, and shows the cumulative proportion of the population on the horizontal axis and the cumulative proportion of expenditure (or income) on the vertical axis.

According to De Maio (2007), one of the most appealing properties of this specific index is due to the fact that it can be used to generate a single summary statistic of the income distribution, which is equivalent to the size of the area between the Lorenz curve and the 45° line of equality divided by the total area under the 45° line of equality. To sum up, a coefficient of zero reflects a perfectly equal society, where the income is equally shared, while a coefficient of one reflects a perfectly unequal society, where all the income is only earned by one individual. The Gini coefficient measures the dispersion of income taking as a reference a hypothetical scenario of perfect equality. Considering  $x_i$  a point on the x axis (representing the cumulative percentage of the population) and  $y_i$  a point on the y axis (axis of the accumulated income percentage), the total inequality measured by the Gini coefficient can be expressed as follows:

$$I(y) = G = 1 - \sum_{i=1}^N (x_i - x_{i-1}) (y_i + y_{i-1})$$

Assuming that  $A$  is the area that lies between the line of perfect equality and the Lorenz curve and  $(A + B)$  is the total area under the line of perfect equality, this coefficient can be illustrated with the following figure:

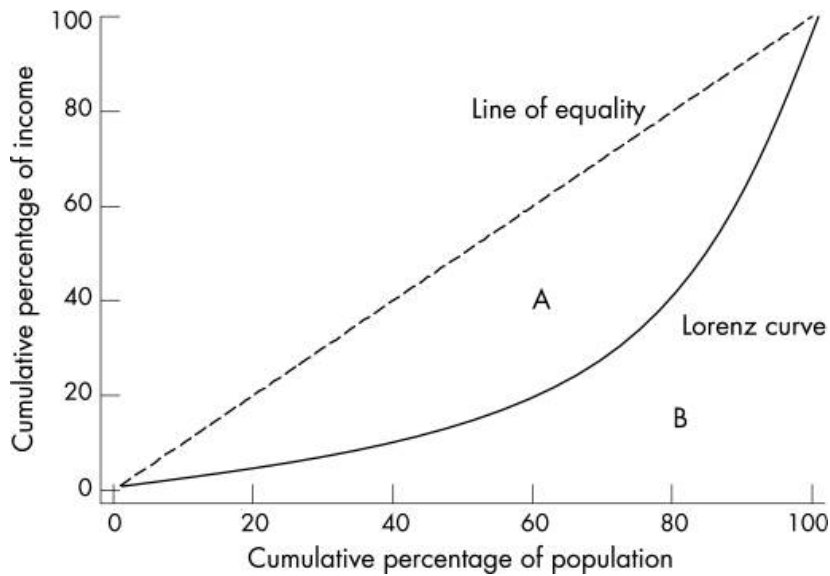


Figure 1 - The Lorenz curve framework (De Maio, 2007, p. 850)

Although this index allows relatively easy interpretations, it has some specific limitations. Firstly, and even though this method can show the existence of income inequality, it does not provide any information about where the income distribution results in income inequality. Secondly, the Gini index is over-sensitive to the changes that occur in the household income towards the middle of the distribution, but not sensitive enough to the changes that occur in the household income at the upper and lower portions of the income distribution (Haughton and Khandker, 2009; Cobham and Sumner, 2013). Thus, this means that the Gini index can be unchanged or even show less inequality, considering that the societal income is becoming more and more polarized (Krozer, 2015).

The generalized entropy measures (GE), such as the Theil index and the Mean Logarithmic Deviation of Income (MLD), are also summary measures of income distribution. The Theil index is more sensitive to changes that affect the upper portion of the income distribution, while the MLD is more sensitive to the changes that affect the lower portion of the income distribution.

These measures have a variable parameter  $\alpha \in [0, +\infty)$  that, depending on its value, gives a higher weight to very high or very low incomes. The parameter  $\alpha$

represents the weight attributed to income differences in different parts of the distribution. When  $\alpha$  is zero, a measure known as mean logarithmic deviation (MLD) is obtained, which gives greater weight to income differences at the bottom of the distribution, highlighting the inequality at the base of the distribution. When  $\alpha$  assumes the value one, this measure corresponds to the Theil index, which gives equal weight to income differences at the upper part of the distribution. The results obtained with the GE measures also vary between zero and infinity, with zero representing an equal distribution and higher values representing a higher level of inequality. The main limitation of these two methods is related to the fact that they both are calculated by using the natural log, meaning that they are undefined for income less than or equal to zero, as it is pointed out by the World Bank Institute (2005). The general equation for the GE measures is:

$$I(y) = GE(\alpha) = \frac{1}{\alpha(\alpha - 1)} \left[ \frac{1}{N} \sum_{i=1}^N \left( \frac{y_i}{\bar{y}} \right)^\alpha - 1 \right]$$

where  $y_i$  represents income,  $\bar{y}$  the average income and  $\alpha$  represents the weight attributed to income differences in the different parts of the distribution.

When  $\alpha$  is zero, we obtain the MLD measure which is represented as follows:

$$I(y) = MLD = \frac{1}{N} \sum_{i=1}^N \log \left( \frac{\bar{y}}{y_i} \right).$$

When  $\alpha$  is one, we obtain the Theil index measure which is given by the following equation:

$$I(y) = T = \frac{1}{N} \sum_{i=1}^N \frac{y_i}{\bar{y}} \log \left( \frac{y_i}{\bar{y}} \right).$$

The second category of the income inequality metrics refers to income ratios. These ratios attempt to explain different parts of the income distribution. The Palma Ratio, for instance, is computed by dividing the total income earned by the richest 10 percent by the poorest 40 percent of the income distribution. This ratio was proposed by Alex Cobham and Andy Sumner in 2013 as an alternative to the Gini coefficient and highlights the differences between those in the upper and lower end of income distribution. It is suggested that this is intuitively easy to understand and more sensitive to changes at the extremes, which may make it more useful for policymakers. Overall, the Palma ratio is used for two specific reasons: 1) changes in income inequality are mainly due to changes in the share of the richest 10 percent and the poorest 40 percent, because the “middle” group were found to have a relatively stable share of income; and 2) this ratio is only sensitive to changes that occur at the top and the bottom of the distribution. Therefore, this ratio provides useful additional information to the one provided by the Gini index.

It is possible to conclude that there are several methods to measure inequality, all of them comprising distinct factors and measures. In fact, and despite the previously mentioned two categories, De Maio (2007) also refers other methods that can be used to measure inequality. Some of them are the Atkinson index, which using an aversion parameter to inequality enables the variation of the sensitivity to transfers at different levels of income and the Coefficient of Variation (CV) that measures income inequality by dividing the standard deviation of the income distribution by its own mean.

The choice of the measure depends on the purpose of the study. However, there are some general principles that should be satisfied when choosing a particular metric: principle of transfers, scale independence, principle of population and decomposability (Cowell, 2009).

For income inequality, which is the type of inequality that this research is about, the principle of transfers is satisfied when, considering any two individuals, transferring a positive amount of income from the richer to the poorer person, the inequality level measured decreases. The principle of population requires that inequality of the income distribution should not depend on the number of individuals. The income scale principle implies that if everyone's income changes by the same proportion, then the value of the inequality measure should remain the same. Finally, to observe the principle of decomposability, there should be a coherent relationship between inequality in the whole population and inequality in its subgroups. According to this principle, this means that it should be possible to find a formula giving total inequality as a function of inequality within the constituent subgroups, and inequality between the subgroups (Cowell, 2009).

## 2.2. Inequality of opportunity

Overall inequality can be decomposed into an ethically acceptable and an ethically offensive component. Inequality of opportunity is related to the unacceptable/offensive part of inequality as it is explained in this section.

Social justice theorists such as Rawls (1971), Dworkin (1981), Arneson (1989) and Cohen (1989) were, among other egalitarian authors, the first ones to introduce a discussion that led to the conception of inequality of opportunity.

Rawls (1971) reported that it is “our natural duty” to identify and remove injustices in our institutions. Dworkin (1981) described two general theories of equality. One of them, called equality of welfare shows that a distributional scheme treats people as equals when it distributes or transfers resources among them until no further transfer would leave them more equal in welfare. The other,

called equality of resources, demonstrates that it treats individuals as equals when it distributes or transfers resources so that no further transfer would leave their share of the total resources more equal. Arneson (1989), for instance, holds that equal opportunity for welfare is the best explanation of the ideal of distributive justice and disagrees that equality of resources captures the ideal of equality. Additionally, Cohen (1989) developed a theory of justice known as “luck egalitarianism”. He stated that the purpose of egalitarianism is to eliminate “involuntary disadvantage”, a concept that he described as a disadvantage for which individuals cannot be considered responsible. Luck egalitarians defend equality in “access to advantage”, meaning that a situation of advantage must depend on choice and not on luck.

The conceptual basis of the definition of inequality of opportunity is provided by Checchi and Peragine (2010), who make a distinction between individual efforts and pre-determined circumstances. In more detail, inequality of opportunity due to individual efforts is not ethically offensive, while inequality of opportunity due to pre-determined circumstances truly represents a violation of the principle of equality of opportunity. Therefore, inequality of opportunity is understood as the part of inequality that exists due to circumstances beyond the control of the individual and is the part of inequality that should be compensated by society. A situation of equality of opportunity exists when society compensates individuals with disadvantageous circumstances, so that their outcomes depend only on factors that they can be accountable for.

