LIS
Working Paper Series

No. 850

What Contributes to Rising Inequality in Large Cities?

Luis Ayala, Javier Martín-Román, Juan Vicente

January 2023

Luxembourg Income Study (LIS), asbl

What contributes to rising inequality in large cities?

Luis Ayala[*]
(*UNED, Equalitas*)

Javier Martín-Román[**]
(*UNED*)

Juan Vicente[***]
(*Universidad de Valladolid*)

[*] *Corresponding author.* Facultad de Derecho, UNED, C/Obispo Trejo, 28040 Madrid, Spain. Email: layala@uned.es

[**] Facultad de Derecho, UNED, C/Obispo Trejo, 28040 Madrid, Spain. Email: jmartin@cee.uned.es

[***] Facultad de Ciencias Económicas y Empresariales, Universidad de Valladolid, Avenida Valle Esgueva 6, 47011 Valladolid, Spain. Email: perdiz@eco.uva.es

Declarations of interest: none.

The authors acknowledge financial support received from from Ministerio de Ciencia e Innovación (PID2021-125710OB-C21) and Comunidad de Madrid H2019/HUM-5793).

Abstract

This paper aims to analyze the trends in income inequality in large cities within a selected sample of OECD countries. Specifically, we consider a set of determinants that account for changes in the income distribution and estimate their contributions to inequality by developing both a dynamic approach —differences in inequality in large cities over the last two decades— and a static approach —differences in inequality between large cities and other areas. We use a combination of reweighting techniques and recentered influence functions (RIF) to detect both an upward trend in inequality within large cities and higher levels of inequality with respect to other areas. These results are mainly driven by changes in the returns to endowments rather than by changes in its distribution. We also find that the results are not of the same magnitude across the countries analysed. The contribution to inequality of the skill premium is considerably higher in the US than in European countries.

JEL classification: D31; P52; R12

Key words: large cities, income inequality, counterfactual analysis, RIF-OLS decomposition

1. INTRODUCTION

According to data from the United Nations, between 1950 and 2018, the world's urban population grew more than four times (United Nations, 2018). The urbanization process will continue for decades, with an increasing proportion of the world's population concentrated in large metropolitan areas. Among the various questions raised by this growing concentration of population in large cities, its potential effects on inequality will undoubtedly be a major focus of policy research for years to come. Inequality and city size are complex and multifaceted concepts, and their relationship is not always well defined.

New scenarios linked to the processes through which large cities developed have emerged, and many in the policy and research communities have speculated about their effects on inequality. These potential effects raise numerous interesting questions that must be addressed: Is inequality greater in large cities than in other areas? Has inequality in large cities increased over time? What factors contribute most to the differences in income inequality? Do all possible variables contribute in the same way?

These questions motivate this paper. Our concern is to contribute to the understanding of the relationship between inequality in disposable income and city size. With this aim, we focus on large cities in a selected sample of OECD countries. The main goal is to determine and quantify the individual contributions of different explanatory factors to differences in inequality both within large cities and between large cities and other areas. We apply the methodology proposed by Firpo et al. (2009, 2018) (FFL henceforth), implementing it using both a *dynamic approach* —focusing on the increase in income inequality in large cities over the last two decades— and a *static approach* —focusing on inequality differences between large cities and other areas. By means of the first approach, we can answer questions such as the following: Is there any common pattern in inequality changes in large cities among the countries analyzed? Which factors account for the changes in the distribution of income in large cities? By means of the static approach, we focus on the most recent data to compare income inequality in large cities and other areas.

To the best of our knowledge, this is the first attempt to put into practice this methodology from a double perspective, which is one of the contributions of the paper. Furthermore, as far as we know, few studies apply the aforementioned methodology to more than one country. The two approaches provide a complementary and comprehensive view of the influence of some of the main factors contributing to income inequality in large cities in the chosen countries. Although this type of methodology only allows us to identify the variables that most contribute to explaining the differences in inequality in the different areas —without delving into the factors that determine them— this identification seems essential to gain a better understanding of the dynamics of inequality in large cities.

Our findings lend support to the thesis that income inequality is higher in large cities than in other areas. This conclusion holds when the relationships are tested with different inequality measures. We also find that inequality increased in large cities during the first two decades of the 21st century. These changes are explained, essentially, by what we call the 'structure effect'. Among the potential factors contributing to these trends, one of the most important is educational attainment. However, the magnitude of the results differs among the countries analyzed. While in the US the skill premium seems more relevant than skill composition, in the other countries studied the role of educational attainment is mostly reflected through changes in endowments. In the European countries examined age, household composition and household size are also relevant in explaining inequality differences within large cities and relative to other areas.

The paper is structured as follows. In section two we review the literature connecting income inequality and city size. In section three we describe the methodology used. In section four we introduce the data sources used and the variables chosen. In section five we present and comment on the main results of the paper. Section six concludes.

2. INCOME INEQUALITY AND CITY SIZE: REVIEW OF THE LITERATURE

An extant literature has tried to quantify the relevance of income inequality in cities and its circumstantial and political drivers. The availability of data and analytical methods has guided the empirical research on inequality in these areas. New or updated data and novel testing tools explain the sequence by which the causes and consequences of inequality in different areas, identified by the theoretical literature, have been tested in the empirical literature. In general, most of the evidence on inequality and city size has referred to US metropolitan areas and the focus has primarily been on the labor market.

This review includes only contributions explicitly addressing the relationship between inequality and city size. Most of them focus on earnings inequalities, with much less evidence on inequality in household disposable income. Garofalo and Fogarty (1979) provided a pioneering theoretical framework for analyzing the urban income distribution based on agglomeration economies and the amenity structure of cities. Assuming that amenities increase until some urban size threshold is reached and then reverse, the authors derived a U-shaped relationship between inequality and city size. However, their empirical results for the US metropolitan areas in the 1970s were not robust and were sensitive to the choice of the inequality measure. Nord (1980a) provided empirical support for the U-shaped hypothesis extending the sample used by previous studies to include smaller cities. He accounted separately for some of the factors outlined by the previous literature (1980b) finding that race reinforced the positive relationship between city size and inequality.

Later evidence for US metropolitan areas from the 1980s is not conclusive. Galster et al. (1988) found only weak support for the hypothesis that population has a direct effect on inequality after controlling for industry and occupational structure. Cloutier (1997) found a positive effect of population size and population growth on inequality after controlling for spatial, demographic and industrial structures, although it was not statistically significant. Under a general equilibrium framework for an open system of cities, Alperovich (1995) proposed different relationships between city size and income inequality. A relevant one was that inequality rises (declines) with city size if the relative preference for non-traded goods increases (decreases) with the level of income. Pooling data on US metropolitan areas from 1970, 1980 and 1990, Wheler (2004a, 2004b) found a negative association between the changes in population density and the 90/10 percentile ratio of wages.

With the turn of the century, the hypothesis of a positive relationship between population size and inequality in cities received increasing empirical support. Using longitudinal US data from the 1979 to 1998 Surveys of Labor Market Experience, Gould (2007) found evidence of a city wage premium, but only for white-collar workers. Glaeser et al. (2009) found a positive link between population size and inequality, which has become stronger since 1980. The causes of household income inequality in cities were a higher skill wage premium in industries such as finance or computing, immigration and, above all, the skill composition within the population.

Baum-Snow and Pavan (2013), using the same datasets than Glaeser et al. (2009) for a similar period but restricting the analysis to working white men aged 25-54, confirmed this increase in inequality with city size. They found that city-size specific factors explained at least a quarter of the overall increase in the variance in wages between 1979 and 2007. City differences in the skill wage premium were more relevant than differences in skill composition for explaining the city size effect on inequality. Baum-Snow et al. (2018), using manufacturing data from core-based statistical areas (CBSAs) from 1980 to 2007, examined some potential causes of the more rapid increase in wage inequality in larger cities over time. The high estimated elasticity of substitution between unskilled labor and capital explains why unskilled wages are much less variable across locations than skilled wages. They also suggest that the increasing complementarity in production between human capital and market scale indicates the growing role of knowledge spillovers in generating agglomeration economies. Finally, Davis and Dingel (2019) developed a spatial equilibrium model to look inside the black box of knowledge spillovers. Their model replicates a range of empirical facts, including that skill premia are higher in larger cities, finding a positive relationship between them and city size.

The evidence for countries other than the US is quite fragmented, and there are very few comparative analyses. In the case of high-income countries, the positive relationship between inequality and city size is generally confirmed, although not all the US results are transferable to other countries. Soroka (1984) found that city size did not have a direct effect on overall urban income distributions in Canada. Lee et al. (2016) found that larger cities were more unequal in Great Britain. This relationship did not hold when the mean wage was included, something that the authors interpreted as the consequence of the abundance of more highly skilled and better paid workers in large cities. Henkel (2017) obtained for Germany results similar to those of Baum-Snow and Pavan (2013) for the US, though the relationship between location size and inequality was stronger and more positive when only the upper part of the income distribution was considered.

Hortas-Rico and Rios (2019) found that population size was a modest determinant of inequality in Spanish cities. De la Roca and Puga (2016) also found a higher mean and greater dispersion of earnings in bigger cities in Spain. Their results show that workers in bigger cities do not have higher initial unobserved ability fixed effects, but they obtain an immediate static premium and accumulate more valuable experience. This additional value of experience was stronger for those with higher initial ability.

In the case of developing and emerging countries, the evidence for the two most populous countries is apparently contradictory. While Chen et al. (2018) found a positive relationship between income inequality and city size in China, Dubey and Mahadevia (2001) found for India that while the Gini coefficient seems to be unrelated to city size the incidence of poverty decreases with it. A possible explanation is that the caste-based segregation in India diminishes with city size (Haque et al., 2019). Another explanation is that the effects of city size on inequality and poverty have opposite signs when poverty lines are defined at the national level.

There are few comparative studies covering different countries. Royuela et al. (2014) used the OECD (2012) metropolitan database and the concept of functional urban areas (FUAs) to find that regional inequality is positively correlated with urbanization. This correlation increases when the definition of 'urban' is restricted to people living in large metropolitan FUAs. Boulant et al. (2016) provided the first estimation of the distribution of household disposable income for 153 metropolitan areas in 11 OECD countries using data from tax records, household surveys or estimates. Using the same data and sample, Castells-Quintana et al. (2020) provided a deeper analysis of the relationship between inequality and city size considering the United States, Canada, Latin America and some European countries. Their estimates suggest that as cities double in size, the Gini index grows by approximately one percentage point.

3. METHODOLOGY

In this paper we try to answer four questions: Has inequality increased in large cities? Is it greater than in other areas? What factors contribute most to explaining differences in inequality in these areas over time? What factors contribute to explaining the difference in inequality with other areas? The empirical strategy we propose consists of implementing an extension of the Oaxaca-Blinder decomposition (Blinder, 1973; Oaxaca, 1973)—OB hereafter—to study recent changes in income inequality trends in a sample of OECD countries. We use recentered influence function (RIF) regressions and analyze four different distributional measures: the Gini coefficient, the P90-P10 ratio (the difference between the 90th and 10th percentiles of equivalent disposable income), and the P90-P50 and P50-P10 ratios[1]. By carrying out a twofold procedure, an 'aggregate decomposition', where the total components linked to the *composition* and *structure effects* are estimated using a reweighting method, and a 'detailed decomposition', where the individual contributions of each attribute considered in the analysis are estimated, we try to identify and quantify the differences between two groups: what we have called large cities —over 500,000 inhabitants— and other areas —those with fewer than 500,000

[1] Although P90/P10, P90/P50 and P50/P10 shares are more frequent in the inequality literature, most studies applying the FFL methodology use the ratio of the upper bounds instead of the percentile shares. We use the former to facilitate comparisons with other studies.

citizens. In particular, we focus on the contributions of a set of covariates to the inequality measures chosen.

This exercise is implemented in each country from a double perspective. On the one hand, we examine the existing differences in inequality between the two groups mentioned above with the most recent data (static approach). On the other hand, we pay attention to the evolution of inequality within large cities during approximately the last two decades (dynamic approach). To the best of our knowledge, this is the first attempt to apply this methodology with the two approaches and using a comparative perspective for more than two countries.

Following Fortin et al. (2011), who presented a review of the main decomposition procedures used to evaluate changes in wage distributions, we chose the FFL proposal (Firpo et al., 2009) with the improvements suggested by Firpo et al. (2018). In this way, we can measure the specific contribution of each covariate included in the model, and some of the main factors contributing to income inequality in large cities can be identified.

This technique consists of a two-stage process that can be illustrated as follows. First, assuming the premises established in OB as the starting point, we conduct an 'aggregate decomposition'. This task allows us to identify the so-called *composition effect*—variation attributable to changes in characteristics—and the *structure effect*—changes associated with the returns to these characteristics. The first stage is performed by means of a reweighting approach. In the second step, to complete the "detailed decomposition", we need to make use of the regression strategy set down in FFL. This method is based on the estimation of a regression where the dependent variable —real equivalent disposable income in our case—is replaced by its transformation through the so-called RIF. This function calculates the effect of small changes on the corresponding distribution on distributional statistics.

Once the previous regressions have been estimated, a standard OB can be developed. Under the assumption of linearity, this method allows us to decompose the income inequality gap between two groups in a straightforward way. On the one hand, we try to identify the variables correlated with the changes in inequality in what we have called large cities over time. On the other hand, we use the latest available data to capture the differences in inequality between the latter compared to the other areas. The RIF regressions implemented in the analysis are quite easy to estimate since they can be performed via ordinary least squares (OLS), just as in OB.[2] As a result, we estimate a RIF-OLS model.

