Working Paper Series

The contribution of the spatial dimension to inequality: A counterfactual analysis for OECD countries

Luis Ayala
Javier Martín-Román
Juan Vicente

ECINEQ WP 2020 - 513

ECINEQ 2020 - 513
January 2020

www.ecineq.org

The contribution of the spatial dimension to inequality: A counterfactual analysis for OECD countries [‡]

Luis Ayala
Universidad Rey Juan Carlos, UNED, Equalitas

Javier Martín-Román
UNED

Juan Vicente
Universidad de Valladolid

Abstract

This paper provides recent evidence on the contribution of the spatial dimension to inequality and more specifically accounts for the impact of the changes in the territorial distribution of the population on the recent dynamics of income inequality. We use LIS harmonized microdata for a selected sample of OECD countries. We provide new evidence over a more varied group of countries and a more recent period than in previous studies. We perform different types of decompositions to isolate the contribution of the changes in the territorial distribution of the population. The results show a generalized increase in income inequality, with an interesting Òreducing effectÓ on this trend due to inter-territorial population movements.

Keywords: income inequality, regional inequality, decomposition methods, counterfactual analysis..

JEL Classification: D31, D63, P52.

[‡]Luis Ayala acknowledges financial support from Comunidad de Madrid (H2019/HUM-5793) and Ministerio de Econom'a y Competitividad (ECO2016-76506-C4-3-R).

Abstract

This paper provides recent evidence on the contribution of the spatial dimension to inequality and more specifically accounts for the impact of the changes in the territorial distribution of the population on the recent dynamics of income inequality. We use LIS harmonized microdata for a selected sample of OECD countries. We provide new evidence over a more varied group of countries and a more recent period than in previous studies. We perform different types of decompositions to isolate the contribution of the changes in the territorial distribution of the population. The results show a generalized increase in income inequality, with an interesting "reducing effect" on this trend due to inter-territorial population movements.

JEL classification: D31; D63; P52.

Key words: income inequality, regional inequality, decomposition methods, counterfactual analysis.

1 INTRODUCTION

"Reduce inequality within and among countries" is one of the 2030 United Nations sustainable development goals. Too high levels of inequality are a menace to equity and can hamper social cohesion. While inequality can influence growth positively by providing incentives for innovation and entrepreneurship and by raising saving and investment, it can also undermine progress in health and education, cause investment-reducing political and economic instability, and undercut the social consensus required to adjust in the face of shocks (Ostry et al. 2014).

The recent increase in income inequality in a large number of countries has made the study of inequality one of the main issues to be explored in economic analysis. Technical progress, changes in the labour market or the limited capacity of some tax-benefit systems to reduce inequality in market income are considered to be common drivers of the current trends in inequality. The contribution of additional dynamics that were important in earlier periods in explaining the evolution of inequality has been somewhat neglected by the recent literature. It seems that the spatial dimension of inequality and the impact of the changes in the territorial distribution of population on income inequality have received less attention than other dimensions. However, population mobility and the spatial dimension of inequality have special significance when aligned with political tensions (Kanbur and Venables, 2005).

At the global level, the between-countries component is still by far the main component of income inequality. By contrast, at the national level inequality between regions is a minor component of national inequality (Jesuit, 2003; Novotný, 2007 and Piaccentini, 2014). Neoclassical models predicted this greater convergence across regions than across countries, because frictions and mobility between-regions are lower than in the between-countries space. At the global level, however, the between-countries component has lost some weight due to the China and India catching-up. At the national level, after a convergence period in many OECD countries (Rodríguez-Pose and Ezcurra, 2009), spatial inequalities have increased again (Lessmann, 2014).

Such recent processes of convergence between countries and divergence between regions seem to fit well with the new models of economic geography that introduce increasing returns and other market failures (Fujita and Krugman, 2004). These models elicit different policy responses in order to prevent agglomeration in the core or to compensate the periphery (Commendatore et al. 2018), being factor mobility an essential element of the analysis. More specifically, the geographical concentration of economic activity will be followed or not by shifts in interregional inequality depending critically on population mobility (Puga, 2002).

Therefore, a central issue is the relationship between fluctuations in the territorial distribution of population and changes in income inequality. There are few recent studies that consider this dimension as a driver of inequality (Martino and Perugini, 2008), and only some papers focus on changes in territorial demographics (Dickey, 2014; Carrillo and Rothbaum, 2016). This paper aims at analysing and measuring the effect of territory on income inequality to determine whether this dimension is still important and how its contribution has evolved from the beginning of the 21st century to the present. Our study makes three contributions. First, we carry out different kinds of decomposition analyses to identify the effect on inequality of both income differences between regions and within regions. Second, we provide new evidence over a more varied group of countries and a more recent period than in previous studies. Third, we add to previous studies the simulation of the impact on inequality that the changes in the inter-territorial distribution of the population may have had.

Using the Luxembourg Income Study (LIS) database, we carry out a comparative analysis using a sample that represents more than two thirds of the OECD population. Our results show a significant and generalized increase in income inequality in most of the countries studied, as well as a revealing (but minor) contribution to this trend of the territorial variables. Despite not having found a strong pattern among the selected countries, our empirical specifications do seem to capture much of the change observed in income inequality during the period explored. In particular, the regional variable exerts an effect that is worth paying attention to.

The rest of the paper is structured as follows. Section 2 provides an overview of the previous relevant literature on the issue under study. In section 3, the data used are described. In section 4, we present our methodological approach. In section 5, we present our main results, and section 6 summarizes the study and presents our conclusions.

2 Literature review

A number of studies have investigated the main sources of inequality to facilitate policy-making at the national level. For developed countries, globalization and technical progress are considered to be common drivers of the current trends in inequality, while the impact of other factors such as regulation, redistributive policies, as well as demographic changes depends more on national idiosyncrasies. In developing countries, different processes of transition also influence these tendencies.

One of the most relevant dimensions in the analysis of changes in inequality is the role that territorial differences have played in amplifying income differences. Perhaps one of the most important attempts to relate space, income and inequality are the different developments of the so called new economic geography (NEG). Since the early 1990s, NEG has provided different strands of theory in understanding the relationships between these dimensions. The overall aim of NEG is to explain regional economic disparities on the basis of spatial agglomeration effects. The most recent NEG models have stressed path dependence and lock-ins as mechanisms to explain the persistence of spatial disparities through time (Hassing and Gong, 2019).

Under this framework, a key issue is labour mobility. NEG models relate labour migrations across regions to the geography of production through real wage differentials. The basic intuition of NEG models highlights the influence of access to markets on location choices of both firms and workers. The cumulative process of agglomeration rests on the complementarity of these two relations: agglomeration may occur only if migrants, like firms, are attracted by high market potential regions (Crozet, 2004). The concentration of manufacturing workers creates a large market, so making the location profitable for firms. And the entry of firms bids up wages, so making the location attractive for workers (Venables, 2016). However, as stressed by Garretsen and Martin (2010), some authors have questioned whether the formal economic models that are the focus of attention within NEG are can adequately capture the full range of factors and forces that help shape the economic landscape, particularly since some of these factors are social, institutional and cultural in nature.

The drivers of inequality have barely been examined at regional level and mostly for EU countries. Martino and Perugini (2008) use LIS and Eurostat data to estimate the influence of different economic, demographic and institutional variables on within-regional inequality finding a non-significant role of the only demographic variable included – the share of the population aged 65 years or over. Castells-Quintana et al. (2015) use data from ECHP and EU-SILC to estimate the relationship between within-regional inequality and regional per capita income and find a significant influence of some control variables, including population density. Mussini (2017) uses Eurostat data to decompose the changes in between-regional inequality and finds a non-negligible role of the change of regional population weights. Finally, Dickey (2014) uses British-HPS data to estimate the impact of migration on the within-regional wage distributions finding a significant impact but with opposite signs across regions.

A number of works have also addressed some methodological issues relevant for the goals of this paper, such as the level of aggregation, the choice of inequality measures or the spatial variation of prices. Regarding the level of spatial aggregation, we are aware that it affects the relative weights of the between and within components for a given distribution (Shorrocks and Wan, 2005). To solve this problem either the between-component can be reinterpreted to discount the effects of the number and size of regions (Elbers et al. 2008) or the classification can be redefined to homogenize the number of regions across countries (Novotný, 2007).

Regarding the choice of the inequality measure, each one implies not only a different concept of inequality but also a different rule of decomposition (Cowell, 2011). The decomposition of the Theil measures is more user friendly than that of the Gini due to their simpler structure, but such simplicity is achieved at the cost of losing useful information. The extra term added to the inter and intra components in the standard decomposition of the Gini index, accounts for the amount of overlapping between subgroups which can be interpreted in terms of stratification and other relevant concepts (Yitzhaki and Schechtman, 2013). Alternative decompositions of the Gini index also allow to analyze other concepts such as the spatial autocorrelation (Rey and Smith, 2013) that is both a nuisance for the analysis and a relevant feature of the topic at hand.

Lastly, regarding the spatial variation of prices, the correlation between price levels and living standards produces an overestimation of the between- component (Shorrocks and Wan, 2005). This is a recurring issue of concern to practically all researchers, but the lack of data makes it difficult to measure properly. In countries that have tried to implement an alternative methodology based on relative region-specific poverty lines, such as France (Insee, 1997) or Spain (Ayala et al. 2014), it has been shown that these alternatives also have problems, such that they end up mixing the disparities in the cost of living with those related to the level of economic development (Brandolini, 2007).

From all the different possible approaches, we chose to focus on the role that changes in the regional population distribution play regarding income inequality. On the one hand, in two-thirds of OECD countries the share of population in predominantly urban regions has increased in the past 15 years (OECD, 2018). On the other hand, as anticipated by NEG theories, regional migration does not affect all regions of a country equally. Distance to labour market and services seems to explain migration within OECD countries. These flows may lead to persistent regional economic disparities also causing effects on the personal income distribution.

Our paper contributes to the existing literature by considering a sample of OECD countries that have very different administrative divisions of their national territory. The paper also contributes to the integration of existing methods by using alternative ways to decompose inequality, which is a promising issue in the research agenda of the dynamics of regional disparities (Rey and Janikas, 2005).

3 Data

In this paper, we have chosen the LIS database for three basic reasons: (1) it has a wider spatial scope than the EU-SILC, (2) it allows access to deeper content than the OECD-IDD, and (3) it allows us to identify the relative weight of the spatial dimension of income inequality and its changes over time. As far as we know, this is the first attempt to conduct such a study implementing a counterfactual analysis.

The LIS is the largest available income database of harmonized microdata and includes approximately 50 years of information on more than fifty countries all over the world. It currently gathers data from very different countries and continues to expand. LIS datasets contain household-level and person-level data on labour income, capital income, social security and private transfers, taxes and contributions, demography, employment, and expenditures. Its use and influence have been steadily increasing, although only a few papers report estimates of income inequality at the subnational level.

The period considered in this exercise is covered by the datasets available for the 21st century. We focus on the OECD area, including some of the most populous ones in 2016.[2] The selected sample represents more than two thirds of the OECD population (see Table A.1 in the Appendix).