For this reason, the theory about this subject emphasizes two principles of equality of opportunity. The first one, called the principle of compensation, points out that inequalities due to circumstances beyond the individual responsibility are inequitable and therefore, should be compensated by society. The second one, known as the principle of natural reward, denotes that distinct achievements, which can be attributed to factors that individuals can be held

accountable for, like effort, are equitable and don't require an act of compensation. Hence, the theory of equality of opportunity emphasizes individual responsibility when analyzing situations of economic disparities (Checchi and Peragine, 2010).

It is discussed that outcomes like income level, education level or health status, are determined by variables that are beyond individuals' responsibility (circumstances) and by variables for which individuals are responsible. Inequalities caused by these so-called circumstances (e.g. race, family background, gender) are assumed to be ethically unacceptable, while those that arise from effort are not classified as offensive (Ramos and Van de gaer, 2012).

Finally, authors such as Checchi and Peragine (2010), Ferreira and Gignoux (2011), Ramos and Van de gaer (2012), Juárez and Soloaga (2014) identified gender, race and family background as the main circumstances beyond the control of the individual. Ferreira and Gignoux (2011) also classified some additional variables like language spoken at home, migration status, access to books at home and the location of the school attended as relevant factors when it comes to measuring inequality of educational opportunity.

There are some empirical studies on this subject. Checchi and Peragine (2010) developed a study to measure inequality in Italy, having concluded that inequality of opportunity is accountable for about 20% of overall income inequality in this country. These authors proposed a non-parametric approach to measure inequality of opportunity and to decompose overall income inequality into an ethically acceptable component and an ethically unacceptable component. The actual earnings were used as individual objective and, for the circumstances, the family background was measured by the level of parents' education. The degree of opportunity inequality was first computed for the entire country distribution and then a partition of the total population was made based on two characteristics: gender and geographic location. The conclusions are that

wages are positively correlated with parental education, they are higher for male subsamples when compared to female ones and, finally, Northern earnings are higher and less dispersed than Southern ones. Complementary, and for the same population subgroups, the authors analyzed the earning distribution according to parental education and individual effort, using the relative income position as a proxy for the extent of effort. They have shown how parental education, taken as a circumstance out of individual control, significantly affect the equality of opportunities especially when considering population subgroups (by gender and by region of residence).

Ferreira and Gignoux (2011) wrote a paper about the of measurement of inequality of opportunity in Latin America and, by comparing six countries in this region, they were able to find similarities in profiles across countries that allows them to conclude about the correlation among certain circumstance variables. The measure of advantage used was the household per capita income (and consumption expenditure per capita, when available) and the variables considered for the family background were the gender, race or ethnicity, birthplace, mother's and father's education, and father's occupation. These authors aim to understand how unequal is the distribution of opportunity for economic advantage in the set of countries for which data was available. An absolute version of the index as inequality of opportunity level as well as a relative version as inequality of opportunity ratio were calculated for both household per capita income and consumption expenditure. The authors have concluded that the entire set of the opportunity-deprived consisted of ethnic or racial minorities in three out of six countries.

More recently, and exploring more in detail inequality of opportunity at the regional level, a study by Melo, et al. (2019) provided empirical evidence regarding the relationship between local area income deprivation and individual

socio-economic outcomes, using information about individuals' characteristics and family background in several European countries.

In Portugal's particular case there are approaches to analyze the concept of inequality of opportunity but, to the best of my knowledge, a deep analysis to explore the impact of the regional dimension was not done yet. The regional dimension is critical since regional disparities could affect beliefs about social fairness and thus become a threat to economic progress and even political stability. The knowledge and explanation of regional differences is important for the design of taxation and public expenditure policies that could lead to a better social justice environment.

### 2.2.1 Approaches to inequality of opportunity

The theoretical literature has proposed two different approaches to measure inequality of opportunity. The ex-ante approach associated with Van de gaer (1993) and the ex-post approach related to Roemer (1998).

The ex-ante approach is related to the measurement of inequality between individuals' opportunity sets, which are determined by their own circumstances. Hence, this approach attaches the exact same value to the opportunity set of those individuals that belong to the same type, measuring inequality in the values of those specific sets. In turn, the ex-post approach is related to the measurement of inequality in individuals' income that present the exact same level of effort. Also, in this approach, all inequalities between individuals must be due to their own circumstances, consisting in the measure of inequality of opportunity (Peragine 2004).

These two approaches are equally valid, but the ex-post approach requires at least an estimate of effort whereas the ex-ante approach can be estimated without it, which makes it easier to implement. To measure inequality of opportunity it is

important to understand both approaches. However, most studies use the ex-ante approach since the ex-post approach demands strong assumptions for effort and luck, which are two characteristics not easily observable and difficult to distinguish empirically. These approaches use information on circumstances, efforts and outcomes, and will be discussed in detail in the next section of this chapter.

### 2.2.1.1 The ex-post approach

The ex-post approach, related to Roemer (1998), focuses on the variable effort and explains that there is a situation of inequality of opportunity when people making the same degree of effort independently of circumstances achieve different outcomes.

Roemer (1998) states that this approach requires the aggregation of outcome differences among individuals at the same effort quantile across types, for each quantile. For this reason, it analyses the inequality within types or circumstance-homogenous groups. Equality of opportunity would imply equality at each quantile and, hence, equality for the all distribution (Ferreira and Gignoux, 2011).

In this study, the ex-post approach will not be utilized since the collection of effort data is difficult and can be quite subjective. This study will instead be based on the ex-ante approach, described in the following section.

### 2.2.1.2 The ex-ante approach

The ex-ante approach, related to Van de gaer (1993), supposes that “types” are groups of people who share the same circumstances and that inequality of

opportunity is based on a comparison of the outcome distributions of different types. Thus, this approach focuses on inequality between types, and is neutral with respect to inequality within types (Checchi and Peragine, 2010).

As explained before, this approach has one empirical advantage: it measures inequality based only on information related to the type to which individuals belong to, without reflecting the impact of effort. This means that it is not necessary to compare individuals based on their level of effort.

Considering this information, inequality of opportunity is defined as the inequality that exists between groups of individuals who share the same circumstances (types) (Ferreira and Gignoux, 2011).

It is crucial to establish a relation between the dependent variable and the independent ones (circumstances). The regression approach makes it possible to define this relation by applying different estimation methods that are adequate to different types of outcome variables.

One of these methods is the ordinary least-squares (OLS) used by Ferreira and Gignoux (2011) to estimate the optimal parameter values of a regression by minimizing the sum of squared residuals. Given that income is a continuous variable with inherent scale, the OLS estimation is the most suitable approach since it computes the average amount for groups of individuals with the same circumstances.

Considering  $y$  as the outcome variable and  $C$  as the vector of circumstances beyond the control of the individuals, Ferreira and Gignoux (2011) compute the expected conditional outcome  $\hat{y}$  using an OLS estimation (i):

$$(i) \quad \hat{y} = E(y|C)$$

For dichotomous and ordered variables such as access to schooling, other researchers like Paes de Barros et al. (2008) and Soloaga and Juárez (2013b) use

logit and probit models to calculate  $\hat{y}$ . These models will not be further discussed in this study since this research will be centralized on a continuous outcome variable, namely income.

# Chapter 3

## Study Development

The main objective of this study is to estimate income inequality of opportunity in Portugal, by degree of urbanization, as well as identify the main circumstances that influence it the most. However, inequality of opportunity is only considered an issue because overall inequality exists. Inequality of opportunity cannot be studied without understanding and quantifying total inequality and the way it is evolving. Thus, this research starts with the measurement of overall income inequality. There are different methods proposed in the literature that can be used to measure total inequality. In this research, it is more convenient to use summary measures so that it is possible to analyze the entire distribution. Although the Gini index is the most commonly used statistical measure, it is not sensitive enough to the changes at the upper and lower portions of the income distribution. The fact that income is becoming more and more polarized, demonstrates that the Gini index might not reflect some relevant changes in income that occur at the extremes of the distribution. For this reason, the mean logarithmic deviation was the one selected in this study, highlighting the inequality at the base of the distribution. Besides that, this option was also based on the fact that the unique measure that satisfies all the required principles to estimate IOp is the mean logarithmic deviation, so this study also applies this measure to compute total inequality for a matter of coherence.

Once total inequality for the variable of interest was calculated, then a decomposition between the part that is due to circumstances that the individual cannot control and the part that is “ethically acceptable” was computed. In order to correlate the circumstance variables with income, this study focuses on the regression approach. For this reason, it was estimated a linear regression that

relates the income variable with a set of circumstances variables. Once this regression was obtained, the estimates of the coefficients for the different circumstance variables were used to calculate an estimate of the income variable based only on the circumstance variables. Given that the outcome variable is continuous, to estimate the expected conditional outcome  $\hat{y}$ , the most adequate method for this calculation is the OLS estimation used by Ferreira and Gignoux (2011). The absolute measure of inequality of opportunity where all variations in income are justified as being only due to circumstances, is presented by the following index:

$$\theta_a = I(\hat{y}_i)$$

It is important to highlight that the mean logarithmic deviation (Ferreira and Gignoux, 2011) is the only suitable method for this purpose because it is the only possible solution that satisfies all the required principles to estimate IOp. First, it satisfies the Pigou-Dalton transfer principle (a transfer in income from a richer individual to a poorer one decreases measured inequality (Haughton and Khandker, 2009)), within-type transfer insensitivity (the measure of inequality remains unchanged to any mean-preserving spread in advantages within a type), the principle of population (the measure of inequality should not change with a replication of the population), scale invariance (the inequality index remains unchanged if every individual's income changes by the same proportion) and the principle of symmetry (if two individuals swap incomes within a type, there should be no change in the measure of inequality). By considering these principles we reduce the number of inequality measures suitable to the measures of the Generalised Entropy class.

However, there is an additional requirement to measure inequality of opportunity that reduces the number of appropriate measures to a unique one.