Compared to other methodologies developed in the literature that focus on decomposition exercises to estimate the differences between distributions, the FFL proposal has a remarkable advantage. While the semiparametric approach of DiNardo et al. (1996) based on the reweighting of samples, the parametric approximation of Juhn et al. (1993) involving the distributions of the residuals, and the conditional quantile regressions (CQR)[3] of Machado and Mata (2005) or Melly (2006) only allow for the calculation of the aggregate effects of characteristics and their returns, the FFL scheme provides a detailed decomposition. In this sense, it identifies the individual contributions of each explanatory factor considered in the model through the profile of characteristics and their corresponding returns. Another advantage is that in contrast to the classical OB approach, FFL allows us to consider the entire distribution, not only the mean. In other words, we can not only distinguish the importance of individual contributions to income inequality in average, but can also delve into which factors have been the most relevant sources of change in the different segments of the distribution by decomposing the variation at different percentiles. However, the approximations of certain nonlinear functions obtained with RIF regressions may not be as rigorous and precise as might be expected in some particular cases (Rothe, 2015).

a) *Aggregate decomposition*

Let us suppose there is a joint distribution function defining all relationships between three variables: the real equivalent disposable income (Y), the regressors or exogenous characteristics (X), and a categorical variable (T) indicating the group to which each observation belongs: $f_{Y,X,T}(y_i, x_i, T_i)$. Bearing in mind that we have only two groups,[4] the

[2] OLS provides a transparent estimate of the average marginal effects of the explanatory variables on the dependent variable, an interpretation that does not apply equally in the case of conditional quantile regressions. The OLS method allows to estimate $E(Y|X) = X\beta$. In turn, according to the law of iterated expectations (LIE), this result implies that $E(Y) = E(X)\beta$. In contrast, conditional quantile regression models postulate $Q_\tau(X) = X\beta_\tau$. In this case, β_τ cannot be interpreted as the marginal effect of X on the unconditional τ-th quantile of the distribution, so they cannot be used to measure the impact on the τ percentile of the distribution of a marginal change in X, which is a major restriction. RIF regressions can solve this econometric problem.

[3] These proposals, additionally, are *path dependent*, as the result of the decomposition is affected by the order in which the mentioned decomposition is implemented. The FFL methodology applied here and derived from RIF regressions is *path independent*.

[4] In the static approach, $T = 0$ represents units living in areas with less than 500,000 inhabitants and $T = 1$ denotes units residing in areas with more than 500,000 inhabitants. In the dynamic approach, $T = 0$ stands for the initial year and $T = 1$ refers to the last year.

joint probability distribution function and the cumulative distribution of real equivalent disposable income conditional on *T* can be described as follows:

$$f_{Y,X}^{k}(y,x) = f_{Y|X}^{k}(Y|X) \cdot f_X^{k}(X) \qquad [1]$$

$$F_Y^{k}(y) = \int F_{Y|X}^{k}(Y|X) \cdot dF_X^{k}(X) \qquad [2]$$

The superscript *k* indicates that the density is conditional on $T = k$, with $k \in [0,1]$. The way to compute the gap between the two groups, given a distributional statistic *v*, such as the median, would be:

$$\Delta_O^{v} = v_1 - v_0 = v(F_Y^{1}) - v(F_Y^{0}) \qquad [3]$$

$$\Delta_O^{v} = v\left(\int F_{Y|X}^{1}(Y|X) \cdot dF_X^{1}(X)\right) - v\left(\int F_{Y|X}^{0}(Y|X) \cdot dF_X^{0}(X)\right) \qquad [4]$$

To assess the relevance of the differences in characteristics $(dF_X^{1}(X) \neq dF_X^{0}(X))$ and their returns $\left(F_{Y|X}^{1}(Y|X) \neq F_{Y|X}^{0}(Y|X)\right)$ when determining the overall differential between both groups, we need to design a hypothetical scenario.[5] The counterfactual statistic can then be denoted as:

$$v_C = v(F_Y^{C}) = v\left(\int F_{Y|X}^{0}(Y|X) \cdot dF_X^{1}(X)\right) \qquad [5]$$

Finally, the 'aggregate decomposition' can be expressed as the difference between the two groups (*v-overall income gap*):

$$\Delta_O^{v} = (v_1 - v_C) + (v_C - v_0) = \Delta_S^{v} + \Delta_X^{v} \qquad [6]$$

The *structure effect* (Δ_S^{v}) is given by $v_1 - v_C$, while $v_C - v_0$ captures the *composition effect* (Δ_X^{v}).

b) *Detailed decomposition*

The influence function for the τ-th quantile can be defined as follows:

$$IF(y; q_\tau, F) = \frac{\tau - l(y \leq q_\tau)}{f_Y(q_\tau)} \qquad [7]$$

[5] To recreate the counterfactual scenario, a situation that cannot be checked in the available data, we apply a reweighting approach similar to those described in DiNardo et al. (1996) or Barsky et al. (2002). The alternative proposed by these authors is to multiply the distribution of characteristics $dF_X^{0}(X)$ with a reweighting factor $\psi(X)$ so that it provides a distribution similar to $dF_X^{1}(X)$.

where $l(y \leq q_\tau)$ is an indicator function showing whether the value of real equivalent disposable income is below q_τ, and $f_Y(q_\tau)$ is the marginal density of the same outcome of interest at q_τ, which is determined by kernel estimation.

For operational reasons, it seems appropriate to center the influence function on the statistic of interest—the Gini coefficient, for instance. All we have to do is adding this statistic to the influence function. The RIF formula becomes:

$$RIF(y; q_\tau, F) = q_\tau + IF(y; q_\tau, F) \quad [8]$$

Once the RIF function has been calculated[6], we obtain the value of the transformed variable for each observation in the sample. As stated by Firpo et al. (2011), the key point is to assume that the conditional expectation of the RIF function can be modeled as a linear function of the explanatory variables. This assumption translates into the fact that RIF regressions can be estimated by simply running OLS.

The RIF regressions (*unconditional quantile regressions,* UQR) provide estimates of the marginal impact of the explanatory variables on the chosen statistic. In other words, the $\hat{\gamma}$ estimated coefficients can be interpreted as the (average partial) effect of an increase in the average value of an explanatory variable on the corresponding statistic—Gini coefficient, variance, percentile, etc.

The detailed decomposition[7] embodies a RIF-OLS decomposition combined with a semiparametric reweighting estimator[8], again applying DiNardo et al. (1996). This decomposition can be disaggregated into four terms as follows:

$$\hat{\Delta}_O^v = \bar{X}_1' \cdot (\hat{\gamma}_1^v - \hat{\gamma}_C^v) + (\bar{X}_1 - \bar{X}_0^C)' \cdot \hat{\gamma}_C^v + (\bar{X}_0^C - \bar{X}_0)' \cdot \hat{\gamma}_0^v + \bar{X}_0^{C'} \cdot (\hat{\gamma}_C^v - \hat{\gamma}_0^v) \quad [9]$$

Where:

$$\hat{\Delta}_{S,p}^v = \bar{X}_1' \cdot (\hat{\gamma}_1^v - \hat{\gamma}_C^v) \quad [10]$$

$$\hat{\Delta}_{S,e}^v = (\bar{X}_1 - \bar{X}_0^C)' \cdot \hat{\gamma}_C^v \quad [11]$$

$$\hat{\Delta}_{X,p}^v = (\bar{X}_0^C - \bar{X}_0)' \cdot \hat{\gamma}_0^v \quad [12]$$

[6] The four RIF functions used here can be checked in the Appendix (see Table A.1).
[7] A very common problem here is the choice of a specific reference for the dummy explanatory variables used in the analysis, since this decision can have an impact on the results (Oaxaca and Ransom, 1999). For this reason, following Yun (2005) we have applied a normalization strategy that allows us to overcome this identification problem and an adequate estimation of the real contribution of each covariate.
[8] The OB-type decomposition, without reweighting, would be: $\hat{\Delta}_{OB}^v = \bar{X}_1' \cdot (\hat{\gamma}_1^v - \hat{\gamma}_0^v) + (\bar{X}_1 - \bar{X}_0)' \cdot \hat{\gamma}_0^v$.

$$\hat{\Delta}^{v}_{X,e} = \bar{X}_0^{C\prime} \cdot (\hat{\gamma}^{v}_{C} - \hat{\gamma}^{v}_{0}) \quad [13]$$

This model, an improved version of the original FFL, corrects some misspecification and reweighting problems existing in the model without reweighting. $\hat{\Delta}^{v}_{S,p}$ represents the *pure structure effect*, and $\hat{\Delta}^{v}_{X,p}$ reflects the *pure composition effect*. As for the two error terms, only included in Firpo et al. (2018), $\hat{\Delta}^{v}_{S,e}$ is the *reweighting error*, used to evaluate the quality of the reweighting strategy, and $\hat{\Delta}^{v}_{X,e}$ denotes the *specification error*, due to misspecifications in the model (i.e., nonlinearities). $\hat{\Delta}^{v}_{S,e}$ tends toward zero when the samples managed are large; $\hat{\Delta}^{v}_{X,e}$ equals zero if the model is truly linear. The sum of $\hat{\Delta}^{v}_{X,p} + \hat{\Delta}^{v}_{X,e}$, on the other hand, shows the aggregate *composition effect* of the detailed decomposition, whereas $\hat{\Delta}^{v}_{S,p} + \hat{\Delta}^{v}_{S,e}$ reproduces its counterpart for the *structure effect*.

4. DATA

To explore the relationships described above, we focus on some of the most populous countries in the OECD area with available data. In particular, the selected sample represents almost 50% of the total OECD population.[9] We use the Luxembourg Income Study Database (LIS) for the following set of countries: Canada, Germany, Italy, Poland, the United States and Spain.[10]

With regard to income inequality and poverty analysis and compared to other data sources, such as the World Income Inequality Database (WIID) of the World Institute for Development Economics Research (WIDER) or the World Bank's PovcalNet, the LIS database includes a characteristic feature that is its greatest value-added. It provides access to a set of harmonized microdata files generated from survey data at the country level, and the information is provided by the national statistical agencies. These datasets permit us to handle standardized income distributions in several countries, allowing us to implement international comparisons with homogenous data. On the other hand, it should be emphasized that one of the original weaknesses of this dataset was its limited coverage. Nonetheless, several new countries have recently been included in the database.

[9] The latest information available can be checked here: https://data.oecd.org/pop/population.htm

[10] The criteria followed for the selection of countries was as follows: we identified the 20 most populous countries in the OECD based on the most recent data available and chose those in which the LIS variable *size of the locality of residence* was defined. This is the reason why important countries such as France and the United Kingdom, among others, are not analyzed in this study. Data for Spain are taken from the Spanish Family Budget Survey (EPF).

Therefore, the range of and possibilities for analysis have expanded not only to high-income countries but also to middle-income countries.

Regarding the variables included, the fundamental one is *size of the locality of residence*. We use it to delimit and distinguish what we have called large cities (areas with over 500,000 inhabitants) from the other territorial areas (those with fewer than 500,000 inhabitants). The definition of what a large city represents is complex, with a wide range of possibilities. In this paper we use the criterion established jointly by the OECD and the European Commission (EC) to define cities based on FUAs (Dijkstra and Poelman, 2012). Thus, we consider large cities to be those urban centers with the size labels 'XL' (500,000–1,000,000 inhabitants), 'XXL' (1,000,000–5,000,000 inhabitants) and 'Global city' (more than 5,000,000 inhabitants). The other group is made up of the rest of the categories: areas with less than 50,000 inhabitants and urban centers with the size labels 'S' (50,000–100,000 inhabitants), 'M' (100,000–250,000 inhabitants) and 'L' (250,000–500,000 inhabitants). Using FUAs as a reference has many advantages. Being composed of a city and its commuting zone, they encompass the economic and functional extent of cities based on daily movements of the people (OECD, 2012). Furthermore, the definition of FUAs aims at providing a functional definition of cities and their area of influence by maximising international comparability and overcoming the limitation of using purely administrative approaches. At the same time, the concept of FUAs, unlike other approaches, ensures a minimum link to the government level of the city or metropolitan area.

One advantage of this way of classifying large cities is that it allows the methodology described above to be approached with a binary variable: large cities and other areas. The main limit is that there can be a high degree of heterogeneity between large cities with less than 1,000,000 inhabitants, those between 1,000,000 and 5,000,000, or what we can call global cities, with more than 5,000,000. It is possible that the contribution to inequality of different potential factors is different in each of these urban areas. Addressing this with the available data is not possible, however, due to the small sample size of our own large cities variable in some countries, such as Germany (less than 3,000 observations) or Italy (less than 1,000 observations).

We use real equivalent disposable income, with all income values expressed in PPP 2017 USD. This variable is derived by dividing household disposable income by the square root of household size.[11] As controls we consider a set of variables related to different characteristics: geography and housing (tenure), household composition and living arrangements (household composition and the number of household members), sociodemographic characteristics (age, sex, marital status, immigration status, disability status, health status, and education) and labor market information (employment status and part-time employment status). A comprehensive definition of each variable can be found in the Appendix (see Table A.3).

Finally, we use two waves of LIS data for the countries mentioned above: one corresponding to the early years of the 21st century and the other corresponding to the most recently available data.[12] These two years are used for the implementation of the *dynamic approach* set out in the preceding section. For the *static approach,* we consider the most recent year available, making it possible to identify, at a more recent moment and for each country, the contributions of the different variables. Whereas the *dynamic approach* focuses exclusively on large cities, the static approach is estimated for both large cities and the other areas. In this way, the two methods applied provide a complementary and comprehensive view of the influence of some of the main factors contributing to income inequality in the chosen countries.

