Inequality and city size: an analysis for OECD functional urban areas

David Castells-Quintana[1]
Vicente Royuela[2]
Paolo Veneri[3]

ABSTRACT:

As cities grow, both the productivity of their inhabitants and the income distribution among them is expected to change. While the empirical literature has widely shown how productivity (and income) changes with city size, the empirical evidence on the effects on income inequality remains very limited. The few papers that study the relationship between city size and city-level inequality focus on a single country and do not provide international comparative evidence. In this paper, we study the relationship between city size and income inequality at city level for a sample of 153 Functional Urban Areas (FUAs) across 11 OECD countries.

Key words:

Cities; FUAs; city size; inequality; development

JEL Codes:

O18, R12, R13

Acknowledgements:

We are grateful for comments received by Lewis Dijkstra and participants at the 57th European Regional Science Association Congress – 2017. We also thank Roberta Capello and two anonymous reviewers for the comments. We also acknowledge Monica Brezzi and Joaquim Oliveira Martins for comments on earlier drafts. David Castells-Quintana acknowledges the support of ECO2016-75805-R and ECO2016-76855-P. Vicente Royuela acknowledges the support of ECO2016-75805-R.

---

[1] Department of Applied Economics. Univ Autonoma de Barcelona. 08193 Bellaterra, Barcelona, Spain. AQR-IREA Resarch Group, Universitat de Barcelona. 08034, Barcelona, Spain. Corresponding author: David.Castells.Quintana@uab.cat

[2] AQR-IREA Resarch Group, Universitat de Barcelona. 08034, Barcelona, Spain. vroyuela@ub.edu

[3] OECD Centre for Entrepreneurship, SMEs, Regions and Cities, 75775 Paris CEDEX 16, France. paolo.veneri@oecd.org

1. INTRODUCTION

Cities naturally change as they grow in population size. Changes include transformations in the industrial structure, in socio-economic dynamics, in the opportunities available, and even in the productivity of urban workers. As a result, not only the average income of cities but also its distribution is expected to change. However, how income inequality changes with city size is not unambiguous: from a theoretical point of view, there are reasons to expect that larger cities become less unequal, but there are also reasons to expect the opposite. Moreover, the size-inequality relationship is likely to depend on many characteristics of cities, like their initial size, income levels, and other contextual factors.

The relationship between city size and city-level inequality is today of special relevance. Metropolitan areas around the world are reaching a size than in many cases exceeds a population of 10, 20 or even 30 million inhabitants. The increasing expansion of cities, both in physical and functional terms, as well as their changing spatial organisation, has pushed an entire research agenda on defining what the actual boundaries of cities are and even what should be considered as 'urban'. At the same time, there is a renovated concern about the soaring inequalities that we see around the world and the fact that in a rapidly urbanising world many of these inequalities are mostly and increasingly explained by what happens within cities (see for instance Weeks et al. 2006; Ravalion and Chen 2007).

While the empirical literature has widely studied how productivity (and income) changes with city size, the empirical evidence on the effects on income inequality remains very limited. In this paper, we study the association between city size and income inequality at city level for a large international sample of cities, relying on panel data for more than 150 Functional Urban Areas (FUAs) across 11 OECD countries.[4] Our main aim is twofold: to describe the main facts of city-level inequality, and to investigate whether larger cities experience higher levels of income inequality. To do so, we implement several estimations

[4] Throughout the paper, when may refer to cities, we are mainly referring to Functional Urban Areas.

techniques (from pooled-OLS to IV estimations), controlling for several city and country characteristics, including demographic, socioeconomic and geographical variables, and perform several robustness checks.

Our paper contributes to the literature in different ways. First, papers to date studying the relationship between city size and city-level inequality focus on a single country, with no paper (to the best of our knowledge) providing evidence for cities across different countries. Relying on data for one single country can result in unreliable generalisations (i.e., the issue of external validity). Furthermore, analysis for single countries restricts the possibilities for comparative analysis. Our paper considers cities in 11 countries using an international consistent definition of cities and a measure of city-level inequality comparable across countries. Second, papers to date usually focus on large cities. The focus on large cities reduces the possibility to study heterogeneities in the size-inequality relationship. We consider both small and large cities, in countries with different income levels. Third, in doing so, we are able to explore in more depth, and explain, differences in the city size-inequality relationship across cities of different sizes, different income levels, and different geographical areas. Finally, by using different inequality measures, we are also able to study the size-inequality relationship along different parts of the income distribution.

Our results show that larger city size is associated with higher inequality. However, our results also suggest that this positive association is mainly driven by relatively rich and large cities (i.e., with an income per capita level above around 20 thousand 2010-PPP US dollars, and an initial population higher than 1.5 million inhabitants). Looking at different parts of the income distribution, we also find that it is mainly inequality among the relatively rich which increases as large cities grow.

The remainder of the paper is organised as follows. Section 2 (briefly) reviews the literature and sets our theoretical framework. Section 3 describes the data and presents some basic stylised facts on the city size-inequality relationship. In Section 4, we perform our econometric analysis and present our results. Finally, in Section 5 we discuss, conclude and provide policy implications from our results and avenues for further research.