This requirement named path-independent decomposability axiom was defined by Foster and Shneyerov (2000). The path-independent decomposability axiom demonstrates that  $I(\{\mu_i^k\}) = I(y) - I(\{v_i^k\})$ , where  $\{\mu_i^k\}$  is the smoothed distribution obtained from a distribution of income where each individual income  $y_i^k$  is replaced by the group-specific mean  $\mu^k(y)$ , with the elimination of all within-group inequality by construction, and  $\{v_i^k\}$  is the standardized distribution obtained from a distribution of income where each individual income  $y_i^k$  is replaced by  $y_i^k \frac{\mu}{\mu^k}$  where  $\mu$  is the grand mean, with the elimination of all between-group inequality by appropriately rescaling all subgroup means (Checchi and Peragine, 2010). Which means that when the set of inequality indices  $I(.)$  is focused on those that use the arithmetic mean as the reference income, and that fulfil the Pigou–Dalton transfer axiom, we restrict these measures to a unique one, the mean logarithmic deviation (Ferreira and Gignoux, 2011).

The relative measure of IOp is then obtained dividing the absolute measure of inequality of opportunity by the total inequality. Knowing that  $I(y)$  represents total inequality, the relative measure of this index, also known as inequality of opportunity ratio, which is relative to overall income inequality can be illustrated as follows:

$$\theta_r = \frac{I(\hat{y}_i)}{I(y)}$$

Considering that the research question of this dissertation is to quantify the percentage of inequality due to circumstances beyond the control of the individual in Portugal and identify differences at a regional level, the inequality opportunity ratio will be calculated at a national level and by degree of urbanization.

Finally, the Shapley decomposition was used to highlight the contribution of each circumstance to the inequality of opportunity estimated. The Shapley decomposition shows how each circumstance contributes to the inequality of opportunity ratio, decomposing inequality exactly into its contributory factors (Shorrocks 1999). To compute this decomposition, it is first necessary to estimate the relative inequality measure ( $\theta_r$ ) for all the possible permutations of each circumstance variables. Then, we calculate the average marginal effect of each circumstance variable on the measure of inequality of opportunity ratio. As an advantage, we identify the Shapley decomposition as a method that provides flexibility to choose different inequality indexes with no violation of the path-independent decomposability axiom. Another advantage is that the sum of the value obtained for each of the different components equal the relative inequality value ( $\theta_r$ ). However, since this method needs to measure inequality for all the possible combinations of variables, when a large number of variables exists it can be extremely intensive in terms of computing time, data processing and storage capacity (Gunatilaka and Chotikapanich, 2006). Considering  $K$  to be the number of circumstances, this method requires a computationally intensive procedure because  $2^K$  inequality estimations must be computed, one for each possible combination of variables. It is important to be aware that because most circumstances are highly correlated, the coefficients might suffer from multicollinearity so the decomposition should be used only to have an idea of the relative importance of each circumstance and not to be seen as a causal decomposition (Ferreira and Gignoux, 2014).

# Chapter 4

## Empirical Analysis

In the present chapter, an extensive description of all the steps taken until the results will be presented, starting with a description of the data and with an in-depth analysis of the estimation process.

### 4.1 Data Description

The European Community produces every year statistics on income, living conditions and social exclusion. This information is available in a European statistical database called EU-SILC (Statistics on Income and Living Conditions) that allows the comparison between countries through several indicators.

Since 2004, Portugal implemented a survey called *Inquérito às Condições de Vida e Rendimento* (ICOR) in order to produce relevant information for EU-SILC about the country. This statistical operation is conducted annually by INE (*Instituto Nacional de Estatística*) to analyze the distribution of income, living conditions and social exclusion in Portugal. This data is obtained by interviewing several individuals in order to form a representative sample of the households in Portugal's national territory.

The data used in this research was collected from ICOR, based on the survey carried out for the years 2005 and 2011. The reference sample for the analyses consists of working age people, aged between 16 and 65 years old<sup>1</sup>.

---

<sup>1</sup> In order to have access to this microdata, a request for scientific research was submitted and the required terms of use and individual confidentiality declarations were accepted.

In addition to the information collected from ICOR, every year INE produces specific ad-hoc modules with data related to living conditions, social exclusion and other complementary subjects. The years selected for the sample were meant to capture the information about family background. In fact, for 2005 a module of Intergenerational Transmission of Poverty is available, with information about the highest level of education attained by father and by mother, their activity status and also their main occupation. For 2011, a module of Intergenerational Transmission of Disadvantages is available with similar information.

The data contained in those ad-hoc modules is included in ICOR database cross-sectional data. This database is composed by four different files but only two of them were used in this research: the D-file that contains data that was known prior to the survey, such as the household ID and the degree of urbanisation and the P-file that contains personal data collected in a survey at the individual level, including information about net and gross income values for employed individuals or self-employed individuals. For this purpose, employed individual's income is defined as the total remuneration paid by an employer to an employee in return for work done by the latter during the income reference period while self-employed individual's income is defined as the income received during the income reference period by an individual, as a result of his/her current involvement in his/her own business. The D-file and the P-file were merged using the household ID variable to aggregate the information. The remaining files, R-file and H-file, didn't contain information regarding the individual's background or any circumstance that is not controlled by the individual and thus were removed.

The qualitative variables (the activity status and the job of the father and the mother) were remapped since they were originally displayed as quantitative. To do so, new binary variables were created. For the activity status of the father and the mother, one new variable was mapped, assuming the value of one when the status is "Self-employed" or "Employee" and zero for the remaining status

(unpaid family worker, unemployed, retired, full time housework and other). As far as the occupations go, three new variables were mapped, one for each sector of activity. For the primary sector, “Skilled Agricultural, Forestry and Fishery Workers” and “Elementary Occupations”, for the secondary sector “Craft and Related Trades Workers” and “Plant and Machine Operators and Assemblers”. Lastly, for the tertiary sector, “Managers”, “Professionals”, “Technicians and Associate Professionals”, “Clerical Support Workers” and “Services and Sales Workers”.

In what concerns the degree of urbanization, the population is grouped according to three different types: 1 - Densely-populated area: contiguous grid cells of 1km<sup>2</sup> with a density of at least 1 500 inhabitants per km<sup>2</sup> and a minimum population of 50 000; 2 - Intermediate area: clusters of contiguous grid cells of 1km<sup>2</sup> with a density of at least 300 inhabitants per km<sup>2</sup> and a minimum population of 5 000; 3 - Thinly-populate area: grid cells outside urban clusters.

Table 1 below describes the original sample obtained for each year with the total observations at a national level and by degree of urbanization.

*Table 1 - Original number of observations*

<b>Number of observations:</b>	2005	2011
<b>National level</b>	10706	12489
<b>Degree of urbanisation:</b>		
Large urban areas	3746	4215
Small urban areas	3681	4234
Rural areas	3279	4040

Then, the cleaning procedures and harmonization process were undertaken in order to have a database prepared for the empirical analysis. The first step was the analysis of the output variable, where all the observations that had a null value for this variable were dropped. In this step both the employed net income as well as the self-employed net income were analyzed. A new variable called total income was then generated at this step. This new variable is the sum of the

self-employed income as well as the employed income. This procedure enables the analyzes of the dataset to be done in the perspective of the output variable total income but could also be done separately for the self-employed income as well as the employed income.

The circumstance variables were then analyzed and all the observations that didn't have the information about the activity status of father or the mother, the highest level of education achieved by the father or the mother and main occupation of the father or the mother fulfilled were removed.

After cleaning the dataset, the next step taken was the harmonization of the variables, in order to be able to compare the results from one year with the other. When both years were compared, some variables had distinct names and possible results, although they were the same. To homogenize the variable names and possible values, their descriptions and values were recoded in order to standardize the samples in 2005 and 2011. The complete description of variables pertaining to individual characteristics and family background is presented in Tables 1 and 2 of Appendix A.

Finally, the last step was to account and remove the presence of outliers in the income distribution. Following the method of Alperin and Van Kerm (2013), the process used for this was to remove observations smaller than 75% of the lowest percentile or higher than 125% of the highest income percentile.

The final number of observations used for the estimation process is presented on Table 2 below, with the total observations at a national level and by degree of urbanization.

*Table 2 - Final number of observations*

<b>Number of observations:</b>	2005	2011
<b>National level</b>	2744	2301
<b>Degree of urbanisation</b>		
Large urban areas	987	814
Small urban areas	964	812
Rural areas	793	675

As stated previously, the selected output variables for this analysis are the net monetary income for employees and the net monetary income for self-employed individuals. The total net monetary income obtained by the sum of these two variables was also analyzed. The net income component corresponds to the remuneration after deduction at source of any taxes or social contributions. Unfortunately, no information regarding gross income was reported in 2005 so only the net income was analyzed in order to be able to extract conclusions regarding the evolution of IOp between 2005 and 2011. In Table 3 the number of observations per year, population density and type of income can be found.

*Table 3 - Number of observations per type of income*

	<b>2005</b>	<b>2011</b>
	Number observations	Number observations
<b>National level</b>		
Total income	2744	2301
Employed income	2280	2034
Self employed income	545	345
<b>Large urban areas</b>		
Total income	987	814
Employed income	866	751
Self employed income	143	98
<b>Small urban areas</b>		
Total income	964	812
Employed income	770	718
Self employed income	221	117
<b>Rural areas</b>		
Total income	793	675
Employed income	644	565
Self employed income	181	130

As observed in Table 3, employed individuals are much more representative than the self-employed ones, which only represent 19,86% of the total observations in 2005 and 14,99% in 2011. Considering the small quantity of self-employed data, this study will be centralized on the employed individuals' income. The results for self-employed income as well as the total income will still be represented Appendix B and C but not analyzed in depth.