Another noteworthy issue addressed by this paper concerns the definition and division of the territorial units used in each country. Bearing in mind that this is a somewhat arbitrary issue, we decided to proceed to use an administrative division as the main classification criterion. This proposal coincides with the sorting provided by the LIS database for the regional variable, and respects the Eurostat recommendations of using the classification closer to the framework adopted by the countries for their regional policy. It also clarifies the subsequent interpretation and justification of the results, being more comprehensible and straightforward

Our administrative criterium gives priority to institutional boundaries. In this manner, the different units used in the study are the following: 7 regions in Australia (the 6 federated states and Canberra); in Canada, the 10 administrative divisions that are responsible for

[2] According to 2016 OECD data, the eleven most populous OECD member countries were the United States, Japan, Mexico, Germany, Turkey, France, United Kingdom, Italy, South Korea, Spain and Poland. We had to exclude Japan (only had data for one year), Turkey (no data) and South Korea (no regional data). Instead, we were compelled to add the OECD countries ranked 12th, 13th, and 14th in population: Canada, Australia and Chile.

sub-national governance (the provinces); in Chile, we have data for 13 regions (all of the country's administrative units); in France the data available provide information on 8 regions; in Germany, the 16 *Länder*; in Italy, the 20 administrative regions; in Mexico, the 32 federal entities of the United Mexican States; in Poland, the 16 *voivodeships* (the largest unit of the Polish administrative political system); in Spain, the 18[3] *Comunidades Autónomas* (the Spanish territorial administrative entities established by the country's Constitution that are endowed with a certain legislative autonomy); in the United Kingdom, we have considered 12 divisions (the 9 regions of England, also known as *Government Office Regions*, Scotland, Wales and Northern Ireland); in the United States 51 territories were analysed, i.e., the states that share sovereignty with the federal government.

Additionally, it is important to mention that another interesting alternative is the criteria proposed by Novotný (2004, 2007), Hoffmeister (2009), and others. Their fundamental premise is to make comparisons among entities with a similar number of divisions, making relevant groupings according to a specific convention when needed. This procedure presents a clear advantage: it allows us to eliminate the discrepancies and biases that would otherwise occur if we calculate the *between* and *within* components in an inequality decomposition by population subgroups and if the units to be examined were not of a similar size. However, there is a major disadvantage in considering divisions of the same size: to assume that the productive structures of different regions are the same regardless of their dimension is a very restrictive assumption. Nonetheless, to test the sensitivity of our results to the criterion for aggregation, we also re-estimated all the calculations following the recommended guidelines of these authors.[4] The new estimates

[3] The 17 Spanish *Comunidades Autónomas* and Ceuta and Melilla considered as a single unit.
[4] Since the LIS database does not allow a lower level of disaggregation, we can only test the possible effects of a higher aggregation level. More precisely, in the case of the two EU countries with more regions we move from NUTS2 to NUTS 1 level (in Italy the change is from 20 to 5 regions, and in Spain from 18 to 7

can be tested in Table A.2 of the Appendix. In general terms, regional aggregation slightly reduces the contribution to national inequality of the between-regions component, but it hardly affects the trends of that contribution.

Regarding the variables considered in the study, the key one is the *real equivalent household disposable income*. The *disposable income* includes both primary (labour and capital/market) and secondary (tax and transfers/non-market) incomes. Following the standard criteria found in the LIS database, we choose the *equivalent household disposable income*, obtained by dividing the disposable income by the square root of the household size. It is also important to clarify two ideas regarding the handling of the data. First, negative and zero income values have been replaced with 1/100 of the mean. In this manner, we can prevent relevant observations from being dropped by default. Second, observations with missing values for the regional variable have been removed to ensure consistency with all the results presented here.

The application of the aggregation criteria and the methodological options previously described to the LIS data allow us to calculate the extent of inequality both in each country of the sample and in each of the territorial units that we have defined. Table A.3 provides a thumbnail sketch of the general picture of inequality within each region. The Table gives general support to the notion of a very wide range in the inequality indicators within each region in all countries, with large differences between the highest and the lowest values of the mean logarithmic deviation.

4 Empirical strategy

regions). In the two non-EU countries with more regions, we take into consideration the regional division used by the United States Census Bureau (grouping the states into 9 divisions), and we group the Mexican states into 8 regions or conglomerations.

The methodology used in this paper follows the proposal by Cowell and Fiorio (2011) to reconcile the conventional theoretical schemes and the most recent regression techniques. This integration of existing methods has been also outlined as a promising issue in the research agenda of the dynamics of regional inequalities (Rey and Janikas, 2005). First, we carry out a standard subgroup decomposition to identify the corresponding weights of the within and between regional components. Second, we develop a dynamic decomposition to identify the weight of population changes. Finally, we apply counterfactual analyses to estimate the effect of these population changes on income inequality.

Among the many ways to quantify inequality in the distribution of income, the two classical measures are the coefficient of variation (*C*) and the Gini index (*G*). They can be expressed in terms of the ratios (λ_i) between income (q_i) and population (p_i) shares of the *i*=1… *n* receivers:

$$0 \leq C^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{q_i}{p_i} - 1\right)^2 = \sum_{i=1}^{n}\frac{1}{n}(\lambda_i - 1)^2 \leq n - 1 \tag{1}$$

$$0 \leq G = \frac{1}{n}\sum_{i=1}^{n}\frac{i}{\frac{n}{2}}\left(\frac{q_i}{p_i} - 1\right) = \sum_{i=1}^{n}\frac{2i}{n^2}(\lambda_i - 1) \leq \frac{n-1}{n}, \quad q_i \leq q_{i+1} \tag{2}$$

A third commonly used measure is the Theil index (*T*), which can be derived from the mean logarithmic deviation (*L*) by exchanging the population and income shares:

$$0 \leq T = \textstyle\sum_{i=1}^{n} q_i log(\lambda_i) \leq \log n \tag{3}$$

$$0 \leq L = \textstyle\sum_{i=1}^{n} p_i log(1/\lambda_i) \leq \infty \tag{4}$$

Cowell reported on other measures of inequality and extensively discussed a measure initially called the ‘generalized information measure’, which was later renamed *generalized entropy measure* (E_θ) following modifications to allow the fulfilment of additional properties (Cowell, 2011):

$$0 \leq E_\theta = (\sum_{i=1}^{n} p_i \lambda_i^\theta - 1)/(\theta^2 - \theta) \leq \infty \tag{5}$$

The members of the entropy family with a parameter θ <2 concentrate their transfer sensitivity more at the lower end the further away parameter θ is from 2, while those with a θ>2 exhibit a transfer sensitivity more towards the top of the distribution the greater the parameter θ value. When θ=2, the transfer sensitivity increases symmetrically at two tails towards both sides.

In our analysis, and in order to determine the explanatory power of the territorial variable in the recent evolution of inequality, we have chosen the mean log deviation as the index to be examined. The main reasons that support this decision are presented in the following section.

4.1 Inequality decomposition by population subgroups

The analysis of regional inequality is mainly related to the decomposition by groups. The standard procedure to implement such a decomposition consists of defining the inequality between-groups (*B*) as that which remains after removing the within-groups inequality (*W*) by replacing individual incomes with their group mean. The *W* component is computed later by subtracting the *B* component from total inequality.

For the generalized entropy family, the weights are a first-order homogeneous function of the population and income shares of the groups: $w_g = p_g \lambda_g^\theta = p_g^{1-\theta} q_g^\theta$.

$$E_\theta = BE_\theta + WE_\theta = (\theta^2 - \theta)^{-1}\left[\sum_{g=1}^{k} w_g - 1\right] + \sum_g w_g E_{\theta g} \tag{6}$$

The weights w_g sum to unity when θ=0 (the income weighted T) or θ=1 (the population weighted L):

$$T = BT + WT = \sum_g q_g \log(\lambda_g) + \sum_g q_g T_g \qquad (7)$$

$$L = BL + WL = \sum_g p_g log(1/\lambda_g) + \sum_g p_g L_g \qquad (8)$$

In the case of the squared coefficient of variation, the weights sum to unity if all subgroup distributions have the same mean:

$$C^2 = BC^2 + WC^2 = \sum_g p_g (\lambda_g - 1)^2 + \sum_g p_g \lambda_g^2 C_g^2 \qquad (9)$$

The mean logarithmic deviation (L) is the only member of the entropy family that produces the same results with both approaches and generates assignments similar to the Shapley decomposition (Shorrocks, 2013). This "path independent" property may explain the preference for this measure in recent empirical work in this area and is also one the motives for choosing this index in the decompositions. Additionally, due to its simplicity, the decomposition of the mean log deviation (L) has been the most widely used in the literature. The results for other indices are not as easily interpreted (Shorrocks and Wan, 2005).

4.2 Dynamic decomposition

This decomposition allows observing the changes produced for a given period instead of for each specific year. This helps us know the importance of each component of the decomposition in explaining the general evolution of the index. The dynamic decomposition was initially proposed by Mookherjee and Shorrocks (1982) and, given its greater simplicity, this decomposition has been developed primarily for L. It can be described as follows:

$$\Delta L = \Delta\left[\sum\nolimits_g p_g L_g - \sum\nolimits_g p_g \lambda_g\right] = \sum\nolimits_g \bar{p}_g \Delta L_g + \sum\nolimits_g \bar{L}_g \,\Delta p_g - \sum\nolimits_g \overline{\lambda_g}\,\Delta p_g - \sum\nolimits_g \bar{p}_g \Delta\,\lambda_g$$

$$\simeq \sum\nolimits_g \bar{p}_g \Delta L_g + \sum\nolimits_g \bar{L}_g \,\Delta p_g + \sum\nolimits_g \left[\overline{\left(\frac{\mu_g}{\mu}\right)} - \overline{\log\left(\frac{\mu_g}{\mu}\right)}\right]\Delta p_g + \sum\nolimits_g (\bar{q}_g - \bar{p}_g)\,\Delta \log \mu_g \quad (10)$$

where Δ shows the variation in the variables of interest from the initial year (t_0) to the final year (t_1). Following those authors, we can express the four terms of expression (10) as follows:

$$\sum\nolimits_g \bar{p}_g \Delta L_g \quad (10a)$$

$$\sum\nolimits_g \bar{L}_g \,\Delta p_g \quad (10b)$$

$$\sum\nolimits_g \left[\overline{\left(\frac{\mu_g}{\mu}\right)} - \overline{\log\left(\frac{\mu_g}{\mu}\right)}\right]\Delta p_g \quad (10c)$$

$$\sum\nolimits_g (\bar{q}_g - \bar{p}_g)\,\Delta \log \mu_g \quad (10d)$$

The first term (10a) denotes the changes in within-subgroup[5] inequality; the second (10b) reflects the variations in the population shares of the "within group" component; expression (10c) reveals the same as the previous one, but for the case of "between-group" inequality; and the last expression (10d) displays the effect of changes attributable to differences in relative incomes for the groups of interest. This decomposition allows us to recognize the influence of each one of these four elements on the *trend* in aggregate inequality[6] as well as link this section with the following one, where the contribution of the territorial dimension is presented through the methodology of counterfactual analysis.

[5] In our case, the different groups are the territorial units for each country selected from the LIS database.

[6] See Table 2.

Once again, the exercise can be understood in terms of the aforementioned counterfactual analysis. In the context of this paper, by taking regions as groups, the sum of the second and the third terms can be interpreted as the change in inequality that would have occurred if the relative incomes *within* and *between* regions had not changed.

4.3 Counterfactual analysis

A growing number of scholars are investigating the drivers of inequality using alternative decompositions to those previously reviewed. Among the methods that go beyond the mean, the *reweighting* approach has been one of the most applied in practice and was first introduced in the decomposition literature by DiNardo, Fortin and Lemieux (DFL) (1996). DFL decomposition is a semi-parametric approach that enables us to extend the results to the whole distribution of income, since it works with the entire population; not only with the mean. Although this is not the first time that the DFL methodology is being applied to the analysis of income inequality focusing on the spatial dimension (Dickey, 2014; Carrillo and Rothbaum, 2016), as far as we know, an analysis similar to the one we describe below has not yet been conducted. We focus on the particular contribution of territory to income inequality not only from a semi-parametric perspective, but linking it to the traditional and more theoretical proposals.

