OECD Regional Development Working Papers 2016/06

# Income Levels And Inequality in Metropolitan Areas: A Comparative Approach in OECD Countries

Justine Boulant

Monica Brezzi

Paolo Veneri

https://dx.doi.org/10.1787/5jlwj02zz4mr-en

WORKING PAPERS

OECD

**OECD REGIONAL DEVELOPMENT WORKING PAPERS**

This series is designed to make available to a wider readership selected studies on regional development issues prepared for use within the OECD. Authorship is usually collective, but principal authors are named. The papers are generally available only in their original language English or French with a summary in the other if available.

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the author(s).

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcomed, and may be sent to either gov.contact@oecd.org or the Public Governance and Territorial Development Directorate, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

Authorised for publication by Rolf Alter, Director, Public Governance and Territorial Development Directorate, OECD.



-----------------------------------------------------------------------------

OECD Regional Development Working Papers are published on
http://www.oecd.org/gov/regional/workingpapers

-----------------------------------------------------------------------------

**INCOME LEVELS AND INEQUALITY IN METROPOLITAN AREAS: A COMPARATIVE APPROACH IN OECD COUNTRIES**[1]

Justine Boulant, Monica Brezzi and Paolo Veneri

Regional Development Policy Division, GOV

OECD

## Abstract

This paper assesses levels and distribution of household disposable income in OECD metropolitan areas. All indicators were produced through a dedicated data collection, which, for most countries, uses administrative data from tax records available at detailed local scale (i.e. municipalities, local authorities, counties, etc.). Using different estimation techniques, we provide internationally comparable figures for 216 OECD metropolitan areas. The results highlight stark differences in both income levels and inequality within metropolitan areas, even for those belonging to the same country. Larger metropolitan areas feature, on average, higher levels of household disposable income but also higher income inequality. The paper then provides a measure of spatial segregation, or the extent to which households with similar incomes concentrate within a metropolitan area. On the governance side, the paper finds a stable and positive relationship between administratively fragmented metropolitan areas and spatial segregation by income.

1 This paper was prepared in the context of the OECD project "Making Inclusive Growth Happen in Cities and Regions" for which the financial support by Ford Foundation is kindly acknowledged. The authors wish to thank Luiz De Mello and Marco Mira D'Ercole (OECD) for initial discussions on the ideas and methods that led to this paper. Valuable comments were also received by Orsetta Causa, Marco Diaz Ramirez, Karen Maguire, Marco Mira D'Ercole, Fabrice Murtin and Joaquim Oliveira Martins (OECD) and participants to the OECD seminar on 7 April, 2016. This work would have not been possible without the help and support by experts from Statistical Offices and other national agencies who provided us the information needed to produce the estimates of income levels and inequality for metropolitan areas. In particular, authors would like to thank Marcos Diaz Ramirez (OECD), Emiko Araki (OECD), Johannes Biricz and Regina Fuchs (Statistics Austria), Lisa Conolly and Lisa Moutzouris (Australian Bureau of Statistics), Harry Estay Jara (INE, Chile), Olaf Foss (Norwegian Institute for Urban and Regional Research), Oscar Gasca Brito (INEGI, Mexico), Nigel Henretty (UK Office for National Statistics), Natalie Holmes (Brookings Institution), Anneli Korpelainen and Pekka Ruotsalainen (Statistics Finland), Johanna Lundström, Kresten Nordestgaard Laursen and Jarl Quitzau (Statistics Denmark), Anett Madarász (Hungarian Ministry for National Economy), Brian Murphy (Statistics Canada), Robert Reynard (Insee), Eedi Sepp (Estonian Ministry of the Interior), Erik Vloeberghs (Statistics Belgium), Tommy Wallster (Statistics Sweden), Ruby de Wilde (Statistics Netherlands).

# TABLE OF CONTENTS



**Tables**



**Figures**

**Boxes**

## 1. Introduction

Cities play an important role in providing opportunities for individuals to prosper and in contributing to national economic performance. In OECD countries, metropolitan areas – cities larger than 500 000 people – host 50% of the population and have contributed to 60% of the total employment creation in the past fifteen years (OECD, 2016). The high concentration of highly productive firms, dense and thick labour markets and learning opportunities allow workers located in metropolitan areas to benefit from higher wages and job quality than those living in other places. At the same time, metropolitan areas tend to be more unequal places than the country as a whole in several domains of life. Differences in living conditions and levels of inequality often have a strong spatial component within cities, where people with similar incomes tend to live close to each other and where poverty is often concentrated in certain neighbourhoods or localities. In a highly unequal city, it may be more difficult to maintain mixed-income school environments that produce better outcomes for low-income children or decent housing accessible across all social groups. Moreover, the spatial concentration of households by income or by other socio-economic or cultural characteristics, can affect people's opportunities. Upward economic mobility, for example, was found to be lower in more unequal, spatially segregated cities with low-quality schools in the United States (Chetty et al., 2014). The opportunities related to economic mobility depend also in the way cities are organised and coordination of sector policies – for example housing, transport and land use – can help make the city more inclusive.

A systematic measurement of income levels, inequality and spatial segregation is a necessary step to understand the factors that allow – or prevent – cities to provide the conditions for people to prosper in the present and in the future. This paper provides, for the first time, comparable statistics on income levels and inequality for metropolitan areas in a set of OECD countries. It addresses the following three questions. First, whether income is significantly higher in the metropolitan areas than in other parts of a country and whether there are important differences in the income levels of households living in different metropolitan areas within the same country. Secondly, whether levels of income inequality differ systematically across cities, and how these differences relate to their average income and population size. Thirdly, the assessment of spatial segregation of households by income levels within metropolitan areas and the exploration of the main factors associated to such phenomenon. The spatial segregation is investigated here by looking at local jurisdictions within a metropolitan area rather than at neighbourhoods. This implies a focus on issues of metropolitan governance and on the different capacity of local jurisdictions to provide public goods that might shape the localisation decisions of households and thus hinder or spur spatial segregation.

Despite the importance of the issue, comparative assessments of income levels and inequality in metropolitan areas across different countries are rare, notably because of the lack of robust and comparable data. Country-level measures of the distribution of household disposable income typically come from household surveys, which are generally not representative at the regional or metropolitan level. Other sources of income data, such as tax records, available at a fine geographic detail are based on different definitions of incomes and are hardly comparable across countries. The paper overcomes these data limitation, by relying on a common method to estimate household disposable income levels and income inequalities in OECD metropolitan areas based on tax records. Income levels were made comparable across countries by benchmarking income values for metropolitan areas coming from the different sources to the regional estimates of household disposable income from the OECD Regional Well-being Database (OECD, 2014). Such a method is applied to 18 OECD countries (11 countries for income inequality), covering 216 of the 281 OECD metropolitan areas. This method could be expanded to other countries in the future.

The main findings of the paper are the following:

- OECD metropolitan areas vary greatly by their average income levels and income inequalities, also within the same country. On average, people living in metropolitan areas have higher income than their non-metropolitan counterparts, though differences in price levels might offset part/most of this difference.

- In almost all OECD countries considered, income inequality is higher in metropolitan areas than in the rest of the country. Average income levels and income inequality tend to be higher in metropolitan area with larger population.

- The concentration of households by income in the various local jurisdictions within a metropolitan area– also called "spatial segregation" in this paper – tends to be higher in more administratively fragmented metropolitan areas. This suggests that the way public services are provided by local authorities can play a role in shaping the localisation of households within the metropolitan space, with disadvantages for those households who end up in the locations with the lowest accessibility or provision of public services.

The paper is organised as follows: Section 2 describes the measures of household income for metropolitan areas that have been constructed, the methods adopted, the main sources of data and the units of analysis. Section 3 presents the results on the levels of average household disposable income for metropolitan areas, while Section 4 shows the results in terms of income inequality. Section 5 presents a method to assess income segregation within metropolitan areas, by decomposing the index of income inequality (Theil or Gini), while Section 6 provides evidence on the relationship between administrative fragmentation and spatial segregation through a regression analysis on metropolitan areas. Section 7 discusses the results and proposes possible extension of the paper. The Appendixes include countries' data sources and details of the estimation techniques employed.

## 2. Data sources and method to estimate household disposable income in metropolitan areas

### *Data sources, unit of analysis and territorial level of aggregation*

Consistently with the approach used in the OECD Income Distribution Database (OECD, 2014), the paper focuses on household disposable income, which represents the amount of income available to households after payment of taxes, social security contributions paid by workers and private current transfers. Disposable income provides a close approximation of the economic resources available to satisfy household consumption (United Nation, 2011). This concept is generally preferred to market (or taxable) income in international comparisons since it accounts for the effects of taxes, which differ across countries and localities. It has to be acknowledged that the adopted definition of disposable income includes cash transfers to households but not in-kind transfers, such as education and health public services. These transfers are often provided at the local level, thus they can have significant distributional effects. Their omissions from the analysis here is unavoidable for reasons of data availability.

National statistics of household income are usually produced using household surveys and administrative records, which provide consistent definitions and maximise international comparisons. Household surveys, however, are usually designed to be representative only at national or, at most, regional scales; hence, they are hardly usable when it comes to assessing income levels and distribution at smaller geographical scales, such as municipalities or metropolitan areas. For this paper, estimates of average household disposable income and its distribution for metropolitan areas are drawn from tax registers in most of the countries considered.

Statistical offices in OECD countries are increasingly using administrative data for measurement purposes in different domains. Recent research analysing income inequality, spatial segregation and income dynamics in cities also uses administrative data (Chetty et al., 2014; Tammaru et al., 2016). Tax records have been used at national level for several purposes, for example to build income statistics over long periods (Piketty, 2014); they are used in this paper as the source of metropolitan income data for most countries. In the case of Chile, Mexico, the United Kingdom and the United States, however, income levels for different local units are based on data collected by national statistical offices using survey data. In the case of Mexico and the United Kingdom, the national survey data on household income are projected to smaller spatial scales through the use of other variables that are highly correlated to household income and available at a more detailed scale, for example from the Population Census (known as small area estimation techniques).[2] For the United States, Census Bureau data are based on households participating in the American Community Survey (ACS) in years 2010 through 2014 (the oldest and most recent data available for the desired measures, respectively), whose sample is representative of every county: these data are pooled from 5 years to allow for more accurate measurement in less populous counties.[3]

Table 1 provides the source and type of income data available in each country, the number of local units considered (municipalities, counties, etc.) and the available years. These data have been used to compute income levels in the OECD metropolitan areas and, for a subset of countries, income inequality (Gini coefficient, top-bottom quintile ratio) and spatial segregation (decomposition of Theil indexes and Gini coefficients). Definitions of income from each source have been carefully considered in order to improve the harmonisation of data and increase international comparability (Appendix 1 for details).

All income indicators are presented at the level of metropolitan areas, according to a harmonised definition of functional urban areas jointly developed by the OECD and the EU. According to this definition, functional urban areas in OECD countries are identified as densely populated local units (urban centres) and surrounding local units connected to the urban centres by high travel-to-work flows (OECD, 2012). Metropolitan areas are the functional urban areas with at least 500 000 inhabitants. Table 1 includes the number of metropolitan areas in each country. Indicators of average/total household income are computed by aggregating data from the smaller spatial scale (local units) up to the metropolitan scale. In most of the countries considered, local units correspond to municipalities, with a few exceptions. In the United States, for example, local units correspond to counties, and in the United Kingdom, they correspond to Middle-Layer Super Output Area (MSOA).

2 In the case of Mexico, data are provided by the National Council for the Evaluation of Social Development Policy (CONEVAL), and follow a collaboration agreement with the National Institute of Statistics and Geography (INEGI) that made it possible to incorporate questions into the 2010 Population Census in order to obtain information on household income.

3 For a brief explanation of both small area estimation techniques and on administrative data available in OECD countries, see OECD (2015a).

**Table 1. Income data at the local unit scale provided by countries**

| Country | Source | Type of data | Nr. of local units | Nr. of metro areas | Information on income distribution | Years |
|---|---|---|---|---|---|---|
| Australia | Australia Bureau of Statistics | Tax records | 1125 | 6 | No | 2006-11 |
| Austria | Statistics Austria | Tax records | 649 | 3 | Yes | 2004, 2007, 2012 |
| Belgium | Statistics Belgium | Tax records | 200 | 4 | Yes | 2005, 2007, 2013 |
| Canada | Statistics Canada | Tax records | - | 34 | Yes | 2006, 2013 |
| Chile | CASEN – Ministry of Social Development | Household income survey | 62 | 3 | Yes | 2009, 2013 |
| Denmark | Statistics Denmark | Register | 49 | 1 | Yes | 2000-14 |
| Estonia | Estonian Tax and Customs Board | Tax records | 28 | 1 | No | 2003-14 |
| Finland | Statistics Finland | Register | 22 | 1 | No | 2000-14 |
| France | INSEE | Tax records | 1409 | 15 | Yes | 2001-11 |
| Hungary | National Tax and Customs Administration of Hungary | Tax records | 183 | 1 | No | 2000-13 |
| Italy | Ministry of Economy and Finance | Tax records | 775 | 11 | Yes | 2008-13 |
| Japan | Ministry of Internal Affairs and Communication | Tax records | 570 | 36 | No | 1990, 1995, 2000, 2005, 2013 |
| Mexico | CONEVAL | Small areas estimation | 296 | 33 | Yes | 2010 |
| Netherlands | CBS (Regional Income Research) | Register & Tax records | 130 | 5 | No | 2006-13 |
| Norway | Statistics Norway | Tax records | 30 | 1 | Yes | 2006-13 |
| Sweden | Statistics Sweden | Register | 54 | 3 | Yes | 2000-13 |
| United Kingdom | ONS | Small area estimations | 2974 | 13 | No | 2008, 2012 |
| United States | ACS web platform | Household survey | 380 | 70 | Yes | 2010-14 |

Note: See Appendix 1 for details on the definition of income from each source.