2. THEORETICAL FRAMEWORK AND REVIEW OF THE LITERATURE

The relationship between city size (in terms of population) and income inequality at city level is not unequivocal. On the one hand, there are theoretical reasons to expect that inequality goes down as cities grow. First, larger cities provide more opportunities, which may more strongly benefit low-income workers (see insights in Harris and Todaro 1976). Second, increasing city size can lead to rising average incomes, which in some cases can favour the low-income urban residents more than the high-income ones (Duncan and Reiss 1956; and Richardson 1973). Third, as cities grow, financial markets develop and this can lead to broader investment in human capital, lower average rate of return of such investments and lower inequality (Frech and Burns 1971). On the other hand, there are also theoretical reasons to expect that city growth leads to higher inequality. First, as cities grow, we can expect a widening of the distribution of skills (Mathur 1970; Farbman 1975). Second, increases in city size raise monopoly rents, which usually favour the relatively rich (Haworth et al. 1978). Third, as cities grow, their industrial structure is likely to change, with potential effects on the distribution of income (see for instance Henderson 2010). Finally, larger (and usually richer) cities also allow for agglomeration economies that tend to increase more the productivity of the high-skilled workers, as well as that of already more productive firms and sectors. In turn, the latter tend to sort into larger cities (Baum-Snow and Pavan 2013; and Robert-Nicoud and Beherens 2014).

Insights on how inequalities increase as cities grow can also be derived from the literature on urban scaling. According to this literature, larger (and richer) cities disproportionally concentrate patent registrations (Bettencourt et al. 2007), innovations and wealth (Bettencourt et al. 2010), human interactions (Schläpfer et al. 2014), and the highest income earners: due to the concentrations of high income and high housing costs, largest cities may have a resulting housing market structure that will push out lower and medium income earners (Sarkar 2018). Because there are usually more people living on lower incomes this translates into a scaling of inequality: the larger the city the greater the inequality (Sarkar 2016).

Whether inequality-decreasing or inequality-increasing dynamics dominate is basically an empirical question. However, the timing and context of these opposing dynamics is relevant. According to Nord (1980), in smaller cities the inequality-decreasing forces tend to dominate and therefore, as cities grow, inequality is expected to decrease. But when large cities grow, inequality-increasing forces dominate and inequality is expected to increase. This suggests a U-shaped relationship between city size and inequality. International evidence looking at average city size and income inequality at country level supports this U-shaped relationship between city size and income inequality (Castells-Quintana 2018).[5]

While the literature has long studied the evolution and the determinants of income inequality at country level, empirical research in this regard at the city level has been much more limited.[6] With cities gaining importance, as they concentrate a higher share of total population, and with an increased data availability, several recent papers from the urban economics literature have empirically studied the relationship between city size and inequality at the city level (Levernier et al. 1998, Baum-Snow and Pavan 2013, Glaeser et al. 2015, Florida and Mellander 2016, Ma and Tang 2016, for the US; Sarkar et al. 2016, for Australia; Lee et al. 2016, for the UK; Chen et al. 2018, for China; and Hortas-Rico and Ríos 2019, for

[5] This type of association is different to the inverted-U shape known as the Williamson hypothesis (Williamson 1965), which explains regional inequality: at early stages of development, growth takes place in specific regions of a country, usually large cities, increasing horizontal inequality. As the country develops, and by means of migration, wages adjustment and redistribution, inequality decreases. Horizontal inequality captures differences between regions, while the focus of our analysis, inequality within cities, calls for vertical inequality (Stewart 2005).

[6] A related strand in the literature has studied vertical inequality *at the regional level* across different countries (i.e., Perugini and Martino 2008; Tselios 2008, 2014; Rodríguez-Pose and Tselios 2009; Castells-Quintana et al. 2015; Royuela et al. 2019).

Spain). Most of these papers suggest that inequality increases with city size. However, these papers show evidence for single countries and have usually relied on the study of large metropolitan areas in rich countries (like the US). To date, to the best of our knowledge, studies for smaller cities and in lower-income countries are missing.[7] According to the theory, for small cities the size-inequality relationship could be different from that in larger counterparts. Furthermore, the growth (and type of growth) that cities experience is likely to depend of on their economic development (Castells-Quintana 2018). Consequently, the size-inequality relationship may differ according to the income level.

In sum, for small cities one can expect income inequality to go down as cities grow in size. By contrast, in large cities one can expect the opposite: inequality to increase as cities continue to grow. And this may further depend on income levels. However, the literature to date lacks empirical evidence provided from an international setting, considering small as well as larger cities across countries with different income levels. Our paper aims to fill this gap using an international consistent definition of cities and comparable data across 153 cities of very different sizes, in 11 countries with significant income differences.

3. DATA AND STYLIZED FACTS:

The units of analysis of our work are metropolitan areas, which are defined as Functional Urban Areas (FUAs) of at least 500,000 inhabitants. The delineation of FUAs starts from the identification of urban centres through gridded data on resident population. An urban centre is a cluster of contiguous grid cells having at least 1,500 inhabitants per km^2 and a total population of at least 50,000 inhabitants. Subsequently, urban centres are adapted to the local (administrative or statistical) units (i.e. municipalities) for which commuting data is available. The "city" or "urban core" is made up of contiguous units that have more than 50% of their population living within an urban centre. After having defined the urban core,

[7] See Castells-Quintana and Rodriguez (2017) for an analysis of Colombian cities.