The non-parametric estimation of density functions by means of kernel methods is a subtle and complementary approach to use when the model followed by our data is unknown. It consists of building a function based on the sample values. If we have several samples, one for each different population, and we do not know the function they describe, we could create a density function for each sample by classifying the new individuals through a simple assignment to the population that has more values. In particular, kernel density functions allow us to estimate the counterfactual density distributions that we intend to study by putting into practice the DFL methodology. They can be defined as follows:

$$\hat{f}_n(t) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{t - X_i}{h_n}\right) \quad (11)$$

where n refers to the number of observations, K is a predetermined density named kernel, and h_n is a chain of smoothing parameters (*bandwidths*) that must slowly tend to zero.

DFL approach

We propose to analyse a selected sample of OECD countries to determine what would have happened to the distribution of the equivalent disposable household income in the final period ($t = 1$) if the territorial distribution of the population had remained constant and equal to that of the analysis starting point ($t = 0$). The individual observations of the income y, a vector of individual attributes x, and a date t belong to a joint distribution F (y, x, t) that, at a given date, becomes the conditional distribution F (y, $x|t$). Thus, at date $t = 1$, the actual density of incomes can be written as:

$$f(y; t_y = 1, t_x = 1) = \int_x f(y|x, t_y = 1) dF(x|t_x = 1) \quad (12)$$

We handled eight representative attributes of the households of interest in addition to the regional variable: one housing variable (owned/rented); one variable related to household composition and living arrangements (number of household members); five socio-demographic characteristics (age, marital status, immigration, health, and education); and one variable reporting on labour market activity (employment). The main reasons for choosing these variables and not others are the large number of countries selected and the different LIS waves covered by the analysis.[7]

[7] Data are lacking for some periods in certain countries.

[Table A.3 around here]

Assuming independence between the income structures and the distribution of attributes, the following represents the hypothetical or counterfactual density of incomes that would have prevailed if the distribution of attributes had remained the same as on the initial date:

$$f(y|t_y = 1, t_x = 0) = \int f(y|x,\, t_y = 1)dF(x|\, t_x = 0)$$

$$=\int f(y|x,\, t_y = 1)\Psi(x)dF(x|t_x = 1) \quad (13)$$

The counterfactual distribution (13) is similar to the actual one (12), except that it introduces a "reweighting" function:

$$\Psi(x) = dF(x|t_x = 0)/dF(x|t_x = 1) \quad (14)$$

Estimating Ψ is not straightforward, but DFL solves the implementation problem with the application of Bayes' rule in order to obtain:

$$\widehat{\Psi}(x) = \frac{Pr(t_x{=}0|x)}{Pr(t_x{=}1|x)}\frac{Pr(t_x{=}1)}{Pr(t_x{=}0)} \quad (15)$$

Unlike (12), equation (13) can be readily estimated by first pooling the individual observations from the two dates and then estimating a probit model for the likelihood that an observation is from date *t* given *x*. The estimates allow us to determine $\widehat{\Psi}(x)$ for each observation, which can be used to obtain the counterfactual density through weighted kernel methods.

In expression (15), $\Pr(t_x = 0|x)$ represents the probability that a randomly chosen individual with characteristics x (variables we have considered relevant in our analysis) belongs to the starting year when all the individuals in the sample are pooled together. $\Pr(t_x = 1|x)$ would reflect the same idea, but for the second period.

Once the pool of data is created, we have to estimate two probit models. The first model is an estimation considering all the attributes of interest, while the second would include all the explanatory variables of the previous estimation, except for the territory variable. The contribution of the territorial variable to income inequality is determined by the difference between the two counterfactual distributions generated.

5 Main results

The first analysis carried out is descriptive and allows us to illustrate the heterogeneity existing in the territorial units under examination. Table A.3 in the Appendix includes, for all the territories studied, the initial and final values of the mean logarithmic deviation (L), the population shares, and the number of observations used. As a complement, Figure A.1 in the Appendix provides an overview of data distribution through some of the most relevant measures of position and dispersion. These figures make it easy to check the symmetry in each particular case and to identify the presence of outliers.

5.1 Between and within regional inequality

In this subsection, we present a set of results corresponding to the standard decomposition by population subgroups. This analysis (see Table 1) provides us a first picture of income inequality in two specific moments in time (2000 and 2016 or nearest year available), and the figures are presented in both absolute and relative terms according to the original regional groupings provided by the LIS database. The general trend reveals a significant increase in total inequality in most countries, as well as a relevant growth for the within-region component in almost all of them.

Table 1. Spatial decomposition of the Mean Log Deviation (L) in the OECD: 1999-2016

Country	Units	Year	Total inequality (L)	Between (B)		Within (W)	
*Australia**	7	2001	0.20674	0.00255	1.2%	0.20419	98.8%
	7	2014	0.20977	0.00318	1.5%	0.20660	98.5%

*Canada****	10	2000	0.17509	0.00497	2.8%	0.17012	97.2%
	10	2013	0.19453	0.00587	3.0%	0.18866	97.0%
Chile[8]***	13	2000	0.51730	0.03214	6.2%	0.48516	93.8%
	13	2015	0.41412	0.02398	5.8%	0.39015	94.2%
France[9]**	8	2000	0.13408	0.00821	6.1%	0.12587	93.9%
	8	2010	0.15161	0.00381	2.5%	0.14780	97.5%
*Germany****	16	2000	0.11965	0.00316	2.6%	0.11649	97.4%
	16	2015	0.15813	0.00402	2.5%	0.15411	97.5%
*Italy***	20	2000	0.22779	0.03144	13.8%	0.19635	86.2%
	20	2014	0.26626	0.02342	8.8%	0.24284	91.2%
Mexico	32	2000	0.45041	0.06910	15.3%	0.38131	84.7%
	32	2012	0.42132	0.04450	10.6%	0.37682	89.4%
Poland	16	1999	0.16887	0.00336	2.0%	0.16551	98.0%
	16	2016	0.17442	0.00343	2.0%	0.17099	98.0%
Spain[10]***	18	2004	0.21216	0.01007	4.7%	0.20209	95.3%
	18	2013	0.24624	0.01713	7.0%	0.22911	93.0%
*United Kingdom***	12	2004	0.23035	0.00910	4.0%	0.22124	96.0%
	12	2013	0.20528	0.00666	3.2%	0.19863	96.8%
*United States****	51	2000	0.27129	0.00458	1.7%	0.26672	98.3%
	51	2016	0.29662	0.00398	1.3%	0.29263	98.7%

Note: Asterisks indicate that initial and final values are significantly different at the 90 (*), 95 (**) and 99 (***) % confidence levels.
Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (multiple countries; 1999-2016). Luxembourg: LIS.

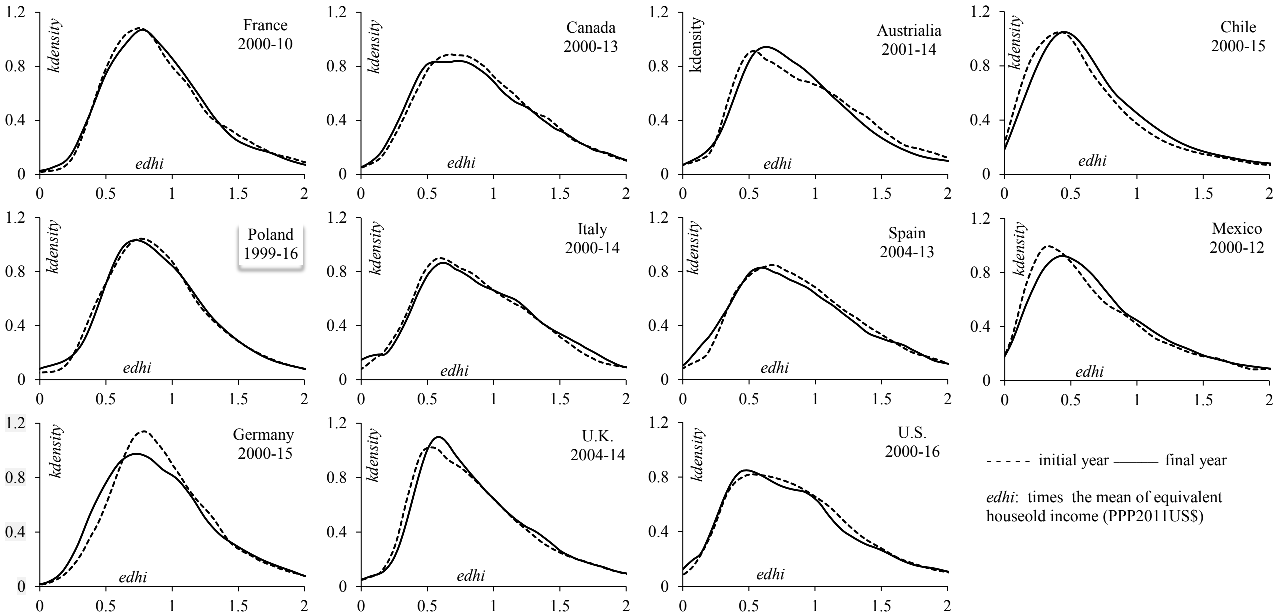
Equally striking from a general perspective is the process of convergence that took place during the first years of the 21st century. This fact is supported by the main figures of the

[8] Arica y Parinacota and Tarapaca have been considered as a single region (the data appear disaggregated in 2015, but not in 2000). For the same reason, Los Lagos and Los Ríos have been analysed as a single territory.
[9] The data corresponding to the Départements d'Outre-Mer (Guadeloupe, Martinique, French Guiana, Réunion and Mayotte) are only available for 2010 and have been deleted. In this way, we can perform a homogeneous analysis between the 8 territories that correspond to the NUTS-1 level.
[10] The Spanish constitutional legal system divides the country into 17 regions and 2 *Autonomous Cities* (Ceuta and Melilla). For the purposes of this exercise, and because of the disaggregation of the region variable in the LIS database, Ceuta and Melilla have been considered as a single entity.

population subgroups decompositions and by reading-through the actual density functions[11] corresponding to the first and last years of interest. Figure 1 allows us to clearly see that the 2015 distribution in Germany is flatter and slightly to the left of the 2000 distribution. Conversely, the comparison for Chile between the two actual distributions provides very clear evidence of the decrease in income inequality from 2000-2015.

[11] In these graphs, the outcome variable has been relativized to the median to obtain a straightforward interpretation of the results.

Figure 1. Initial and final actual density functions in the most populated OECD countries: 1999-2016



Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (multiple countries; 1999-2016). Luxembourg: LIS.

Regarding the specific countries examined, it is necessary to provide some clarifications. Chile experienced the greatest reduction in income inequality. This fall in income inequality since the beginning of the 21st century is explained, on one hand, by the implementation of three inclusive policies that benefited the most precarious households: *Chile solidario*, the gradual health reform called AUGE, and a noteworthy Social Security reform (Contreras and Ffrench-Davis, 2014). On the other hand, Parro and Reyes (2017) attributed it to the link between the factors that determine economic growth, and the focus on education that prompted individuals to invest in higher education.

Mexico has also experienced a remarkable decrease in inequality. The main drivers for this reduction during the last decade seem to be the labour market forces (a remarkable decrease in the wage ratio between skilled and unskilled workers), and institutional factors (minimum wage and unionization rate agreements) (Esquivel et al. 2010). Another noteworthy fact pertaining to Mexico is that it had the greatest weight of the *between* component among the eleven countries analysed.