The number of observations per dataset, as well as the mean, median, standard deviation and total inequality of the employed net income are represented in Table 4. These analysis for the other income categories can be found in Table 1 of Appendix B.

*Table 4 – Employed Net Income descriptive statistics*

	2005			2011		
	Mean	Median	Total Inequality	Mean	Median	Total Inequality
National level	9020.24 (6688.25)	7000.00	0.22	10591.24 (6655.54)	8610.00	0.18
Large urban areas	10316.60 (7738.73)	7996.81	0.23	11530.79 (7264.17)	9178.88	0.18
Small urban areas	8389.80 (6111.97)	6407.11	0.20	10465.94 (6654.72)	8444.46	0.17
Rural areas	8030.80 (5439.34)	6720.00	0.20	9705.40 (5947.65)	8159.00	0.18

Note: standard deviation in parenthesis.

The descriptive statistics in Table 4 show that, in average, the employed net income per individual in 2005 at a national level was € 9020.24 per year. Large urban areas are the ones where the average net income is higher (€ 10316.60 per year) and the lowest amount was observed in rural area with an average annual income of 8030.80 per individual. From 2005 to 2011, the net monetary income increased, in average, 17% at a national level, with a higher increase in small urban areas and rural areas (25% and 21%, respectively). One of the possible reasons for this evolution could be the attractiveness for new investments in these areas due to the lower costs observed in comparison with large urban areas. On the contrary, in large urban areas is where the increase in total income was the lowest (12%). A possible reason for this result is that there are more young people living in these locations, some of them occupying trainee positions with low wages to start their careers. For the year of 2005, total income inequality as measured by the mean log deviation (MLD) was 0.22 at a national level.

This value is higher in large urban areas than in small urban and rural, 0.23 and 0.20, respectively. From 2005 to 2011, there is a decrease of 18% on the national total inequality level that indicates that the distribution of income is more equal in 2011 than in 2005. The greater impact of this evolution was observed in large urban areas (22% decrease) compared with small urban and rural areas (15% and 10% respectively).

For the employed income section of the dataset, the detail and evolution of the individual's characteristics and family background is presented on Table 5.

*Table 5 - Detail and evolution of individual's characteristics and family background*

<i>National Level</i>			2005			2011		
<b>Variable</b>	<b>Min</b>	<b>Max</b>	<b>Mean</b>	<b>Median</b>	<b>St. Deviation</b>	<b>Mean</b>	<b>Median</b>	<b>St. Deviation</b>
female	0	1	0.49	0	0.50	0.51	1	0.50
father's education	0	3	0.72	1	0.65	0.94	1	0.61
father's occupational status	0	1	0.98	1	0.13	1.00	1	0.04
father's job primary sector	0	1	0.39	0	0.49	0.31	0	0.46
father's job secondary sector	0	1	0.38	0	0.49	0.38	0	0.49
mother's education	0	3	0.61	1	0.68	0.93	1	0.69
mother's occupational status	0	1	0.78	1	0.41	1.00	1	0.05
mother's job primary sector	0	1	0.58	1	0.49	0.50	0	0.50
mother's job secondary sector	0	1	0.20	0	0.40	0.20	0	0.40
<b>Number of observations:</b>			<b>2280</b>			<b>2034</b>		

It should be noted that mother's and father's job tertiary sector variables, as well as the male gender and the unemployed occupational status of both parents are omitted in the analysis because they would lead to perfect multicollinearity. In a regression with multicollinear variables, it is impossible to compute the OLS estimates.

At a national level, the distribution of gender is balanced, with a slight increase of the female gender between 2005 and 2011, from 49% to 51%.

The father and mother's education are both, on average a low-level education, primary or lower secondary education, although the father's is, on average, higher than the mother's, but both trended positively from 2005 to 2011.

For the parents' occupational status, both are in majority "employed", averaging, for 2011, 100% employment for both parents. Even though the mother's major activity status in 2005 was already "employed", there was a big increase from 2005 to 2011, from 78% to 100%. This increase came with a decrease of mother's doing "full time housework" and "unpaid family worker", which can be due to the fact that women are increasingly working outside of their homes.

Taking a deeper look into the parents' occupation, it is interesting to observe that there was a trend between the years of 2005 and 2011, where there was a decrease in the number of people working in skilled agricultural, forestry and fishery jobs as well as elementary occupations and an increase in tertiary sector jobs. For the father's occupation, this led to a reduction of 8 pp in the primary sector, from 39% to 31% and the increase of tertiary activities from 23% to 31%. As far as the mother goes, the same 8 pp variation is present, a reduction from 58% to 50% in the primary sector and an increase from 22% to 30% in the tertiary sector.

In large urban areas, as observed in Table 6, for the year of 2005, both the father's and the mother's education was higher than at the national level, other than that, there were also less people working at the primary sector and more at the tertiary sector, when compared with the national average, with 24% of men and 41% of women working in the primary sector. The trend between the years of 2005 and 2011 is very similar to the national level, being positive for the education levels and, for the employment status and occupation, more women were employed and there was an increase of tertiary sector jobs.

Table 6 - Detail and evolution of individual's characteristics and family background for large urban areas

<i>Large urban areas</i>			2005			2011		
Variable	Min	Max	Mean	Median	St. Deviation	Mean	Median	St. Deviation
female	0	1	0.51	1	0.50	0.52	1	0.50
father's education	0	3	0.89	1	0.68	1.07	1	0.69
father's occupational status	0	1	0.99	1	0.10	1.00	1	0.04
father's job primary sector	0	1	0.24	0	0.42	0.20	0	0.40
father's job secondary sector	0	1	0.45	0	0.50	0.38	0	0.49
mother's education	0	3	0.74	1	0.72	1.04	1	0.77
mother's occupational status	0	1	0.80	1	0.40	1.00	1	0.04
mother's job primary sector	0	1	0.41	0	0.49	0.36	0	0.48
mother's job secondary sector	0	1	0.28	0	0.45	0.26	0	0.44
<b>Number of observations:</b>			<b>866</b>			<b>751</b>		

In small urban areas, as observed in Table 7, the main difference to the national level in 2005 was in the education levels, where the father's average education was 63% and the mother's 55%, compared to the national 72% and 61% average. When looking at the trend, the numbers are extremely similar to the national averages.

Table 7 - Detail and evolution of individual's characteristics and family background for small urban areas

<i>Small urban areas</i>			2005			2011		
Variable	Min	Max	Mean	Median	St. Deviation	Mean	Median	St. Deviation
female	0	1	0.50	0	0.50	0.50	1	0.50
father's education	0	3	0.63	1	0.62	0.90	1	0.58
father's occupational status	0	1	0.98	1	0.15	1.00	1	0.04
father's job primary sector	0	1	0.42	0	0.49	0.32	0	0.47
father's job secondary sector	0	1	0.40	0	0.49	0.40	0	0.49
mother's education	0	3	0.55	0	0.67	0.90	1	0.65
mother's occupational status	0	1	0.79	1	0.41	1.00	1	0.04
mother's job primary sector	0	1	0.63	1	0.48	0.52	1	0.50
mother's job secondary sector	0	1	0.19	0	0.39	0.21	0	0.41
<b>Number of observations:</b>			<b>770</b>			<b>718</b>		

In rural areas, as observed in Table 8, for the year of 2005, there were two big differences in relation to the national level. The first being the educational level, that was lower, 60% and 50% for the father and the mother, respectively and the number of people working on the primary sector, with 55% of the men, compared to the 39% national average, and 76% of the women, compared with the 58%

national average. There are no notable trends specifically for the rural areas, they follow the same as the national values, with an increase in education levels and a decrease of primary sector jobs.

*Table 8 - Detail and evolution of individual's characteristics and family background for large rural areas*

<i>Rural areas</i>			2005			2011		
<b>Variable</b>	<b>Min</b>	<b>Max</b>	<b>Mean</b>	<b>Median</b>	<b>St. Deviation</b>	<b>Mean</b>	<b>Median</b>	<b>St. Deviation</b>
female	0	1	0.47	0	0.50	0.49	0	0.50
father's education	0	3	0.60	1	0.60	0.82	1	0.50
father's occupational status	0	1	0.98	1	0.14	1.00	1	0.04
father's job primary sector	0	1	0.55	1	0.50	0.43	0	0.50
father's job secondary sector	0	1	0.27	0	0.44	0.36	0	0.48
mother's education	0	3	0.50	0	0.61	0.80	1	0.58
mother's occupational status	0	1	0.76	1	0.43	0.99	1	0.07
mother's job primary sector	0	1	0.76	1	0.43	0.65	1	0.48
mother's job secondary sector	0	1	0.10	0	0.31	0.11	0	0.32
<b>Number of observations:</b>			<b>644</b>			<b>565</b>		

Regarding the father's occupation, there is a trend towards more craft workers and machine operators (secondary sector) in detriment of elementary occupations and agricultural jobs. For the mother's occupation there is a propensity to increase tertiary sector jobs and to reduce primary sector jobs.

In the next section, it will be estimated the inequality of opportunity level.

## 4.2 Estimation of Inequality of Opportunity

The estimation process starts with a linear regression of total net income on the circumstance variables. With this regression, it is possible to correlate the dependent variable with the circumstances and thus estimate the net monetary income based only on the circumstance variables. The dependent variable considered is the net monetary income for employed individuals and the

independent variables are the gender, parent’s education, parent’s occupational status and parent’s job.

The output of the regression can be seen in Tables 9 and 10, with the regression coefficients and the predicted income statistics, respectively.