5. MAIN RESULTS

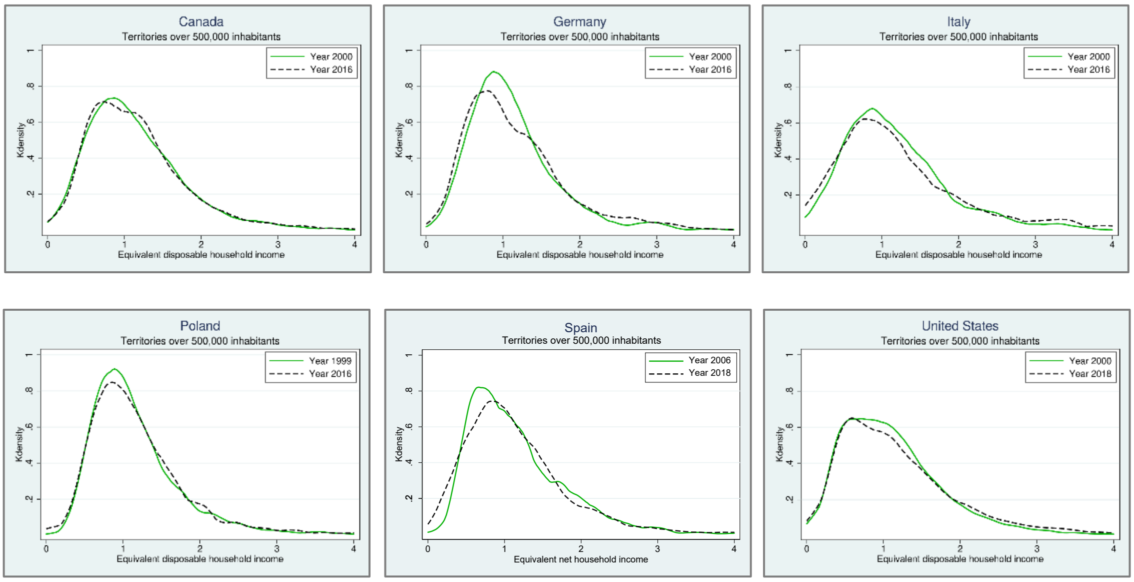
5.1. Descriptive analysis

Figure 1 and Figure 2 present general and descriptive evidence on the shape of the income distribution in large cities in the set of countries examined. The first figure presents the evolution of equivalent disposable income within large cities in the selected countries. The time period analyzed covers approximately 20 years, with one curve reflecting the disposable income distributions at the beginning of the 21st century and the other curve reflecting the distribution corresponding to the most recent year available.

Notably, the distributions in which a greater probability mass is concentrated near the median are those of the initial wave. These profiles anticipate an increase in inequality in large cities.

[11] Negative and zero incomes have been replaced by 1/100 of the mean to prevent relevant observations from being dropped by default.
[12] In the case of Spain, we use the EPF waves of 2006 and 2018. In 2006 a number of methodological improvements were incorporated, such as the change of periodicity (from quarterly to annual) as well as a notable increase in the sample size (up to 24,000 households).

Figure 1. Density of equivalent disposable household income in large cities (PPP 2017 USD)
Dynamic approach



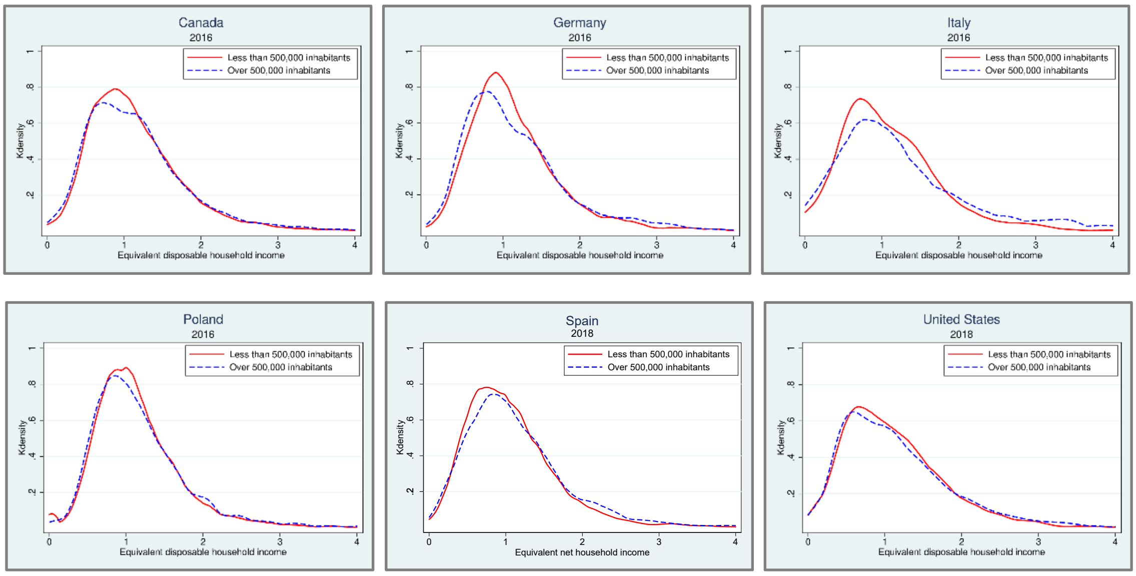
Note: Income values are expressed as a percentage of the median.
Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (multiple countries; 1999-2018). Luxembourg: LIS and EPF.

Figure 2 shows the divergence between the density of disposable income in large cities compared to the other areas. In general, the probability mass appears to be more concentrated around the median in areas with fewer than 500,000 inhabitants than in large cities, where the curves are flatter. The shape of the upper tail in the different distributions also seems to point in the same direction of higher inequality in large cities.

Other additional descriptive results that yield interesting information are, on the one hand, those related to the sample means of the socioeconomic characteristics examined and, on the other hand, several inequality measures (see Tables A3 to A8 in the Appendix).[13] First, it is worth noting that in four of the six countries under study, the percentage of areas with a population over 500,000 inhabitants has increased since the early 2000s. In the United States, which accounts for the greatest number of large cities, this percentage was close to 60% in 2018. Second, the reported mean differences in observable characteristics are significant for most variables in the different countries. The same can be said regarding the differences in the income distribution by percentile.[14]

[13] Other descriptive measures, such as skewness or kurtosis coefficients, are available upon request.
[14] The only exception is Italy with the *dynamic approach*.

Figure 2. Density of equivalent disposable household income in large cities and in other areas (PPP 2017 USD)
Static approach



Note: Income values are expressed as a percentage of the median.
Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (multiple countries; 1999-2018). Luxembourg: LIS and EPF.

In most countries inequality is higher in large cities than in areas with fewer than 500,000 inhabitants. This is true for all the inequality indicators considered except Poland (Gini index), although the difference is not significant. Within large cities, the values are also higher in the most recent year than at the beginning of the time period analyzed, although the differences are not significant in some cases.

5.2. Reweighted RIF-OLS decomposition: Dynamic approach

As stated before, income inequality as measured by the Gini index increased in large cities in all the countries analyzed except Poland. The differences in Gini values in the two years considered are determined, basically, by the *structure effect*. In the countries with significant results in the aggregate decomposition when using the Gini index, Germany and the United States, this effect has a relative importance of 92.2% and 71.3%, respectively.

Similar to the changes in the Gini coefficient, the changes in the three percentile ratios during the period analyzed are also mainly explained by the *structure effect*. The magnitude of this effect is noticeably higher in the bottom of the distribution in countries such as Germany and Poland. In others, such as the United States and Canada, the relative weight of the changes assigned to returns to characteristics is greater in the upper tail of the distribution.

[*Table 1a, 2a, 3a, 4a, 5a and 6a*]

As shown by the *unexplained* part of the model, which is attributable to differences in prices associated with characteristics, the most important factor contributing to inequality dynamics in large cities is the education premium. In keeping with the literature summarized above, changes in the returns to education (the skill premium) account for a remarkable amount of the rise in inequality in Canada, especially in the top half of the income distribution, and in the United States. In Germany, the trend observed is just the opposite, while in other European countries the contribution is lower than in North American countries. This difference warns against the possible generalisation of the results found in the literature —very focused on the US— to all OECD countries.

However, the changes in the returns to the other variables analyzed make it difficult to find a clear pattern among countries. Heterogeneity can be appreciated not only regarding the relative importance of the diverse controls but also when considering other results, such as the identification of the impact over the entire distribution or the sign of the direction of change. In three countries (Canada, Poland and Spain), changes in the returns to one-person households have a common feature: a reduction in income differences in the lower tail of the distribution. Additionally, the results for the variations in the returns to marital status are associated with a significant spread of income differences in Spain and the United States, especially at the half bottom of the distribution. As expected, the effect of the reweighting errors is practically imperceptible.

For the detailed decomposition of endowments, and as far as the *pure explained* component is involved, changes in educational attainment seem to have a great impact on increases in inequality differences. This result is found in Canada, Poland, Spain and the United States throughout the entire distribution and in Germany at the bottom half of the distribution. The magnitude of this impact—the extent to which inequality differences increase—is the largest of all the control variables. In relative terms, the impact of changes in educational attainment is greater when considering the lower part of the income distribution (P50-P10 ratio) in Spain and the United States, while in Canada it is slightly greater at the half top of the distribution. At the same time, reductions in the number of household members have a smaller effect on inequality in large cities than they did two decades ago.

5.3. Reweighted RIF-OLS decomposition: Static approach

We also analyze the differences between the contributions to income inequality in large cities and in the other areas—those with fewer than 500,000 inhabitants. Similar to the previous findings with the dynamic approach, the results here also confirm the widespread growth in Gini coefficient differences in most countries. As in the previous analysis, divergences among the main factors contributing to inequality in both areas are mainly explained by the *structure effect*. The two components of the aggregate decomposition are now significant in practically all countries. Germany (59.9% vs. 40.1%) and Spain (66.4% vs. 33.6%) are, in this order, the two countries with the most balanced percentages. On the other hand, the *structure effect* has a higher relative importance in the bottom half of the distribution in Canada, Germany, Italy and the United States.

[*Table 1b, 2b, 3b, 4b, 5b and 6b*]

Considering the other three inequality measures (P90-P10, P90-P50 and P50-P10 ratios), and beginning with the *structure effect*, the first relevant result is the change in the returns to education (skill premium), which accounts for an important amount of the increase in inequality in the United States. In Germany, the trend is the opposite again. The results, therefore, corroborate the differential effect of the wage premium on inequality in large cities in the United States and its smaller contribution in European countries. The partial contributions of the rest of the potential drivers, in contrast to what is observed for the *composition effect*, barely allow us to establish a minimum correspondence or similarity among countries. In contrast, they are characterized by a remarkable lack of uniformity as well as by differences that are less significant.

Regarding the *composition effect*, changes in education have a notable effect on the income gaps between large cities and the rest of the areas. The contribution of the maximum level of education achieved is again the key driver of the *composition effect*. This is so not only because of the magnitude of the effect, which is the greatest, but also because of its significance throughout the whole distribution for each one of the six countries. Likewise, the impact is much more pronounced in the lower tail of the distribution (the P50-P10 ratio) for all countries. This result is in accordance with the literature reviewed (Bacolod et al., 2009; Baum-Snow and Pavan, 2013), being the consequence of the growth of more highly skilled and better paid workers in large cities—the 'paradox of progress' (Bourguignon et al., 2005). The wages and the number of skilled workers have increased more in large cities, while the wage of unskilled workers have remained constant or decreased in those areas.

Another variable playing a prominent role in inequality differences between large cities and other areas with regard to endowments is age. In large cities from countries such as Canada, Germany, Spain and the United States, aging has a strong negative effect on inequality. This result is consistent with the findings of Alimi et al. (2018), who found that metropolitan areas experienced rapid growth in inequality but slower rates of aging, which is mainly attributable to net inward migration rather than greater fertility, while nonmetropolitan areas had slower growth in inequality and faster rates of aging.

In short, our results show that during the first decades of the 21st century inequality increased in large cities and its level remained higher than in other areas. In most countries, the most important variables determining these results are the evolution of the wage premium and changes in the composition of the population by educational level. In any case, there are singularities specific to each country that determine the relationship between city size and inequality, the most relevant being changes in the demographic characteristics of each country.

In a country-by-country analysis, our results for Germany are in accordance with those of Biewen and Juhasz (2012) for an earlier period (2000–2006), who attributed the increase in income inequality to changes in the restructuration of household organization and to other variations in some relevant socioeconomic characteristics such as age or education. Czyż and Hauke (2011) checked that the development-activating elements have not managed to reduce inter-territorial differences in Poland and could explain the increase we see in our results. In Canada, some factors driving the recent evolution of inequality within Canadian provinces could be the roles played by human capital and the life cycle (inter-temporal dynamics) (Gray et al., 2004). Regarding Spain, Tirado et al. (2016) confirmed the existence of great disparities between the most prosperous Spanish areas, those of the North-east, and the poorest territories, located in the South. Finally, in the United States spatial-specific income dynamics characterized by segmented income classes of neighbours, among others, could be the reason for the increase in income inequality from 2000 to 2016 (Lee and Rodríguez-Pose, 2013).

6. CONCLUSIONS

The comparative study of the relationships between income inequality and city size has traditionally been constrained by several limitations. The lack of empirically well-versed models, the inadequate understanding of variable interactions and data restrictions have been severe barriers to the construction of suitable measurements and an understanding of their interactions.

This paper contributes to a better understanding of this relationship. We have conducted empirical research focused on inequality differences in large cities and other areas within a selected sample of OECD countries. The goal has been to identify and quantify the individual contributions of different factors to these differences. For this purpose, we have used a methodological approach that allows to look at these differences both from a dynamic and static perspective.

One of the main findings of this paper is the remarkable increase in income inequality in large cities in the selected countries. This is found with both approaches and with different inequality measures. These differences are mainly determined by a *structure effect* rather than changes in the distribution of endowments. The evidence obtained sheds light on the relevance of education as one of the most important factors contributing to income inequality in these areas.

We have examined income inequality trends in large cities for a period covering close to two decades. Changes in the skill premium account for a large amount of the increase in inequality in Canada—noticeably, in the top half of the income distribution—and in the United States, but in some European countries the trend is the opposite. This result cautions against generalisations of the US results, which might invite us to focus on the effect of the skill premium on inequality dynamics in large cities. According to our results, in European countries there are other relevant factors contributing to the changes of inequality in these areas. The returns to the geographical and housing control variables, as well as those related to labor market status, seem to describe a more idiosyncratic behavior in each country. Regarding endowments, changes in educational attainment seem to also play a central role in shaping income inequality in large cities over time. Household size is also observed to increase inequality in several countries.