In the European context, the United Kingdom deserves special mention, as it is the main exception among European countries regarding the evolution of the mean log deviation from 2004 to 2013. In Germany, the results are in accordance with the ones obtained by Biewen and Juhasz (2012) for an earlier period (2000-2006), who attributed the increase in income inequality to changes in the tax system, restructuring of household organization and to other variations in some highly important socioeconomic characteristics (such as age or education). The results observed in Italy could find justification in the works of Torrisi et al. (2015), who mainly attributed the reduction of regional disparities between the northern regions and the southern ones to changes in population dynamics, and Usseglio (2016), whose main findings detected that human capital has not played such an important role as employment opportunities. In a similar way, Tirado et al. (2016) confirmed the existence of great disparities between the most prosperous Spanish regions, those of the North-east, and the poorest territories, located in the South.[12] As regards France, Combes et al. (2011) underscored the recent convergence across their territories in labor productivity and drew special attention to the influence of the agglomeration economies. To conclude with the European scanning, Czyż and Hauke (2011) observed that the development-activating elements have not managed to reduce inter-regional differences in Poland and could explain the increase we see in our results.

Concerning Canada, some factors driving the recent evolution of family inequality *within* Canadian provinces could be the roles played by human capital and the life cycle (inter-temporal dynamics) (Gray et al. 2004).

In Australia, the discrepancies noticed around the between-region and within-region component do seem to lay on the effects of the 'mining boom decade' of 2001–2011 (Fleming and Measham, 2015).[13] Miranti et al. (2013) revealed, on the other hand, that income inequality varied noticeably when focusing on small spatial areas and justified the relevance of exploring territorial characteristics for a better understanding of both *between* and *within* regional complexity.

Lastly, 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. On the opposite side, there is some evidence

[12] They highlight that "as NEG literature stresses, the presence of economic activities characterized by the emergence of economies of scale makes accessibility to the nodes act as a catalyst in boosting regional development".
[13] Athanasopoulous and Vahid (2003) had previously confirmed the predominance of the within-region component in accounting for total income inequality. They also pointed out that this country's increase in income inequality during the 1990s had been more pronounced around major metropolitan areas.

indicating that innovation is not one of the main drivers in explaining increasing income inequality in the United States (Lee and Rodríguez-Pose, 2013).[14]

In general, the results obtained are consistent with the evidence reported in much of the literature, and they corroborate that inequality of disposable income not only differs between countries but also between regions within countries (Jesuit, 2003). In the same manner, the results lead us to confirm some stylized facts, including the very high weight of the within-region component (Piacentini, 2014).

These results are also confirmed when a spatial perspective is added to the distributive analysis taking stratification into account as well. In Table A.5 of the Appendix we present an analysis of the Gini index (ANOGI) that jointly accounts for the contribution of inter- and intra- regional inequality to total inequality and the contribution of the overlapping of both components' regional distributions (Frick et al. 2006). Perfect stratification (the inverse of overlapping) occurs when the incomes of each region belong to a specific range and the ranges of the regions do not overlap (see Yitzhaky and Schetchetman, 2013). The measure employed here is $I=Gb/Gbp$, where G_b and G_{pb} are the between group component of the Gini decomposition in Yitzhaki (1994) and Pyatt (1967), respectively.[15]

The last column presents the stratification index, which increases as inter-regional inequality rises and also when intra-regional inequality falls. The two more stratified countries in our sample are Italy and Mexico, followed by Chile and Spain. During the

[14] The relationship between innovation and income inequality is another interesting variant in the study of spatial inequality. These authors found a strong link in the case of European regions. However, their findings only provided inconclusive results for the territories of the United States. Differences in labor market flexibility as well as dissimilarities in the levels of migration look to be behind these discrepancies.
[15] It is identified in Milanovic and Yitzhaki (2002) as the loss of between group inequality due to overlapping and latter reinterpreted by Monti and Santoro (2011) for the two groups case, and by Allanson (2014) for the general case, as an index of non-overlapping or stratification.

analysis period, France registered a significant improvement due to the fall in interregional inequality, enhanced by an increase in intraregional inequalities. In Spain stratification worsened due to the rise of interregional disparities barely compensated by the increase in inequality between regions.

Explaining the varied country specific causes for this change is beyond the scope of this paper, as it would require bearing in mind not only the common drivers of the evolution of regional disparities, but also a variety of idiosyncratic factors. Among the most relevant ones, the dominant role played by some big cities like Paris (Dormard, 2004), the isolation of some of the poorest regions of East-Germany (Frick and Goebel, 2008) or the mining boom of 2000`s in Australia (Fleming and Measham, 2015) stand out. The evidence shown in this section not only makes updated results on the inter- and intra-regional components available, but also provides information on regional cohesion measured through the overlapping degree.

5.2 Decomposition of the *trend* of inequality

The static decomposition offers us a “picture” of inequality at one specific moment. This second decomposition will provide us with an idea of the evolution of the index. More importantly, it allows us to know and quantify the contribution of territory to income inequality through two of its four components.

This method is based on the mean log deviation (*L*) because of the difficulties involved in implementing it with other inequality indices. It breaks down income inequality into two blocks that also correspond to the four identities listed in equation (10). The extensive and generalizable increase in aggregate inequality observed can be disaggregated into these two blocks, as detailed below.

Table 2. Decomposition of the *trend* in aggregate inequality (Lx10[2]): 1999-2016

	Change in aggregate inequality	Contribution to ΔL attributable to variations in[16]							
		Within group inequality		Population shares				Mean group incomes	
	= [11a] + [11b] + [11c] + [11d]	[11a]		[11b]		[11c]		[11d]	
Countries and time	$\Delta L = L_f - L_i$	$\sum_g \bar{p}_g \Delta L_g$		$\sum_g \bar{L}_g \Delta p_g$		$\sum_g \left[\overline{\left(\frac{\mu_g}{\mu}\right)} - \overline{\log\left(\frac{\mu_g}{\mu}\right)}\right] \Delta p_g$		$\sum_g (\bar{q}_g - \bar{p}_g)\,\Delta \log \mu_g$	
Australia 2001-2014	0.303	0.224	(71.7%)	0.017	(5.3%)	0.000	(0.0%)	0.072	(23.0%)
Canada 2000-2013	1.945	1.809	(92.1%)	0.046	(2.3%)	0.010	(0.5%)	0.099	(5.0%)
Chile 2000-2015	− 10.317	− 9.479	(91.8%)	− 0.022	(0.2%)	− 0.017	(0.2%)	− 0.810	(7.8%)
France 2000-2010	1.753	2.141	(122.2%)	0.051	(2.9%)	− 0.007	− (0.4%)	− 0.433	− (24.7%)
Germany 2000-2015	3.848	3.690	(96.0%)	0.072	(1.9%)	− 0.019	− (0.5%)	0.102	(2.6%)
Italy 2000-2014	3.847	5.567	(144.7%)	− 0.919	− (23.9%)	− 0.123	− (3.2%)	− 0.678	− (17.6%)
Mexico 2000-2012	− 2.909	− 0.329	(11.3%)	− 0.120	(4.1%)	− 0.136	(4.7%)	− 2.316	(79.8%)
Poland 1999-2016	0.555	0.324	(58.4%)	0.225	(40.5%)	0.019	(3.3%)	− 0.012	− (2.2%)
Spain 2004-2013	3.408	2.670	(78.4%)	0.032	(0.9%)	0.006	(0.2%)	0.699	(20.5%)
United Kingdom 2004-2013	− 2.507	− 2.378	(94.9%)	0.116	− (4.6%)	0.010	− (0.4%)	− 0.255	(10.2%)
United States 2000-2016	2.532	2.514	(99.3%)	0.078	(3.1%)	− 0.010	− (0.4%)	− 0.050	− (2.0%)

Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (multiple countries; 1999-2016). Luxembourg: LIS.

[16] See equation (11).

One effect would be due to inequality within the regions and corresponds entirely to component [a]. It gathers the effect of the changes in the intra-group component (*pure within effect*) and is the most important component in our decomposition. However, even more significant is the fact that this first component is greater than the observed change in *L* for many countries, being especially notable in Italy. Therefore, variations in relative means, or group compositions, can hardly account for the rise in inequality during the period of reference.

The second effect refers to inequality between the regions and can be divided into the *income effect* and the *allocation effect*. The *income effect* (component [d]) denotes the effect on the variation of the relative income of the regions considered. Also noteworthy is its negative contribution (in most of the cases), which represents the changes attributable to shifting relative income between groups. This contribution is positive in two of the three countries where inequality grew the most (Germany and Spain) and in Australia and Canada.

The *allocation effect* (components [b] and [c]) incorporates modifications attributable to changing numbers in the different regions (variations in their populations). Components [b] and [c] acquire special relevance concerning the question of population shares. They show a very revealing fact: territory exerts some influence on income inequality, but the magnitude of change as well as the sign of variation differ markedly depending on the country. In this sense, the evolution of country disparities is driven fundamentally by the role of population movements within the regions themselves [b], rather than by the effect of movements between territories [c], which is the decomposition component that contributes the least.

Of the three countries where a reduction in aggregate inequality is observed, Chile and Mexico prove how population shares contributed to reduce inequality. In the United Kingdom, the opposite happens. Likewise, it is important to clarify that in Italy, the country (along with Germany) where aggregate inequality increases the most, territory has the greatest reducing effect, mainly through component [b].

The opposite behaviour of two of the countries analysed also draws our attention: Spain and the United Kingdom. In Spain, where inequality increased meaningfully, the four elements of the decomposition contribute with a positive sign. The opposite is true for the United Kingdom, where there was a remarkable decrease in income inequality, with all components contributing to this reduction.

All these dissimilar effects lead us to ask the following questions. What would have happened in the levels of income inequality if the population weights of the different regions had remained constant during the period investigated? This is precisely what we try to elucidate and measure in the simulation exercise implemented through the counterfactual analysis.

5.3 How much inequality is explained by territory?

We carry out a decomposition of the mean log deviation index into three components (see Table 3). On the one hand, we can identify an *unexplained inequality* that would be included in the variation attributable to the "other characteristics" column. It would show the contribution to inequality of those characteristics or explanatory variables not taken into account in the analysis. On the other hand, we find two other elements (last two columns) that we group together and call *explained inequality.* The first element represents the contribution of the region variable to inequality (variation attributable to "territory"), our main objective in this paper. The second element of *explained inequality* is the contribution to inequality of the other attributes relevant to the analysis – variation attributable to “control variables”. The latter correspond to the variables described in the data section and are further detailed in Table A.4.

Table 3. Estimation results of the DFL decomposition (Lx10^2): 1999-2016

Countries chosen and time period	*L* change: observed	*L* change: estimated (contribution of the decomposition components)					
		Unexplained inequality		*Explained* inequality			
	= [1] + [2] + [3]	[1] Other characteristics		[2] Territory		[3] Control variables	
	$\Delta L = L_f - L_i$	$L_f(c2) - L_i$		$L_f(c1) - L_f(c2)$		$L_f - L_f(c1)$	
Australia 2001-2014	0.303	1.632	(538.5%)	– 0.006	– (2.0%)	– 1.323	– (436.5%)
Canada 2000-2013	1.945	3.106	(159.7%)	0.088	(4.5%)	– 1.250	– (64.3%)
Chile 2000-2015	– 10.317	– 10.762	(104.3%)	– 0.166	(1.6%)	0.610	– (5.9%)
France 2000-2010	1.753	0.888	(50.6%)	– 0.063	– (3.6%)	0.928	(52.9%)
Germany 2000-2015	3.848	3.453	(89.7%)	– 0.028	– (0.7%)	0.423	(11.0%)
Italy 2000-2014	3.847	3.266	(84.9%)	– 0.479	– (12,5%)	1.060	(27.5%)
Mexico 2000-2012	– 2.909	– 4.331	(148.9%)	– 0.192	(6.6%)	1.614	– (55.5%)
Poland 1999-2016	0.555	0.406	(73.1%)	0.093	(16.8%)	0.056	(10.1%)
Spain 2004-2013	3.408	0.117	(3.4%)	0.057	(1.7%)	3.235	(94.9%)
United Kingdom 2004-2013	– 2.507	– 4.621	(184.3%)	– 0.200	(8.0%)	2.314	– (92.3%)
United States 2000-2016	2.532	0.785	(31.0%)	0.059	(2.3%)	1.689	(66.7%)

Notes: (1) L_f and L_i are the observed inequality in the last and first year, respectively; (2) $L_f(c1)$ is the inequality estimated for the first counterfactual income distribution (the one that would have prevailed in the final year if the eight selected attributes and the regional variable had stayed at the same values of the first year); (3) $L_f(c2)$ is the inequality estimated for the second counterfactual income distribution (remaining constant the eight selected attributes, with the figures of the first year, but not the regional variable).
Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (multiple countries; 1999-2016). Luxembourg: LIS.