Table 9 - Regression coefficients for employed income at a national level

<b>National Level</b>		2005		2011	
<b>Variable</b>	<b>Coefficient</b>	<b>Standard error</b>	<b>Coefficient</b>	<b>Standard error</b>	
female	-2119.60***	256.16	-1776.74***	282.18	
father's education	1477.75***	274.66	958.65***	315.50	
father's occupational status	597.81	983.77	-7472.18**	3675.65	
father's job primary sector	-2262.86***	431.81	-1428.83***	446.84	
father's job secondary sector	-2264.81***	383.71	-1269.81***	379.88	
mother's education	934.38***	260.30	1147.63***	282.35	
mother's occupational status	-598.28*	316.38	4885.04*	2848.67	
mother's job primary sector	-1620.63***	407.59	-1602.50***	412.22	
mother's job secondary sector	-1694.18***	440.43	-944.22**	451.13	
constant	11334.67***	1088.25	14082.01***	4685.85	
<b>Number of observations:</b>	<b>2280</b>		<b>2034</b>		
<b>R-squared:</b>	<b>0.17</b>		<b>0.12</b>		

Note: \*10% significance level; \*\* 5% significance level; \*\*\* 1% significance level.

When analysing the estimates of the regression coefficients of female, for both years they are negative, although there is a slight decrease from 2005 to 2011. This means that the income level is lower for females. When it comes to the father’s and the mother’s education, both are positively correlated with the income level. For the year of 2005, the father’s education has more weight than the mother’s but, when looking at 2011, the inverse is true, meaning that the mother’s education is more influential. Both the parent’s jobs are negatively correlated with income for both years, when they are on the primary or secondary sectors, meaning that a job on the tertiary sector is associated with a higher income level. Given that, there is also a trend in these variables that is starting to mitigate these differences, in 2011, jobs on the primary and secondary sectors didn’t have as

negative of a influence as in 2005. Lastly, the occupational status of the parents is not significant to explain income.

The R-squared observed is 0.17 in 2005 which means that around 17% of the individuals' income variance is explained by the variance of the collected variables. In 2011, the R-squared reduced to 0.12.

After the regression coefficients were calculated, it was possible to estimate the individual's income and, with that, the inequality of opportunity, using an OLS estimation.

Dividing the absolute measure of inequality of opportunity by the total inequality, it is possible to obtain the relative measure of inequality of opportunity or, in other words, the part of overall inequality that is caused by inequality of opportunity.

Table 10 - Estimated income statistics at a national level

	2005				2011			
	Mean	Median	Absolute IOp	Relative IOp	Mean	Median	Absolute IOp	Relative IOp
National level	9020.24 (2765.04)	8341.53	0.05	0.21	10647.84 (2295.9)	10569.82	0.02	0.13
Large urban areas	10316.60 (3175.94)	10139.01	0.05	0.20	11530.79 (2424.03)	11635.60	0.02	0.13
Small urban areas	8389.80 (2393.32)	7863.95	0.04	0.19	10465.94 (2047.31)	10297.47	0.02	0.11
Rural areas	8030.80 (2391.79)	7276.15	0.04	0.21	9705.40 (2533.67)	9707.93	0.03	0.18

Note: standard deviation in parenthesis.

Table 10 shows that the absolute values of IOp from 2005 to 2011 decreased and both the mean and median values of income increased. At the national level, the variation in net income due to circumstances beyond the control of individuals was 21% in 2005. However, this percentage decreases to a level of 13% in 2011 which seems to be a positive trend. When comparing the R-squared value of the regression in the relative inequality they are very similar, 17% and

21% for 2005 and 12% and 13% for 2011, respectively. Given that, with the variables collected, it is possible to explain a percentage of the income that is very close to the value that is dependent on circumstances and thus, it is possible to conclude that the variables collected are able to explain the segment of income that is not dependent on the individual's effort.

It is interesting to compare these results with the ones obtained in Spain and France in a study by Melo et al. (2019), also focused on income of employed individuals. The results for Spain were only available for 2011 and show that, at the national level, 19% of the variation in labour income is due to circumstances uncontrollable by the individual. In France, the level of inequality of opportunity was slightly lower in 2011 compared to 2005 (21% vs 22%) but higher compared to Spain. It should be noted that the variables used in this study are very similar with the ones used by Melo et al. (2019), and thus the relevance of these comparisons.

#### 4.2.1 Analysis by Degree of Urbanization

Looking at the breakdown by degree of urbanization, the regression coefficients and estimated income for large urban areas can be analyzed in Tables 11 and 12.

Table 11 - Regression coefficients for employed income for large urban areas

<b>Large urban areas</b>		2005		2011	
<b>Variable</b>	<b>Coefficient</b>	<b>Standard error</b>	<b>Coefficient</b>	<b>Standard error</b>	
female	-2857.31***	480.26	-1492.26***	505.92	
father's education	1362.72***	502.07	166.10	504.00	
father's occupational status	1766.70	2390.25	2528.47	6923.90	
father's job primary sector	-2165.16***	810.04	-1255.06	807.58	
father's job secondary sector	-3110.26***	639.44	-600.61	627.93	
mother's education	993.54**	452.01	2086.52***	459.46	
mother's occupational status	-1875.50***	597.49	173.79	6933.97	
mother's job primary sector	-1090.93	696.88	-1170.68	725.14	
mother's job secondary sector	-1285.53*	707.27	-1138.57	734.42	
constant	12287.36***	2543.69	8452.01	9830.61	
<b>Number of observations:</b>	<b>866</b>		<b>751</b>		
<b>R-squared:</b>	<b>0.17</b>		<b>0.11</b>		

Note: \*10% significance level; \*\* 5% significance level; \*\*\* 1% significance level.

Analysing Table 11, for large urban areas, there are some differences in terms of variable significance, especially on 2011, where only female and the mother's education are significant for the analysis. The gender continues to be one of the most influential factors and having well educated parents as well.

The R-squared observed is 0.17 in 2005 which means that around 17% of the individuals' income variance is explained by the variance of the collected variables. In 2011, the R-squared reduced to 0.11.

Comparing the absolute IOp in Table 12 with the total inequality showed in Table 4, the same pattern occurs as already described for the national level. This means that the absolute IOp decreased and both the mean and median of the estimated income increased. Comparing the R-squared with the relative IOp, the values, as with the national level, are very similar, leading to the same conclusions as the ones stated previously.

Table 12 - Estimated income statistics for large urban areas

	2005				2011			
	Mean	Median	Absolute IOp	Relative IOp	Mean	Median	Absolute IOp	Relative IOp
Large urban areas	10316.60 (3175.94)	10139.01	0.05	0.20	11530.79 (2424.03)	11635.60	0.02	0.13

Note: standard deviation in parenthesis.

Tables 13 and 14 contain the same analysis for small urban areas.

Table 13 - Regression coefficients for employed income for small urban areas

<i>Small urban areas</i>	2005		2011	
	Coefficient	Standard error	Coefficient	Standard error
female	-1629.91***	410.16	-1983.93***	477.59
father's education	1177.51***	442.76	1549.42***	557.06
father's occupational status	-240.14	1378.02	1774.68	6407.27
father's job primary sector	-2690.13***	714.99	-1045.81	750.98
father's job secondary sector	-2420.54***	652.42	-1645.86**	661.31
mother's education	1239.55***	423.41	158.25	488.14
mother's occupational status	79.10	515.72	1856.04	6394.57
mother's job primary sector	-779.54	673.42	-1790.24**	697.80
mother's job secondary sector	-1277.10*	741.82	-801.40	783.93
constant	10786.61***	1562.87	8395.19	9120.02
<b>Number of observations:</b>	<b>770</b>		<b>718</b>	
<b>R-squared:</b>	<b>0.15</b>		<b>0.09</b>	

Note: \*10% significance level; \*\* 5% significance level; \*\*\* 1% significance level.

The biggest difference from large to small urban areas is that, for the year of 2011, instead of the mother's education being the most relevant variable after gender it's the father's education.

The R-squared observed is 0.15 in 2005 which means that around 15% of the individuals' income variance is explained by the variance of the collected variables. In 2011, the R-squared reduced to 0.09.

Table 14 - Estimated income statistics for small urban areas

	2005				2011			
	Mean	Median	Absolute IOp	Relative IOp	Mean	Median	Absolute IOp	Relative IOp
Small urban areas	8389.80 (2393.32)	7863.95	0.04	0.19	10465.94 (2047.31)	10297.47	0.02	0.11

Note: standard deviation in parenthesis.

The absolute IOp in Table 14 decreases for employed income, following the same trend showed in Table 4 for total inequality. The R-squared for the small urban areas regression is also very similar to the relative IOp.

Table 15 contains the same coefficient analysis for rural areas.

Table 15 - Regression coefficients for employed income for rural areas

<i>Rural areas</i>	2005		2011	
	Coefficient	Standard error	Coefficient	Standard error
female	-1932.41***	393.18	-1795.96***	460.05
father's education	1666.88***	434.71	1819.01***	598.40
father's occupational status	-157.61	1420.94	-27550.74***	5463.33
father's job primary sector	-476.25	680.14	-1931.32**	798.65
father's job secondary sector	-214.70	673.39	-1720.41**	690.92
mother's education	471.27	435.90	646.25	526.77
mother's occupational status	38.42	472.19	6383.49**	3158.17
mother's job primary sector	-3653.86***	738.55	-1587.82**	700.59
mother's job secondary sector	-3871.25***	860.35	-1242.02	858.10
constant	11308.15***	1615.07	32364.40***	6377.70
<b>Number of observations:</b>	<b>644</b>		<b>565</b>	
<b>R-squared:</b>	<b>0.19</b>		<b>0.18</b>	

Note: \*10% significance level; \*\* 5% significance level; \*\*\* 1% significance level.