*Method to estimate total household disposable income at metropolitan level*

The available information on household income at local unit scale generally refers to taxable income, gross income or other definitions that are not comparable across countries.[4] In absence of detailed information on public transfers and taxes, which would be needed to move from taxable to disposable income, we opted for a simple method that makes use of the estimates of mean equivalised household disposable income available for larger subnational regions, via the OECD Regional Database.[5] The latter database provides, for OECD territorial level 2 (TL2) regions, the overall mean, the mean by quintile, and the Gini coefficient for equivalised household disposable income, as well as the total number of households in each region. These indicators are sourced from the same national household surveys and administrative records on household income that are used by the OECD for its national-level reporting.[6] The method used in this paper to derive estimates of total household disposable income for metropolitan areas consists in two simple steps.

- First, regional household income from tax records is computed by aggregating values of all local units up to the TL2 region. The resulting total income measures are compared with the regional values of household equivalised disposable income provided by the OECD Regional Database.

- Second, a coefficient — computed as the ratio between the two income values mentioned above at regional level — is applied to re-scale the income value of each local unit to a disposable income definition. The sum of these rescaled values, across all local units within the same region, is hence equal to the total mean equivalised household disposable income for the region considered as available in the OECD Regional database.

The method can be illustrated by the example of household disposable income in the metropolitan areas of California, US.

- First, values of total household incomes across California's local units are summed to obtain the total household income of the state of California.

- Second, the level of California's equivalised disposable income, as given by the OECD Regional Database is divided by the value described above: this gives a coefficient of 0.9 for 2010 (Table 2).

- Third, this coefficient is then applied to the income values of all local units of California, in order to approximate the levels of equivalised household disposable income at the local unit level.

This method is applied to the total income value for local units available for 18 OECD countries, thus allowing 'proxies' of the mean equivalised disposable income of each units (and, by aggregation, of metropolitan areas) to be obtained. Two potential limitations of the above described method should be acknowledged. The first is that moving from pre-tax to disposable income requires making the hypothesis that tax and transfer do not vary within the same region, thus differences in the taxation and transfers provided across the municipalities of the same region are not taken into account. Second, the use of tax

4 Disposable income is defined as the sum of income deriving from employment (both paid and self-employment), property, production of household services for own consumption, and current transfers received (i.e. pensions, social benefits, etc.) minus current transfers paid (taxes, fees, social contributions, etc.).

5 http://stats.oecd.org/Index.aspx?DataSetCode=RWB.

6 See Piacentini (2014) for details.

records implies that households with no income are excluded in the analysis, even if they might receive positive transfers. In addition, the use of tax records makes it difficult to account for tax avoidance by the rich, which could be another potential source of bias.

**Table 2. Examples of the coefficients used for estimating household disposable income in metropolitan areas in California, US**

| Levels | Coefficient for California (U.S.) |
|---|---|
| Overall mean | 0.9 |
| 1st quintile levels | 1.3 |
| 2nd quintile levels | 1.0 |
| 3rd quintile levels | 0.9 |
| 4th quintile levels | 0.9 |
| 5th quintile levels | 0.8 |

Source: Authors' elaborations.

***Method to estimate the distribution of household disposable income within metropolitan level***

Estimates of the Gini coefficient – or other indicators of inequality – for household disposable income are not available directly for most metropolitan areas in OECD countries. For a subset of countries, some information is available on the distribution of income within local units. While the form of this information differ somewhat across countries (e.g. average or total income by quantiles of households), this allows producing 'proxies' of income distribution within local units and metropolitan areas. Proxy measures were computed by first 'simulating' the entire income distribution in each metropolitan area, based on the assumption that this income distribution has the lognormal functional form and by fitting this model through the upper-bound of each population quintiles (Box 1). This method has the advantage of avoiding the dependence of the Gini coefficient on the granularity of measurement (using income quintiles, for example, would result in a lower Gini coefficient than using income deciles taken from the same distribution) and on the number of local units included in a metropolitan area.

The method can again be illustrated by the example of metropolitan areas in California:

- First, measures of total taxable income across quintiles of households for each local unit are obtained from the national source used.

- Second, regional coefficients for each income quintile (computed as the ratio between the level of California's equivalised disposable income for each quintile from the OECD Regional Database and the corresponding values from the national sources used, see Table 2) are applied to the income values of each quintile in each local units, in order to approximate the equivalised household disposable income across quintiles of each local unit.

- Third, the full distribution of income in each local unit was generated by fitting the quintile data-points previously obtained. The number of observations generated corresponded to the total number of households in each local unit.

The distribution of household disposable income in each metropolitan area is modelled as a mixture of various log-normal distributions. Once a sample of households whose incomes fit with the income quintiles is generated, the Gini coefficient can be computed for each metropolitan area. The hypothesis of lognormal income distribution has long been used and tested in the literature (Steyn, 1959; Balintfy and Goodman, 1973; Lopez and Servén, 2006). For OECD countries, Díaz Ramírez and Murtin (2016) simulate the national disposable income distribution as a mixture of group-specific log-normal income distributions, where groups are defined in terms of educational attainment. Similarly to what done by Diaz Ramírez and Murtin, log-normal distributions are assumed here for municipalities (or local units at which income data is available). The hypothesis of log-normal distribution of income within municipalities has also been empirically tested for municipalities in Chile, the only country where micro-data on the entire distribution were available, with results suggesting a good fit of the log normal hypothesis (see Appendix 2).

**Box 1. How to generate income samples of lognormal distribution in local units?**

A lognormal distribution is determined by two parameters, the mean and the variance of the related normal distribution. These two values, $\mu$ and $\sigma$ have to be estimated in order to generate random samples from this distribution that suit the municipal datasets. One should note that inequality depends only on $\sigma$, which uniquely determines the shape of the Lorenz curves. For any income variant following a lognormal distribution, the Gini coefficient depends uniquely on this parameter:

$$G = 2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1$$

with $\Phi$ being the density function of the normal distribution.

A criterion of minimum sum of absolute errors has been used to fit the quantiles shares of total income to shares of a lognormal sample for these quantiles. This allows generating a sample that fits observed data – quintiles of income in local units – and a theoretical distribution (lognormal). Based on this approach, one obtains the 2 parameters $\mu_1$, $\sigma_1$ that minimize the function below:

$$criteria = \sum_{d=1}^{D} \left| y_d - y_{sort,d} \right|$$

where $y_d$ is the vector of the quantiles of the household disposable income in each local unit and $y_{sort,\ d}$ is the vector of the quantiles from the generated observations. The optimization algorithm estimates the parameters $\mu$ and $\sigma$ for a lognormal sample that minimize these sum of absolute errors. This method implies a somewhat arbitrary element: providing different intervals (deciles, quintiles…) would adapt the fit to precisely these intervals. Though, it still fulfils its purpose as the goal is not to generate random numbers that obey some distribution but to accurately calculate the Gini coefficients in metropolitan areas. Thus, the two lognormal parameters have been estimated for each municipality.

Subsequently, synthetic data obeying to a lognormal distribution with the estimated parameters were generated by using the software R; the size of each sample generated by this method equals the number of resident households of the municipality considered; while these income values approximately fit the real disposable income data. The quantiles shares of total income from the random samples have been compared to the primary values of quantiles shares and the overall fit appears very satisfactory, except for the first quantile share which is always overestimated due to the poor fit tails. On computing the metropolitan areas Gini coefficients, all municipal samples are gathered together in order to reconstitute the metropolitan population.

The generation of the income distribution with the method explained above was applied to 8 OECD countries, thus allowing 'proxies' of the distribution of household disposable income in each units (and, by aggregation, of metropolitan areas) to be obtained. In the case of other three countries, indicators of income distribution within metropolitan areas were provided by the respective National Statistical Offices (Canada and Mexico) or estimated directly from the existing micro-data (Chile) without the need to generate a simulated distribution. While this method allows comparison of income inequalities in 216 cities across OECD countries for the first time, it should be acknowledged that these estimates are based on simulations of the income distributions.

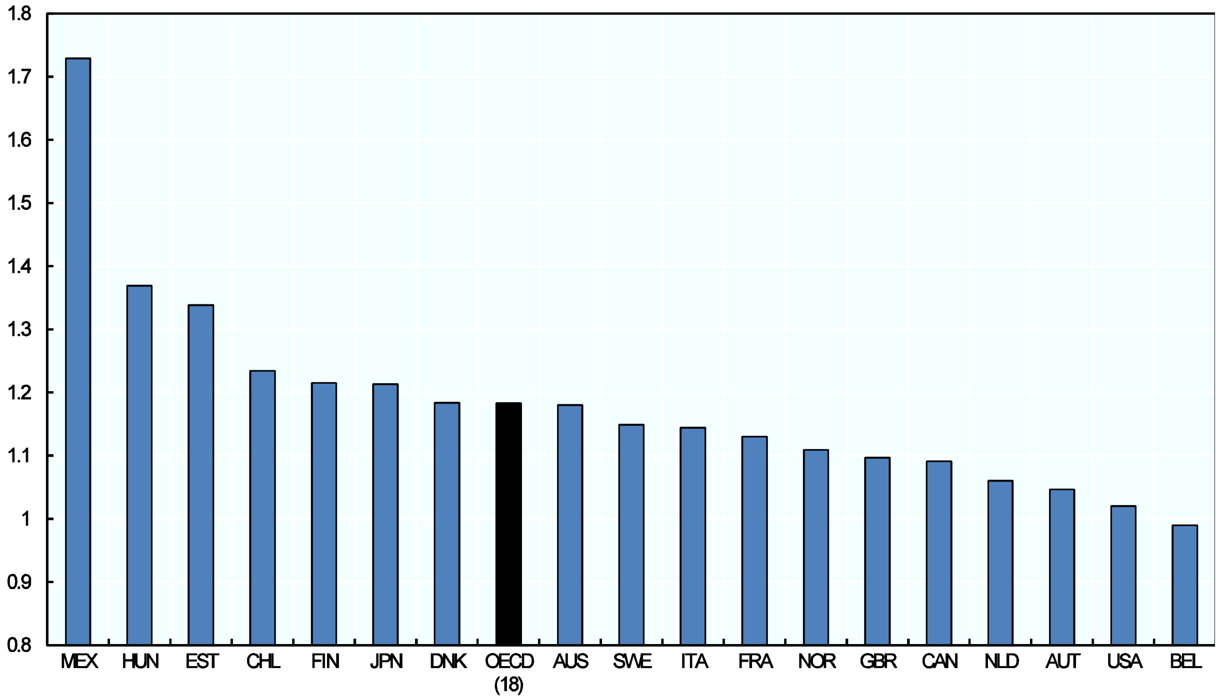
Various calibrations, or equivalence scales, have been devised to adjust the incomes of households in a way that reflects differences in the needs of individuals living in each household, and the economies that flow from sharing resources (Atkinson et al., 1995). For the sake of simplicity, the equivalence scale used in this paper consists in dividing household income by the square root of the household size, a method used by the OECD when comparing household income across countries (OECD, 2011). When household income is adjusted according to an equivalence scale, the resulting "equivalised" income can be viewed as an indicator of the economic resources available to people in each household.

## 3. Average income levels in metropolitan areas

Overall in 18 OECD countries, the equivalised disposable income of households living in metropolitan areas is, on average, 18% higher than that of households living in other parts of the country. The income *premium* in metropolitan areas with respect to the national average is always positive, with the exceptions of Belgium, but its size differs significantly across countries (Figure 1). Mexico is the country where the income premium is the highest (73%), followed by Hungary (37%), Estonia (34%) and Chile (23%). It should be noted however that higher levels of average income in metropolitan areas do not necessarily imply a higher purchasing power available to metropolitan residents, since differences in living costs between locations can partially offset earning differences across urban and rural places (World Bank, 2015). Due to the lack of regional or metropolitan purchasing power parities, differences in living cost among metropolitan areas are not taken into account in this paper.