First, territory has a “reducing effect” on inequality in most of the countries examined (7 of 11). That is, if the demographic weights of the regions examined had not changed in recent years, income inequality would have been higher. Although the magnitude of this effect varies significantly among the countries analyzed, it is necessary to highlight this as the most striking result. This simulation enhances the results presented in the dynamic decomposition and confirms the importance of this variable in countries such as Italy and Mexico. On the other hand, it is equally remarkable to note that this “reducing effect” is not only found in countries where aggregate inequality has diminished but also in those where it has increased considerably.

A second and more specific comment relates to the relevance of the attributes chosen for the analysis. As expected, the weight assigned to "other characteristics" is very high in this decomposition. However, the results obtained also show that the control variables play an important role in explaining income inequality, and the influence of territory is not at all negligible.

To test the sensitivity of the results to the measure chosen, we also performed the three decompositions with other inequality indices. The results change very slightly, and the general conclusions do not vary significantly with other measures. In general terms, our results confirm and update those of previous works, as well as some of the theoretical premises reviewed in the previous sections. First, income differences between regions have a small explanatory capacity of inequality in the selected countries. As mentioned earlier, considering only a single explanatory dimension significantly reduces the weight of this component, as previous studies showed for other periods and countries.

Our results are in keeping with empirical evidence on the determinants of inequalities in market income in several of the countries included in the sample. As stressed by different authors, wages and unemployment are by far the most important channels of adjustment to macro-economic shocks in these countries being also the main drivers of inequality, while labour migration has played a secondary role (Brandsma et al. 2013). Second, we confirm that the changes in the territorial distribution of population affect inequality. Among other possible causes, these flows are related to spatial agglomeration effects -as OECD (2018) data seems to confirm-, which, in turn, are at the root of regional economic disparities that can affect inequality in income distribution.

6 Conclusions

Our study makes three contributions. First, we carry out different kinds of decomposition analyses to identify the effect on inequality of both income differences between regions and within regions. Second, we provide new evidence over a more varied group of countries and a more recent period than in previous studies. Third, we add to previous studies the simulation of the impact on inequality that the changes in the inter-territorial distribution of the population may have had.

We have implemented several complementary techniques in using the LIS database and by relying on the potential of the decompositions of inequality, useful tools for the correct design of redistributive policies. The idea was not only to provide a particular perspective on the same problem (the relevance of territory in the description of income inequality) but to offer a global and joint perspective by aggregating all of them. The question we tried to answer in this paper was concise: To what extent does territory drive the main changes observed in the recent evolution of income inequality? Despite not having found a strong pattern among the countries selected, the empirical specification applied for the DFL decomposition does seem to capture much of the change observed in income inequality during the years analysed. The region variable, in particular, exerts a noteworthy effect. In particular, the evidence found reveals a generalized and interesting "reducing effect" in income inequality that is directly linked to territory. All this takes on great importance because high levels of spatial inequality are undesirable for the development and economic growth of any society. They can constrain progress, are often associated with crime problems, reveal institutional weaknesses, and even hinder social cohesion among their regions (Atkinson, 2015).

In closing, we have shed some light on the transcendence of the territorial dimension in inequality by implementing an application not developed thus far: a counterfactual analysis for population weights in the context of international comparative analysis. In general terms, our results confirm and update those of previous works, as well as some of the theoretical premises of the related literature. Income differences between regions have a small explanatory capacity of inequality in the selected countries while the changes in the territorial distribution of population affect inequality. Even with the challenges of providing conclusive evidence -the influence of social and institutional factors is very diverse in each country- we have obtained some interesting results that underscore the relevance of this determining factor and that could encourage the pursuit of future research on spatial inequality.

These results are important both to inform the theory and to contribute to the design of public policies related to social cohesion. Regarding the first issue, we have confirmed that changes in the territorial distribution of population affect inequality. As stated by NEG models, these flows are related to spatial agglomeration effects, which, in turn, are at the root of regional economic disparities that can affect inequality in the income distribution.

A political implication of our results is that if population mobility is conceived as a possible tool to influence inter- and intra-inequality, it is necessary to coordinate this type of specific policies with other redistributive measures in order to optimize their joint impact. In any case, we need to remain cautious when drawing other policy implications, given the difficulty of going from an accounting exercise to the design of specific policies. As stated by Fujita and Krugman (2005), "because geography is such a crucial factor in development, and there are undoubtedly strong policy implications of some sort, it is an important subject for further research". In this regard, some of the limitations of our work may be interpreted as further promising extensions.

One of these caveats is the very high weight of "other characteristics" suggesting that some relevant variables might have been omitted. Since we focus exclusively on the information included in LIS data, relevant issues such as technology, trade, or decentralization are not considered here. Second, a larger and more varied sample might enrich the analysis. Third, most of our analyses are accounting exercises of how much the changes in the territorial distribution of population may influence recent inequality trends. A more detailed analysis of population changes at a higher level of disaggregation should be a promising extension of this work.

APPENDIX

Table A.1. Basic descriptive statistics for the most populated OECD countries: 1999-2016

Wave	Years	AU 2001-2014	CA 2000-2013	CL 2000-2015	FR 2000-2010	DE 2000-2015	IT 2000-2014	MX 2000-2012	PL 1999-2016	ES 2004-2013	UK 2004-2013	US 2000-2016
		Countries and time										
(a) Sample size (number of household respondents)												
V	99/01	6,786	28,881	65,036	10,305	11,796	8,000	10,108	31,428	-	-	78,054
VI	03/05	11,361	27,665	68,153	10,240	11,294	8,012	22,595	32,214	12,996	27,753	76,447
VII	06/08	9,345	26,560	73,720	-	10,921	7,977	29,468	37,366	13,014	24,977	75,872
VIII	09/11	18,008	24,826	71,460	10,342	16,703	7,941	27,655	37,412	13,109	25,350	75,188
IX	12/14	14,115	23,014	66,725	-	15,946	8,151	9,002	37,181	11,965	20,135	51,498
X	15/16	-	-	83,887	-	14,426	-	-	36,886	-	-	69,957
(b) Population size (inhabitants: Ix10^3)												
V	99/01	18,747	29,798	15,039	59,329	83,150	56,635	98,163	38,666	-	-	275,662
VI	03/05	19,521	30,851	15,571	59,500	83,086	57,208	102,989	37,784	42,874	57,945	286,674
VII	06/08	20,508	31,889	16,115	-	82,739	58,360	111,612	37,708	45,109	59,829	292,009
VIII	09/11	21,343	32,945	16,583	61,781	80,579	58,854	114,560	37,726	45,900	61,041	296,992
IX	12/14	22,389	34,017	17,256	-	81,351	60,439	117,284	38,101	45,977	62,853	305,234
X	15/16	-	-	17,530	-	83,157	-	-	38,004	-	-	310,964
(c) Equivalent household disposable income (PPP$US2011)												
V	99/01	23,692	27,602	9,417	23,683	27,464	21,334	6,664	9,538	-	-	37,803
VI	03/05	26,504	29,335	9,217	24,408	27,760	28,844	6,950	9,647	21,109	26,612	38,226
VII	06/08	33,481	31,904	10,057	-	28,225	22,247	8,072	11,708	23,086	28,747	39,040
VIII	09/11	33,524	32,690	10,681	27,982	28,295	21,527	7,038	13,637	21,069	27,617	37,557
IX	12/14	39,977	35,044	13,121	-	28,229	19,270	7,520	13,684	21,161	27,128	37,882
X	15/16	-	-	13,080	-	28,870	-	-	15,925	-	-	41,209

Notes: (1) AU=Australia, CA=Canada, CL=Chile, FR=France, DE=Germany, IT=Italy, MX=Mexico, PL=Poland, ES=Spain, UK=United Kingdom, US=United States; (2) To ensure the consistency of all the tables, data from 1999 in the United Kingdom and from 2000 in Spain, relative to Wave V, have been excluded. There are no data for Northern Ireland in 1999 and there is no regional information for Spain at a NUTS-2 level of disaggregation in 2000; (3) Overseas regions are not taken into account in France.
Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (multiple countries; 1999-2016). Luxembourg: LIS.

Table A.2. Checking aggregation effects

Table A.2.1. Spatial decomposition of the Mean Log Deviation (L) in the OECD: 1999-2016

Country	Units	Year	Total inequality (L)	Between (B)		Within (W)	
*Italy***	5	2000	0.22779	0.02580	11.3%	0.20199	88.7%
	5	2014	0.26626	0.01713	6.4%	0.24914	93.6%
*Mexico****	8	2000	0.45041	0.04727	10.5%	0.40314	89.5%
	8	2012	0.42132	0.02756	6.5%	0.39376	93.5%
*Spain****	7	2004	0.21216	0.00794	3.7%	0.20422	96.3%
	7	2013	0.24624	0.01308	5.3%	0.23316	94.7%
*United States***	9	2000	0.27129	0.00171	0.6%	0.26958	99.4%
	9	2016	0.29662	0.00174	0.6%	0.29488	99.4%

Note: Asterisks indicate that initial and final values are significantly different at the 90 (*), 95 (**) and 99 (***) % confidence levels.

Table A.2.2. Decomposition of the *trend* in aggregate inequality (Lx10^2): 1999-2016

	Change in aggregate inequality	Contribution to ΔL attributable to variations in							
		Within group inequality		Population shares				Mean group incomes	
		[11a]		[11b]		[11c]		[11d]	
Countries and time	$\Delta L = L_f - L_i$	$\sum_g \bar{p}_g \Delta L_g$		$\sum_g \bar{L}_g \Delta p_g$		$\sum_g \left[\overline{\left(\frac{\mu_g}{\mu}\right)} - \overline{\log\left(\frac{\mu_g}{\mu}\right)}\right] \Delta p$		$\sum_g (\bar{q}_g - \bar{p}_g)\, \Delta \log \mu_g$	
Italy (5) 2000-2014	3.847	5.076	(132.1%)	– 0.362	– (9.4%)	– 0.054	– (1.4%)	– 0.816	– (21.2%)
Mexico (8) 2000-2012	– 2.909	– 0.835	(28.8%)	– 0.103	(3.6%)	– 0.054	(1.9%)	– 1.903	(65.8%)
Spain (7) 2004-2013	3.408	2.868	(84.1%)	0.026	(0.8%)	0.007	(0.2%)	0.508	(14.9%)
United States (9) 2000-2016	2.532	2.441	(96.4%)	0.088	(3.5%)	– 0.003	– (0.1%)	0.006	(0.2%)

Table A.2.3. Estimation results of the DFL decomposition (Lx10^2): 1999-2016

Countries chosen and time period	L change: observed	L change: estimated (contribution of the decomposition components)					
		Unexplained inequality		*Explained* inequality			
	= [1] + [2] + [3]	[1] Other characteristics		[2] Territory		[3] Control variables	
	$\Delta L = L_f - L_i$	$L_f(c2) - L_i$		$L_f(c1) - L_f(c2)$		$L_f - L_f(c1)$	
Italy (5) 2000-2014	3.847	2.542	(66.1%)	0.245	(6.4%)	1.060	(27.5%)
Mexico (8) 2000-2012	– 2.909	– 4.470	(153.7%)	– 0.054	(1.8%)	1.614	– (55.5%)
Spain (7) 2004-2013	3.408	0.106	(3.1%)	0.066	(1.9%)	3.235	(94.9%)
United States (9) 2000-2016	2.532	0.744	(29.4%)	0.100	(3.9%)	1.689	(66.7%)

Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (multiple countries; 1999-2016). Luxembourg: LIS.