We have also compared large cities with other areas. In keeping with previous studies, we find that inequality is higher in the former and the returns to educational skills are also higher in larger agglomerations of the United States. However, the opposite result is observed in other countries, with larger cities exhibiting smaller returns for high-skilled workers. In the case of composition effects, changes in educational attainment play a predominant role in explaining the growth in income inequality. This ‘paradox of progress’ reveals “a rightward movement in the distribution of years of schooling shifts population density to steeper segments of the earnings-education profile, leading to wider earnings gaps” (Ferreira et al., 2017). Regarding demography, the evidence shown reveals a negative age-composition effect in most of the countries examined. Household composition and household size also play a relevant role in almost all European countries studied.

In short, we contribute to a growing body of empirical evidence with a more accurate measurement of the effects of different factors on income inequality in large cities by considering a larger number of countries and a dual static and dynamic perspective. It is especially relevant to clarify the reasons behind the recent rise in income inequality in areas where an increasing percentage of the population resides. Although what we have done is primarily an identification of the factors that contribute to inequality in large cities, our work can contribute to future research in which the main drivers explaining these contributions can be analysed in sufficient depth.

REFERENCES

Alimi, O. B., Maré, D. C., Poot, J. (2018). More pensioners, less income inequality? The impact of changing age composition on inequality in big cities and elsewhere." In Stough, R.R., Kourtit, K., Nijkamp, P., Blien, U. (Eds.), Modelling aging and migration effects on spatial labor markets, Springer, Cham., pp. 133-159.

Alperovich, G., 1995. The Relationship between Income Inequality and City Size: A General Equilibrium Model of an Open System of Cities Approach, Urban Studies 32, 853-862.

Bairoch, P., 1988. Cities and economic development. From the dawn of history to the present, The University Chicago Press.

Barsky, R., Bound, J., Charles, K. K., Lupton, J. P., 2002. Accounting for the black–white wealth gap: a nonparametric approach, Journal of the American statistical Association 97, 663-673.

Baum-Snow, N., Pavan, R., 2013. Inequality and city size, Review of Economics and Statistics 95, 1535-1548.

Baum-Snow, N., Freedman, M., Pavan, R., 2018. Why has urban inequality increased, American Economic Journal: Applied Economics 10, 1–42.

Bacolod, M., Blum, B. S., Strange, W.C., 2009. Skills in the city, Journal of Urban Economics 65, 136-153.

Biewen, M., Juhasz, A., 2012. Understanding rising income inequality in Germany, 1999/2000–2005/2006. Review of Income and Wealth 58, 622–647.

Blinder, A. S., 1973. Wage discrimination: reduced form and structural estimates, Journal of Human Resources 8, 436-455.

Boulant, J., Brezzi, M., Veneri, P., 2016. Income levels and Inequality in Metropolitan Areas: A comparative Approach in OECD Countries, OECD Publishing, Paris.

Bourguignon, F., Ferreira, F., Lustig, N., 2005. The Microeconomics of Income Distribution Dynamics in East Asia and Latin America. Oxford University Press and World Bank, Washington DC.

Castells-Quintana, D., Royuela, V., Veneri, P., 2020. Inequality and city size: An analysis for OECD functional urban areas, Papers in Regional Science 99, 1045-1064.

Chen, B., Liu, D., Lu, M., 2018. City size, migration and urban inequality in China, China Economic Review 51, 42-58.

Czyż, T., Hauke, J., 2011. Evolution of regional disparities in Poland. Quaestiones Geographicae 30, 35–48.

Cloutier, N.R., 1997. Metropolitan income inequality during the 1980s: The impact of urban development, industrial mix an, and family structure, Journal of Regional Science 37, 459-478.

Davis, D.R., Dingel, J.I., 2019. A spatial knowledge economy, American Economic Review 109, 153-170.

De la Roca, J., Puga, D., 2016. Learning by Working in Big Cities. Review of Economic Studies 84, 106–142.

Dijkstra, L., Poelman, H., 2012. Cities in Europe: the new OECD-EC definition, Regional Focus 1, 1-13.

DiNardo, J. M. Fortin N., Lemieux, T., 1996. Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. Econometrica 64, 1001-1044.

Dubey, A., Mahadevia, D., 2001. Poverty and inequality in Indian metropolises, Indian Journal of Labour Economics 44, 223-242.

Ferreira, F. H., Firpo, S. P., Messina, J., 2017. Ageing Poorly? Accounting for the decline in earnings inequality in Brazil, 1995–2012, The World Bank, Washington DC.

Firpo, S. P., Fortin, N. M., Lemieux, T., 2009. Unconditional quantile regressions, Econometrica 77, 953-973.

Firpo, S. P., Fortin, N. M., Lemieux, T., 2018. Decomposing wage distributions using recentered influence function regressions, Econometrics 6, 1-40.

Fortin, N., Lemieux, T., Firpo, S., 2011. Decomposition methods in economics. In Ashenfelter, O., Card, D. (Eds.), Handbook of labor economics (Vol. 4, part A), Elsevier, Amsterdam, pp. 1-102.

Galster, G., McCorkhill, G., Gopalan, S., 1988. The Determinants of Income Inequality, In Metropolitan Areas, Review of Business 10, 17-22.

Garofalo, G., Fogarty, M.S., 1979. Urban Income Distribution and the Urban Hierarchy-Equality Hypothesis, Review of Economics and Statistics 61, 381-388.

Glaeser, E., Resseger, M., Tobio, C., 2009. Inequality in cities, Journal of Regional Science 49, 617–646.

Gould, E.D., 2007. Cities, Workers, and Wages: A Structural Analysis of the Urban Wage Premium, Review of Economic Studies 74, 477-506.

Gradín, C., 2016. Why is Income Inequality so High in Spain? In Income Inequality Around the World (Research in Labor Economics, Vol. 44), Emerald Group Publishing Limited, pp. 109-177.

Gray, D., Mills, J. A., Zandvakili, S., 2004. An analysis of differential provincial income inequality trends in Canada. In J. A. Bishop, & Y. Amiel (Eds.), Studies on economic well-being: Essays in the honor of John P. Formby. Bingley: Emerald Group Publishing.

Haque, I., Das, D. N., Patel, P. P., 2018. Spatial segregation in Indian cities- does the city size matter? Environment and Urbanization Asia 9, 52–68.

Henkel, M., 2017. Decomposing the role of urbanization in wage inequality. In Three essays on the causes and consequences of spatial inequality, Doktors dissertation. Universität Duisburg-Essen.

Hortas-Rico, M., Rios, V., 2019. The drivers of local income inequality: a spatial Bayesian model-averaging approach, Regional Studies 53, 1207–1220.

Juhn, C., Murphy, K. M., Pierce, B. 1993. Wage inequality and the rise in returns to skill, Journal of Political Economy 101, 410-442.

Lee, N., Rodríguez-Pose, A., 2013. Innovation and spatial inequality in Europe and USA. Journal of Economic Geography 13, 1–22.

Lee N, Sissons P, Jones, K, 2016. The Geography of Wage Inequality in British Cities. Regional Studies 50, 1714-1727.

Machado, J., Mata, J.A.F., 2005. Counterfactual decomposition of changes in wage distributions using quantile regression. Journal of Applied Econometrics 20, 445-465.

Melly, B., 2006. Estimation of counterfactual distributions using quantile regression (mimeo), Swiss Institute for International Economics and Applied Economic Research, University of St. Gallen.

Nord, S., 1980a. An Empirical Analysis of Income Inequality and City Size. Southern Economic Journal 46, 863-872.

Nord, S., 1980b. Income inequality and city size: An examination of alternative hypothesis for large and small cities. Review of Economics and Statistics 62, 502-508.

Oaxaca, R. L., 1973. Male-female wage differentials in urban labor markets. International Economic Review 14, 693-709.

Oaxaca, R. L., Ransom, M. R., 1999. Identification in detailed wage decompositions, Review of Economics and Statistics 81, 154-157.

OECD, 2012. Redefining "Urban": A New Way to Measure Metropolitan Areas. OECD Publishing, Paris.

OECD, 2020. Cities Policy Responses, Tackling Coronavirus (Covid-19): Contributing to a Global Effort, OECD Publishing, Paris.

Rothe, C., 2015. Decomposing the composition effect: the role of covariates in determining between-group differences in economic outcomes. Journal of Business & Economic Statistics 33, 323-337.

Royuela, V., Veneri, P., Ramos, R., 2014. Income Inequality, Urban Size and Economic Growth in OECD Regions. OECD Regional Development Working Papers, 2014/10, OECD Publishing.

Soroka, L.A., 1984. City size and income distributions: The Canadian experience. Urban Studies 21, 359-365.

Tirado, D. A., Díez-Minguela, A., Martínez-Galarraga, J., 2016. Regional inequality and economic development in Spain, 1860–2010. Journal of Historical Geography 54, 87–98.

United Nations, 2018. The World's Cities in 2018. (ST/ESA/ SER.A/417). Department of Economic and Social Affairs, Population Division, New York: United Nations.

United Nations, 2019. World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420). Department of Economic and Social Affairs, Population Division, New York: United Nations.

Wheeler, C., 2004a. On the distributional aspects of urban growth. Journal of Urban Economics 55, 371-397.

Wheeler, C., 2004b. Wage inequality and urban density, Journal of Economic Geography 4,421-437.

Yun, M., 2005. A Simple Solution to the Identification Problem in Detailed Wage Decompositions, Economic Inquiry 43,766-772.

Table 1a. RIF-OLS decomposition results with reweighting

CANADA

	Dynamic approach Areas with over 500,000 inhabitants			
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
Overall				
Year 2016	53,038.79***	33,793.29***	19,245.50***	29.718***
Year 2000	41,469.61***	25,318.18***	16,151.43***	29.199***
Difference/Total change	11,569.18***	8,475.11***	3,094.07***	0.519
[1] Composition effect	1,417.65*** (12.25%)	659.83 (7.79%)	757.82*** (24,49%)	0.571*** (110.02%)
[2] Structure effect	10,151.53*** (87.75%)	7,815.28*** (92.21%)	2,336.25*** (75,51%)	-0.052 -(10.02%)
[1] *Composition effect (Explained)*				
$\hat{\Delta}^{v}_{X,p}$= Pure explained	1,953.84***	999.50***	954.35***	0.730***
Tenure	20.13	-4.89	25.03	-0.031
Household composition	15.57	-33.15	48.71	0.052
Household size	118.31	37.80	80.51*	-0.020
Age	665.28***	432.88**	232.39**	0.171
Sex	11.26	43.22	-31.97	0.063
Marital status	-98.37	-56.43	-41.94	-0.020
Immigrant	-	-	-	-
Disabled	-	-	-	-
Health status	-	-	-	-
Education	853.59***	449.63***	403.96***	0.158*
Employed	-	-	-	-
Part-time employment	368.07***	130.42**	237.65***	0.358***
$\hat{\Delta}^{v}_{X,e}$ = Specification error	-536.19*	-339.66	-196.52	-0.159***
[2] *Structure effect (Unexplained)*				
$\hat{\Delta}^{v}_{S,p}$ = Pure unexplained	10,117.77***	7,808.37***	2,309.40***	-0.041
Tenure	2,616.15	264.15	2,351.99***	-0.874
Household composition	-1,488.86*	-363.15	-1,125.71***	-0.371
Household size	-10,568.61***	-7,046.02*	-3,522.04***	-2.568
Age	4,053.91	-3,799.41	7,853.33	10.072
Sex	-9,630.90***	-7,686.69**	-1,944.21	-3.172*
Marital status	3,094.94**	2,953.20*	141.74	0.757
Immigrant	-	-	-	-
Disabled	-	-	-	-
Health status	-	-	-	-
Education	2,870.80**	2,535.59**	336.21	0.318
Employed	-	-	-	-
Part-time employment	-367.61	-210.88	-156.73	-0.033
Constant	19,536.96	21,161.59	-1,624.63	-4.172
$\hat{\Delta}^{v}_{S,e}$ = Reweighting error	33.75	6.91	26.84	-0.011

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age aggregates the original age variable and age squared. Education includes all three education categories: low, medium and high. (4) Some sums may not match perfectly due to rounding. (5) Empty cells are due to a lack of data.
Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (Canada; 2000-2016). Luxembourg: LIS.