**Figure 1. Income ratio between metropolitan and non-metropolitan areas by country**

Average income per equivalent household member; 2014 or latest available year



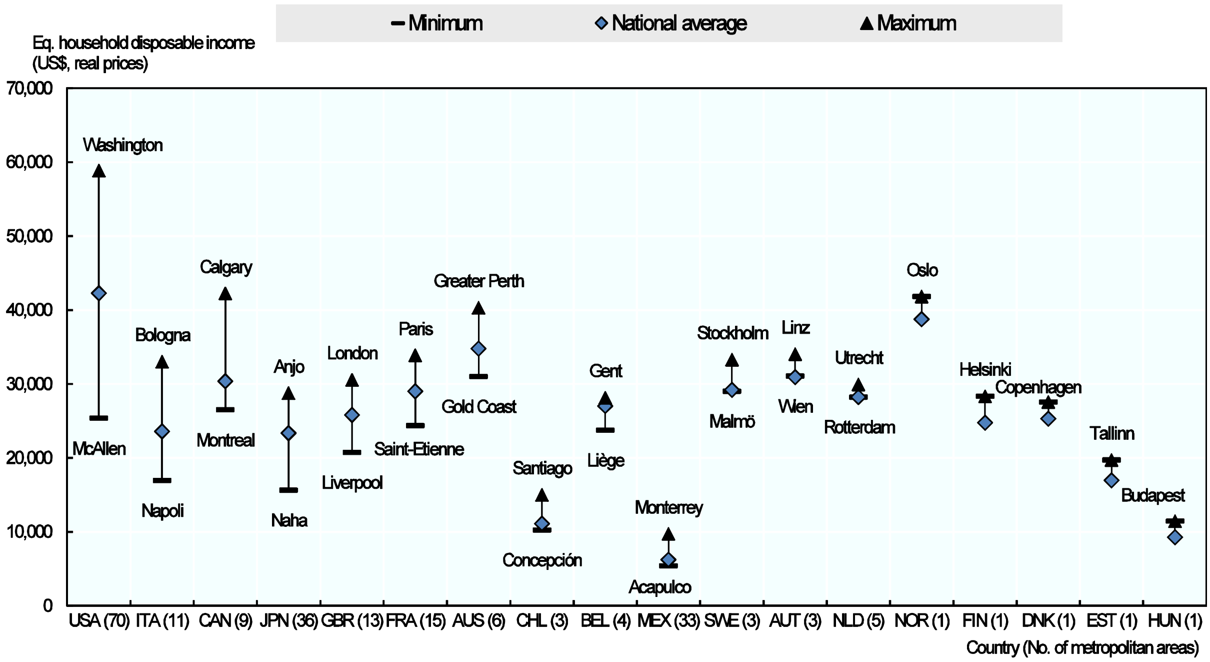

Note: The graph plots the ratio between household disposable income per equivalent household in metropolitan areas over that in the rest of the national territory. Countries are ordered by decreasing value of that ratio.

*Source*: Authors' elaborations on national income data (Appendix 1).

While metropolitan households have on average higher income than their non-metropolitan counterparts, large differences can also be observed among metropolitan areas in the same country (Figure 2). The largest disparities across metropolitan areas in the disposable income of households are observed in the United States, where the income of those living in Washington, DC is 2.3 times higher on average than that of households living in McAllen, Texas. Large differences are also observed in Italy and Japan, where the average income in the richest metropolitan areas (i.e. Bologna and Anjo, respectively) is almost two times higher than that in the metropolitan areas with the lowest income (Naples and Naha, respectively). Similarly, in Mexico the average income of households in Monterrey is 80% higher than those living in Acapulco. In other countries, the pattern is different. In Austria for example, the difference is only 1% between Linz and Vienna.

**Figure 2. Average household disposable income in metropolitan areas**

Differences between maximum and minimum metropolitan values; 2014 or latest available year, per equivalent household




Note: Last year available was 2013 for Belgium, Canada, Chile, Italy, Japan, Sweden, Netherlands, Hungary and Norway; 2012 for Austria and United Kingdom; 2011 for France and Australia; 2010 for Mexico

*Source*: Authors' elaborations on national income data (Appendix 1).

For most countries, it was possible to assess average income levels at more than one point in time. However, the availability of income data over time differs across countries, and it is hard to make comparisons on the growth of income across all cities. Over the period between the mid-2000s and 2013, average income levels increased in most of metropolitan areas suggesting that, on average, households have higher incomes that those they had before the economic crisis of 2008.[7] The growth rate of income was particularly high in Norway and in Australia, where it exceeded 2% annually. In Hungary, Italy and the United States, household income slightly declined, although the period under consideration is shorter for the latter two countries (2008-13 and 2010-14, respectively).

Among the factors that are correlated with different levels of average household incomes across metropolitan areas, population size plays an important role. On average, people living in larger cities have higher levels of income, with the exception of Mexico (Figure 3).[8] The positive relationship between urban size and average income is well documented in the literature and has several possible explanations. First,

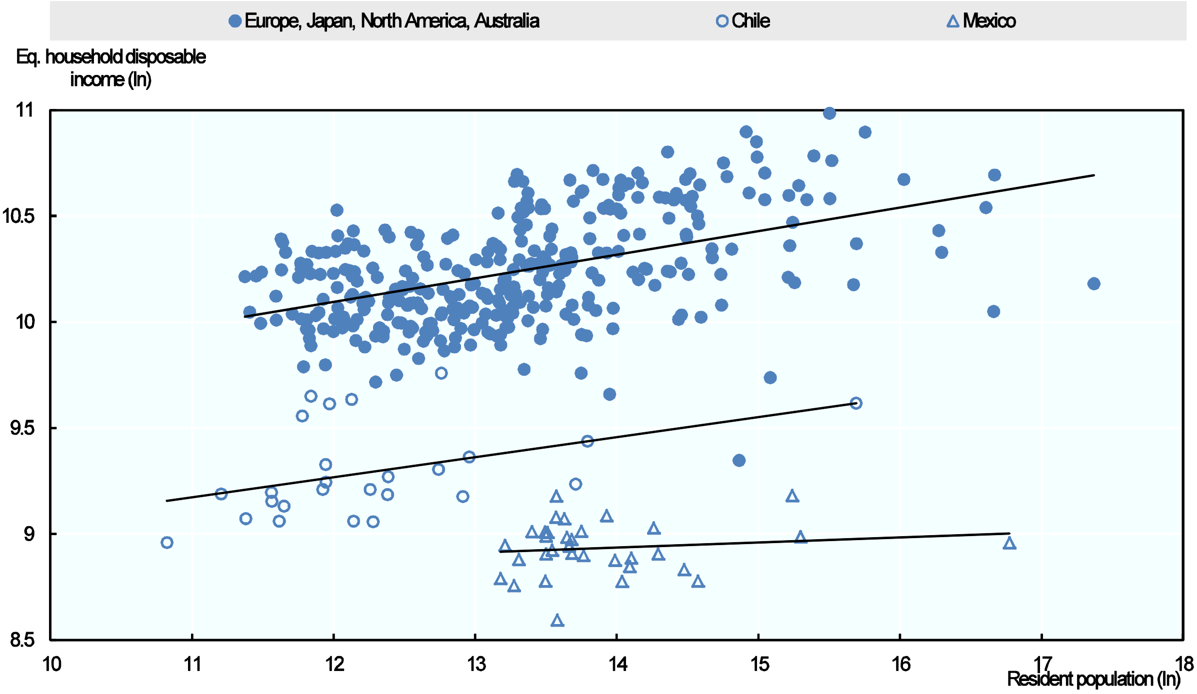
7 Data on income changes for metropolitan areas are not available for Mexico.

8 Chilean and Mexican cities are displayed separately from the other cities because their income levels are significantly lower than in the other countries.

more talented individuals tend to move to large cities, where returns to talent are higher and higher wages will be paid to talented workers (Behrens et al., 2014). Second, larger metropolitan areas can benefit from agglomeration economies that allow them to be more productive and pay higher wages; these agglomeration effects include productive advantages coming from the location of firms close to other firms, thanks to faster circulation of ideas (learning), thicker labour markets (matching) and sharing of indivisible facilities (sharing) (Duranton and Puga, 2004). Finally, in cities that are highly specialised in sectors with relatively high value added per worker, such as finance, information technology or advanced manufacturing, wages are on average higher than in other locations (Florida and Mellander, 2016).

**Figure 3. Population size and average household disposable income across metropolitan areas**

2014 or latest available year, per equivalent household




*Source*: Authors' elaborations on national income data (Appendix 1).

## 4. Income inequality in metropolitan areas

Measures of income inequality within metropolitan areas were estimated for eleven OECD countries: Austria, Belgium, Canada, Chile, Denmark, France, Italy, Mexico, Norway, Sweden and the United States. The main indicator of income inequality considered here is the Gini coefficient for equivalised household disposable income, which is the most used and well-known measure of inequality.[9] The Gini coefficient satisfies some important properties such as the comparability over time, the invariance to any deflator measures and the transfer principle[10].

In all countries considered income inequality in metropolitan areas is higher than the national average, with the exception of Canada. Among the 153 metropolitan areas in the 11 countries considered, the Gini

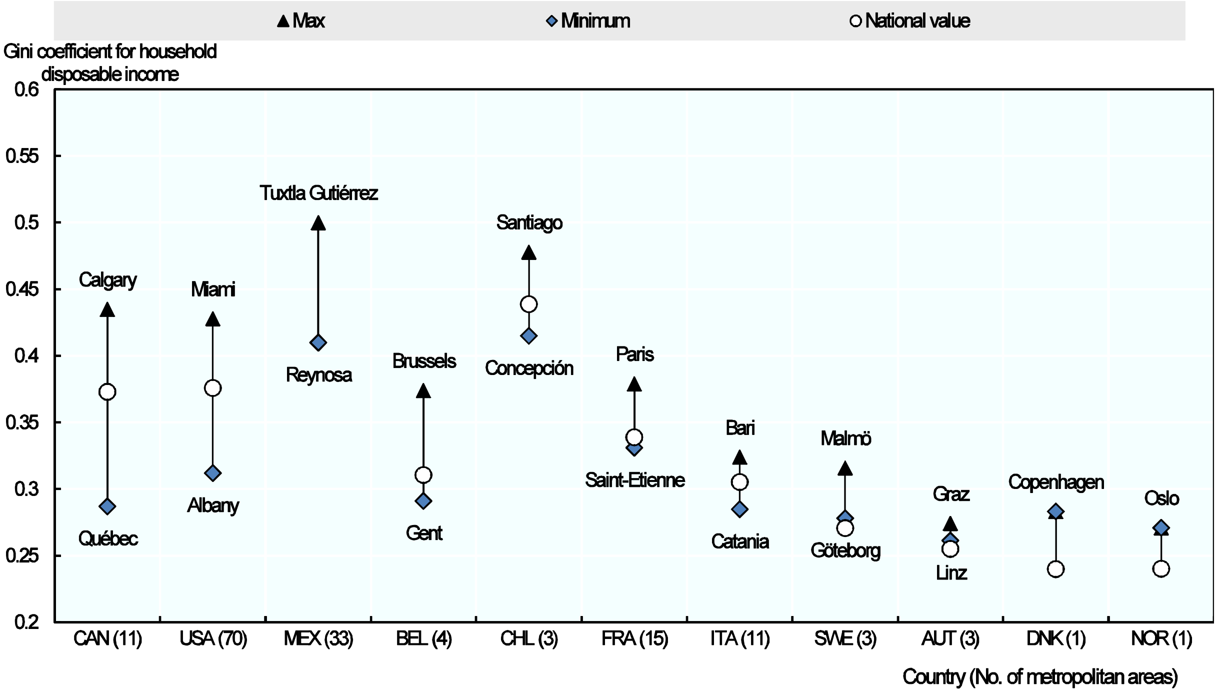
9 The Gini coefficient is based on the comparison of cumulative proportions of the population against cumulative proportions of income they receive. It ranges from 0 to 1, with 1 indicating maximum concentration of income (all income accrues to one individual only).

10 The Pigou-Dalton transfer principle states that if an amount of income is transferred from a rich individual to a poor one, while still preserving their ranking by income, then the measured inequality should decrease.

coefficients of disposable income vary between 0.26 in Linz (Austria) to 0.5 in Tuxtla Gutiérrez (Mexico). High and low levels of income inequality are observed in the metropolitan areas of Canada, United States, Mexico and Belgium: for example, while the Gini coefficient in Calgary (Canada) is 0.43, it is less than 0.3 in Québec (Figure 4).

**Figure 4. Gini coefficients for household income in metropolitan areas, circa 2014**

Metropolitan areas with minimum and maximum Gini coefficients, by country

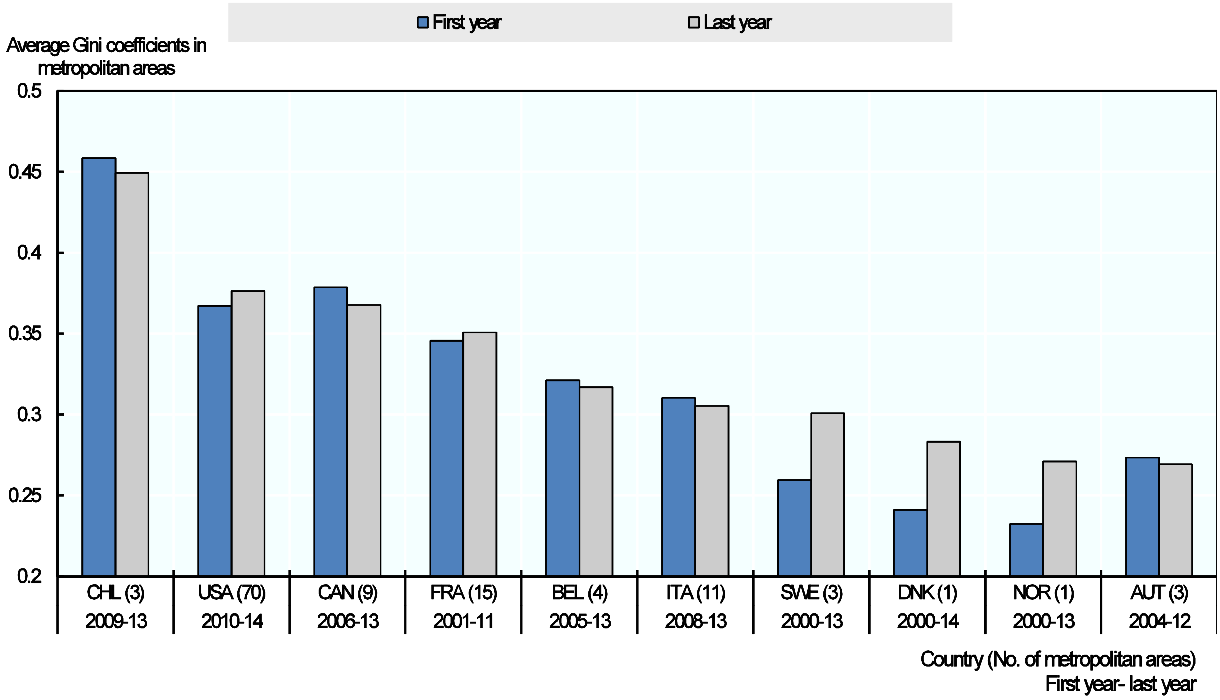



Note: The national values of Gini index are estimated using the same source of data employed for the metropolitan areas. They might be slightly different from values provided by national surveys. Data do not allow the national Gini index for Mexico to be provided.