Table A.3. Descriptive statistics in the OECD countries by territories: 1999-2016

Territories of Australia: 2001-2014	Initial year			Final year		
6 Federated States and Canberra	L_i	% Pop.	N	L_f	% Pop.	N
Canberra region and Northern T.	0.18079	2.32	381	0.16821	2.42	1,395
New South Wales	0.20682	32.75	1,530	0.23913	31.82	2,532
Queensland	0.19940	19.08	1,218	0.22140	20.17	2,272
South Australia	0.16315	8.33	889	0.19465	7.66	2,131
Tasmania	0.13893	2.60	482	0.15753	2.39	1,704
Victoria	0.20055	25.30	1,400	0.21083	25.10	2,287
Western Australia	0.16391	9.62	886	0.28533	10.42	1,841

Territories of Canada: 2000-2013	Initial year			Final year		
10 Provinces	L_i	% Pop.	N	L_f	% Pop.	N
Alberta	0.17309	9.85	2,342	0.19393	11.49	2,213
British Columbia	0.19719	13.15	2,514	0.20585	12.94	2,434
Manitoba	0.16173	3.61	2,172	0.17011	3.46	2,049
New Brunswick	0.15330	2.45	1,728	0.14225	2.14	1,278
Newfoundland	0.14688	1.72	1,177	0.16608	1.52	884
Nova Scotia	0.15204	3.04	1,954	0.17474	2.66	1,388
Ontario	0.17213	38.47	8,384	0.20684	38.93	5,968
Prince Edward Island	0.14949	0.45	821	0.14001	0.42	622
Quebec	0.15434	24.11	5,755	0.15791	23.44	4,510
Saskatchewan	0.16554	3.15	2,034	0.19219	3.01	1,718

Territories of Chile: 2000-2015	Initial year			Final year		
13 Regions	L_i	% Pop.	N	L_f	% Pop.	N
Antofagasta	0.43746	3.05	1,804	0.34863	3.06	2,025
Arica y Parinacota and Tarapacá	0.40927	2.59	2,278	0.40753	2.73	3,419
Atacama	0.36747	1.59	1,930	0.30354	1.49	3,989
Aysén	0.42563	0.62	878	0.34652	0.64	1,152
Biobío	0.49256	12.38	11,432	0.34151	12.05	11,490
Coquimbo	0.40592	3.89	3,146	0.32915	4.17	3,745
La Araucanía	0.57609	5.56	6,434	0.36124	5.71	7,040
Libertador Bernardo O'Higgins	0.34134	5.20	4,746	0.31847	5.29	7,165
Los Lagos and Los Ríos	0.44695	7.12	5,880	0.34265	7.24	9,562
Magallanes and La Antártica Chile	0.57442	1.03	845	0.38106	0.94	1,892
Maule	0.51325	6.04	6,370	0.32077	6.13	5,687
Región Metropolitana Santiago	0.51783	40.20	13,100	0.43930	39.74	17,723
Valparaíso	0.36132	10.69	6,193	0.34529	10.82	8,998

Territories of France: 2000-2010	Initial year			Final year		
NUTS-1 (8 ZEAT[17])	L_i	% Pop.	N	L_f	% Pop.	N
Bassin parisien	0.13464	17.39	1,842	0.12641	17.28	1,922
Centre-est	0.12448	11.83	1,221	0.13911	12.04	1,138
Est	0.10225	9.19	1,081	0.12760	8.63	936
Méditerranée	0.16017	12.19	1,224	0.15202	13.00	1,193
Nord	0.13920	6.47	675	0.13915	6.14	761
Ouest	0.12266	13.49	1,535	0.14600	13.35	1,568
Région parisienne	0.14956	18.06	1,609	0.19289	18.18	1,625
Sud-ouest	0.13501	11.38	1,118	0.15531	11.37	1,199

Territories of Germany: 2000-2015	Initial year			Final year		
NUTS-1 (16 Länder)	L_i	% Pop.	N	L_f	% Pop.	N
Baden-Wuerttemberg	0.12017	12.46	1,381	0.15616	12.81	1,645
Bavaria	0.13369	14.54	1,634	0.16694	15.48	2,322
Berlin	0.15679	4.81	487	0.15221	4.81	614

[17] *Zones d'études et d'aménagement du territoire* (Research and National Development Zones).

Brandenburg	0.08048	3.08	503	0.15835	2.98	569
Bremen	0.23559	1.07	100	0.18832	0.87	105
Hamburg	0.22795	2.21	182	0.14440	2.40	265
Hesse	0.15632	7.35	789	0.21360	7.39	989
Lower Saxony	0.13515	9.32	967	0.15217	9.54	1,357
Mecklenburg-Western Pomerania	0.10637	2.18	289	0.15525	2.03	330
North Rhine-Westphalia	0.13366	21.77	2,512	0.21261	21.36	2,930
Rhineland-Palatinate	0.11541	4.46	551	0.12869	4.70	713
Saarland	0.08532	1.56	171	0.13989	1.21	140
Saxony	0.08323	5.52	857	0.11861	5.42	878
Saxony-Anhalt	0.09106	3.19	508	0.13457	2.85	516
Schleswig-Holstein	0.15073	3.63	367	0.13038	3.43	505
Thuringia	0.08210	2.87	498	0.16015	2.72	548

Territories of Italy: 2000-2014	Initial year			Final year		
NUTS-2 (20 Regioni)	L_i	% Pop.	N	L_f	% Pop.	N
Abruzzo	0.23960	2.02	228	0.14706	2.04	204
Basilicata	0.11683	1.99	95	0.15293	3.12	128
Calabria	0.15522	2.61	210	0.16736	3.62	217
Campania	0.21439	8.82	815	0.22106	7.63	716
Emilia Romagna	0.14742	7.66	751	0.16686	7.78	677
Friuli	0.15542	2.17	255	0.14613	3.49	214
Lazio	0.12401	9.74	425	0.34521	9.48	452
Liguria	0.14654	3.25	316	0.10185	3.66	347
Lombardia	0.18214	16.62	860	0.18865	12.17	944
Marche	0.18889	2.36	328	0.15291	3.16	345
Molise	0.22230	0.68	83	0.27921	1.78	111
Piemonte	0.13844	8.02	732	0.21700	9.02	725
Puglia	0.20749	6.17	471	0.24561	5.72	453
Sardegna	0.18040	2.69	308	0.21700	2.45	343
Sicilia	0.28245	8.14	630	0.26316	5.50	618
Toscana	0.13941	6.12	598	0.15180	6.01	605
Trentino	0.12557	1.52	161	0.13015	4.90	238
Umbria	0.10067	1.42	271	0.24910	1.55	277
Valle d'Aosta	0.06396	0.17	25	0.10929	0.71	43
Veneto	0.20700	7.84	439	0.14882	6.22	499

Territories of Poland: 1999-2016	Initial year			Final year		
NUTS-2 (16 Voivodeships)	L_i	% Pop.	N	L_f	% Pop.	N
Dolnoslaskie	0.13657	7.93	2,489	0.13035	8.10	2,899
Kujawsko-Pomorskie	0.12881	5.46	1,728	0.14503	5.24	1,968
Lodzkie	0.13947	8.47	2,665	0.13559	7.02	2,515
Lubelskie	0.14980	5.71	1,791	0.16936	5.33	2,062
Lubuskie	0.10497	2.81	889	0.12231	2.60	999
Malopolskie	0.13171	7.36	2,307	0.12876	8.26	3,117
Mazowieckie	0.17919	13.03	4,059	0.19126	14.99	5,535
Opolskie	0.15615	2.69	838	0.11968	2.49	973
Podkarpackie	0.12130	5.06	1,592	0.10656	4.85	1,878
Podlaskie	0.13883	2.80	883	0.15070	3.12	1,152
Pomorskie	0.13809	5.36	1,683	0.20087	5.92	2,230
Slaskie	0.10355	13.40	4,228	0.11558	12.42	4,297
Swietokrzyskie	0.14022	3.07	960	0.12758	3.13	1,193
Warminsko-Mazurskie	0.12267	4.23	1,332	0.14271	3.65	1,478
Wielkopolskie	0.16812	8.44	2,661	0.12619	8.41	3,066
Zachodnio-Pomorskie	0.14879	4.19	1,323	0.10365	4.49	1,524

Territories of Mexico: 2000-2012	Initial year			Final year		
32 Federal Entities	L_i	% Pop.	N	L_f	% Pop.	N
Aguascalientes	0.21190	0.99	226	0.31265	1.06	266
Baja California Norte	0.24744	2.59	264	0.31505	2.85	267
Baja California Sur	0.18527	0.44	209	0.29350	0.60	264

Campeche	0.37318	0.72	218	0.42726	0.74	287
Chiapas	0.55252	4.15	325	0.44411	4.31	296
Chihuahua	0.21598	2.91	247	0.28289	3.08	266
Ciudad de México	0.40714	8.44	456	0.36397	7.54	365
Coahuila de Zaragoza	0.24660	2.23	289	0.21898	2.44	266
Colima	0.35353	0.56	203	0.29144	0.59	275
Durango	0.29121	1.49	294	0.44399	1.46	272
Guanajuato	0.58894	4.79	288	0.36621	4.85	266
Guerrero	0.49396	3.16	311	0.39153	2.99	254
Hidalgo	0.35198	2.31	254	0.34716	2.36	280
Jalisco	0.23771	6.53	345	0.45714	6.53	334
México	0.39081	13.58	402	0.43785	13.80	372
Michoacán de Ocampo	0.35990	4.12	353	0.28050	3.82	288
Morelos	0.45057	1.61	197	0.35123	1.58	265
Nayarit	0.31086	0.94	213	0.49357	0.99	281
Nuevo León	0.23969	3.94	367	0.30357	4.17	243
Oaxaca	0.52580	3.55	238	0.52159	3.35	280
Puebla	0.40518	5.28	308	0.34095	5.13	275
Querétaro	0.61300	1.45	214	0.46545	1.63	280
Quintana Roo	0.41285	0.92	194	0.29955	1.24	277
San Luis Potosí	0.28936	2.36	275	0.37650	2.29	263
Sinaloa	0.32482	2.61	252	0.39448	2.48	275
Sonora	0.44931	2.29	251	0.30667	2.41	261
Tabasco	0.47245	1.95	224	0.42890	1.97	263
Tamaulipas	0.25368	2.85	279	0.32535	2.93	243
Tlaxcala	0.28028	1.01	232	0.22711	1.05	273
Veracruz-Llave	0.37690	7.13	1,735	0.42836	6.72	360
Yucatán	0.53801	1.71	234	0.26250	1.74	267
Zacatecas	0.33491	1.38	211	0.42849	1.31	278