Table 1b. RIF-OLS decomposition results with reweighting
CANADA

	Static approach Year 2016			
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
Overall				
Over 500,000 inhabitants	53,038.79***	33,793.29***	19,245.50***	29.718***
Fewer than 500,000 inhabitants	48,525.05***	30,840.31***	17,684.73***	28.022***
Difference/Total change	4,513.74***	2,952.98***	1,560.76***	1.696***
[1] Composition effect	-1,004.67***	-486.34*	-518.32***	0.177*
	-(22.26%)	-(16.47%)	-(33,21%)	(10.44%)
[2] Structure effect	5,518.41***	3,439.32***	2,079.09***	1.519***
	(122.26%)	(116.47%)	(133,21%)	(89.56%)
[1] *Composition effect (Explained)*				
$\hat{\Delta}^{v}_{X,p}$= Pure explained	-930.95***	-552.15***	-378.79***	0.222**
Tenure	-134.91	130.43	-265.34***	0.454***
Household composition	-55.02	-31.65	-23.37	0.016
Household size	-736.38***	-379.93***	-356.46***	-0.119***
Age	-378.92***	-466.64***	87.72**	-0.054
Sex	-198.73***	-155.55***	-43.18**	6.90e-05
Marital status	-14.09	-5.06	-9.03	-4.56e-04
Immigrant	-	-	-	-
Disabled	-	-	-	-
Health status	-	-	-	-
Education	623.52***	361.62***	261.89***	-0.026
Employed	-	-	-	-
Part-time employment	-36.40	143.99	-31.02	-0.048
$\hat{\Delta}^{v}_{X,e}$ = Specification error	-73.72	65.81	-139.53	-0.046***
[2] *Structure effect (Unexplained)*				
$\hat{\Delta}^{v}_{S,p}$ = Pure unexplained	5,438.63***	3,396.69***	2,041.94***	1.531***
Tenure	-3,516.69	-1,150.14	-2,366.55	1.845
Household composition	274.67	286.53	-11.85	-0.066
Household size	-2,856.18	-1,632.13	-1,224.04	-0.187
Age	-12,834.48	-9,923.35	-2,911.13	-2.218
Sex	-1,141.62	-462.90	-678.72	-2.304*
Marital status	3,004.44**	3,306.38**	-301.95	1.385**
Immigrant	-	-	-	-
Disabled	-	-	-	-
Health status	-	-	-	-
Education	1,231.37	902.20	-329.17	0.656
Employed	-	-	-	-
Part-time employment	-240.09	-254.15	14.07	0.092
Constant	21,517.19	12,324.25	9,192.95	2.326
$\hat{\Delta}^{v}_{S,e}$ = Reweighting error	79.78***	42.63***	37.15***	-0.012

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age aggregates the original age variable and age squared. Education includes all three education categories: low, medium and high. (4) Some sums may not match perfectly due to rounding. (5) Empty cells are due to a lack of data.
Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (Canada; 2000-2016). Luxembourg: LIS.

Table 2a. RIF-OLS decomposition results with reweighting
GERMANY

	Dynamic approach Areas with over 500,000 inhabitants			
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
Overall				
Year 2016	49,221.64***	32,506.90***	16,714.74***	32.241***
Year 2000	38,770.78***	23,420.82***	15,349.96***	26.495***
Difference/Total change	10,450.86***	9,086.08***	1,364.78***	5.746***
[1] Composition effect	2,376.09 (22.74%)	2,324.87 (25.59%)	51.22 (3.75%)	0.448 (7.80%)
[2] Structure effect	8,074.78*** (77.26%)	6,761.21** (74.41%)	1,313.56*** (96.25%)	5.298*** (92.20%)
[1] *Composition effect (Explained)*				
$\hat{\Delta}^{v}_{X,p}$ = Pure explained	693.23	919.20	-225.97	0.258
Tenure	323.58	183.37	140.21	0.034
Household composition	618.43	577.15	41.28	-0.135
Household size	-1,801.45**	-1,190.70	-610.75*	-0.054
Age	130.75	169.16	-38.41	-0.235
Sex	-360.66	-303.87	-56.78	-0.188
Marital status	95.15	89.78	5.37	-0.013
Immigrant	787.90	1,228.86*	-440.95	0.273
Disabled	31.20	35.00	-3.80	0.023
Health status	-498.78	-380.21	-118.57	-0.038
Education	648.46	270.29	378.17*	0.031
Employed	95.54	-7.37	102.91	0.073
Part-time employment	623.11*	247.74	375.37*	0.486**
$\hat{\Delta}^{v}_{X,e}$ = Specification error	1,682.86	1,405.67	277.19	0.190
[2] *Structure effect (Unexplained)*				
$\hat{\Delta}^{v}_{S,p}$ = Pure unexplained	7,399.95**	6,251.51**	1,148.44	5.220***
Tenure	2,139.80	1,361.48	778.31	1.042
Household composition	850,94	1,122.00	-271.06	1.345
Household size	2,866.67	5,599.46	-2,732.79	3.507
Age	-33,216.85	-45,325.33	12,108.49	-10.923
Sex	3,204.68	2,275.28	929.40	1.328
Marital status	2,340.14	160.37	2,179.77	1.151
Immigrant	-6,567.80***	-5,357.84**	-1,209.96	-1.919**
Disabled	372.61	297.38	75.23	0.124
Health status	-182.83	603.67	-786.50	2.896
Education	-3,938.43*	-2,274.55	-1,163.89	-2.016*
Employed	-4,462.70	4,284.55	-8,747.25	-9.380
Part-time employment	-2,415.16	-379.18	-2,035.98*	-0.722**
Constant	46,408.88	44,384.22	2,024.66	18.785
$\hat{\Delta}^{v}_{S,e}$ = Reweighting error	674.82	509.70	165.12	0.077

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age aggregates the original age variable and age squared. Education includes all three education categories: low, medium and high. Health status: gathers all five health status groupings: very good, good, satisfactory, poor and bad. (4) Some sums may not match perfectly due to rounding.
Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (Germany; 2000-2016). Luxembourg: LIS.

Table 2b. RIF-OLS decomposition results with reweighting GERMANY

	Static approach Year 2016			
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
Overall				
Over 500,000 inhabitants	49,221.64***	32,506.90***	16,714.74***	32.241***
Fewer than 500,000 inhabitants	39,309.42***	25,247.03***	14,062.39***	28.289***
Difference/Total change	9,912.22***	7,259.87***	2,652.35***	3.952***
[1] Composition effect	691.50 (6.98%)	568.62 (7.83%)	122.88 (4.63%)	1.586*** (40.13%)
[2] Structure effect	9,220.72*** (93.02%)	6,691.25*** (92.17%)	2,529.47*** (95.37%)	2.366* (59.87%)
[1] *Composition effect (Explained)*				
$\hat{\Delta}^{v}_{X,p}$ = Pure explained	1,501.26***	1,375.66***	125.60	2.003***
Tenure	-1,806.68***	-790.79***	-1,015.89***	0.185
Household composition	-700.41***	-354.92*	-345.49***	0.067
Household size	1,402.20***	713.40***	688.80***	0.388***
Age	-654.95***	-699.54***	44.59	-0.190*
Sex	11.87	8.69	3.17	0.009
Marital status	-36.33	-7.64	-28.70	0.391***
Immigrant	-208.12*	82.19	-290.32***	0.105
Disabled	65.99	57.96	8.03	0.060*
Health status	309.35***	266.11**	43.24	0.025
Education	3,037.69***	2,083.01***	954.69***	1.012***
Employed	101.60	33.73	67.88	0.091
Part-time employment	-20.95	-16.55	-4.40	-0.141**
$\hat{\Delta}^{v}_{X,e}$ = Specification error	-809.76**	-807.04***	-2.72	-0.417***
[2] *Structure effect (Unexplained)*				
$\hat{\Delta}^{v}_{S,p}$ = Pure unexplained	9,339.22***	6,795.53***	2,543.69***	2.405*
Tenure	2,443.89*	2,252.14	191.76	1.692**
Household composition	-1,201.38	-1,308.75	107.37	0.935
Household size	-3,855.40	-2,749.64	-1,105.77	7.931
Age	-51,254.37**	-45,941.39**	-5,312.98	-31.813*
Sex	2,844.87	-511.55	3,396.42	5.225
Marital status	5,133.14**	3,540.31	1,592.83*	1.785
Immigrant	-2,634.06**	-1,966.11	-667.95	-0.503
Disabled	-112.51	-89.38	-23.13	0.152
Health status	-748.40	-817.10	68.70	-0.013
Education	-467.08	-670.12	203.04	-1.236*
Employed	2,341.17	3,472.65	-1,131.47	2.224
Part-time employment	-39.19	723.73	-762.91	0.226
Constant	56,848.54**	50,860.76**	5,987.78	15.799***
$\hat{\Delta}^{v}_{S,e}$ = Reweighting error	-118.50*	-104.28**	-14.22	-0.039

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age aggregates the original age variable and age squared. Education includes all three education categories: low, medium and high. Health status: gathers all five health status groupings: very good, good, satisfactory, poor and bad. (4) Some sums may not match perfectly due to rounding.

Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (Germany; 2000-2016). Luxembourg: LIS.

Table 3a. RIF-OLS decomposition results with reweighting
ITALY

	Dynamic approach Areas with over 500,000 inhabitants			
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
Overall				
Year 2016	45,265.33***	31,587.85***	13,677.48***	36.234***
Year 2000	40,516.49***	27,391.27***	13,125.22***	31.791***
Difference/Total change	4,748.84	4,196.57	552.26	4.443*
[1] Composition effect	10,545.85 (222.07%)	10,598.16 (252.54%)	-52.31 -(9.47%)	1.759 (39.59%)
[2] Structure effect	-5,797.01 -(122.07%)	-6,401.58 -(152.54%)	604.57 (109.47%)	2.684 (60.41%)
[1] *Composition effect (Explained)*				
$\hat{\Delta}^{v}_{X,p}$ = Pure explained	3,509.39	3,125.31	384.08	-0.018
Tenure	516.19	569.29	-53.10	-0.397
Household composition	-2,951.46	-2,288.95	-662.54	-0.067
Household size	2,980.92	1,833.38	1,147.53	-1.575
Age	2,536.38	1,992.39	543.98	2.784
Sex	-2,002.36	-884.87	-1,117.48*	-0.959
Marital status	-152,94	11.63	-164.56	0.341
Immigrant	-30.89	75.55	-106.44	0.029
Disabled	-	-	-	-
Health status	-	-	-	-
Education	3,156.53	2,115.39	1,041.13	0.087
Employed	-	-	-	-
Part-time employment	-542.98	-298.52	-244.46	-0.259
$\hat{\Delta}^{v}_{X,e}$ = Specification error	7,036.46	7,472.85	-436.39	1.777*
[2] *Structure effect (Unexplained)*				
$\hat{\Delta}^{v}_{S,p}$ = Pure unexplained	-13,243.65	-14,472.03	1,228.38	0.501
Tenure	8,230.72	3,661.64	4,569.07	2.669
Household composition	-26,538.89	-29,700.45*	3,161.57	-8.051
Household size	-19,811.01	-26,851.01	7,040.00	-5.869
Age	193,930.70	146,042.30	47,888.40	100.654
Sex	26,163.07	14,189.26	11,973.81*	-16.016*
Marital status	-2,080.97	-1,026.83	-1,054.14	-7.962
Immigrant	2,589.95	1,916.04	673.91	0.360
Disabled	-	-	-	-
Health status	-	-	-	-
Education	-532.63	-332.31	-200.31	-0.251
Employed	-	-	-	-
Part-time employment	-456.79	-684.46	227.67	1.029
Constant	-194.737.80	-121,686.20	-73,051.59**	-66.059
$\hat{\Delta}^{v}_{S,e}$ = Reweighting error	7,446.64	8,070.45	-623.81	2.182*

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age aggregates the original age variable and age squared. Education includes all three education categories: low, medium and high. Health status: gathers all five health status groupings: very good, good, satisfactory, poor and bad. (4) Some sums may not match perfectly due to rounding. (5) Empty cells are due to a lack of data.
Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (Italy; 2000-2016). Luxembourg: LIS.

Table 3b. RIF-OLS decomposition results with reweighting
ITALY

	Static approach Year 2016			
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
Overall				
Over 500,000 inhabitants	45,265.33***	31,587.85***	13,677.48***	36.234***
Fewer than 500,000 inhabitants	30,310.16***	18,317.70***	11,992.46***	29.296***
Difference/Total change	14,955.17***	13,270.15***	1,685.01**	6.938***
[1] Composition effect	3,911.50*** (26.15%)	3,870.15*** (29.16%)	41.36 (2.45%)	0.812 (11.70%)
[2] Structure effect	11,043.67** (73.85%)	9,400.00** (70.84%)	1,643.66** (97.55%)	6.126** (88.30%)
[1] *Composition effect (Explained)*				
$\hat{\Delta}^{v}_{X,p}$ = Pure explained	2,798.55***	2,326.53***	472.03	0.837
Tenure	7.55	-2.57	10.12	-0.019
Household composition	-456.34	-374.99	-81.34	-0.177
Household size	142.42	210.87	-68.44	-0.539*
Age	766.20	707.62	58.58	0.672
Sex	140.24	133.16	7.07	0.161
Marital status	-154.50	-155.95	1.44	0.013
Immigrant	5.06	-15.20	20.26	-0.021
Disabled	-	-	-	-
Health status	-	-	-	-
Education	2,398.49***	1,781.14***	617.35***	1.100**
Employed	-	-	-	-
Part-time employment	-50.55	42.46	-93.01	-0.353***
$\hat{\Delta}^{v}_{X,e}$ = Specification error	1,112.95	1,543.62*	-430.67	-0.025
[2] *Structure effect (Unexplained)*				
$\hat{\Delta}^{v}_{S,p}$ = Pure unexplained	10,955.83**	9,330.58**	1,625.25**	6.146**
Tenure	-557.44	-2,779.79	2,222.35	2.908
Household composition	-1,301.26	-1,903.00	871.75	-2.107
Household size	14,320.88	10,691.86	3,629.03	5.506
Age	-5,532.37	19,006.97	-24,559.34	42.669
Sex	-10,833.86	-12,480.40	1,646.54	-11.772
Marital status	-9,140.04	-8,587.29	-552.75	-9.816*
Immigrant	-522.54	73.78	-596.32	0.014
Disabled	-	-	-	-
Health status	-	-	-	-
Education	-185.63	-70.47	-115.15	-0.302
Employed	-	-	-	-
Part-time employment	381.79	230.25	151.54	0.862
Constant	24,076.29	5,148.68	18,927.60	-21.816
$\hat{\Delta}^{v}_{S,e}$ = Reweighting error	87.83	69.42	18.41	-0.020

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age aggregates the original age variable and age squared. Education includes all three education categories: low, medium and high. Health status: gathers all five health status groupings: very good, good, satisfactory, poor and bad. (4) Some sums may not match perfectly due to rounding. (5) Empty cells are due to a lack of data.
Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (Italy; 2000-2016). Luxembourg: LIS.