*Source*: Authors' elaborations on national income data (Appendix 1).

Income inequality changes slowly over time and its evolution is generally monitored over longer time spans. With the available information, it is only possible to assess changes in income inequality in metropolitan areas over the course of the last decade, thus results should be interpreted with caution. Figure 5 shows the average values of the Gini coefficients for household incomes in metropolitan areas in the first and final years available. In several European countries – where levels of inequality are on average relatively low – metropolitan areas have experienced an increase of inequality: this is the case of Denmark, Norway and Sweden. A slight increase in inequality is also observed in metropolitan areas in France and the US during the last 4 years. On the other hand, metropolitan areas in Chile and Austria show a slight decrease in their income inequality.

**Figure 5. Average Gini coefficient of household income in metropolitan areas, by country and year**



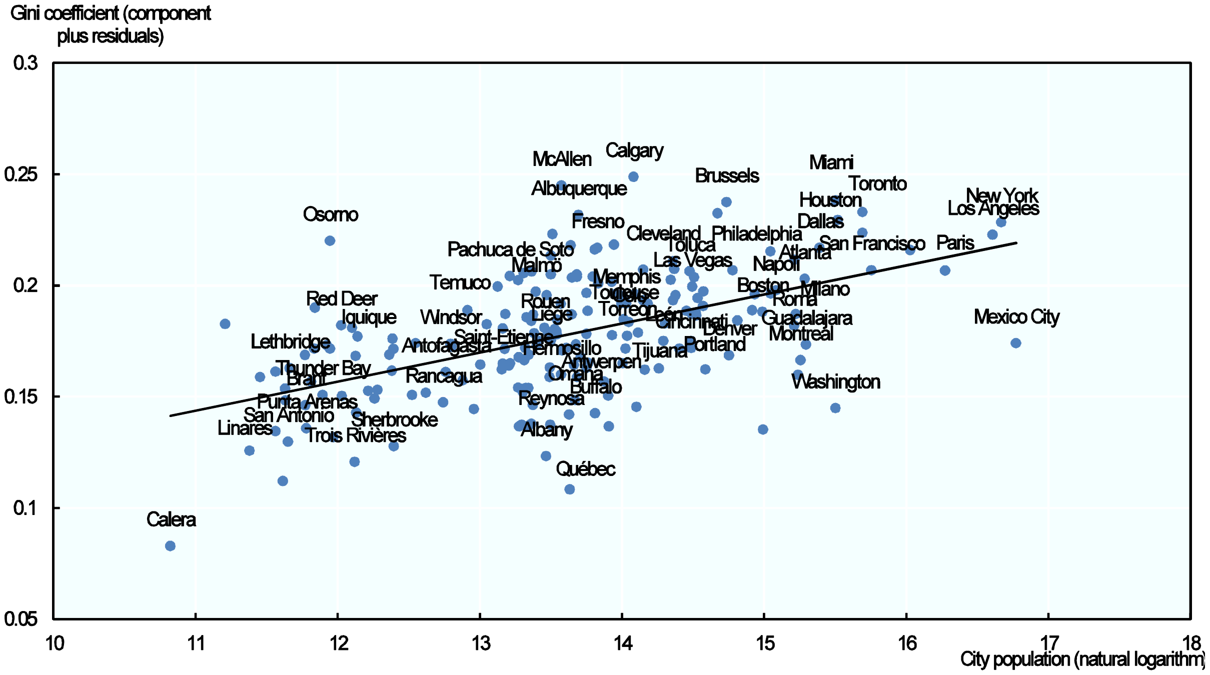

*Source*: Authors' elaborations on national income data (Appendix 1).

Large cities are on average more unequal than smaller ones. The Gini coefficients for the metropolitan areas considered in this paper are positively associated with the metropolitan population, even after controlling for the initial level of income and for the country to which each metropolitan area belongs (Figure 6).[11] Several arguments have been put forward to explain this evidence. Among these factors, agglomeration economies and firm selection play a role. The latter lead the most productive firms to concentrate in large cities, and foster rural-urban migration of people looking for opportunities. This, in turn, increases productivity, but also income inequality, as the returns to skills of urban residents increase, pushing earning differences up (Behrens and Robert-Nicoud, 2014). Education and migration patterns are drivers of inequality at local level. In the metropolitan areas of the United States, in 2000 the share of inequality due to the inequality in skills was around 33%, according to Glaeser et al., 2009. Cities and neighbourhoods with lower incomes typically have poorer schools and local amenities and often suffer from poorer access to several types of important public services.

11 The regression coefficient of total population (in natural logarithm) is equal to 0.013 and it is statistically significant at 99% confidence level. This result is obtained from a cross section of metropolitan areas in the last year available. Similar results are obtained using more points in time (both with and without time and metropolitan fixed effects).

**Figure 6. Metropolitan population and income inequality, circa 2014**

Metropolitan size and inequality, once controlled for income levels and country effect




Note: The figure is generally known as component plus residual plot. It represents the relationships between metropolitan population (in natural logarithm) and the Gini coefficient for household income, after having controlled for the initial level of income in the metropolitan area and for the country to which each metropolitan area belong. Thus, the vertical axis does not report the raw Gini coefficients, but a proxy (component plus residuals) obtained by summing the residuals of a regression of the Gini coefficient on the logarithm of metropolitan population, income levels and country dummies with the product between the logarithm of metropolitan population and its relative coefficient estimated with the linear regression.

*Source*: Authors' elaborations on national income data (Appendix 1).

The Gini coefficient provides an overall indication of disparities in the distribution of household disposable income in each metropolitan area it does not capture where in the distribution the inequality occurs. Moreover, the Gini coefficient is more sensitive to changes around the middle of the income distribution than at the extremes. This sensitivity arises because the Gini coefficient reflects the ranking of the population. Ranking is most likely to change at the densest part of the income distribution, which is likely to be around the middle. Consequently, the ratio between the disposable income of the top 20% and that of the bottom 20% (referred as S80/S20) can be more revealing of inequality in the household populations as it reduces the effect on the statistics of outliers at the top and bottom and prevent the middle 60% hiding inequality.

According to the S80/S20 income ratio, differences observed across cities within the same country are even starker than when looking at the Gini coefficient. In Canada, this ratio is above 10 in Calgary, Vancouver and Toronto in 2013, while it is below 6 in Quebec City. In Belgium, this ratio was 7.6 in Brussels in 2013 but only 4.4 in Gent. Similarly, in Chile the same ratio varies between 11.1 in the Santiago metropolitan areas and 8.7 in Concepción.

## 5. Spatial segregation of households by income within metropolitan areas

Income inequality has a clear spatial dimension in cities, where rich and poor people often live in different neighbourhoods or areas. The spatial organization of metropolitan areas and the differences in living conditions within different parts of a city have been long studied. Spatial segregation is a particular situation in which people within the metropolitan space are concentrated along specific socio-economic lines, such as income, race or ethnicity. Since residential location is the result of a competitive bidding process for housing, land markets play a crucial role in the spatial distribution of socio-economic groups (Mills and Hamilton, 1994). As cities grow, land values become increasingly differentiated due to differences in commuting costs and in the mix of public services and amenities (Monkkonen and Zhang, 2014).

While a certain degree of sorting within metropolitan areas is natural, as households tend to choose neighbourhoods with people who are similar to themselves in terms of income, class, ethnicity or religion (Feijten and van Ham, 2009; Schelling 1969, 1971; Clark, 1991; Morrison, 2015), income segregation becomes a problem when it prevents segments of the population from accessing opportunities and services that would enable them to fully participate in the political and economic process. Segregation may increase the spatial mismatches between affordable housing for low-income households and the jobs they can find (McKenzie, 2016), and reduce the economic opportunities of people living in poorer areas because of inefficient use of public assets or the unavailability of public services (Moretti, 2014, p. 368).

Reseach suggests that living in neighbourhoods with high poverty is associated to worse individual outcomes such as health, income, education, security and general well-being (van Ham et al. 2012). Residential segregation may harm the educational attainment of lower income groups without increasing that of higher income ones (Quillian, 2014). Moreover, the negative effects are passed on to the next generation: in more segregated cities, upward economic mobility for children growing in poor neighbourhoods tends to be lower than in less segregated cities (Chetty, 2014; OECD, 2016). Similarly, concentrated neighbourhood poverty has been found to be associated to higher crime rates and limited social mobility for the people –and especially the children – who live in these neighbourhoods (Sharkey, 2008; Sampson and Sharkey, 2008).

The spatial segregation of people within the metropolitan area reflects several factors such as housing, land-use and transportation. Municipal governments may have different tax and spending capacities, resulting in unequal quantity and quality of services provided or unequal attractiveness to firms. The degree to which different local jurisdictions within metropolitan areas can provide public goods and services of a certain quality is important for increasing the opportunities for those who start from more disadvantaged conditions.

Spatial income segregation within a metropolitan area can be assessed at different geographical scales (neighbourhoods, school districts, municipalities, jurisdictions, etc.). This section focuses on income segregation across municipalities (or local units) that constitute a metropolitan area. In general, municipalities within metropolitan areas are endowed with some governmental/administrative duties and various service provision responsibilities. With this choice of geography we can assess two issues. First, the extent to which households tends to concentrate spatially in different locations (i.e. municipalities) according to their levels of income, thus generating concentration of advantages and disadvantages. Second, whether such concentration is associated to the characteristics of metropolitan governance, such as the degree of administrative fragmentation.

***Decomposing income inequality to measure spatial segregation of metropolitan residents***

Using a generalized entropy measure of inequality such as the Theil index, the income inequality of a metropolitan area can expressed as the sum of a "within-group" component and a "between-group" one, where groups are identified based on the location of households in the local units that constitute the metropolitan area (Theil, 1967; Bourguignon, 1979; Atkinson, 1970; Shorrocks, 1980). The ratio of the "between-group" component and the Theil inequality index can be interpreted as the amount of spatial segregation. The higher the income inequality explained by differences in income between locations, the higher the extent to which households live concentrated, or "segregated", in different municipalities according to their levels of income.[12]

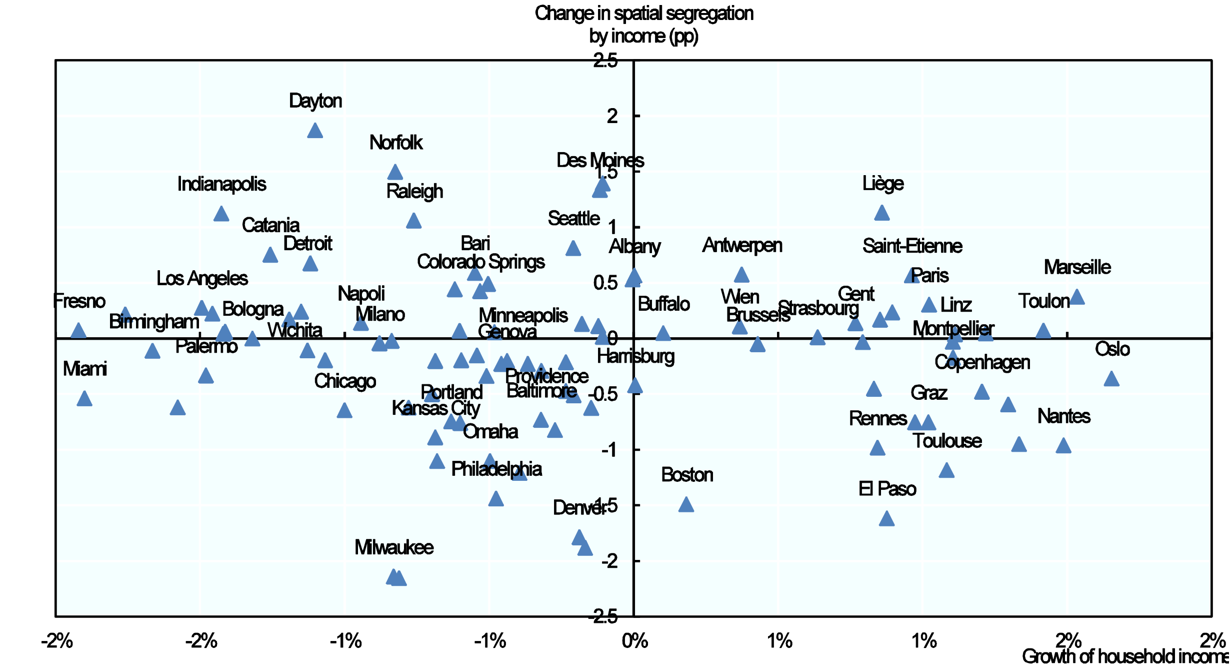
International comparisons of metropolitan areas in terms of spatial (between-groups) inequality are not straightforward. The value of the "between group" component will in fact increase with the number of sub-groups considered, and decrease with their relative size (Cowell and Jenkins, 1995; Shorrocks and Wan, 2005). Metropolitan areas are composed by municipalities (or local units) with different population and in different numbers, both across and within countries, which could bias comparisons across metropolitan areas through conventional decompositions. To overcome this problem, following Elbers et al. (2008), the *between-group* component of inequality is normalised by dividing it by the maximum *between-group* component obtainable given the number and size of groups under analysis. In other words, instead of using the conventional ratio between the between-group inequality ($I_b$) and total inequality, the denominator of this ratio is replaced by the maximum between-group inequality that could be obtained if the number and size of groups were restricted to be the same as for the numerator ($I_{elmo}$). This means re-allocating all individuals in the groups so as to maximise the between-group inequality. This strategy provides a complementary measure of spatial inequality (or segregation, $I_b/I_{elmo}$) which can be used to make comparisons across metropolitan areas with different size and numbers of local jurisdictions (groups).