Territories of Spain: 2004-2013	Initial year			Final year		
NUTS-2 (17 Regiones + CA[18])	L_i	% Pop.	N	L_f	% Pop.	N
Andalucía	0.17243	16.80	1,610	0.24659	17.25	1,478
Aragón	0.21117	3.06	582	0.18065	2.95	541
Canarias	0.20308	4.11	645	0.23558	4.39	501
Cantabria	0.20838	1.24	344	0.14411	1.32	295
Castilla La Mancha	0.23397	4.23	680	0.19186	4.31	564
Castilla y León	0.22080	5.96	913	0.21670	5.68	845
Cataluña	0.17737	16.49	1,376	0.19581	16.18	1,264
Ceuta and Melilla	0.34812	0.29	265	0.31958	0.27	258
Comunidad de Madrid	0.18062	13.26	801	0.22300	13.85	1,134
Comunidad Foral de Navarra	0.19075	1.38	429	0.15204	1.39	426
Comunidad Valenciana	0.17656	10.90	1,089	0.18500	10.87	891
Extremadura	0.20687	2.48	554	0.19476	2.35	506
Galicia	0.18238	6.31	911	0.16736	5.86	811
Illes Balears	0.19178	2.31	508	0.23154	2.38	373
La Rioja	0.17093	0.69	411	0.17045	0.70	385
País Vasco	0.17090	5.11	731	0.15756	4.86	696
Principado de Asturias	0.20210	2.59	593	0.18328	2.50	498
Región de Murcia	0.18230	2.79	557	0.20979	2.91	499

Territories of UK: 2004-2013	Initial year			Final year		
NUTS-1 (12 Regions)	L_i	% Pop.	N	L_f	% Pop.	N
East Midlands	0.19447	7.00	1,849	0.17747	7.17	1,294
East of England	0.21671	8.96	2,216	0.19318	9.19	1,674
London	0.31422	12.70	2,548	0.26678	12.44	1,748
North East (including Cumbria)	0.17741	4.32	1,103	0.16176	4.26	743
North West	0.18978	11.52	2,996	0.17316	11.29	1,967
Northern Ireland	0.18873	2.60	1,927	0.16218	2.74	1,965
Scotland	0.18281	8.87	4,523	0.18119	8.90	3,000
South East (excluding London)	0.26375	13.28	3,294	0.20641	13.43	2,421

[18] CA = *Ciudades Autónomas* (Ceuta and Melilla).

South West	0.20921	8.69	2,163	0.18056	8.60	1,448
Wales	0.17993	4.86	1,234	0.16364	4.92	875
West Midlands	0.18559	8.73	2,172	0.16094	8.65	1,519
Yorkshire and Humber	0.16903	8.47	2,016	0.19046	8.41	1,483

Territories of US: 2000-2016	Initial year			Final year		
50 States and Distric of Columbia	L_i	% Pop.	N	L_f	% Pop.	N
Alabama	0.25001	1.60	1,255	0.36027	1.54	1,487
Alaska	0.24050	0.21	1,081	0.27291	0.21	769
Arizona	0.22902	1.78	1,047	0.32178	2.15	1,123
Arkansas	0.24447	1.00	903	0.34505	0.96	1,294
California	0.28731	11.03	5,611	0.35227	11.04	6,226
Colorado	0.25764	1.57	1,522	0.25882	1.83	895
Connecticut	0.26327	1.25	1,339	0.31166	1.13	685
Delaware	0.24366	0.28	943	0.25084	0.31	746
District of Columbia	0.33622	0.23	1,015	0.42728	0.25	1,468
Florida	0.28213	6.03	3,455	0.30774	6.70	3,367
Georgia	0.23652	2.86	1,094	0.33338	3.14	1,566
Hawaii	0.19624	0.38	1,031	0.27041	0.38	1,094
Idaho	0.24410	0.46	996	0.27659	0.50	1,031
Illinois	0.27631	4.35	2,721	0.30223	4.01	1,982
Indiana	0.27836	2.24	1,389	0.27601	2.08	1,064
Iowa	0.19144	1.08	1,293	0.25753	1.05	758
Kansas	0.22011	1.01	1,276	0.30665	0.92	850
Kentucky	0.24158	1.48	1,062	0.26942	1.45	864
Louisiana	0.26233	1.56	870	0.32899	1.47	1,680
Maine	0.20792	0.50	1,268	0.26116	0.47	545
Maryland	0.27886	1.92	1,274	0.27643	1.84	885
Massachusetts	0.28009	2.37	1,433	0.29150	2.19	1,374
Michigan	0.25455	3.56	2,172	0.24912	3.21	1,496
Minnesota	0.25672	1.81	1,406	0.23488	1.76	853
Mississippi	0.27511	1.02	800	0.31658	0.91	1,347
Missouri	0.27557	2.05	1,180	0.27897	1.93	950
Montana	0.21289	0.33	866	0.25419	0.35	1,397
Nebraska	0.21559	0.62	1,217	0.22765	0.60	805
Nevada	0.24943	0.70	1,417	0.25034	0.90	916
New Hampshire	0.23964	0.46	1,281	0.23014	0.43	833
New Jersey	0.26053	2.98	1,963	0.30917	2.70	1,373
New Mexico	0.28593	0.62	1,017	0.37996	0.64	1,482
New York	0.31062	6.70	4,202	0.32928	6.13	2,912
North Carolina	0.24438	2.93	1,652	0.32193	3.19	1,619
North Dakota	0.20654	0.25	1,101	0.26671	0.25	937
Ohio	0.22308	4.19	2,475	0.28857	3.76	1,793
Oklahoma	0.26576	1.26	1,091	0.31001	1.24	1,069
Oregon	0.25090	1.27	1,198	0.25803	1.36	997
Pennsylvania	0.25206	4.48	2,659	0.28285	4.08	1,793
Rhode Island	0.25060	0.40	1,326	0.27223	0.36	613
South Carolina	0.31398	1.45	1,014	0.32509	1.61	1,156
South Dakota	0.24155	0.28	1,210	0.30113	0.28	690
Tennessee	0.31564	2.10	973	0.31969	2.18	1,298
Texas	0.32035	6.95	3,504	0.31745	8.00	4,207
Utah	0.18434	0.68	932	0.26970	0.83	932
Vermont	0.34441	0.24	1,174	0.26549	0.22	821
Virginia	0.27686	2.53	1,331	0.34010	2.56	1,286
Washington	0.31934	2.15	1,350	0.31141	2.26	1,229
West Virginia	0.23042	0.69	1,132	0.30432	0.60	1,465
Wisconsin	0.27140	1.94	1,488	0.26541	1.89	925
Wyoming	0.22652	0.18	1,045	0.27274	0.19	1,010

Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (multiple countries; 1999-2016). Luxembourg: LIS.

Table A.4. Control variables used in the probit regressions

	Variables	Definition
Geography and housing	Owned/Rented housing (h-file) LIS variable: own	Indicator of housing tenure. We have defined it as a dummy variable (1=owned; 0=rented/others).
Household composition and living arrangements	Household members (h-file) LIS variable: *nhhmem*	Number of household members. We have defined it as a dummy variable (1=one member; 0=more than one member).
Socio-demographic characteristics	Age (p-file) LIS variable: *age*	Age in years (min/35=1, 36/55=2, 36/max=3)
	Marital status (p-file) LIS variable: *marital*	Classification of persons according to their marital status, as provided in relation to the marriage laws or customs of the country. We have defined it as a dummy variable (1=married/in consensual union; 0=another status).
	Immigration (p-file) LIS variable: *immigr*	All persons 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).
	Health (p-file) LIS variable: *disabled*	Disabled persons 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).
	Education (p-file) LIS variable: *educ*	Recoding of highest level of education completed into three categories: low: less than secondary education completed (never attended, no completed education or education completed at the ISCED levels 0, 1 or 2); medium: secondary education completed (completed ISCED levels 3 or 4); high: tertiary education completed (completed ISCED levels 5 or 6).
Labour-market information	Employment (p-file) LIS variable: *emp*	Indicator that employment is the status of the main current activity as self-assessed by the respondent. It is defined as a dummy variable (0=not employed; 1=employed).

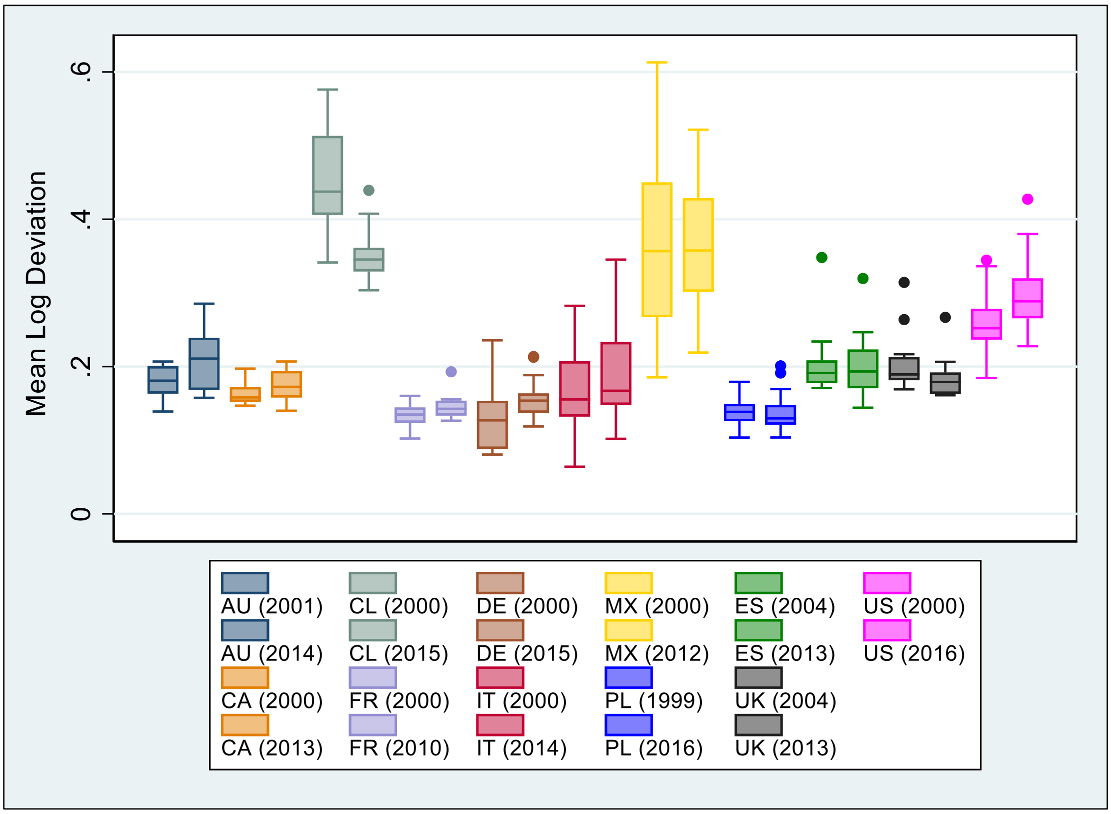
Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (multiple countries; 1999-2016). Luxembourg: LIS.