It should be noted that, in rural areas, the most significant variables are the gender, that maintains the same trend and the father's education, that influences the income positively. Mother's occupation is also a significant variable with a negative correlation for both years and father's occupation has also a negative correlation but is only significant in 2011. Mother's and father's occupational status are only significant in 2011, with a negative correlation for the father and

positive for the mother which means that mother being employed has a positive impact in the individual's income.

The R-squared observed is 0.19 in 2005 which means that around 19% of the individuals' income variance is explained by the variance of the collected variables. In 2011, the R-squared reduced to 0.18.

Table 16 - Estimated income statistics for rural areas

	2005				2011			
	Mean	Median	Absolute IOp	Relative IOp	Mean	Median	Absolute IOp	Relative IOp
Rural areas	8030.80 (2391.79)	7276.15	0.04	0.21	9705.40 (2533.67)	9707.93	0.03	0.18

Note: standard deviation in parenthesis.

Once again, comparing Table 16 with the values of Table 4, the rural area patterns are similar for both total inequality and absolute IOp, with all forms of inequality diminishing between 2005 and 2011. Once again, the value of R-squared match with the values of relative inequality, and thus, it is possible to conclude that, with the selected variables, it is possible to explain a percentage of the income that is very similar to the percentage of income that is out of the individual's control.

The results obtained show that inequality of opportunity is higher in rural areas compared to large urban areas and the gap tends to increase from 2005 to 2011 between these two regions. In 2005, the main contributors for the variation in net income are the gender and the parents' education. The gender tends to be more influential in urban areas than in rural ones.

Comparing these results with the ones obtained in Spain and France in the study by Melo et al. (2019), there are different conclusions. In Spain, the variation in labour income due to circumstances is higher in large urban areas (20%) compared to both small urban areas and rural areas (18%). In France, inequality was also higher in more densely populated areas, however this result shows a

tendency to decrease. In the case of France, rural areas have a pronouncedly lower IOp, particularly in 2005.

As stated before, Appendix D contains the results for a similar analysis for self-employed income and total income. These results show that, at the national level, the impact in net income due to circumstances not controlled by the individual is lower for self-employed income with a tendency to decrease from 2005 to 2011. In what concerns the degree of urbanization, the inequality of opportunity appears to be higher for self-employed income in small urban areas than in large or even in rural areas. This leads us to conclude that possibly there are few opportunities for people to create their own business in small urban areas in comparison with large urban areas.

The Shapley decomposition will be applied to measure the contribution of each circumstance to the level of inequality of opportunity estimated.

### 4.3 Shapley Decomposition

The Shapley decomposition was used to measure the contribution of each circumstance to the inequality of opportunity computed above, showing how each circumstance contributes to the inequality of opportunity ratio, decomposing inequality exactly into its contributory factors.

The results obtained for the total income as well as the results for the self-employed income are presented on Tables 1D to 8D of Appendix D.

Table 17 below shows the results achieved for the employed income at a national level.

Table 17 – Shapley decomposition results in 2005 and 2011 at the national level for employed income

TOTAL - National level		2005		2011	
Variable	Value	In percentage	Value	In percentage	
gender	0.03	16.46%	0.02	15.84%	
father's education	0.06	27.62%	0.03	20.17%	
mother's education	0.04	20.05%	0.03	24.16%	
father's occupational status	0.00	0.23%	0.00	1.19%	
father's job primary sector	0.03	12.76%	0.02	12.09%	
father's job secondary sector	0.01	6.65%	0.01	4.19%	
mother's occupational status	0.00	1.62%	0.00	3.38%	
mother's job primary sector	0.02	12.05%	0.02	17.34%	
mother's job secondary sector	0.00	2.56%	0.00	1.64%	
TOTAL	0.21	100.00%	0.13	100.00%	

The main contributors for these results are the parents' education and the gender, although the weight of gender was slight less relevant in 2011. Father's education is more influential than mother's education in 2005 but, in 2011, mother's education begins to be more influential which shows an increase of the impact of mother's education on inequality of opportunity. The mother's occupational status was more influential than the father's activity status in 2011 but in 2005 the contribution of these variables to the inequality of opportunity ratio is extremely low.

### 4.3.1 Analysis by Degree of Urbanization

The same analysis is now presented by degree of urbanization to understand the differences between each region (large urban areas, small urban areas, and rural areas) and the variation in net income due to circumstances beyond the control of the individual.

Table 18 - Shapley decomposition results in 2005 and 2011 for employed income for large urban areas

TOTAL - Large urban areas	2005		2011	
	Variable	Value	In percentage	Value
gender	0.04	23.05%	0.01	11.32%
father's education	0.04	16.29%	0.02	15.45%
mother's education	0.03	13.40%	0.05	43.03%
father's occupational status	0.00	0.69%	0.00	0.37%
father's job primary sector	0.01	3.38%	0.01	10.74%
father's job secondary sector	0.03	9.16%	0.00	2.31%
mother's occupational status	0.02	7.78%	0.00	0.04%
mother's job primary sector	0.01	3.12%	0.01	12.05%
mother's job secondary sector	0.01	1.60%	0.01	4.67%
TOTAL	0.20	100.00%	0.13	100.00%

Table 19 - Shapley decomposition results in 2005 and 2011 for employed income for small urban areas

TOTAL - Small urban areas	2005		2011	
	Variable	Value	In percentage	Value
gender	0.02	11.81%	0.03	24.61%
father's education	0.06	29.99%	0.03	25.98%
mother's education	0.05	25.96%	0.01	10.73%
father's occupational status	0.00	0.01%	0.00	1.01%
father's job primary sector	0.03	14.66%	0.01	7.11%
father's job secondary sector	0.01	5.89%	0.01	9.87%
mother's occupational status	0.00	0.06%	0.00	0.47%
mother's job primary sector	0.01	8.27%	0.02	18.59%
mother's job secondary sector	0.00	3.35%	0.00	1.58%
TOTAL	0.19	100.00%	0.11	100.00%

The father's job is also one important contributor for inequality having more importance in large urban areas while mother's job is particularly influential in rural areas in 2005 and in small urban areas in 2011. From 2005 until 2011, the gender seems to lose importance in large urban areas, being particular influential in small urban areas. The parents' education is also one of the most important contributors.

Table 20 - Shapley decomposition results in 2005 and 2011 for employed income for rural areas

TOTAL - Rural areas	2005		2011	
	Value	In percentage	Value	In percentage
gender	0.04	17.54%	0.02	12.05%
father's education	0.06	27.27%	0.03	13.66%
mother's education	0.03	15.99%	0.02	10.77%
father's occupational status	0.00	0.12%	0.02	7.91%
father's job primary sector	0.02	10.65%	0.02	8.82%
father's job secondary sector	0.00	1.49%	0.01	3.98%
mother's occupational status	0.00	0.09%	0.01	4.91%
mother's job primary sector	0.05	21.57%	0.03	12.94%
mother's job secondary sector	0.01	5.21%	0.00	1.55%
TOTAL	0.21	100.00%	0.18	100.00%

The father's education is more influential in small urban areas and mother's education is much more influential in large urban areas. The main occupation of the father is more relevant in small urban areas as well as the mother's occupation.

# Chapter 5

## Conclusion

This dissertation aims to answer what percentage of the regional differences observed in net monetary income of individuals can be attributed to circumstances beyond the control of the individual in Portugal and what circumstances are responsible for creating IOp between regions with different degrees of urbanization.

It was first important to understand and quantify the level of inequality in the individual's income. The results obtained show that, for employed income, total inequality decreased in the two years analyzed (2005 and 2011) from 22% to 18%. Inequality of opportunity is only considered a relevant topic of study because total inequality exists. Given that inequality of opportunity is the part of inequality that exists due to circumstances not controlled by individuals, it is the fraction of inequality that should be addressed and mitigated by the policy makers.

Using the data collected from ICOR that was based on the survey carried out for the years 2005 and 2011, a global analysis regarding inequality was made for these two years. The analysis was mainly focused on the employed individuals' income because the quantity of data related to self-employed individuals' income was very reduced. It is also more appropriate to analyze this income since the study by Melo et al. (2019) used to compare with the results achieved in this research also use this source of income as the outcome variable of study.

The results show that the degree of inequality of opportunity at the national level represents 21% of the total income inequality in 2005 but there's a positive evolution in 2011, where the IOp decreases to the level of 13%. Compared these results with the ones obtained in Spain and France, the evolution seems to be

more positive in Portugal. In fact, Spain shows that in 2011 at the national level, 19% of the variation in labour income is due to circumstances not controlled by individuals. Identical results were obtained in France, where IOp was 22% in 2005 and 21% in 2011, which shows a slight decrease between these two years. Of all the circumstances beyond the control of the individual considered, parents' education and gender appear as the most determinant to the value obtained in the inequality of opportunity ratio.

By degree of urbanization, the results obtained show that inequality of opportunity in both years is higher in rural areas in comparison with large and small urban areas. It is important to highlight that from 2005 until 2011 inequality of opportunity decreased for all these three degrees of urbanization. Comparing with Spain and France, the results obtained are quite different. In Spain, the variation in labour income due to circumstances is higher in large urban areas compared to small urban areas and rural areas (20% in large urban areas compared with 18% in rural areas). For France, inequality was also higher in more densely populated areas, however it shows a tendency to decrease.