The index of segregation by income ($I_b/I_{elmo}$) is computed for the metropolitan areas in eight OECD countries, where data available made it possible. According to this index, spatial segregation is highest in Philadelphia and Baltimore (United States) and lowest in Little Rock, Baton Rouge (United States) and Genova (Italy). According to the Theil index, the share of total inequality explained by the differences between municipalities ($I_b/ I_{elmo}$), is, on average, around 5 %, much lower compared with the inequality across individuals within each municipality.

Overall, more unequal metropolitan areas tend to have higher levels of income segregation across their municipalities but the correlation is weak (the correlation with the Gini coefficient is 0.21). Since 2007 many metropolitan areas in France – Nantes, Toulouse, Montpellier, Rennes or Grenoble – and in North-Central Europe (Copenhagen, Oslo, Graz, etc.) have experienced both higher average household income and reduced spatial inequality. Other metropolitan areas, especially in Belgium and France, have combined income growth with a slight fall in spatial segregation. Many other metropolitan areas have experienced lower household income since 2007; in the case of Dayton, Indianapolis, Norfolk, Raleigh (United States) or Catania, Bari, Bologna and Naples (Italy), these declines have combined with an increase of spatial segregation (Figure 7).

12 In addition to the Theil index, which has the property of being equal to the sum of its "between" and "within" components, this section also assesses the spatial inequalities in metropolitan areas by decomposing the Gini coefficient in the sum of three components: "between", "within" and "overlap". The latter consists in inter-group inequalities resulting from the overlap between each group's distribution (see Appendix 3).

**Figure 7. Disposable income growth and change in spatial segregation by income across local jurisdictions, 2007-14**




*Source*: Authors' elaborations on national income data (Appendix 1).

## 6. Fragmented governance and income segregation in metropolitan areas

Metropolitan areas are economically integrated units, but they are often divided into a large number of local jurisdictions without adequate mechanisms of co-ordination for public policy. In this respect, metropolitan areas can have different levels of administrative fragmentation, meaning different extents to which their governance is characterised by many and uncoordinated administrative units.

This section explores empirically the hypothesis that fragmented metropolitan governance is associated to spatial segregation by income within the metropolitan area. From a theoretical point of view, two major mechanisms have been put forward in the literature to explain local administrative structures and the link with the way individuals choose their place of residence (Bischoff, 2008; Lens and Monkkonen, 2016). On the one hand, the Tiebout model links individual location choices with the provision of services by different local authorities (Tiebout, 1956): in this model, a fragmented metropolitan area can favor the sorting of people in those local juridisdictions that provide the set of services that best fits with their preferences and budget constraints. However, different municipalities might not be able to deliver public services of comparable quality, generating disadvantages for people living in the least wealthy ones. In this respect, Jimenez (2014a) analysed the budgetary policy of municipal governments in the United States, concluding that provision of public services is sub-optimal in more fragmented metropolitan areas. This relationship may be explained by limited political influence of citizens of the most disadvantaged places, and lead to class-based population sorting within the metropolitan space.

A second model to understand the implications of administrative fragmentation looks at the supply side rather than at the location choices of individuals. Political boundaries shape many important policies such as transport, housing and local taxation in a way that can isolate some residents, especially the most disadvantaged ones (Danielson, 1976). In other words, local public action can induce a certain separation with respect to neighbouring local administrative units through specific policies in sectors such as education, land-use (i.e. zoning laws) or housing. A high adminstrative fragmentation might induce more competition among municipalities for attracting people and activities generating high revenues. This may lead to an underprovision of services for low-income residents and foster spatial segregation.

The empirical evidence on the link between administrative fragmentation and spatial income segregation is not yet clear. The seminal study by Hill (1974) investigated the relationship between the structure of the local public sector and the inequality between local administrative units, finding a positive association. These results might however be biased by the dependence of the chosen measure of spatial income inequality – the standard deviation of the median household income among municipalities – by the size and number of municipalities (Ostrom, 1983). More recently, Jimenez (2014b) did not find any robust relationship between administrative fragmentation and spatial segregation of income in US metropolitan areas, though he measured segregation at the neighbourhood level and not at the local jurisdiction scale.

In order to investigate the major factors associated with spatial segregation of income and the role of administrative fragmentation for OECD metropolitan areas, a regression analysis was performed using metropolitan areas as units of observation. To our knowledge, there are no other works assessing this relationship by comparing metropolitan areas in different countries, and linking the phenomenon of spatial segregation by income at the scale at which public services are effectively provided (i.e. local jurisdictions). The analysis does not account for reverse causality, thus results should be interpreted with care and for descriptive purposes. However, it can help disentangle some of the urban characteristics that tend to be associated with segregation of household by income at the local jurisdiction scale.

The dependent variable is spatial segregation, measured by the between-group inequality divided by the maximum between-group inequality obtainable given the size and number of local jurisdictions in each metropolitan area ($I_b/I_{elmo}$). For reasons of robustness, both the Theil index and the Gini coefficient were used to compute the indicator of spatial segregation. The dependent variable was regressed on a measure of administrative fragmentation and additional controls. Again, for the sake of robustness, administrative fragmentation was measured through three different indicators: first, the number of municipalities per 100 000 inhabitants for each metropolitan area; second, the logarithm of the number of local administrative units in each metropolitan area; finally, the fragmentation index used by Bischoff (2008), based on the share of population in each local unit.[13]

Other controls were included in the regressions in order to account for other factors affecting spatial segregation of households by income. First, the natural logarithm of income has been added to control for the overall level of development of the metropolitan area. Second, the overall levels of income inequality in the metropolitan area, as measured by the Gini index, as suggested by Reardon and Bischoff, 2011). Further controls include the natural logarithm of total metropolitan population and the degree of decentralisation of the resident population from the main centre, as computed in Veneri (2015). The idea underlying this latter control is that metropolitan areas where people are located more towards the periphery might have undergone a suburbanisation process driven by a preference for location choices towards more isolated and socially homogeneous places. In the same spirit, the ratio between the average household income in the core city-centre over that in the commuting zone was added in order to control for the type of suburbanisation patterns characterising each metropolitan area. In previous works, higher

13 This is computed as follows (Bischoff, 2008):

$$Fragmentation = \sum_{i=1}^{k} P_i(1 - P_i) \ ,$$

where $P$ is the proportion of population who lives in the $i$-th local unit within each metropolitan area. The indicator ranges between 0 and 1, with 0 indicating complete amalgamation (one single local government) and 1 indicating complete fragmentation.

income in central cities were found to be associated with lower segregation, while higher incomes in suburban places were associated with higher segregation of households by income (Lewis and Hamilton, 2011). Table 3 summarises all variables used in the regression analysis, providing descriptive statistics for each of them.

**Table 3. Variables used in the regression analysis: basic statistics**

| Variable | Obs. | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Segregation ($I_b/I_{elmo}$) from Theil index | 225 | 4.968 | 4.293 | 0.007 | 19.284 |
| Segregation ($I_b/I_{elmo}$) from Gini coefficient | 202 | 19.23 | 9.645 | 0.828 | 43.972 |
| Income inequality (Gini coefficient) | 225 | 0.346 | 0.041 | 0.232 | 0.428 |
| Household income (ln) | 224 | 10.442 | 0.253 | 9.672 | 10.998 |
| Admin. fragmentation (n. local jurisdictions per 100k inhab.) | 225 | 8.710 | 10.065 | 0.260 | 51.420 |
| Admin. fragmentation (n. of local jurisdictions) | 225 | 79.067 | 178.791 | 2 | 1375 |
| Admin. fragmentation (Bischoff, 2008) | 224 | 0.667 | 0.229 | 0.008 | 0.980 |
| Population (ln) | 225 | 14.141 | 0.803 | 13.144 | 16.690 |
| City-commuting zone income ratio | 210 | 1.027 | 0.171 | 0.635 | 1.530 |
| Decentralisation of population (sprawl) | 225 | 16.581 | 6.791 | 3.296 | 44.592 |

Regression results are shown in Tables 4-5; estimates refer to different specifications, and alternative measures of administrative fragmentation and spatial segregation (between-group inequality). For each of the three indicators of administrative fragmentation (number of local units per 100 000 inhabitatnts, logarithm of the number of local units, and the fragmentation index in [1]), results are reported by considering at least two different model specifications: a pooled regression using the whole sample of metropolitan areas for three points in time (two in the case of the United States); and a pooled regression including also year dummies. Finally, for the first indicator of spatial segregation, results using metropolitan areas fixed effects instead of country dummies are reported when using the first indicator of segregation, since it is the only one that can be sensitive to change over the short time span considered in this analysis.

Table 4 reports the results of the regression analysis when the dependent variable is the ratio between the between component of the Theil index for household disposable income and the maximum between-group component given the size and numbers of local units, according to Elbers et al. (2009). Table 5 presents the results when the dependent variable is computed by decomposing the Gini coefficient instead of the Theil index. All regressions were estimated through OLS with robust standard errors. For robustness purposes, Appendix 4 presents regression results with the conventional decomposition of inequality (both Theil and Gini), without adjusting for the possible bias of comparing between-group inequality across metropolitan areas with different number and size of local units.

**Table 4. Estimation results. Dependent variable: between component of the Theil index for household income over maximum between component given size and number of groups**

OLS with robust standard errors. Alternative measures of administrative fragmentation

| Variable | Mod1 | Mod2 | Mod3 | Mod4 | Mod5 | Mod6 | Mod7 |
|---|---|---|---|---|---|---|---|
| Admin. fragmentation (n. local jurisdictions per 100k inhab.) | **.1027*** ** | **.1032*** ** | **.3854*** ** | | | | |
| Admin. fragmentation (Bischoff, 2008) | | | | **6.978*** ** | **7.05*** ** | | |
| Admin. fragmentation (n. of local jurisdictions) (ln) | | | | | | **1.007* ** | **1.003* ** |
| Household income (ln) | -0.2283 | -0.1432 | -1.535 | -1.06 | -0.9773 | -0.5421 | -0.4508 |
| Income inequality (Gini coefficient) | 8.774 | 10.69 | -2.997 | 17.19 | 19.57 | 8.112 | 9.783 |
| Population (ln) | 2.265*** | 2.258*** | 0.1658 | 1.59*** | 1.577*** | 1.165* | 1.161* |
| City-commuting zone income ratio | -2.293 | -2.284 | 1.391 | -0.9336 | -0.8993 | -2.068 | -2.062 |
| Decentralisation of population (sprawl) | -.2049*** | -.2089*** | | -.2627*** | -.2683*** | -.17** | -.1735** |
| Constant | -25.58 | -26.14 | 14.59 | -8.372 | -8.83 | -10.83 | -11.5 |
| | | | | | | | |
| Observations | 210 | 210 | 210 | 210 | 210 | 210 | 210 |
| Adjusted R-squared | 0.310 | 0.280 | 0.984 | 0.343 | 0.315 | 0.299 | 0.268 |
| N. of cities | 92 | 92 | 92 | 92 | 92 | 92 | 92 |
| N. years | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| aic | 1126 | 1143 | 207 | 1116 | 1133 | 1129 | 1147 |
| Country fixed effects | yes | yes | no | yes | yes | yes | yes |
| Metropolitan fixed effects | no | no | yes | no | no | no | no |
| Year fixed effects | no | yes | yes | no | yes | no | yes |

legend: * p<.1; ** p<.05; *** p<.01

**Table 5. Estimation results. Dependent variable: between component of the Gini coefficient for household income over maximum between component given size and number of groups**