Table 4a. RIF-OLS decomposition results with reweighting
POLAND

	Dynamic approach Areas with over 500,000 inhabitants			
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
Overall				
Year 2016	28,659.98***	18,536.37***	10,123.62***	29.362***
Year 1999	17,867.00***	12,051.07***	5,815.93***	30.843***
Difference/Total change	10,792.98***	6,485.29***	4,307.69***	-1.480
[1] Composition effect	4,000.26*** (37.06%)	2,999.07*** (46.24%)	1,001.19*** (23.24%)	0.127 -(8.58%)
[2] Structure effect	6,792.73*** (62.94%)	3,486.23** (53.76%)	3,306.50*** (76.76%)	-1.607 (108.58%)
[1] *Composition effect (Explained)*				
$\hat{\Delta}^{v}_{X,p}$ = Pure explained	4,202.32***	2,823.80***	1,378.52***	1.376
Tenure	670.90***	443.71**	227.19***	0.704
Household composition	196.19***	271.87	-75.68	1.029
Household size	740.77	501.54*	211.20***	-0.873
Age	-191.44	-287.58	-201.62	-1.066***
Sex	-60.11	-42.09	161.09	-0.147
Marital status	-858.48***	-700.06***	1,100.86*	-0.235
Immigrant	-	-	-	-
Disabled	-3.99	-11.02	7.03	-0.012
Health status	-	-	-	-
Education	3,705.78***	2,653.17***	1,052.60***	1.952*
Employed	-	-	-	-
Part-time employment	2.71	-5.74	8.45	0.026
$\hat{\Delta}^{v}_{X,e}$ = Specification error	-202.06	175.27	-377.33**	-1.249***
[2] *Structure effect (Unexplained)*				
$\hat{\Delta}^{v}_{S,p}$ = Pure unexplained	6,835.48***	3.502.29**	3,333.19***	-1.898
Tenure	429.64	-158.68	588.32	-1.379
Household composition	-2,200.98**	-1,547.12	-653.86**	-1.484*
Household size	287.60	446.67	-159.06	-2.895
Age	8,879.83	5,714.22	3,165.61	6.094
Sex	1,579.49	1,845.81	-266.32	6.838**
Marital status	-3,691.37*	-2,400.66	-1,290.71**	-1.757
Immigrant	-	-	-	-
Disabled	-220.87	-191.76	-29.10	-0.183
Health status	-	-	-	-
Education	-1,732.79	-1,606.39	-126.39	-1.213
Employed	-	-	-	-
Part-time employment	258.18	158.90	99.28	0.050
Constant	3,246.74	1,241.31	2,005.43	-5.966
$\hat{\Delta}^{v}_{S,e}$ = Reweighting error	-42.75	-16.06	-26.69	0.291*

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age aggregates the original age variable and age squared. Education includes all three education categories: low, medium and high. Health status: gathers all five health status groupings: very good, good, satisfactory, poor and bad. (4) Some sums may not match perfectly due to rounding. (5) Empty cells are due to a lack of data.
Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (Poland; 1999-2016). Luxembourg: LIS.

Table 4b. RIF-OLS decomposition results with reweighting
POLAND

	Static approach Year 2016			
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
Overall				
Over 500,000 inhabitants	28,659.98***	18,536.37***	10,123.62***	29.362***
Fewer than 500,000 inhabitants	20,571.21***	13,154.53***	7,416.68***	29.368***
Difference/Total change	8,088.77***	5,381.84***	2,706.93***	-0.006
[1] Composition effect	1,715.54*** (21.21%)	1,054.82*** (19.60%)	660.72*** (24.41%)	-0.917** (15.283%)
[2] Structure effect	6,373.23*** (78.79%)	4,327.02*** (80.40%)	2,046.21*** (75.59%)	0.911 -(15.183%)
[1] *Composition effect (Explained)*				
$\hat{\Delta}^{v}_{X,p}$ = Pure explained	2,203.18***	1,620.82***	582.36***	-0.637
Tenure	-372.56***	-112.81	-259.75***	-0.247***
Household composition	7.74	111.28	-103.54**	0.405*
Household size	357.21***	180.53*	176.68***	-0.317
Age	-169.18**	-23.91	-145.27***	-0.156**
Sex	-391.42***	-253.58***	-137.84***	0.021
Marital status	-339.74***	-243.32***	-96.42***	-0.243*
Immigrant	118.73*	141.75**	-23.02	0.229*
Disabled	8.19*	4.57	3.61*	0.004
Health status	-	-	-	-
Education	2,957.24***	1,848.37***	1,108.87***	-0.443
Employed	-	-	-	-
Part-time employment	26.97	-32.06**	59.03***	0.111***
$\hat{\Delta}^{v}_{X,e}$ = Specification error	-487.64*	-566.00**	78.36***	-0.280
[2] *Structure effect (Unexplained)*				
$\hat{\Delta}^{v}_{S,p}$ = Pure unexplained	6,371.96***	4,301.73***	2,070.23***	0.864
Tenure	-325.25	34.20	-359.46	-1.596
Household composition	-577.37	-339.49	-237.88	-0.716
Household size	-2,519.08	-2,921.45	402.37	-1.448
Age	20,072.80*	24,701.92**	-4,629.11	19.091*
Sex	281.55	-767.36	1,048.91	2.848
Marital status	-943.43	-823.36	-120.06	-1.532
Immigrant	-88.59	-187.78	99.19**	-0.035
Disabled	-63.57	-91.94	28.36	-0.012
Health status	-	-	-	-
Education	-73.99	237.36	-311.35	-0.149
Employed	-	-	-	-
Part-time employment	-30.66	90.57	-121.23	-0.162
Constant	-9,360.43	-15,630.93	6,270.50*	-15.424*
$\hat{\Delta}^{v}_{S,e}$ = Reweighting error	1.27	25.29	-24.02*	0.047

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age aggregates the original age variable and age squared. Education includes all three education categories: low, medium and high. Health status: gathers all five health status groupings: very good, good, satisfactory, poor and bad. (4) Some sums may not match perfectly due to rounding. (5) Empty cells are due to a lack of data.
Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (Poland; 1999-2016). Luxembourg: LIS.

Table 5a. RIF-OLS decomposition results with reweighting
SPAIN

	Dynamic approach Areas with over 500,000 inhabitants			
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
Overall				
Year 2018	2,028.98***	1,301.16***	727.82***	31.585***
Year 2006	1,926.10***	1,292.09***	634.01***	30.407***
Difference/Total change	102.88**	9.06	93.81***	1.177***
[1] Composition effect	34.04 (33.09%)	34.03 (375.61%)	0.01 (0.02%)	0.342 (29.06%)
[2] Structure effect	68.84 (66.91%)	-24.97 -(275.61%)	93.80*** (99.98%)	0.835 (70.94%)
[1] *Composition effect (Explained)*				
$\hat{\Delta}^{v}_{X,p}$ = Pure explained	94.15***	62.16***	31.99***	1.115***
Tenure	-19.43***	-4.89	-14.53***	0.299***
Household composition	0.29	0.46	-0.17	0.255***
Household size	-18.83***	-18.79***	-0.03	-0.438***
Age	20.92***	22.10***	-1.17	0.480***
Sex	3.68	1.39	2.29	0.502***
Marital status	0.30	-1.51	1.81	-0.073*
Immigrant	-8.27	-1.71	-6.55**	0.038
Disabled	-	-	-	-
Health status	-	-	-	-
Education	115.48***	65.13***	50.36***	0.052
Employed	-	-	-	-
Part-time employment	-	-	-	-
$\hat{\Delta}^{v}_{X,e}$ = Specification error	-60.11***	-28.13*	-31.98***	-0.773***
[2] *Structure effect (Unexplained)*				
$\hat{\Delta}^{v}_{S,p}$ = Pure unexplained	64.76	-23.31	88.07***	0.888*
Tenure	-296.46***	-6.72	-289.74***	-2.565***
Household composition	-14.21	12.89	-27.11***	-0.108
Household size	-50.59	-76.31	25.72	-2.276
Age	100.76	-694.05	794.82***	5.582
Sex	59.23	8.15	51.08*	0.574
Marital status	41.65	6.18	35.47**	-0.326
Immigrant	41.49**	32.60*	8.89	0.679***
Disabled	-	-	-	-
Health status	-	-	-	-
Education	13.39	26.11	-12.72	-0.416
Employed	-	-	-	-
Part-time employment	-	-	-	-
Constant	169.49	667.84	-498.35*	-2.276
$\hat{\Delta}^{v}_{S,e}$ = Reweighting error	4.08	-1.65	5.73**	0.053

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age aggregates the original age variable and age squared. Education includes all three education categories: low, medium and high. Health status: gathers all five health status groupings: very good, good, satisfactory, poor and bad. (4) Some sums may not match perfectly due to rounding. (5) Empty cells are due to a lack of data.
Source: Spanish *Encuesta de Presupuestos Familiares* (EPF).

Table 5b. RIF-OLS decomposition results with reweighting SPAIN

	Static approach Year 2018			
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
Overall				
Over 500,000 inhabitants	2,028.98***	1,301.16***	727.82***	31.585***
Fewer than 500,000 inhabitants	1,625.78***	995.22***	630.56***	30.045***
Difference/Total change	403.20***	305.94***	97.26***	1.540***
[1] Composition effect	93.75*** (23.25%)	61.45*** (20.09%)	32.30*** (24.41%)	0.518*** (33.63%)
[2] Structure effect	309.45*** (76.75%)	244.49*** (79.91%)	64.96*** (75.59%)	1.022*** (66.37%)
[1] *Composition effect (Explained)*				
$\hat{\Delta}^{v}_{X,p}$ = Pure explained	110.21***	73.43***	36.77***	0.853***
Tenure	13.67***	10.64***	3.02***	0.621***
Household composition	-1.85***	-0.37	-1.47***	0.075***
Household size	0.32	0.36	-0.04	-0.101***
Age	-11.45***	-6.35***	-5.10***	-0.099***
Sex	8.01***	5.93***	2.14***	0.111***
Marital status	-1.55**	-1.45**	-0.10	-0.008
Immigrant	5.33***	5.61***	-0.28	0.341***
Disabled	-	-	-	-
Health status	-	-	-	-
Education	97.67***	59.06***	38.61***	-0.087**
Employed	-	-	-	-
Part-time employment	-	-	-	-
$\hat{\Delta}^{v}_{X,e}$ = Specification error	-16.45***	-11.97*	-4.47**	-0.335***
[2] *Structure effect (Unexplained)*				
$\hat{\Delta}^{v}_{S,p}$ = Pure unexplained	310.27***	245.33***	64.94***	1.044***
Tenure	31.93	114.15*	-82.22***	1.444**
Household composition	204.42	13.09	-7.74	0.015
Household size	-514.18*	194.61**	9.81	-2.544
Age	520.99	251.67	269.31	1.004
Sex	36.72	22.22	14.50	0.982**
Marital status	12.69	7.31	5.38	0.070
Immigrant	9.76	5.49	4.27	-0.138
Disabled	-	-	-	-
Health status	-	-	-	-
Education	2.57	28.07	-25.50**	-0.366
Employed	-	-	-	-
Part-time employment	-	-	-	-
Constant	-514.18	-391.30	-122.87	-2.544
$\hat{\Delta}^{v}_{S,e}$ = Reweighting error	-0.82***	-0.83***	0.02	0.022***

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age aggregates the original age variable and age squared. Education includes all three education categories: low, medium and high. Health status: gathers all five health status groupings: very good, good, satisfactory, poor and bad. (4) Some sums may not match perfectly due to rounding. (5) Empty cells are due to a lack of data.
Source: Spanish *Encuesta de Presupuestos Familiares* (EPF).

Table 6a. RIF-OLS decomposition results with reweighting UNITED STATES

	Dynamic approach Areas with over 500,000 inhabitants			
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
Overall				
Year 2018	82,118.39***	55,769.74***	26,348.65***	37.074***
Year 2000	61,807.52***	38,573.80***	23,233.72***	34.491***
Difference/Total change	20,310.87***	17,195.94***	3,114.94***	2.582***
[1] Composition effect	3,904.41*** (19.22%)	3,008.79*** (17.50%)	895.62*** (28.75%)	0.739*** (28.62%)
[2] Structure effect	16,406.46*** (80.78%)	14,187.15*** (82.50%)	2,219.31*** (71.25%)	1.842*** (71.34%)
[1] *Composition effect (Explained)*				
$\hat{\Delta}^{v}_{X,p}$ = Pure explained	4,361.57***	3,320.48***	1,041.09***	0.974***
Tenure	-257.27***	-70.75*	-186.53***	0.121***
Household composition	-292.65***	-216.94***	-75.71***	-0.068**
Household size	797.16***	485.03***	312.13***	0.015
Age	1,732.28***	1,509.66***	222.63***	0.316**
Sex	-156.06***	-95.52***	-60.54***	-0.087***
Marital status	-129.78***	-19.85	-109.93***	0.174***
Immigrant	65.13	131.62**	-66.48**	0.095**
Disabled	-29.63*	-25.79*	-3.84	0.001
Health status	-318.29***	-193.04***	-125.25***	-0.013
Education	2,865.66***	1,772.46***	1,093.21***	0.269***
Employed	22.88	6.14	16.74	0.028
Part-time employment	62.15*	37.47	24.69	0.121***
$\hat{\Delta}^{v}_{X,e}$ = Specification error	-457.16	-311.69	-145.47*	-0.234***
[2] *Structure effect (Unexplained)*				
$\hat{\Delta}^{v}_{S,p}$ = Pure unexplained	16,807.93***	14,602.61***	2,205.32***	2.032***
Tenure	-608.24	-846.53	238.29	-0.129
Household composition	-609.40	-467.85	-141.55	0.443
Household size	-1,537.84	-1,439.84	-98.00	-2.729**
Age	26,316.52*	22,115.59	4,200.93	5.679
Sex	-5,651.04*	-5,215.51*	-435.52	3.490**
Marital status	4,182.29***	2,471.24*	1,711.04***	1.208**
Immigrant	-337.72	-113.56	-224.15	0.453*
Disabled	225.29	221.96*	3,33	0.014
Health status	4,109.54**	2,702.85*	1,406.69**	1.199
Education	3,292.56***	2,182.75***	1,109.81***	0.049
Employed	-917.14	608.18	-1,525.32	-0.482
Part-time employment	232.61	159.11	73.50	0.216
Constant	-11,889.50	-7,775.79	-4,113.71	-7.380
$\hat{\Delta}^{v}_{S,e}$ = Reweighting error	-401.47	-415.56*	13.99	-0.189***

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age aggregates the original age variable and age squared. Education includes all three education categories: low, medium and high. Health status: gathers all five health status groupings: very good, good, satisfactory, poor and bad. (4) Some sums may not match perfectly due to rounding.

Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (United States; 2000-2018). Luxembourg: LIS.

Table 6b. RIF-OLS decomposition results with reweighting
UNITED STATES

	Static approach Year 2018			
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
Overall				
Over 500,000 inhabitants	82,118.39***	55,769.74***	26,348.65***	37.074***
Fewer than 500,000 inhabitants	62,049.97***	40,248.00***	21,801.96***	34.129***
Difference/Total change	20.068.43***	15,521.74***	4,546.69***	2.944***
[1] Composition effect	1,548.17*** (7.71%)	1,334.22*** (8.60%)	213.95 (4.71%)	0.497*** (16.88%)
[2] Structure effect	18,520.25*** (92.29%)	14,187.52*** (91.40%)	4,332.74*** (95.29%)	2.447*** (83.12%)
[1] *Composition effect (Explained)*				
$\hat{\Delta}^{v}_{X,p}$ = Pure explained	1,659.75***	1,572.53***	87.22	0.853***
Tenure	-296.03***	16.47	-312.51***	0.228***
Household composition	-134.97**	-108.76**	-26.21*	0.007
Household size	-119.53	-76.09	-43.43	-0.021
Age	-213.45**	-243.32***	29.87	-0.048*
Sex	16.13	13.97	2.16	0.002
Marital status	-160.54***	-67.64**	-92.89***	0.043**
Immigrant	249.77	546.89*	-297.12**	0.339**
Disabled	46.15*	44.53*	1.62	0.029***
Health status	203.16***	88.44	114.71***	-0.029
Education	2,171.08***	1,408.68***	762.40***	0.220***
Employed	-59.01**	-17.14	-41.87**	-0.068**
Part-time employment	-42.98	-33.50	-9.48	-0.086***
$\hat{\Delta}^{v}_{X,e}$ = Specification error	-111.57	-238.30	126.73*	-0.119**
[2] *Structure effect (Unexplained)*				
$\hat{\Delta}^{v}_{S,p}$ = Pure unexplained	18,411.02***	14,123.58***	4,287.43***	2.478***
Tenure	597.99	261.71	336.28	-0.282
Household composition	-775.44	-589.33	-186.10	-0.106
Household size	-1,723.99	-1,219.55	-504.44	0.538
Age	1,270.94	-1,001.50	2,272.45	2.084
Sex	-3,807.38	-2,626.51	-1,180.87	0.819
Marital status	2,346.82*	1,920.21	426.61	0.109
Immigrant	-360.09	-69.62	-290.47	0.308
Disabled	104.92	97.62	7.29	0.123**
Health status	1,620.72	1,460.77	159.94	0.677
Education	1,718.04*	1,772.76**	-54.71	-0.238
Employed	1,345.07	2,017.66	-672.59	0.012
Part-time employment	-256.40	-156.53	-99.86	0.056
Constant	16,329.81	12,255.90	4,073.91	-1.623
$\hat{\Delta}^{v}_{S,e}$ = Reweighting error	109.23	63.93	45.29	-0.031

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age aggregates the original age variable and age squared. Education includes all three education categories: low, medium and high. Health status: gathers all five health status groupings: very good, good, satisfactory, poor and bad. (4) Some sums may not match perfectly due to rounding.
Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (United States; 2000-2018). Luxembourg: LIS.

APPENDIX

Table A.1. Functional statistics and RIF used

Statistic	*Definition*	*RIF*
Interquantile range	$iqr_Y(10,90) = q_{90} - q_{10}$ $iqr_Y(50,90) = q_{90} - q_{50}$ $iqr_Y(10,50) = q_{50} - q_{10}$	$RIF(y; iqr_Y(10,90)) = RIF(y;\ q_{90}, F_Y) - RIF(y; q_{10}, F_Y)$ $RIF(y; iqr_Y(50,90)) = RIF(y;\ q_{90}, F_Y) - RIF(y; q_{50}, F_Y)$ $RIF(y; iqr_Y(10,50)) = RIF(y;\ q_{50}, F_Y) - RIF(y; q_{10}, F_Y)$
Gini	$v^G(F_Y) = 1 - \frac{2}{\mu_y} R(F_Y)$ $R(F_Y) = \int_0^1 GL(p; F_Y)dp$ $GL(p; F_Y) = \int_{-\infty}^{F^{-1}(p)} zdF_y(z)$	$RIF(y; v^G, F_Y) = 2\frac{y}{\mu}\left[F_Y(y) - \frac{(1+v^G)}{2}\right] + 2\left[\frac{(1-v^G)}{2} - GL(p; F_Y)\right] + v^G$

Note: $GL(p; F_Y)$ is the Generalized Lorenz ordinate of F_Y.
Source: Firpo et al. (2018).

Table A.2. Variables

Categorization	Name	Definition
OUTCOME VARIABLE		
Major economic aggregates (Income aggregates)	Equivalent disposable income LIS variable: *dhi*	Sum of cash and non-cash income from labour, income from capital, income from pensions (including private and public pensions) and non-pension public social benefits stemming from insurance, universal or assistance schemes (including in-kind social assistance transfers), as well as cash and non-cash private transfers, less the amount of income taxes and social contributions paid.
CONTROL VARIABLES		
Geography and housing	Tenure LIS variable: *own*	Indicator of housing tenure. We have redefined it as a dummy variable: 1 = owned; 0 = rented/other.
Household composition and living arrangements	Household composition LIS variable: *hhtype*	The composition of the household with respect to the head. We have redefined it as a dummy variable: 1 = one-person household; 0 = other values.
	Household size LIS variable: *nhhmem*	Number of household members.
Socio-demographic Characteristics	Age[15] LIS variable: *age*	Age in years.
	Sex LIS variable: *sex*	Classification of individuals according to their sex. It is defined as a dummy variable: 1 = female; 0 = male.
	Marital status LIS variable: *marital*	Classification of individuals according to their marital status, as provided in relation to the marriage laws or customs of the country. We have redefined it as a dummy variable: 1 = married/in union; 0 = another status.

[15] We have also included in the estimates the variable age squared.

Socio-demographic characteristics	Immigrant LIS variable: *immigr*	Individuals who have that country as their country of usual residence and (in order of priority): whom the data provider identified as immigrants; who self-identify as immigrants; who are a citizen/national of another country; who were born in another country. It is defined as a dummy variable: 1 = immigrant; 0 = not immigrant.
	Disabled LIS variable: *disabled*	Individuals who have a permanent disability condition, defined as a (physical or mental) health condition that permanently limits an individual in his/her basic activity functioning (such as walking or hearing), even if the limitation is ameliorated by the use of assistive devices or a supportive environment. It is defined as a dummy variable: 1 = disabled; 0 = not disabled.
	Health status LIS variable: *health_c*	Subjective evaluation of one own's self-perceived health status, including any dimension as considered appropriate by the individual (physical, emotional, mental, etc.). It is reported in a scale of ratings.
	Education LIS variable: *educ*	Recoding of highest level of education completed into three categories: low: less than upper secondary education completed (never attended, no completed education or education completed at the ISCED 2011 levels 0, 1 or 2); medium: upper secondary education completed or post-secondary non-tertiary education (completed ISCED 2011 levels 3 or 4); high: tertiary education completed (completed ISCED 2011 levels 5 to 8).
Labour market information	Employed LIS variable: *emp*	Indicator that employment is the main current activity status as self-assessed by the respondent. It is defined as a dummy variable: 1 = employed; 0 = not employed.
	Part-time employment LIS variable: *emp*	Time schedule in the first job, as self-reported by the individual or defined by the data provider. It is defined as a dummy variable: 1 = part-time; 0 = full-time.

Notes: (1) The outcome variable used has been obtained dividing the disposable income by the square root of the household size. Negative and zero income values have been replaced with 1/100 of the mean.
Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (multiple countries; 1999-2018). Luxembourg: LIS and EPF.

Table A.3. Descriptive statistics for CANADA

Table A.3a. Socioeconomic characteristics of households. Sample means

	Static approach Year 2016			*Dynamic approach* Areas > 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2016	2000	Difference
Geography and housing						
Tenure	0.651	0.731	-0.080***	0.651	0.641	0.010
Household composition and living arrangements						
Household composition	0.285	0.287	-0.002	0.285	0.271	0.014*
Household size	2.460	2.282	0.178***	2.460	2.504	-0.044**
Socio-demographic characteristics						
Age	47.979	50.283	-2.304***	47.979	46.644	1.335***
Sex	0.431	0.408	0.023***	0.431	0.394	0.037***
Marital status	0.475	0.475	0.000	0.475	0.486	-0.011
Immigrant	-	-	-	-	-	-
Disabled	-	-	-	-	-	-
Health status						
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
Education						
Low	0.108	0.165	-0.057***	0.108	0.194	-0.086***
Medium	0.223	0.234	-0.011*	0.223	0.267	-0.044***
High	0.659	0.590	0.069***	0.659	0.537	0.122***
Labour market information						
Employed	-	-	-	-	-	-
Part-time employment	0.086	0.093	-0.007	0.086	0.057	0.029***
% of sample of the year	26.74	73.26		26.74	26.41	
Sample size (N)	7,133	19,542		7,133	7,650	

Table A.3b. Distribution of the equivalent disposable household income

	Static approach Year 2016			*Dynamic approach* Areas > 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2016	2000	Difference
Mean	38,461.95	35,520.90	2,941.05***	38,461.95	31,496.75	6,965.20***
Income percentiles						
D10	18,898.64	19,071.89	-173.25	18,898.64	15,329.65	3,568.99***
D20	24,120.94	24,265.89	-144.95	24,120.94	20,291.97	3,828.97***
D30	28,747.20	28,354.11	393.09	28,747.20	24,105.80	4,641.39***
D40	33,139.20	32,517.41	621.79	33,139.20	27,721.31	5,417.89***
D50	38,144.13	36,756.59	1,387.54***	38,144.13	31,481.07	6,663.06***
D60	42,989.31	41,676.02	1,313.29***	42,989.31	35,487.75	7,501.56***
D70	48,993.43	47,414.58	1,578.85***	48,993.43	40,351.19	8,642.24***
D80	57,481.91	55,023.84	2,458.06***	57,481.91	46,503.22	10,978.69***
D90	71,937.42	67,596.90	4,340.52***	71,937.42	56,799.25	15,138.17***
Inequality measures						
P90-P10	53,038.79	48,525.05	4,513.74***	53,038.79	41,469.61	11,569.18***
P90-P50	33,793.29	30,840.31	2,952.98***	33,793.29	25,318.18	8,475.11***
P50-P10	19,245.50	17,684.73	1,560.76***	19,245.50	16,151.50	3,094.07***
Gini index	0.297	0.280	0.017***	0.297	0.292	0.005

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Calculations made using sample weights. (3) Income values expressed in PPP 2017 USD. (3) Empty cells due to lack of data.
Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (Canada; 2000-2016). Luxembourg: LIS.

Table A.4. Descriptive statistics for GERMANY

Table A.4a. Socioeconomic characteristics of households. Sample means

	Static approach Year 2016			*Dynamic approach* Areas > 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2016	2000	Difference
Geography and housing						
Tenure	0.255	0.487	-0.232***	0.255	0.228	0.027
Household composition and living arrangements						
Household composition	0.310	0.237	0.073***	0.310	0.386	-0.076***
Household size	2.375	2.542	-0.167***	2.375	2.067	0.308***
Socio-demographic characteristics						
Age	49.316	51.626	-2.310***	49.316	49.233	0.083
Sex	0.502	0.491	0.109	0.502	0.435	0.067***
Marital status	0.453	0.546	-0.093***	0.453	0.422	0.031**
Immigrant	0.263	0.186	0.077***	0.263	0.146	0.117***
Disabled	0.119	0.126	-0.007	0.119	0.143	-0.024**
Health status						
Very Good	0.106	0.085	0.021***	0.106	0.109	-0.003
Good	0.397	0.377	0.020*	0.397	0.390	0.007
Satisfactory	0.305	0.340	-0.035***	0.305	0.313	-0.008
Poor	0.145	0.152	-0.007	0.145	0.136	0.009
Bad	0.038	0.037	0.001	0.038	0.044	-0.006
Education						
Low	0.093	0.114	-0.021***	0.093	0.115	-0.022**
Medium	0.449	0.561	-0.112***	0.449	0.516	-0.067***
High	0.434	0.306	0.128***	0.434	0.331	0.103***
Labour market information						
Employed	0.652	0.651	0.001	0.652	0.596	0.056***
Part-time employment	0.245	0.269	-0.024**	0.245	0.174	0.071***
Percentage of sample size (%)	17.30	82.70		17.30	13.71	
Sample size (N)	2,736	13,080		2,736	1,617	

Table A.4b. Distribution of the equivalent disposable household income

	Static approach Year 2016			*Dynamic approach* Areas > 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2016	2000	Difference
Mean	34,477.69	31,829.29	2,648.40***	34,477.69	31,208.15	3,269.44***
Income percentiles						
D10	16,422.89	16,492.46	-519.57	16,422.89	16,161.52	261.37
D20	20,761.67	21,041.43	-279.76	20,761.67	21,832.96	-1,071.29
D30	24,595.33	24,411.09	184.24	24,595.33	25,094.05	-498.72
D40	28,535.24	27,736.67	798.56	28,535.24	28,250.69	284.55
D50	33,137.63	31,004.77	2,132.86***	33,137.63	31,512.45	1,625.19**
D60	37,774.47	35,025.21	2,749.25***	37,774.47	35,041.66	2,732.81***
D70	43,315.37	39,551.55	3,763.83***	43,315.37	39,671.25	3,644.12***
D80	50,186.77	45,610.59	4,576.18***	50,186.77	45,644.27	4,542.49***
D90	65,644.53	56,251.47	9,393.06***	65,644.53	54,933.04	10,711,49***
Inequality measures						
P90-P10	49,221.64	39,309.42	9,912.22***	49,221.64	38,770.78	10,450.86***
P90-P50	32,506.90	25,247.03	7,259.87***	32,506.90	23,420.82	9,086.08***
P50-P10	16,714.74	14,062.39	2,652.35***	16,714.74	15,349.96	1,364.78
Gini index	0.322	0.282	0.040***	0.322	0.265	0.057***

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Calculations made using sample weights. (3) Income values expressed in PPP 2017 USD.
Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (Germany; 2000-2016). Luxembourg: LIS.