The main limitations found are related to the data available for this research. In fact, there are only two years available with information about family background and not completely comparable. Furthermore, only information about the degree of urbanization was available. It could have been interesting to analyze and compare in detail the regional differences in Portugal but there was no available information in the data for these two years. Besides that, another limitation found was the outcome variable of study. The data obtained had information regarding the net income of the individual for both years but not regarding gross income (only available in 2011). This can be prejudicial to the conclusions obtained since taxes are a tool used by the government with the intent to mitigate income inequality and the net income already reflects the

impact of taxes. Therefore, the real IOp results in this study might be underestimated due to the use of this outcome variable.

The relevance of this research for this field of study is to add additional information for the discussion about the concept of IOp, mainly for the particular case of Portugal, with detail about the differences in regions with different population size. The results achieved might be useful to provide some guidance to policy makers in their search for adequate measures to mitigate the impact of these uneven circumstances. The concept of equality of opportunity is related to social justice and, given that, economic inequalities due to factors that are beyond the individual responsibility should be compensated by society.

Regarding future developments, there are some paths that can be explored in the future in the context of this dissertation. One of them will be to carry out some similar analysis with geographical identifiers for Portuguese NUTS2 regions (Nomenclature of territorial units for statistics that identify basic regions for the application of regional policies on the economic territory of the EU) if INE releases data in the future with this information in ICOR.

# Bibliography

Arneson, R. J. (1989). Equality and equal opportunity for welfare. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 56(1), 77-93.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of economic theory*, 2(3), 244-263.

Checchi, D., & Peragine, V. (2010). Inequality of opportunity in Italy. *The Journal of Economic Inequality*, 8(4), 429-450.

Cohen, G. A. (1989). On the currency of egalitarian justice. *Ethics*, 99(4), 906-944.

Cowell, F. A. (2009). *Measuring Inequality*, Oxford University Press.

De Maio, F. G. (2007). Income inequality measures. *Journal of Epidemiology & Community Health*, 61(10), 849-852.

Dworkin, R. (1981). What is equality? Part 1: Equality of welfare. *Philosophy & public affairs*, 185-246.

Dworkin, R. (1981). What is equality? Part 2: Equality of resources. *Philosophy & public affairs*, 283-345.

Ferreira, F. H., & Gignoux, J. (2008). *The measurement of inequality of opportunity: Theory and an application to Latin America*. The World Bank.

Ferreira, F. H. G., & Gignoux, J. (2011). The measurement of educational inequality: Achievement and opportunity (Discussion paper series, IZA DP No. 6161).

Foster, J. E., & Shneyerov, A. A. (2000). Path independent inequality measures. *Journal of Economic Theory*, 91(2), 199-222.

Gunatilaka, R., & Chotikapanich, D. (2006). Inequality trends and determinants in Sri Lanka 1980-2002: a Shapley approach to decomposition (No. 6/06). Monash University, Department of Econometrics and Business Statistics.

Haughton, J., & Khandker, S. R. (2009). *Handbook on poverty+ inequality*. World Bank Publications.

Juárez, F. W. C., & Soloaga, I. (2014). iop: Estimating ex-ante inequality of opportunity. *The Stata Journal*, 14(4), 830-846.

Krozer, A. (2015). The inequality we want: How much is too much?. *Journal of International Commerce, Economics and Policy*, 6(03), 1550016.

Melo, P., Gaspar, J., Janssen, H. J., van Ham, M., Andersson, E., & Malmberg, B. (2019). Location and context–analysis of spatial inequalities at different geographical scales. Deliverable 5.3. H2020 project Relocal.

Peragine, V. (2004). Ranking income distributions according to equality of opportunity. *The Journal of Economic Inequality*, 2(1), 11-30.

Ramos, X., & Van de Gaer, D. (2012). *Empirical approaches to inequality of opportunity: Principles, measures, and evidence*.

Rawls, John, 1921-2002. (1971). *A theory of justice*. Cambridge, Mass. :Belknap Press of Harvard University Press.

Roemer, J. E. (1993). A pragmatic theory of responsibility for the egalitarian planner. *Philosophy & Public Affairs*, 146-166.

Roemer, J. E., & Trannoy, A. (2015). Equality of opportunity. In *Handbook of income distribution* (Vol. 2, pp. 217-300). Elsevier.

Roemer, J. E. (2003). Defending equality of opportunity. *The Monist*, 86(2), 261-282.

Shorrocks, A. F. (1982). Inequality decomposition by factor components. *Econometrica: Journal of the Econometric Society*, 193-211.

Van Kerm, P., & Alperin, M. N. P. (2013). Inequality, growth and mobility: the intertemporal distribution of income in European countries 2003–2007. *Economic Modelling*, 35, 931-939.

# Appendix A – Variable Description

Table 1.A – Description of individual and family characteristics in 2005

2005				
Variable	Code	Original category	New code	New category
<b>Gender</b>	1	Male	0	Male
	2	Female	1	Female
<b>Highest level of education achieved by the father/mother</b>	0	Less than primary education	0	Father/mother could neither read nor write in any language
	1	Primary education	1	Low level (pre-primary, primary education or lower secondary education)
	2	Lower secondary education	1	
	3	Upper secondary education	2	Medium level (upper secondary education and post-secondary non-tertiary)
	4	Post-secondary non tertiary education	2	
	5	First stage of tertiary education	3	High level (first stage of tertiary education and second stage of tertiary education)
6	Second stage of tertiary education	3		
<b>Activity status of father/mother</b>	1	Employee	1	Employed
	2	Self-employed	1	
	3	Unpaid family-worker	0	Not employed
	4	Unemployed	0	
	5	Retired, early retired	0	
	6	Full time housework	0	
	7	Other	0	
<b>Main occupation of father/mother</b>	1	LEGISLATORS, SENIOR OFFICIALS AND MANAGERS	3	father's or mother's job tertiary sector
	11	LEGISLATORS AND SENIOR OFFICIALS	3	
	21	PHYSICAL, MATHEMATICAL AND ENGINEERING SCIENCE PROFESSIONALS	3	
	22	LIFE SCIENCE AND HEALTH PROFESSIONALS	3	
	23	TEACHING PROFESSIONALS	3	
	24	OTHER PROFESSIONALS	3	
	31	PHYSICAL AND ENGINEERING SCIENCE ASSOCIATE PROFESSIONALS	3	
	32	LIFE SCIENCE AND HEALTH ASSOCIATE PROFESSIONALS	3	
	33	TEACHING ASSOCIATE PROFESSIONALS	3	
	34	OTHER ASSOCIATE PROFESSIONALS	3	
	41	OFFICE CLERKS	3	
	42	CUSTOMER SERVICES CLERKS	3	
	51	PERSONAL AND PROTECTIVE SERVICES WORKERS	3	
	52	MODELS, SALESPERSONS AND DEMONSTRATORS	3	
	61	MARKET-ORIENTED SKILLED AGRICULTURAL AND FISHERY WORKERS	1	father's or mother's job primary sector
	71	EXTRACTION AND BUILDING TRADES WORKERS	2	
	72	METAL, MACHINERY AND RELATED TRADES WORKERS	2	
	73	PRECISION, HANDICRAFT, PRINTING AND RELATED TRADES WORKERS	2	
	74	OTHER CRAFT AND RELATED TRADES WORKERS	2	
	81	STATIONARY-PLANT AND RELATED OPERATORS	2	
82	MACHINE OPERATORS AND ASSEMBLERS	2		
83	DRIVERS AND MOBILE-PLANT OPERATORS	2		
91	SALES AND SERVICES ELEMENTARY OCCUPATIONS	1	father's or mother's job primary sector	
92	AGRICULTURAL, FISHERY AND RELATED LABOURERS	1		
93	LABOURERS IN MINING, CONSTRUCTION, MANUFACTURING AND TRANSPORT	1		

Note: In order to prepare the variables for the analysis, the activity status of the father/mother was remapped into employed or not employed, the main occupation of the father/mother was remapped into the three sectors of activity, and the educational level was remapped in order to match the levels established in 2011.

Table 2.A - Description of individual and family characteristics in 2011

2011				
Variable	Code	Original category	New code	New category
Gender	1	Male	0	Male
	2	Female	1	Female
Highest level of education achieved by the father/mother	-1	Don't know		Not considered
	0	Father/mother could neither read nor write in any language	0	Father/mother could neither read nor write in any language
	1	Low level (pre-primary, primary education or lower secondary education)	1	Low level (pre-primary, primary education or lower secondary education)
	2	Medium level (upper secondary education and post-secondary non-tertiary)	2	Medium level (upper secondary education and post-secondary non-tertiary)
	3	High level (first stage of tertiary education and second stage of tertiary education)	3	High level (first stage of tertiary education and second stage of tertiary education)
Activity status of father/mother	-1	Don't know		Not considered
	1	Employee	1	Employed
	2	Self-employed (including family worker)	1	
	3	Unemployed	0	Not employed
	4	In retirement or in early retirement or had given up business	0	
	5	Fulfilling domestic tasks and care responsibilities	0	
Main occupation of father/mother	6	Other inactive person	0	
	-1	Don't know		Not considered
	1	Managers	3	father's or mother's job tertiary sector
	2	Professionals	3	
	3	Technicians and Associate Professionals	3	
	4	Clerical Support Workers	3	
	5	Services and Sales Workers	3	
	6	Skilled Agricultural, Forestry and Fishery Workers	1	father's or mother's job primary sector
	7	Craft and Related Trades Workers	2	father's or mother's job secondary sector
8	Plant and Machine Operators and Assemblers	2		
9	Elementary Occupations	1	father's or mother's job primary sector	

Note: In order to prepare the variables for the analysis, the activity status of the father/mother was remapped into employed or not employed, and the main occupation of the father/mother was remapped into the three sectors of activity.