OLS with robust standard errors. Alternative measures of administrative fragmentation

| Variable | Mod1 | Mod2 | Mod3 | Mod4 | Mod5 | Mod6 | Mod7 |
|---|---|---|---|---|---|---|---|
| Admin. fragmentation (n. local jurisdictions per 100k inhab.) | **.0014*** | **.0014*** | **.0074**** | | | | |
| Admin. fragmentation (Bischoff, 2008) | | | | **.1363***** | **.1377***** | | |
| Admin. fragmentation (n. of local jurisdictions) (ln) | | | | | | **.0229*** | **0.0229** |
| Household income (ln) | -0.017 | -0.0147 | -0.0546 | -0.0297 | -0.027 | -0.0321 | -0.0307 |
| Income inequality (Gini coefficient) | 0.1723 | 0.2086 | -0.1058 | 0.3716 | 0.4207 | 0.1773 | 0.2078 |
| Population (ln) | .058*** | .0578*** | -0.0069 | .0432*** | .0429*** | .0405** | .0404** |
| City-commuting zone income ratio | 0.026 | 0.026 | 0.0493 | 0.0535 | 0.0541 | 0.0274 | 0.0272 |
| Decentralisation of population (sprawl) | -.0048** | -.0049** | | -.0064*** | -.0065*** | -.0046** | -.0046** |
| Constant | -0.4944 | -0.509 | 0.7728 | -0.3066 | -0.3235 | -0.1897 | -0.1976 |
| | | | | | | | |
| Observations | 187 | 187 | 187 | 187 | 187 | 187 | 187 |
| Adjusted R-squared | 0.2598 | 0.2218 | 0.9774 | 0.2978 | 0.2627 | 0.2639 | 0.226 |
| N. of cities | 87 | 87 | 87 | 87 | 87 | 87 | 87 |
| N. years | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| aic | -394.5 | -377.1 | -1167 | -404.4 | -387.2 | -395.6 | -378.2 |
| Country fixed effects | yes | yes | no | yes | yes | yes | yes |
| Metropolitan fixed effects | no | no | yes | no | no | no | no |
| Year fixed effects | no | yes | yes | no | yes | no | yes |

legend: * p<.1; ** p<.05; *** p<.01

The results showed in Tables 4-5 confirm that more administratively fragmented metropolitan areas have higher spatial segregation of households by income. The sign and statistical significance of the coefficient are robust in all specifications, both in terms of indicator of fragmentation and of segregation. Results are also robust when spatial segregation is measured with the conventional measure of the between-group inequality (Appendix 4). Given that segregation is measured at the scale of local jurisdictions rather than neighbourhoods, these findings are in line with the idea that households might sort in space according to their preferences for public goods provided by the different local jurisdictions and their ability to pay (Tiebout, 1956). The introduction of country fixed effects in the empirical models should address the issue that the number and population size of the local jurisdictions differ across countries, thus enhancing the comparability of the results on the links between administrative fragmentation and economic segregation.

Regarding the other controls, the average household income in metropolitan areas is not strongly associated with spatial segregation. Both the sign and statistical significance are affected by the different model specifications. Overall, income inequality is not significantly associated with spatial segregation, differently from what found by Reardon and Bischoff (2011) when analysing segregation at the neighbourhood level in the United States. Similalry, the ratio between average income in the metropolitan core – the high-density part of the metropolitan area including and surrounding the main city centre – over that in the commuting zone does not show a significant association with our headline measures of segregation, though the association becomes negative and often significant when using as dependent variable the unadjusted spatial decomposition of inequality. A negative association is consistent with results by Lewis and Hamilton (2011), who found that cities with richer suburbs and low-density peripheral areas (i.e. exurban) are on average more spatially segregated. The size of metropolitan areas is positively associated with higher spatial segregation, though the coefficient loses its statistical significance when including metropolitan fixed effects. Finally, the extent to which the resident population is centralised and close to the main centre, rather than being decentralised in the more peripheral areas, is associated to higher spatial segregation. This latter result suggests that the tendency towards segregation is higher when people are physically more concentrated in the most central area of the city, thus in a setting where people live relatively close to other people.

## 7. Conclusions

This paper provides internationally comparable estimates of average income levels and distribution at the scale of metropolitan areas. Due to data availability, indicators of average income levels cover all metropolitan areas in 18 OECD countries, while those on income inequality cover metropolitan areas in 11 countries. In addition, income inequality within metropolitan areas was decomposed in a between and within local units component for 9 countries. This decomposition shed some light on how people and households sort themselves in the metropolitan space according to their levels of income. As such, this represents a way to assess spatial segregation of households by income at the scale of local jurisdictions.

One novel aspect of this paper is that, for most countries, it uses country-specific tax records, instead of household surveys, as the starting point for constructing these income estimates. This allows assessing levels and distribution of income at small spatial scales, but it also required harmonising different definitions of income in order to make meaningful international comparisons. This harmonisation was achieved by adjusting country-specific measures of average household income across local units (typically referring to pre-tax household income) to the comparable measures of equivalised household disposable income that are available through the OECD Regional Well-Being Database. For a subset of 11 countries, the same adjustment was applied to country-specific measures of household income by quantiles across local units. Finally, based on assumptions on the form of the income distribution of income within metropolitan areas, these estimates were used to simulate a continuous distribution for each metropolitan areas in these countries.

Results confirm that, in general, people living in metropolitan areas have higher average income than their non-metropolitan counterparts, although the size of such income gap varies both across countries and across metropolitan areas within the same country. In addition, income inequalities within metropolitan areas, as measured through the Gini coefficient, are generally higher in larger cities.

Another innovative aspect of this paper is that it assesses spatial inequality across local jurisdictions within metropolitan areas, which we call "spatial segregation". The paper explores how the administrative fragmentation of metropolitan areas is linked to such spatial segregation. Results from regression analysis suggest a stable and positive association between the two, with results being robust to several measures of both fragmentation and segregation. Spatial segregation is measured at the level of local jurisdictions rather than at the neighbourhood scale where it is usually assessed. This requires adapting the interpretation of the findings and focusing more on the importance of the set of public goods provided by local authorities. The results in the paper suggest that how public goods and services are provided can shape the localisation decisions by households through a different channel than neighbourhood effects. In any case, households' preferences alone might not be sufficient to produce spatial segregation, which also depend on housing policies that tend to sort people in space by income, through, for example, restrictions on lot sizes, residential density or public housing (Reardon and Bischoff, 2011).

Overall, this paper provides new statistical evidence for OECD metropolitan areas on income-related topics that are important to understand levels and trends in material prosperity and inclusion of metropolitan areas. Further work will be needed to improve the capacity to deal with different sources of data and different definitions of income when it comes to comparing metropolitan areas of different countries. The diversity of the smallest spatial units at which income is assessed in each country might also introduce a "modifiable areal unit problem" (MAUP), a situation that makes results vary in line with the aggregation of data into areal units of different size (Openshaw and Taylor, 1979). This issue should be addressed more thoroughly by moving towards a benchmark scale for the production of statistics at the local level.

**References**

Atkinson, A.B. (1970) "On the measurement of income inequality". *Journal of Economic Theory*; 2:244-63.

Atkinson, A.B., L. Rainwater and T. M. Smeeding (1995), Income Distribution in OECD Countries, OECD Social Policy Studies, No. 18, Paris.

Atkinson, A.B. and F. Bourguignon (Eds.) (2014) *Handbook of Income Distribution*. North Holland, Amsterdam.

Balintfy, L. and Goodman, S. (1973), Socio-Economic Factors in Income Inequality: A Log-Normal Hypothesis. *Zeitschrift fiir Nationalokonomie*, Vol. 33, pp. 389-402.

Behrens, K. and Robert-Nicoud, F. (2014) "The survival of the fittest in cities: urbanisation and inequality", *The Economic Journal*, Vol. 124, pp. 1 371-1 400.

Bischoff, K. (2008), "School District Fragmentation and Racial Residential Segregation. How do Boundaries Matter?," *Urban Affair Review*, Vol. 44(2), pp. 182-217.

Bourguignon, F. (1979) Decomposable Income Inequality Measures. *Econometrica*, 47(4), pp. 901-920.

Chetty, R., N. Hendren, P. Kline and E. Saez (2014), "Where is the land of opportunity? The geography of intergenerational mobility in the United States", *The Quarterly Journal of Economics* 129, No. 4, pp. 1 553-1 623.

Chetty, R., Hendren, N., Katz, L.F. (2015), "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment," *NBER Working Paper Series* 21156, available at: http://www.nber.org/papers/w21156.pdf.

Clark, W.A.V. (1991), "Residential preferences and neighborhood racial segregation - a test of the Schelling segregation model", *Demography*, Vol. 28, pp. 1-19.

Cowell, F.A., Jenkins, S.P. (1995), "How much inequality can we explain? A methodology and an application to the United States", *Economic Journal*, Vol. 105(429), pp. 421-430.

Cutler, D.M., Glaeser, E.L. (1997), "Are Ghettos Good or Bad?," *The Quarterly Journal of Economics*, Vol. 112 (3), pp. 827-872.

Dagum, C. (1997) "A new approach to the decomposition of the Gini income inequality ratio". *Empirical Economics*, 22(4), pp. 515-531.

Diaz, M. and F. Murtin (2016), "Socio-economic Inequality in Living Standards", *OECD Statistics Directorate Working Paper*, No. 71, OECD Publishing, Paris.

Duranton, G. and D. Puga (2004), "Microfoundations of urban agglomeration economies", in J.V. Henderson and J.F. Thisse (eds.), *Handbook of Regional and Urban Economics*, Ed. 1, Vol. 4, Chapter 48, Elsevier, pp. 2 063-2 117.

Elbers, C., Lanjouw, P., Mistiaen, J.A., Özler, B. (2008), "Reinterpreting between-group inequality", *Journal of Economic Inequality*, Vol. 6, pp. 231-245.

Feijten, P.M. and M. van Ham (2009), "Neighbourhood change… reason to leave?", *Urban Studies* Vol. 46(10), pp. 2 103–2 122.

Florida, R. and C. Mellander (2016), "The geography of inequality: Difference and determinants of wage and income inequality across US metros", *Regional Studies*, Vol. 50(1), pp. 79-92.

Glaeser, E.L., Resseger, M. and Tobio, K. (2009), "Inequality in cities", *Journal of Regional Science*, Vol. 49(4), pp. 617-646.

Hedman L., Manley D., van Ham M. and Östh J. (2015), "Cumulative exposure to disadvantage and the intergenerational transmission of neighbourhood effects". *Journal of Economic Geography* 15(1), pp. 195-215

Hill, R.C. (1974), "Separate and unequal: Governmental inequality in the metropolis," *American Political Science Review*, Vol. 68, pp. 1 557-1 568.

James, D.R. and Taeuber, K.E. (1985) Measures of segregation. *Sociological Methodology*, 14, 1-32.

Jargowsky, P.A. and J. Kim (2009) "The information theory of segregation: Using segregation and inequality in a common framework", in Y. Fluckiger, S., F. Reardon and J. Silber (Eds.) *Occupational and residential segregation* (Vol. 17), Research on Economic Inequality, Bingley, UK: Emerald.

Jimenez, B.S. (2014a), "Separate, Unequal, and Ignored? Interjurisdictional Competition and the Budgetary Choices of Poor and Affluent Municipalities," *Public Administration Review*, Vol. 74(2), pp. 246–257.

Jimenez, B.S. (2014b), "Externalities in the Fragmented Metropolis: Local Institutional Choices and the Efficiency-Equity Trade-Off", *American Review of Public Administration*, Published online before print October 6, 2014, doi: 10.1177/0275074014550703.

Kaplov, L. (2005), "Why measuring inequality?", *Journal of Economic Inequality*, 3, pp. 65–79.

Lewis, J.H., Hamilton, D.K. (2011), "Race and Regionalism: The Structure of Local Government and Racial Disparity", *Urban Affairs Review*, Vol. 47(3), pp. 349–384.

Lens, M.C. and Monkkonen, P. (2016), "Do Strict Land Use Regulations Make Metropolitan Areas More Segregated by Income?," *Journal of the American Planning Association*, Vol. 82(1), pp. 6-21.

Lopez, J.H. and L. Servén (2006), "A Normal Relationship? Poverty, Growth, and Inequality", World Bank Policy Research Working Paper 3814, January 2006.

Massey, D.S., and N.A. Denton (1988). The dimensions of residential segregation. *Social Forces*, 67: 281-315.

Mills, E. and Hamilton, B. (1994) *Urban Economics*. HarperCollins College Publishers, New York.

Monkkonen, P. and Zhang, X. (2014) Innovative measurement of spatial segregation: Comparative evidence from Hong Kong and San Francisco. *Regional Science and Urban Economics*, 47, pp. 99-111.

Moretti, E. (2013) *The new geography of jobs*. Houghton Mifflin Harcourt

Morrison, P.S. (2015), "The Inequality Debate. The neglected role of residential sorting", *Policy Quarterly*, Vol. 11(2), pp. 72-79.

OECD (2011), Divided We Stand – Why Inequality Keeps Rising, Paris, www.oecd.org/social/inequality.htm and www.oecd.org/fr/social/inegalite.htm.

OECD (2012), *Redefining "Urban": A New Way to Measure Metropolitan Areas*, OECD Publishing. http://dx.doi.org/10.1787/9789264174108-en.

OECD (2013), *OECD Framework for Statistics on the Distribution of Household Income, Consumption and Wealth*, OECD Publishing, http://dx.doi.org/10.1787/9789264194830-en.

OECD (2014), Terms of Reference. OECD project on the distribution of household incomes. OECD Income Distribution Database, available at: http://www.oecd.org/els/soc/IDD-ToR.pdf.

OECD (2015a), "A Note on measuring income levels and distribution in cities: Data, feasibility and challenges", *Paper presented at the 28th session of the Working Party on Territorial Indicators*, 27 April 2015, Paris.

OECD (2015b), *All on Board. Making Inclusive Growth Happen*. OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264218512-en.