Table A.5. ANOGI decomposition and regional stratification in OECD countries: 1999-2016

Countries	Units	Year	Overall Gini	G_{wo} $\sum q_i G_i O_i$	G_b	IG $\sum q_i G_i$	IGO $\sum q_i G_i (O_i - 1)$	BGp G_{bp}	BGO G_b-G_{bp}	Stratification Index % G_b/Gbp
			1=2+3	2=4+5	3=6+7	4	5	6	7	100*(3/6)
Australia	7	2001	0.3199	0.3153	0.0046	0.3173	-0.0020	0.0342	-0.0296	3.16
(2001-2014)	7	2014	0.3315	0.3265	0.0050	0.3291	-0.0026	0.0404	-0.0354	3.81
Canada	10	2000	0.3029	0.2928	0.0101	0.2977	-0.0048	0.0542	-0.0441	4.89
(2000-2013)	10	2013	0.3212	0.3109	0.0103	0.3159	-0.0050	0.0568	-0.0465	5.16
Chile	13	2000	0.5270	0.4933	0.0337	0.5144	-0.0211	0.1347	-0.1010	12.21
(2000-2015)	13	2015	0.4674	0.4403	0.0271	0.4566	-0.0163	0.1164	-0.0892	10.55
France	8	2000	0.2781	0.2605	0.0176	0.2689	-0.0084	0.0624	-0.0449	5.32
(2000-2010)	8	2010	0.2870	0.2807	0.0063	0.2834	-0.0027	0.0458	-0.0395	4.22
Germany	16	2000	0.2591	0.2521	0.0070	0.2550	-0.0029	0.0394	-0.0324	3.53
(2000-2015)	16	2015	0.2966	0.2889	0.0077	0.2926	-0.0037	0.0473	-0.0396	4.32
Italy	20	2000	0.3347	0.2746	0.0601	0.3049	-0.0304	0.1340	-0.0739	10.43
(2000-2014)	20	2014	0.3322	0.2866	0.0456	0.3073	-0.0207	0.1133	-0.0677	8.84
Mexico	32	2000	0.4998	0.4106	0.0892	0.4587	-0.0481	0.2102	-0.1210	16.91
(2000-2012)	32	2012	0.4720	0.4111	0.0609	0.4463	-0.0352	0.1660	-0.1052	14.03
Poland	16	1999	0.2917	0.2857	0.0059	0.2887	-0.0029	0.0450	-0.0390	4.20
(1999-2016)	16	2016	0.2899	0.2835	0.0064	0.2863	-0.0028	0.0461	-0.0397	4.25
Spain	18	2004	0.3213	0.3015	0.0199	0.3112	-0.0097	0.0799	-0.0601	6.98
(2004-2013)	18	2013	0.3454	0.3138	0.0316	0.3287	-0.0149	0.1043	-0.0728	8.76
United Kingdom	11	2004	0.3543	0.3427	0.0116	0.3486	-0.0060	0.0752	-0.0636	6.95
(2004-2013)	11	2013	0.3336	0.3233	0.0103	0.3285	-0.0052	0.0645	-0.0542	5.93
United States	51	2000	0.3721	0.3648	0.0073	0.3684	-0.0036	0.0517	-0.0444	4.80
(2000-2016)	51	2016	0.3860	0.3796	0.0064	0.3824	-0.0028	0.0498	-0.0435	4.62

Note: The intragroup (2) and between group (3) components of overall inequality (1) are decomposed to account for the effects of overlapping on within (5) and between (7) components. *IG* is an average of regions' Ginis, weighted by income shares (q_i), G_{bp} is the between-groups component (based on Pyatt 1967), O_i is the overlapping index of the group *i* with the entire population. The last column presents the stratification index, which increases as interregional inequality rises and also when intraregional inequality falls. Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (multiple countries; 1999-2016). Luxembourg: LIS.

Figure A.1. *Box-and-whisker plot* for the mean log deviation (L) in OECD countries: 1999-2016



Notes: (1) AU=Australia, CA=Canada, CL=Chile, FR=France, DE=Germany, IT=Italy, MX=Mexico, PL=Poland, ES=Spain, UK=United Kingdom, US=United States; (2) For each country, the *box-and-whisker plot* on the left corresponds to the initial year, while the one on the right is relative to the final year. Source: Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (multiple countries; 1999-2016). Luxembourg: LIS.

References

Allanson P (2014) Income stratification and between-group inequality. *Economics letters* 124: 227–230

Athanasopoulos G, Vahid F (2003) Statistical Inference and Changes in Income Inequality in Australia. *The Economic Record* 79: 412–424

Atkinson AB (2015) *Inequality*. Harvard University Press, Massachusetts

Ayala L, Jurado A, Pérez-Mayo J (2014) Drawing the poverty line: Do regional thresholds make a difference? *Applied Economic Perspectives* & *Policy* 36: 309–322

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

Brandolini A (2007) Measurement of Income Distribution in Supranational Entities: The Case of the European Union. In: Jenkins SP, Micklewright, J (eds) *Inequality and Poverty Reexamined.* Oxford University Press, Oxford

Brandsma A, Kancs DA, Persyn D (2013) Modelling migration and regional labour markets: an application of the new economic geography model rhomolo. European Commission Joint Research Centre, Technical Report

Carrillo PE, Rothbaum JL (2016) Counterfactual spatial distributions. *Journal of Regional Science* 56: 868–894

Castells-Quintana D, Ramos R, Royuela V (2015) Income inequality in European Regions. Recent trends and determinants. *Review of Regional Research* 35: 123–146

Combes PP, Lafourcade M, Thisse JF, Toutain JC (2011) The rise and fall of spatial inequalities in France: A long–run perspective. *Explorations in Economic History* 48: 243–271

Commendatore P, Kubin I, Mossay P (2018) On the new economic geography of a multicone world. *Review of International Economics* 26: 539–554

Contreras D, Ffrench-Davis R (2014) Policy regimes, inequality, poverty and growth: The Chilean experience, 1973–2010. In Cornia G (ed.) *Falling inequality in Latin America: policy changes and lessons*. Oxford University Press, Oxford

Cowell FA (2011) *Measuring Inequality*. Oxford University Press, Oxford

Cowell FA, Fiorio C (2011) Inequality Decompositions. A Reconciliation. *Journal of Economic Inequality* 9: 509–528

Crozet M (2004) Do migrants follow market potentials? An estimation of a new economic geography model. *Journal of Economic Geography* 4: 439–458

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

Dickey H (2014) The Impact of Migration on Regional Wage Inequality: a Semiparametric Approach. *Journal of Regional Science* 54: 893–915

DiNardo J, Fortin N, Lemieux T (1996) Labour Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach. *Econometrica* 64: 1001–1044

Dormard S (2004) Economic development and regional disparities in France". In Helmut, K. (ed.) *Employment and regional development policy: Market efficiency versus policy intervention*. Verlag der ARL, Hannover

Elbers C, Lanjouw P, Mistiaen JA, Özler B (2008) Reinterpreting between-group inequality. *The Journal of Economic Inequality* 6: 231–245

Esquivel G, Lustig N, Scott J (2010) Mexico. A decade of falling inequality: Market forces or state action? In: Lopez-Calva LF, Lustig, N (eds) *Declining inequality in Latin America: A decade of progress*. Brookings Institution Press and UNDP, Washington

Ezcurra R, Rodríguez-Pose A (2009) Measuring the regional divide. In: Capello R, Nijkamp P (eds) *Handbook of regional growth and development theories.* Edward Elgar, Cheltenham

Fleming DA, Measham TG (2015) Income inequality across Australian regions during the mining boom: 2001–11. *Australian Geographer* 46: 203–216

Frick JR, Goebel J (2008) Regional income stratification in unified Germany using a Gini decomposition approach. *Regional Studies* 42: 555–577

Frick JR., Goebel J, Schechtman E, Wagner GG, Yitzhaki S (2006) Using analysis of Gini (ANOGI) for detecting whether two subsamples represent the same universe: The German Socio–Economic Panel Study (SOEP) experience. *Sociological Methods & Research* 34: 427–468

Fujita M, Krugman P (2005) The new economic geography: Past, present and the future. *Papers in Regional Science* 83, 139–164

Garretsen H, Martin R (2010) Rethinking (New) Economic Geography Models: Taking Geography and History More Seriously. *Spatial Economic Analysis* 5: 127–160

Gray D, Mills JA, Zandvakili S (2004) An analysis of differential provincial income inequality trends in Canada. In: Bishop JA, Amiel Y (eds) *Studies on Economic Well–Being: Essays in the Honor of John P. Formby (Research on Economic Inequality, Vol. 12).* Emerald Group Publishing Limited, Bingley

Hassing R, Gong H (2019) New Economic Geography. In: Orum AM (ed.) *The Wiley Blackwell Encyclopedia of Urban and Regional Studies*. Wiley Online Library

Hoffmeister O (2009) The spatial structure of income inequality in the enlarged EU. *Review of Income and Wealth* 55: 101–127

Insee (1997) *Revenus et patrimoine des ménages,* Édition 1997. Insee. Paris

Jesuit (2003) The regional Dynamics of European Electoral Politics. *European Journal Politics* 4: 139–164

Kanbur R, Venables AJ (eds) (2005) *Spatial inequality and development*. Oxford University Press

Lee N, Rodríguez-Pose A. (2013) Innovation and spatial inequality in Europe and USA. *Journal of economic geography* 13, 1–22

Lessmann C (2014) Spatial inequality and development - Is there an inverted-U relationship? *Journal of Development Economics* 106: 35–51

Martino G, Perugini C (2008) Income inequality within European regions: determinants and effects on growth. *Review of Income and Wealth* 54: 337–406

Milanovic B, Yitzhaki S (2002) Decomposing World Income Distribution: Does the World have a Middle Class? *Review of Income and Wealth* 48: 155–78

Miranti R, Cassells R, Vidyattama Y, McNamara J (2013) Measuring Inequality for Small Areas Using Spatial Microsimulation: An Australian Case Study. In *Proceedings of the The 42nd Australian Conference of Economists, Murdoch University,* Perth

Monti M, Santoro A (2011) Stratification and between–group inequality: A new interpretation. *Review of Income and Wealth* 57: 412–427

Mookherjee D, Shorrocks A (1982) A Decomposition Analysis of the Trend in UK Income Inequality. *The Economic Journal* 92: 886–902

Mussini M (2017) Decomposing changes in inequality and welfare between the EU regions. *Social Indicators Research* 130:455–478

Novotný J (2004) Decomposition of global and European socio-economic inequalities with attention to their regional dimensions. *Regional and Sectoral Economic Studies* 4: 59–78

Novotný J (2007) On the measurement of regional inequality: does spatial dimension of income inequality matter? *The Annals of Regional Science* 41: 563–580

OECD (2018) *Job creation and local economic development 2018: Preparing for the future of work*. OECD Publishing, Paris

Ostry MJD, Berg MA, Tsangarides MCG (2014) Redistribution, Inequality, and Growth. IMF Staff Discussion Note SDN/14/02

Parro F, Reyes L (2017) The rise and fall of income inequality in Chile. *Latin American Economic Review* 26: 1–31

Puga D (2002) European regional policies in light of recent location theories. *Journal of economic geography* 2: 373–406

Piaccentini M (2014) Measuring Income Inequality and Poverty at the Regional Level in OECD Countries. OECD Statistics Working Papers, 2014/03

Pyatt G (1976) On the Interpretation and Disaggregation of Gini Coefficient. *Economic Journal* 86: 243–55

Rey SJ, Janikas MV (2005) Regional convergence, inequality and space. *Journal of Economic Geography* 5: 155–176

Rey SJ, Smith RJ (2013) A spatial decomposition of the Gini coefficient. *Letters in Spatial and Resource Sciences* 6: 55–70

Rodríguez-Pose A, Ezcurra R (2009) Does decentralization matter for regional disparities? A cross-country analysis. *Journal of Economic Geography* 10: 619–644

Shorrocks AF (2013) Decomposition procedures for distributional analysis: a unified framework based on the Shapley value. *Journal of Economic Inequality* 11: 1–28

Shorrocks AF, Wan G (2005) Spatial Decomposition of Inequality. *Journal of Economic Geography* 5: 59–81

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

Torrisi G, Pike A, Tomaney J, Tselios V (2015) (Re-) exploring the link between decentralization and regional disparities in Italy. *Regional Studies, Regional Science* 2: 123–140

Usseglio AN (2016) The North–South Divide: Regional Economic Inequality in Contemporary Italy. UVM Honors College Senior Theses, 122

Venables AJ (2016) New Economic Geography. In *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, London

Yitzhaki S (1994) Economic Distance and Overlapping of Distributions. *Journal of Econometrics* 61: 147–59

Yitzhaki S, Schechtman E (2013) More than a dozen alternative ways of spelling Gini. In *The Gini Methodology* (pp. 11–31). New York: Springer