## Appendix B – Net Income Descriptive Statistics

Table 1.B – Net Income Descriptive Statistics

	2005			2011		
	Mean	Median	Total Inequality	Mean	Median	Total Inequality
<b>National level</b>						
Total income	8892.78 (7198.56)	6853.00	0.27	10338.11 (7296.98)	8293.30	0.22
Self employed income	6384.55 (6237.34)	4800.00	0.53	7125.11 (11482.88)	3598.92	0.56
<b>Large urban areas</b>						
Total income	10258.79 (8144.95)	7869.18	0.27	11533.67 (7898.37)	8870.27	0.21
Self employed income	7303.63 (6916.23)	5520.00	0.50	8095.64 (11445.95)	4547.28	0.72
<b>Small urban areas</b>						
Total income	8225.78 (6697.53)	6267.29	0.26	10007.75 (7073.75)	7895.59	0.22
Self employed income	6167.76 (5653.89)	4500.00	0.50	6570.05 (11446.68)	3592.20	0.60
<b>Rural areas</b>						
Total income	8003.41 (6219.23)	6479.20	0.26	9293.78 (6579.61)	7697.75	0.23
Self employed income	5923.02 (6311.00)	4734.94	0.58	6070.37 (6747.25)	3598.92	0.38

Note: standard deviation in parenthesis.

## Appendix C – Predicted income statistics

Table 1.C - Income statistics for total and self-employed income

	2005				2011			
	Mean	Median	Absolute IOp	Relative IOp	Mean	Median	Absolute IOp	Relative IOp
<b>National level</b>								
Total income	8892.78 (2967.14)	8199.21	0.05	0.20	10338.11 (2328.36)	10249.81	0.02	0.11
Self employed income	6384.52 (2235.10)	6177.23	0.07	0.13	6815.12 (2035.4)	6282.72	0.05	0.08
<b>Large urban areas</b>								
Total income	10258.79 (3184.46)	10153.52	0.05	0.18	11533.67 (2475.17)	11289.60	0.02	0.11
Self employed income	7303.63 (1945.55)	7725.59	0.04	0.08	8095.64 (2219.42)	7323.97	0.04	0.05
<b>Small urban areas</b>								
Total income	8225.78 (2668.88)	7710.22	0.05	0.18	10007.75 (2260.27)	10034.33	0.02	0.11
Self employed income	6167.76 (2635.24)	5986.36	0.12	0.23	6570.05 (3004.31)	6551.11	0.12	0.19
<b>Rural areas</b>								
Total income	8003.41 (2866.50)	7121.31	0.06	0.22	9293.78 (2181.27)	9125.12	0.02	0.11
Self employed income	5923.02 (2773.42)	5022.26	0.10	0.18	6070.37 (1629.95)	6380.19	0.05	0.12

Note: standard deviation in parenthesis.

## Appendix D – Detailed Shapley decomposition results

Table 1.D – Shapley decomposition results in 2005 and 2011 at the national level for total income

TOTAL - National level	2005		2011	
	Value	In percentage	Value	In percentage
female	0.04	20.16%	0.01	12.10%
father's education	0.05	27.21%	0.03	23.44%
father's occupational status	0.00	1.28%	0.00	0.61%
father's job primary sector	0.03	12.79%	0.01	13.02%
father's job secondary sector	0.01	3.64%	0.00	3.76%
mother's education	0.04	19.28%	0.03	26.95%
mother's occupational status	0.00	1.25%	0.00	2.66%
mother's job primary sector	0.02	11.53%	0.02	16.26%
mother's job secondary sector	0.01	2.84%	0.00	1.19%
TOTAL	0.20	100.00%	0.11	100.00%

Table 2.D - Shapley decomposition results in 2005 and 2011 for Large urban areas for total income

TOTAL - Large urban areas	2005		2011	
	Value	In percentage	Value	In percentage
female	0.04	24.64%	0.01	4.97%
father's education	0.04	22.17%	0.02	15.26%
father's occupational status	0.00	1.23%	0.00	0.57%
father's job primary sector	0.01	6.99%	0.02	17.75%
father's job secondary sector	0.02	12.51%	0.00	4.18%
mother's education	0.03	15.44%	0.04	38.95%
mother's occupational status	0.01	6.67%	0.00	0.04%
mother's job primary sector	0.01	6.01%	0.01	13.25%
mother's job secondary sector	0.01	4.35%	0.01	5.01%
TOTAL	0.18	100.00%	0.11	100.00%

Table 3.D - Shapley decomposition results in 2005 and 2011 for Small urban areas for total income

<b>TOTAL - Small urban areas</b>		<b>2005</b>		<b>2011</b>	
<i>Variable</i>	<i>Value</i>	<i>In percentage</i>	<i>Value</i>	<i>In percentage</i>	
female	0.03	19.63%	0.02	21.71%	
father's education	0.05	29.18%	0.03	31.12%	
father's occupational status	0.00	1.57%	0.00	0.52%	
father's job primary sector	0.03	15.05%	0.01	7.65%	
father's job secondary sector	0.01	2.69%	0.01	6.54%	
mother's education	0.04	22.79%	0.02	15.12%	
mother's occupational status	0.00	0.04%	0.00	0.25%	
mother's job primary sector	0.01	6.61%	0.02	15.94%	
mother's job secondary sector	0.00	2.35%	0.00	1.16%	
<b>TOTAL</b>	<b>0.18</b>	<b>100.00%</b>	<b>0.11</b>	<b>100.00%</b>	

Table 4.D - Shapley decomposition results in 2005 and 2011 for Rural areas for total income

<b>TOTAL - Rural areas</b>		<b>2005</b>		<b>2011</b>	
<i>Variable</i>	<i>Value</i>	<i>In percentage</i>	<i>Value</i>	<i>In percentage</i>	
female	0.04	20.18%	0.02	15.16%	
father's education	0.05	24.08%	0.02	17.64%	
father's occupational status	0.00	0.32%	0.01	9.28%	
father's job primary sector	0.02	8.46%	0.01	5.83%	
father's job secondary sector	0.00	1.08%	0.00	1.10%	
mother's education	0.04	18.47%	0.02	15.73%	
mother's occupational status	0.00	0.50%	0.00	5.59%	
mother's job primary sector	0.04	20.12%	0.01	9.49%	
mother's job secondary sector	0.01	6.77%	0.00	0.45%	
<b>TOTAL</b>	<b>0.22</b>	<b>100.00%</b>	<b>0.11</b>	<b>100.00%</b>	

Table 5.D - Shapley decomposition results in 2005 and 2011 at the national level for self-employed income

TOTAL - National level	2005		2011	
	Value	In percentage	Value	In percentage
female	0.06	44.25%	0.01	13.39%
father's education	0.02	17.62%	0.03	33.74%
father's occupational status	0.01	4.28%	0.00	0.40%
father's job primary sector	0.01	9.53%	0.00	4.30%
father's job secondary sector	0.00	1.07%	0.00	3.82%
mother's education	0.01	8.87%	0.03	34.08%
mother's occupational status	0.00	0.85%	0.00	0.00%
mother's job primary sector	0.02	12.04%	0.01	7.46%
mother's job secondary sector	0.00	1.42%	0.00	2.76%
<b>TOTAL</b>	<b>0.13</b>	<b>100.00%</b>	<b>0.08</b>	<b>100.00%</b>

Table 6.D - Shapley decomposition results in 2005 and 2011 for Large urban areas for self-employed income

TOTAL - Large urban areas	2005		2011	
	Value	In percentage	Value	In percentage
female	0.06	67.85%	0.00	1.83%
father's education	0.02	19.06%	0.02	40.35%
father's occupational status	0.00	0.15%	0.00	0.25%
father's job primary sector	0.00	1.23%	0.00	2.29%
father's job secondary sector	0.00	0.52%	0.00	2.14%
mother's education	0.00	6.41%	0.01	19.86%
mother's occupational status	0.00	0.59%	0.00	0.00%
mother's job primary sector	0.00	2.06%	0.01	14.88%
mother's job secondary sector	0.00	1.84%	0.01	18.39%
<b>TOTAL</b>	<b>0.08</b>	<b>100.00%</b>	<b>0.05</b>	<b>100.00%</b>

Table 7.D - Shapley decomposition results in 2005 and 2011 for Small urban areas for self-employed income

TOTAL - Small urban areas	2005		2011	
	Value	In percentage	Value	In percentage
female	0.09	36.25%	0.06	27.94%
father's education	0.06	25.21%	0.03	16.31%
father's occupational status	0.02	8.02%	0.00	0.00%
father's job primary sector	0.03	11.71%	0.00	2.21%
father's job secondary sector	0.01	3.75%	0.01	3.76%
mother's education	0.02	6.87%	0.07	32.33%
mother's occupational status	0.00	0.80%	0.00	0.00%
mother's job primary sector	0.01	4.41%	0.01	3.78%
mother's job secondary sector	0.00	0.06%	0.01	4.09%
TOTAL	0.23	100.00%	0.19	100.00%

Table 8.D – Shapley decomposition results in 2005 and 2011 for Rural areas for self-employed income

TOTAL - Rural areas	2005		2011	
	Value	In percentage	Value	In percentage
female	0.04	22.02%	0.01	5.71%
father's education	0.03	14.59%	0.00	2.94%
father's occupational status	0.00	0.36%	0.00	0.00%
father's job primary sector	0.01	5.59%	0.05	38.92%
father's job secondary sector	0.01	3.59%	0.01	10.56%
mother's education	0.05	27.51%	0.02	13.74%
mother's occupational status	0.00	0.50%	0.00	0.00%
mother's job primary sector	0.04	21.07%	0.02	19.34%
mother's job secondary sector	0.01	4.75%	0.01	8.80%
TOTAL	0.18	100.00%	0.12	100.00%