Table A.5. Descriptive statistics for ITALY

Table A.5a. Socioeconomic characteristics of households. Sample means

	Static approach Year 2016			*Dynamic approach* Areas > 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2016	2000	Difference
Geography and housing						
Tenure	0.702	0.720	-0.018	0.702	0.590	0.112***
Household composition and living arrangements						
Household composition	0.404	0.337	0.067***	0.404	0.214	0.190***
Household size	2.138	2.226	-0.088*	2.138	2.647	-0.509***
Socio-demographic characteristics						
Age	61.976	62.180	-0.204	61.976	55.143	6.833***
Sex	0.441	0.447	-0.036*	0.441	0.329	0.112***
Marital status	0.486	0.534	-0.048**	0.486	0.638	-0.152***
Immigrant	0.069	0.071	-0.002	0.069	0.025	0.044***
Disabled	-	-	-	-	-	-
Health status						
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
Education						
Low	0.500	0.556	-0.056***	0.500	0.552	-0.052*
Medium	0.296	0.336	-0.040**	0.296	0.297	-0.001
High	0.203	0.107	0.096***	0.203	0.151	0.052**
Labour market information						
Employed	-	-	-	-	-	-
Part-time employment	0.071	0.128	-0.057***	0.071	0.086	-0.015
Percentage of sample size	8.42	91.58		8.42	8.64	
Sample size (N)	625	6,795		625	691	

Table A.5b. Distribution of the equivalent disposable household income

	Static approach Year 2016			*Dynamic approach* Areas > 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2016	2000	Difference
Mean	26,146.63	22,078.65	4,067.73***	26,146.63	25,153.25	993.38
Income percentiles						
D10	11,953.69	11,123.39	830.30	11,953.69	11,910.95	42.74
D20	15,998.34	14,399.22	1,599.11	15,998.34	14,826.22	1,172.11
D30	19,771.48	17,254.34	2,517.14	19,771.48	18,343.68	1,427.79
D40	22,812.93	20,304.76	2,508.18**	22,812.93	22,238.14	574.79
D50	25,631.17	23,115.70	2,515.47***	25,631.17	25,036.43	594.75
D60	28,031.66	25,708.84	2,322.82	28,031.66	28,941.64	-909.98
D70	34,050.51	28,867.72	5,182.79***	34,050.51	33,146.51	904.00
D80	42,802.20	33,223.67	9,578.53***	42,802.20	39,017.34	3,784.86
D90	57,219.02	41,431.76	15,787.26***	57,219.02	52,427.64	4,791.38
Inequality measures						
P90-P10	45,265.33	30,310.16	14,955.17***	45,265.33	40,516.49	4,748.84
P90-P50	31,587.85	18,317.70	13,270.15***	31,587.85	27,391.27	4,196.58
P50-P10	13,677.48	11,992.46	1,685.02**	13,677.48	13,125.22	552.26
Gini index	0.362	0.293	0.069***	0.362	0.318	0.044*

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Calculations made using sample weights. (3) Income values expressed in PPP 2017 USD. (3) Empty cells due to lack of data.
Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (Italy; 2000-2016). Luxembourg: LIS.

Table A.6. Descriptive statistics for POLAND

Table A.6a. Socioeconomic characteristics of households. Sample means

	Static approach Year 2016			*Dynamic approach* Areas > 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2016	1999	Difference
Geography and housing						
Tenure	0.702	0.837	-0.135***	0.702	0.490	0.212***
Household composition and living arrangements						
Household composition	0.318	0.189	0.129***	0.318	0.229	0.089***
Household size	2.231	2.758	-0.527***	2.231	2.584	-0.353***
Socio-demographic characteristics						
Age	50.619	53.254	-2.634***	50.619	50.234	0.385
Sex	0.469	0.375	0.094***	0.469	0.437	0.032***
Marital status	0.506	0.644	-0.138***	0.506	0.628	-0.122***
Immigrant	0.014	0.006	0.008***	0.014	-	-
Disabled	0.078	0.099	0.021***	0.078	0.127	-0.049***
Health status						
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
Education						
Low	0.055	0.148	-0.093***	0.055	0.134	-0.079***
Medium	0.482	0.660	-0.178***	0.482	0.629	-0.147***
High	0.463	0.191	0.272***	0.463	0.237	0.226***
Labour market information						
Employed	-	-	-	-	-	-
Part-time employment	0.065	0.047	0.018***	0.065	0.074	-0.009***
Percentage of sample size	12.91	87.09		12.91	13.96	
Sample size (N)	4,762	32,124		4,762	4,388	

Table A.6b. Distribution of the equivalent disposable household income

	Static approach Year 2016			*Dynamic approach* Areas > 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2016	1999	Difference
Mean	21,424.85	15,903.10	5,521.75***	21,424.85	13,674.88	7,749.97***
Income percentiles						
D10	11,502.36	8,372.74	3,129.62***	11,502.36	7,305.44	4,196.92***
D20	14,433.93	10,625.04	3,808.89***	14,433.93	8,926.55	5,507.38***
D30	16,953.51	12,437.87	4,515.64***	16,953.51	10,370.43	6,583.08***
D40	19,379.97	14,092.52	5,287.45***	19,379.97	11,650.38	7,729.59***
D50	21,625.97	15,789.41	5,836.56***	21,625.97	13,121.15	8,504.82***
D60	24,449.82	17,687.58	6,762.24***	24,449.82	14,852.06	9,597.76***
D70	27,702.15	20,039.24	7,662.91***	27,702.15	16,670.74	10,941.41***
D80	31,879.78	23,140.58	8,739.20***	31,879.78	19,703.14	12,176.64***
D90	40,162.34	28,943.91	11,218.44***	40,162.34	25,172.12	14,990.22***
Inequality measures						
P90-P10	28,659.98	20,571.21	8,088.77***	28,659.98	17,687.00	10,792.98***
P90-P50	18,536.37	13,154.53	5,381.84***	18,536.37	12,051.07	6,485.30***
P50-P10	10,123.62	7,416.68	2,706.93***	10,123.62	5,815.93	4,307.69***
Gini index	0.293	0.293	0.000	0.293	0.308	-0.015

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Calculations made using sample weights. (3) Income values expressed in PPP 2017 USD. (3) Empty cells due to lack of data.
Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (Poland; 1999-2016). Luxembourg: LIS.

Table A.7. Descriptive statistics for SPAIN

Table A.7a. Socioeconomic characteristics of households. Sample means

	Static approach Year 2018			*Dynamic approach* Areas > 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2018	2006	Difference
Geography and housing						
Tenure	0.714	0.813	-0.098***	0.714	0.815	-0.101***
Household composition and living arrangements						
Household composition	0.106	0.091	0.015***	0.106	0.073	0.033***
Household size	3.110	3.192	-0.082***	3.110	3.325	-0.215***
Socio-demographic characteristics						
Age	53.646	54.066	-0.420***	53.646	51.394	2.252***
Sex	0.629	0.602	0.027***	0.629	0.587	0.042***
Marital status	0.435	0.470	-0.035***	0.435	0.475	-0.040***
Immigrant	0.139	0.087	0.052***	0.139	0.083	0.056***
Disabled	-	-	-	-	-	-
Health status						
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
Education						
Low	0.121	0.175	-0.054***	0.121	0.231	-0.109***
Medium	0.421	0.495	-0.074***	0.421	0.403	0.018**
High	0.456	0.328	0.127***	0.456	0.365	0.091***
Labour market information						
Employed	-	-	-	-	-	-
Part-time employment	-	-	-	-	-	-
Percentage of sample size	10.72	89.28		10.72	10.49	
Sample size (N)	5,974	49,751		5,974	5,847	

Table A.7b. Distribution of the equivalent net household income

	Static approach Year 2018			*Dynamic approach* Areas > 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2018	2006	Difference
Mean	1,515.12	1,291.80	223.32***	1,515.12	1,455.89	59.23***
Income percentiles						
D10	572.60	539.17	33.43***	572.60	618.85	-46.25***
D20	810.59	708.81	101.78***	810.59	789.71	20.88*
D30	973.62	842.73	130.88***	973.62	941.59	32.03*
D40	1,161.73	998.38	163.34***	1,161.73	1,088.22	73.51***
D50	1,300.43	1,169.75	130.67***	1,300.43	1,252.86	47.57***
D60	1,520.55	1,303.59	216.95***	1,520.55	1,441.93	78.62***
D70	1,776.50	1,512.88	263.61***	1,776.50	1,656.45	120.05***
D80	2,046.80	1,775.24	271.56***	2,046.80	2,008.06	38.74
D90	2,601.58	2,164.98	436.60***	2,601.58	2,544.95	56.63
Inequality measures						
P90-P10	2,028.98	1,625.78	403.20***	2,028.98	1,926.10	102.88***
P90-P50	1,301.15	995,21	305.94***	1,301.15	1,292.09	9.06
P50-P10	727.82	630.56	97.26***	727.82	634.01	93.81***
Gini index	0.315	0.300	0.015***	0.315	0.304	0.011**

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Calculations made using sample weights. (3) Income values expressed in constant euros (reference year 2016 = 100). (4) Empty cells due to lack of data.
Source: Spanish *Encuesta de Presupuestos Familiares* (EPF).

Table A.8. Descriptive statistics for UNITED STATES

Table A.8a. Socioeconomic characteristics of households. Sample means

	Static approach Year 2018			*Dynamic approach* Areas > 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2018	2000	Difference
Geography and housing						
Tenure	0.622	0.700	-0.078***	0.622	0.651	-0.029***
Household composition and living arrangements						
Household composition	0.254	0.255	-0.001	0.254	0.226	0.028***
Household size	2.660	2.581	0.078***	2.660	2.809	-0.149***
Socio-demographic characteristics						
Age	50.492	51.832	-1.339***	50.492	46.864	3.628***
Sex	0.500	0.498	0.002	0.500	0.470	0.030***
Marital status	0.520	0.527	-0.007*	0.520	0.551	-0.031***
Immigrant	0.239	0.083	0.156***	0.239	0.191	0.048***
Disabled	0.097	0.128	-0.031***	0.097	0.102	-0.005***
Health status						
Excellent	0.251	0.212	0.039***	0.251	0.294	-0.043***
Very good	0.328	0.312	0.016***	0.328	0.320	0.008**
Good	0.280	0.291	-0.011***	0.280	0.243	0.037***
Fair	0.105	0.132	-0.027***	0.105	0.099	0.006***
Poor	0.034	0.051	-0.017***	0.034	0.042	-0.008***
Education						
Low	0.099	0.101	-0.002	0.099	0.155	-0.056***
Medium	0.403	0.500	-0.097***	0.403	0.463	-0.060***
High	0.497	0.397	0.100***	0.497	0.381	0.116***
Labour market information						
Employed	0.657	0.600	0.057***	0.657	0.706	-0.049***
Part-time employment	0.133	0.159	-0.026***	0.133	0.116	0.017***
Percentage of sample size	59.75	40.25		59.75	57.56	
Sample size (N)	40,835	27,510		40,835	44,927	

Table A.8b. Distribution of the equivalent disposable household income

	Static approach Year 2018			*Dynamic approach* Areas > 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2018	2000	Difference
Mean	49,655.28	40,592.91	9,062.38***	49,655.28	43,312.54	6,342.74***
Income percentiles						
D10	19,599.46	17,721.03	1,878.43***	19,599.46	17,870.59	1,728.87***
D20	26,144.83	23,562.47	2,582.35***	26,144.83	24,277.80	1,867.03***
D30	32,882.96	28,710.61	4,172.35***	32,882.96	30,191.69	2,691.27***
D40	39,412.65	33,973.13	5,439.52***	39,412.65	35,609.69	3,802.96***
D50	45,948.12	39,522.99	6,425.13***	45,948.12	41,104.30	4,843.82***
D60	53,750.22	45,349.90	8,400.32***	53,750.22	46,817.26	6,932.96***
D70	62,955.50	52,306.12	10,649.38***	62,955.50	53,485.38	9,470.12***
D80	76,876.32	61,735.61	15,140.72***	76,876.32	62,924.31	13,952.02***
D90	101,717.90	79,770.98	21,946.88***	101,717.90	79,678.09	22,039.77***
Inequality measures						
P90-P10	82,118.39	62,049.97	20,068.43***	82,118.39	61,807.52	20,310.87***
P90-P50	55,769.74	40,248.00	15,521.74***	55,769.74	38,573.80	17,195.94***
P50-P10	26,348.65	21,801.96	4,546.69***	26,348.65	23,233.72	3,114.93***
Gini index	0.370	0.341	0.029***	0.370	0.344	0.026***

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Calculations made using sample weights. (3) Income values expressed in PPP 2017 USD.
Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (United States; 2000-2018). Luxembourg: LIS.