OECD (2016), *OECD Regions at a Glance 2016*, OECD Publishing, Paris, http://dx.doi.org/10.1787/reg_glance-2016-en.

Openshaw, S. and Taylor, P.J. (1979), "A Million Or So Correlation Coefficients: Three experiments on the Modifiable Areal Unit Problem". In N. Wrigley (Eds.), *Statistical Applications in the Spatial Sciences* (pp. 127–144). Pion, London.

Ostrom, E. (1983), "The social stratification-government inequality thesis explored," *Urban Affairs Quarterly*, Vol. 19, pp. 91-112.

Piacentini, M. (2014), "Measuring Income Inequality and Poverty at the Regional Level in OECD Countries", *OECD Statistics Working Papers*, 2014/03, OECD Publishing, http://dx.doi.org/10.1787/5jxzf5khtg9t-en.

Piketty, T. (2015), *Capital in the Twenty-First Century*, Harvard University Press: Paris.

Quillian, L. (2014), "Does Segregation Create Winners and Losers? Residential Segregation and Inequality in Educational Attainment", *Social Problems*, Vol. 61(3), pp. 402–426.

Reardon, S.F., Bischoff, K. (2011), "Income Inequality and Income Segregation", *American Journal of Sociology*, Vol. 116(4), pp. 1 092-1 153.

Rothwell, J.T. and D.S. Massey (2015), “Geographic Effects on Intergenerational Income Mobility”, *Economic Geography*, Vol. 91, No. 1, pp. 83-106.

Sampson, R.J., Sharkey, P. (2008), “Neighborhood selection and the social reproduction of concentrated racial inequality”, *Demography*, 45(1), pp. 1-29.

Schelling, T.C. (1969), “Models of segregation”, *The American Economic Review*, Vol. 59, pp. 488-493.

Schelling, T.C. (1971), “Dynamic models of segregation”, *Journal of mathematical sociology*, Vol. 1, pp. 143-186.

Sen, A.K. (1973), *On Economic Inequality*. Clarendon Press, Oxford.

Sharkey, P. (2008), “The intergenerational transmission of context”, *American Journal of Sociology*, 113(4), pp. 931-969.

Shorrocks, A. (1980), *Inequality decomposition by factor components*. Institute for Economic Research, Queen’s University, Kingston, Ontario.

Shorrocks, A., Wan, G. (2005), “Spatial decomposition of inequality”, *Journal of Economic Geography*, Vol. 5(1), pp. 59-81.

Steyn, H.S. (1959), “A model for the distribution of incomes”, *South African Journal of Economics*, Vol. 27(2), 149-156.

Tammaru T., S. Marcińczak, M. van Ham and S. Musterd (eds) (forthcoming) *Socio-Economic Segregation in European Capital Cities: East Meets West*. Routledge: Oxford.

Theil, H. (1967). *Economics and Information Theory*. North Holland, Amsterdam.

United Nations (2011) *Camberra Group Handbook on Household Income Statistics. Second Editions*. United Nation Publishing, Geneva, available at: http://www.unece.org/fileadmin/DAM/stats/groups/cgh/Canbera_Handbook_2011_WEB.pdf.

Veneri, P. (2015), Urban Spatial Structure. Is Urban Population Decentralising or Clustering? *OECD Regional Development Working Papers* 2015/1, OECD Publishing, http://dx.doi.org/10.1787/5js3d834r3q7-en.

World Bank (2015), “A measured approach to ending poverty and boosting shared prosperity: concepts, data, and the Twin Goals”, Policy Research Report, World Bank, Washington, DC.

Zhang, X. and R. Kanbur (2001) “What Difference Do Polarisation Measures Make? An Application To China,” *Journal of Development Studies*, 37(3), pp. 85-98.

## APPENDIX 1: SOURCES OF DATA FOR COMPUTING HOUSEHOLD INCOME LEVELS IN METROPOLITAN AREAS

**Table A1.1. Sources of data**

| Country | Source | Link | Original definition of income | Households estimation |
|---|---|---|---|---|
| Australia | Australian Bureau of Statistics | http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/6524.0.55.0022005-06%20to%202010-11?OpenDocument | Total income from all sources: Wages and salaries, Own unincorporated business, Superannuation and annuities, Investments and Other income (excl. Government pensions & allowances) | Estimated from official population |
| Austria | Statistics Austria | Sent by Statistics Austria | Net income = total income including transfer payments – tax paid | Estimated from official population |
| Belgium | Statistics Belgium | http://statbel.fgov.be/fr/modules/publications/statistiques/marche_du_travail_et_conditions_de_vie/Statistique_fiscale_des_revenus.jsp | Taxable net total income (all net income after subtracting deductible expenses) - Total income tax (amount of state taxes, local taxes and agglomeration taxes) | Official |
| Canada | Statistics Canada | Sent by Statistics Canada | Household disposable income, defined as employment income + semi-employment income + investment income + transfers received – transfers paid. | Official from Census |
| Chile | CASEN – Ministry of Social Development | http://www.ministeriodesarrollosocial.gob.cl/basededatoscasen.php | Household disposable income | Estimated from the micro-data |
| Denmark | Statistics Denmark | http://statistikbanken.dk/statbank5a/default.asp?w=1920 | Disposable income per fiscal household excluding imputed rent | Official |
| Estonia | Statistics Estonia | http://pub.stat.ee/px-web.2001/Dialog/varval.asp?ma=IM005&ti=AVERAGE+MONTHLY+GROSS+INCOME+PER+EMPLOYEE+AND+RECIPIENTS+OF+GROSS+INCOME+BY+REGION%2F++ADMINISTRATIVE+UNIT%2C+SEX+AND+AGE+GROUP&path=../I_Databas/Social_life/06Income/&lang=1 | Gross income – remuneration subject to social tax, paid to the employee or public servant; scholarship, allowance and pension paid in relation to the employment or service relationship; remuneration paid for the performance of work paid pursuant to a legal act or other legislation; remuneration paid to a person after the end of employment or service relationship (excl. benefit paid to the employee or public servant upon the termination of contract or upon removal from post) according to the Estonian Tax and Customs Board declaration of income and social tax, unemployment insurance premiums and contributions to mandatory funded pension. | Estimated from official population |
| Finland | Statistics Finland | http://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/StatFin__tul__tjkt/010_tjkt_tau_101.px/?rxid=10c8d752-b029-42d2-8e14-58949da5daf0 | Households' disposable money income includes monetary income items and benefits in kind connected to employment relationships. Money income does not include imputed income items, of which the main one is imputed rent. The formation of disposable money income can be described as follows: wages and salaries + entrepreneurial income + property income (without imputed rent) + current transfers received (without imputed rent) – current transfers paid. | Official |

**Table A1.1. Sources of data** *(cont.)*

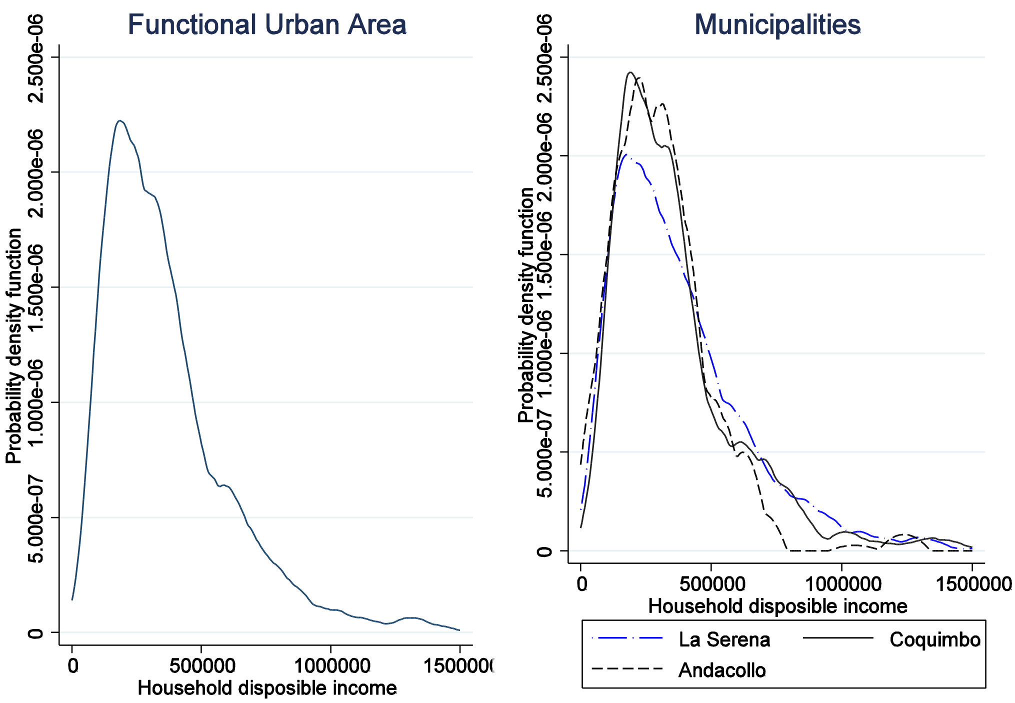
| Country | Source | Link | Original definition of income | Households estimation |
|---|---|---|---|---|
| France | Insee | http://www.insee.fr/fr/bases-de-donnees/default.asp?page=statistiques-locales/revenu-niveau-vie.htm | Tax income corresponds to the sum of the resources declared by taxpayers in their income tax return prior to any deduction. It does not correspond to disposable income. Tax income thus includes the income from salaried activity and self-employment, disability and retirement pensions (excluding the minimum for old age), alimony received (with alimony paid deducted), certain income from household assets, and taxable social income: sickness and unemployment benefits (excluding the RSA). Tax income is broken down into four main categories : salaried income ; income from non-salaried professions (profits) ; pensions and annuities ; other income (essentially from assets). | Official |
| Hungary | Regional Development and Spatial Planning Information System | https://www.teir.hu/ | Net personal income. The net personal income data is the income after tax per capita (for one year). The net income is equivalent the domestic income minus tax per population. | Estimated from official population |
| Italy | Ministry of Economy and Finance - Dept. of Finance | http://www1.finanze.gov.it/finanze2/pagina_dichiarazioni/dichiarazioni.php | Total taxable income from fiscal declarations | Estimated from official population |
| Japan | Ministry of Internal Affairs and Communications | http://www5.cao.go.jp/keizai-shimon/kaigi/special/future/keizai-jinkou_data.html | Taxable income per taxpayer from fiscal declaration | Official. Estimated from population for 2013. |
| Mexico | CONEVAL | Sent by INEGI | Household total income is equal to monetary income and non-monetary income: work-related income (remuneration for subordinate work and independent work income), property rental income and transfers (including in kind transfers). | Estimated by the states numbers of households |
| Netherlands | Statistics Netherlands | http://www.cbs.nl/nl-NL/menu/themas/inkomen-bestedingen/cijfers/inkomen-van-huishoudens/default.htm | Disposable income is gross income minus current transfers paid as alimony of the ex-spouse(s), income insurance premium such as premiums paid for social/national/private insurance in related to unemployment/disability/old-age/next-of-kin, health insurance premiums. | Official |
| Norway | Statistics Norway | https://www.ssb.no/en/statistikkbanken | Ordinary income after special deductions is the equivalent of net income. Ordinary income after special deductions is the basis for municipal income tax, county income tax and community tax. Special deductions are given due to age, disabilities or reduced ability to earn an income, unusual high expenses due to illness, and parents' deductions. | Official |
| Sweden | Statistics Sweden | http://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__HE__HE0110__HE0110A/SamForvInk1/?rxid=55325ff2-4a5e-48e6-b8a4-a102bdd8c16d | Income from employment and business. It also includes income from pensions, sick pay, and unemployment benefits. | Estimated from official population |
| UK | ONS | http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-416744 | Weekly household income, which is the sum of the gross income of every member of the household plus any income from taxes/benefits such as Working Families Tax Credit | Estimated from official population |
| US | American Community Survey | http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t#none | Total income is the sum of the amounts reported separately for wage or salary income; net self-employment income; interest, dividends, or net rental or royalty income or income from estates and trusts; Social Security or Railroad Retirement income; Supplemental Security Income (SSI); public assistance or welfare payments; retirement, survivor, or disability pensions; and all other income. | Official |

**APPENDIX 2. IS HOUSEHOLD DISPOSABLE INCOME AT THE MUNICIPALITY LEVEL LOG-NORMALLY DISTRIBUTED?**

This paper estimates the complete income distributions of several municipalities and metropolitan areas by making use of available information on the income distributions (e.g. the mean, the quantiles, the deciles, etc.) and by assuming that household disposable income follows a log-normal distribution. The latter hypothesis can be tested for Chile, the only country for which micro data on income at the municipality level is available.

The Figure below shows the estimated probability density function (PDF) of household disposable income in the FUA of Coquimbo-La Serena, as well as in the three municipalities that compose it (i.e. La Serena, Coquimbo and Andacollo). Just by looking at Figure A3.1 one can reasonably assume that these PDFs can be well fitted by a log-normal density function; however, more formal tests are required to claim the log-normality of income.

**Figure A2.1. Household income distribution of Coquimbo-La Serena**



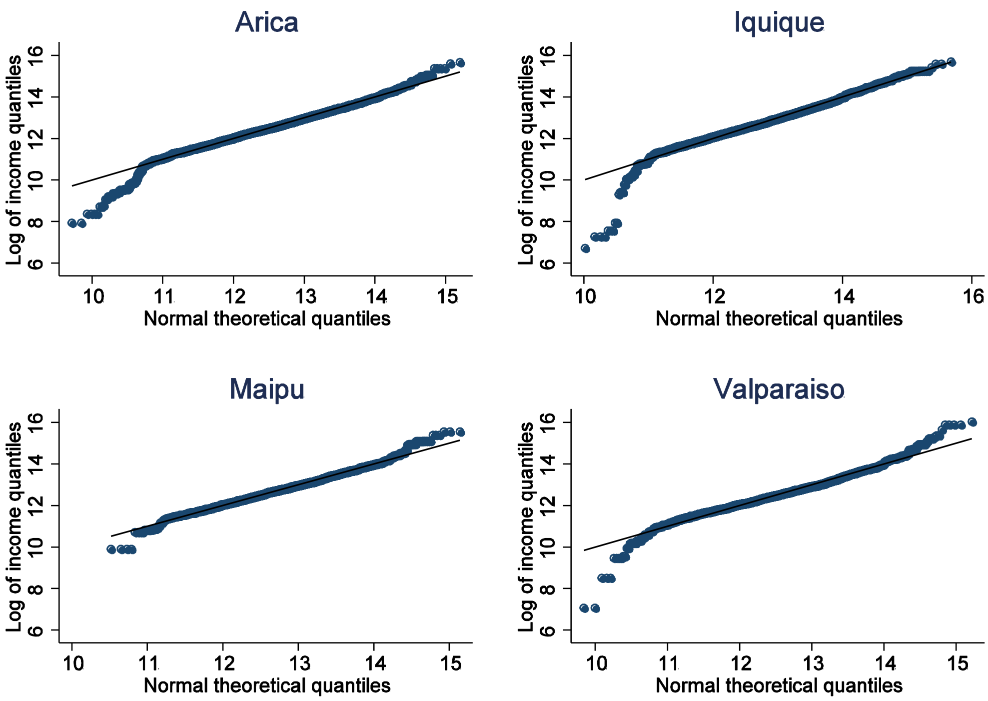

Source: Own elaboration.

Figure A3.2 presents the quantiles generated by the distribution of the log of household disposable income[14] against the theoretical quantiles of a normal distribution for four Chilean municipalities (or *comunas*). From these plots it is possible to see that the quantiles of the log of household disposable

[14] Remind that if a variable X is log-normally distributed then the log of this variable X (i.e. ln(X)) is normally distributed. Thus, to verify that household disposable income is log-normally distributed one can proceed to apply tests of normality to the log of household disposable income.

income are very close to the theoretical quantiles of normal distribution, with the exception of some deviations in the left tail of the distribution.

**Figure A2.2. Normal quantile-quantile plots for different Chilean municipalities**




Source: Own elaboration.

Three standard tests are performed to verify the normality of the log of household disposable income, namely the Shapiro-Wilk normality test, the Shapiro-Francia normality test and the Skewness-Kurtosis test for normality; since these tests are not suitable with large samples (with large samples the smallest deviation from perfect normality will lead to the significantly rejection of the normality hypothesis), instead of looking at the distribution in the whole municipality one can break down the municipality into municipality-gender-education groups, in total there are 28 gender-education groups for the 101 municipalities (with different sample sizes but all below the 1500 observations).

For different sample sizes and confidence intervals, Table A2.1 shows the percent of municipality-gender-education groups for which the null hypothesis that the distribution is normal could not be rejected. As expected, the higher the sample size the larger the probability of rejecting the null hypothesis; however, at the 99% and 95% of confidence, even when the samples can go up to 1500 observations, at least 75% and 67% of the distributions respectively behave like normal distributions. Similarly, Table A2.2 shows the percentage of not-rejection of the normal hypothesis for samples of different sizes randomly selected. Again, when samples are sufficiently small, the test does not reject such hypothesis for small samples.

**Table A2.1. Percentages of non-rejection of normality test for the log of household disposable income**

Samples identified using groups by education-gender

| Sample size | Confidence intervals | Number groups | Shapiro Wilk | Shapiro Francia | Skewness Kurtosis |
|---|---|---|---|---|---|
| 35 | 99% | 1317 | 85.6 | 84.7 | 85.6 |
| | 95% | | 76.3 | 74.6 | 76.9 |
| 50 | 99% | 1515 | 81.4 | 80.1 | 81.7 |
| | 95% | | 71.7 | 69.7 | 72.3 |
| 75 | 99% | 1644 | 77.0 | 75.8 | 77.6 |
| | 95% | | 67.6 | 65.6 | 68.2 |
| 100 | 99% | 1673 | 75.8 | 74.6 | 76.3 |
| | 95% | | 66.5 | 64.6 | 67.1 |

Source: CASEN 2013.

**Table A2.2. Percentages of non-rejection of normality test for the log of household disposable income**

Randomly selected samples of different sizes

| Sample size | Confidence level | Number of samples | Shapiro Wilk | Shapiro Francia | Skewness Kurtosis |
|---|---|---|---|---|---|
| 35 | 99% | 100 | 82 | 81 | 80 |
| | 95% | | 77 | 73 | 73 |
| 50 | 99% | 100 | 75 | 69 | 69 |
| | 95% | | 62 | 56 | 59 |
| 100 | 99% | 100 | 60 | 57 | 59 |
| | 95% | | 49 | 43 | 50 |
| 200 | 99% | 100 | 40 | 38 | 41 |
| | 95% | | 35 | 31 | 32 |

Source: CASEN 2013.

## APPENDIX 3: ASSESSING SPATIAL INEQUALITY USING THE GINI COEFFICIENT: PROPERTIES AND LIMITATIONS

The decomposition of the Gini coefficient to measure the economic segregation of metropolitan areas is particularly data demanding since the Gini coefficient is the sum of three components: the within inequality, the net between inequality and the residual term which measures the overlap among subpopulations. While the entropy indicators have the advantage of being neatly decomposable into the sum of a between and a within component, the Gini coefficient and the Atkinson index satisfy four criteria considered relevant to measure segregation or dissimilarity among sub-groups: principle of transfer, size invariance, organizational equivalence and compositional invariance (James and Taeuber, 1985). The transfer principal property states that the segregation measures should be affected by changes of units from the minority group to another area; size invariance means that a segregation measure should not change if the number of people in each group is multiplied by a constant; the organizational equivalence states that the segregation measure is unaffected by combining units with the same minority composition (Massey and Denton, 1988). The final property on the compositional invariance can be explained as follows. The Gini coefficient is a summary of the differences between each individual's income and every other individual in the population. Mathematically, it is calculated as the arithmetic mean of the absolute value of differences between all pairs of incomes, divided by the average income. In the entropy index, instead, the income of each person is compared with the mean income of the population. In this way, the between groups entropy index is only dependent on the income average of each subpopulation ($j = 1,2, \dots, k$):

$$I_{\beta b} = \frac{1}{\beta(\beta+1)} \sum_{j=1}^{k} \frac{n_j}{n} \frac{\mu_j}{\mu} \left[ \left( \frac{\mu_j}{\mu} \right)^{\beta} - 1 \right]$$

As emphasized by Dagum (1997), the indexes whose intergroup components are calculated solely in accordance with the averages of the sub-groups do not account for the asymmetry and variance because it is as if the subpopulations were normally and equally distributed, of common variance, and statistically independent. Instead the residual component of the Gini index quantifies the weight of inter-group inequalities resulting from the overlap between the k distributions. It means that some individuals in the lowest part of the income distributions have higher incomes than some of those in the highest part.

In the assessment of spatial segregation through the decomposition of the Gini coefficient, the main idea is to measure the dissimilarity in income distribution between local units (municipalities) in a metropolitan area. The Gini coefficient of the metropolitan area can be written as

$$G = G_w + G_{gb} = G_w + G_b + G_t$$

where $G$ is the Gini coefficient of disposable income of a metropolitan area; $G_w$ measures the within-group (municipality) inequality and $G_{gb}$ is the gross between-group (municipality) component. The $G_{gb}$ component is given by the sum of $G_b$ and $G_t$, with $G_b$ being the net contribution of inequality of the between-group municipalities component. While $G_b$ captures the averages inequalities between groups (and thus $G_b = 0$ if the means of the sub-populations are all the same), $G_t$ measures the contribution of the income intensity of trans variation between sub-populations. It quantifies the weight of inter-group inequalities resulting from the overlap between the $k$ distributions. These are particular inequalities: the overlap means that, in a given local unit, some individuals in the low part of the distributions have higher incomes than some individuals in the high part of the distribution who are located in another unit. If instead the $k$ sub-populations do not overlap ($G_t = 0$), the relative position of each individual in a sub-population is the same as in the total income distribution.

Previous work suggested using the components of the Gini coefficient as measures of residential segregation: for example, Zhang and Kanbur (2001) use the ratio of Gini between-group inequality to within-group inequality as a socioeconomic polarization index, although Atkinson and Bourguignon (2014) question the use of within-group inequality to capture internal homogeneity in the income polarization framework. Jargowsky and Kim (2009) suggest to measure income segregation via the ratio of between-group inequality over total inequality ($G$).

## APPENDIX 4. REGRESSION ANALYSIS: ROBUSTNESS CHECKS

**Table A4.1. Estimation results. Dependent variable: between component of the Theil index**

OLS with robust standard errors. Alternative measures of administrative fragmentation

| Variable | Mod1 | Mod2 | Mod3 | Mod4 | Mod5 | Mod6 | Mod7 |
|---|---|---|---|---|---|---|---|
| Admin. fragmentation (n. local jurisdictions per 100k inhab.) | **3.3e-04*** ** | **3.3e-04*** ** | **7.4e-04*** ** | | | | |
| Admin. fragmentation (Bischoff, 2008) | | | | **.014*** ** | **.0141*** ** | | |
| Admin. fragmentation (n. of local jurisdictions) (ln) | | | | | | **.0038*** ** | **.0038*** ** |
| Household income (ln) | 0.0068 | 0.0075 | -0.0065 | 0.0071 | 0.0079 | 0.0055 | 0.0061 |
| Income inequality (Gini coefficient) | .0578* | .0628* | .0314*** | .0704** | .0765** | .0581* | .0624* |
| Population (ln) | .0073*** | .0072*** | -0.0022 | .0051*** | .0051*** | .0034** | .0033** |
| City-commuting zone income ratio | -0.0092 | -0.0092 | 0.0024 | -0.0047 | -0.0047 | -0.0086 | -0.0086 |
| Decentralisation of population (sprawl) | -4.4e-04* | -4.5e-04* | | -5.5e-04*** | -5.7e-04*** | -3.40E-04 | -3.50E-04 |
| Constant | -.175** | -.1794** | 0.0875 | -.1595** | -.1655** | -.122* | -0.1269 |
| | | | | | | | |
| Observations | 213 | 213 | 213 | 213 | 213 | 213 | 213 |
| Adjusted R-squared | 0.434 | 0.409 | 0.997 | 0.434 | 0.408 | 0.429 | 0.403 |
| aic | -1324 | -1307 | -2565 | -1324 | -1307 | -1322 | -1305 |
| Country fixed effects | yes | yes | no | yes | yes | yes | yes |
| Metropolitan fixed effects | no | no | yes | no | no | no | no |
| Year fixed effects | no | yes | yes | no | yes | no | yes |

legend: * p<.1; ** p<.05; *** p<.01

**Table A4.2. Estimation results. Dependent variable: between component of the Gini coefficient**

OLS with robust standard errors. Alternative measures of administrative fragmentation

| Variable | Mod1 | Mod2 | Mod3 | Mod4 | Mod5 | Mod6 | Mod7 |
|---|---|---|---|---|---|---|---|
| Admin. fragmentation (n. local jurisdictions per 100k inhab.) | **8.9e-04*** ** | **8.9e-04*** ** | **.0027*** ** | | | | |
| Admin. fragmentation (Bischoff, 2008) | | | | **.0917*** ** | **.0917*** ** | | |
| Admin. fragmentation (n. of local jurisdictions) (ln) | | | | | | **.015*** ** | **.015*** ** |
| Household income (ln) | 0.0149 | 0.0155 | -.0346** | 0.0014 | 0.0012 | 0.0074 | 0.0075 |
| Population (ln) | .0231*** | .0231*** | -0.0058 | .0181*** | .0182*** | .0097** | .0097** |
| City-commuting zone income ratio | -.0515*** | -.0519*** | 0.0141 | -.0379** | -.0383** | -.0511*** | -.0515*** |
| Decentralisation of population (sprawl) | -1.60E-05 | -1.70E-05 | | -9.0e-04* | -9.0e-04* | 2.20E-04 | 2.20E-04 |
| Constant | -.3968*** | -.4013*** | 0.4576 | -0.1714 | -0.1709 | -0.1979 | -0.2004 |
| | | | | | | | |
| Observations | 213 | 213 | 213 | 213 | 213 | 213 | 213 |
| Adjusted R-squared | 0.5436 | 0.523 | 0.9934 | 0.6483 | 0.6326 | 0.5576 | 0.5377 |
| aic | -947.2 | -929.6 | -1986 | -1003 | -985.2 | -953.8 | -936.3 |
| Country fixed effects | yes | yes | no | yes | yes | yes | yes |
| Metropolitan fixed effects | no | no | yes | no | no | no | no |
| Year fixed effects | no | yes | yes | no | yes | no | yes |

legend: * p<.1; ** p<.05; *** p<.01