FUAs' commuting zone is made up of the surrounding local units that are linked to the urban core by the commuting pattern of their workforce. Any municipality with at least 15% of its employed residents working in a certain urban core is considered part of the same FUA. FUAs have the advantage of at least partially overcoming a list of problems associated with de *jure spatial* units (Stimson 2016) and to maximise comparability across countries and jurisdictions. FUAs have been used in several studies analysing different issues of urban development, including city size distribution (Schmidheiny and Suedekum 2015; Veneri 2016), urban spatial structure (Veneri 2018; OECD 2018), or agglomeration economies (Ahrend et al. 2017), among others.

Our key dependent variable is income inequality at the city (FUA) level. We rely on a study by Boulant et al. (2016), who estimated levels and distribution of household disposable income within FUAs with at least 500,000 inhabitants in 18 and 11 countries, respectively. Such estimations are based on available income data aggregated at very small spatial scale (i.e. census tracts, municipalities, etc.). Depending on the country, such information either comes from administrative records, such as tax records, or from surveys. Sources of data, income definition and building block units of analysis are reported in Annex 1 (Table A.1).[8] We rely on several measures. First, we use the Gini coefficient, as the most commonly used indicator of income distribution. Second, as our dataset includes information on the levels of income for different percentiles along the income distribution (except for Mexico), we complement with percentile ratios, as common in the inequality literature. The 90/10 percentile ratio compares the income of a rich individual (i.e., in 90th percentile of the income distribution) with that of a poor individual (i.e., in the 10th percentile of the

[8] The final income levels and inequality indicators were estimated by simulating the income distribution within each building block unit based on the assumption that such distribution followed a lognormal functional form, and using the available information on income quantiles to fit such distribution. The lognormal hypothesis for the distribution of income has been long tested in the literature (Balintfy and Goodman 1973; Diaz-Ramirez and Murtin 2016). After having generated this 'modelled' income distribution, inequality measures for each metropolitan area were computed with the standard approach.

distribution). We also calculate the 50/10 percentile ratio, which compares a median-income individual with a low-income one, and can therefore be understood as providing information of inequality in the bottom part of the distribution (i.e., among the relatively poor). Finally, we calculate the 90/50 percentile ratio, which compares a high-income individual with a median-income one, and can therefore be understood as providing information of inequality in the top part of the distribution (i.e., among the relatively rich).

Our data on inequality gives as a sample of 153 FUAs in 11 OECD countries.[9] Seven of these countries are European: Austria (with 3 FUAs in the sample), Belgium (4), Denmark (1), France (15), Italy (11), Norway (1), and Sweden (3); and four American countries: Canada (with 9 FUAs in the sample), Chile (3), Mexico (33), and USA (70). Thus, we work here with a sample where 25% of the FUAs are in Europe, 46% are in the USA, and the remaining 29% in other countries in the Americas. For these 153 FUAs, we have information for our inequality measures for at least two years between the years 2000 and 2014 (except for Mexican cities, for which we only have information for 2010).

Together with the inequality measures, our data set includes, for every city, information on size, GDP per capita, labour participation and the share of population over 65 years. All these indicators are publicly available from the OECD Metropolitan Database, and the full list of variables and their respective definition is reported in Table A.2 in the Appendix.[10] In addition, in some regressions we also control for the industrial structure of FUAs. However, at this spatial scale, comparable data at sectoral level is not available for all FUAs. Where data was missing, we considered the industrial structure of the smallest region

[9] Our 153 FUAs are all the FUAs above 500,000 inhabitants that the OECD has identified as such and for which a number of statistical indicators are made available. For more information on this see https://www.oecd.org/cfe/regional-policy/functionalurbanareasbycountry.htm

[10] GDP and labour participation indicators at FUA level are aggregated from official GDP and labour participation statistics at small regional level (i.e. NUTS3 in Europe). A population grid was used to adjust for the differences in the boundaries between FUAs and small regions based on the amount of population living within FUA borders. Other demographic indicators are directly aggregated from official statistics available at building block unit level (e.g. share of elderly population).

with available data in which the FUA is located.[11] Table A.3 provides the details of the data sources for industrial structure.

The countries in our sample are varied in several dimensions.[12] In terms of income inequality, according to 2015 World Bank estimates, Norway is the country with the lowest Gini index (27.5) and Chile with the highest (47.7). The average Gini coefficient of the seven European countries considered is 30.2, while the average of the four American countries is 41.6. In terms of income, Norway and the United States are the countries with the highest GDP per capita, while Chile and Mexico are the ones with the lowest. In terms of population size, the United States and Mexico are the largest countries, with more than 100 million inhabitants, while Norway and Denmark have less than 6 million each. Finally, in terms of average FUA size, and for the FUAs in our sample, cities in the Americas tend to be about 50% larger than European counterparts.

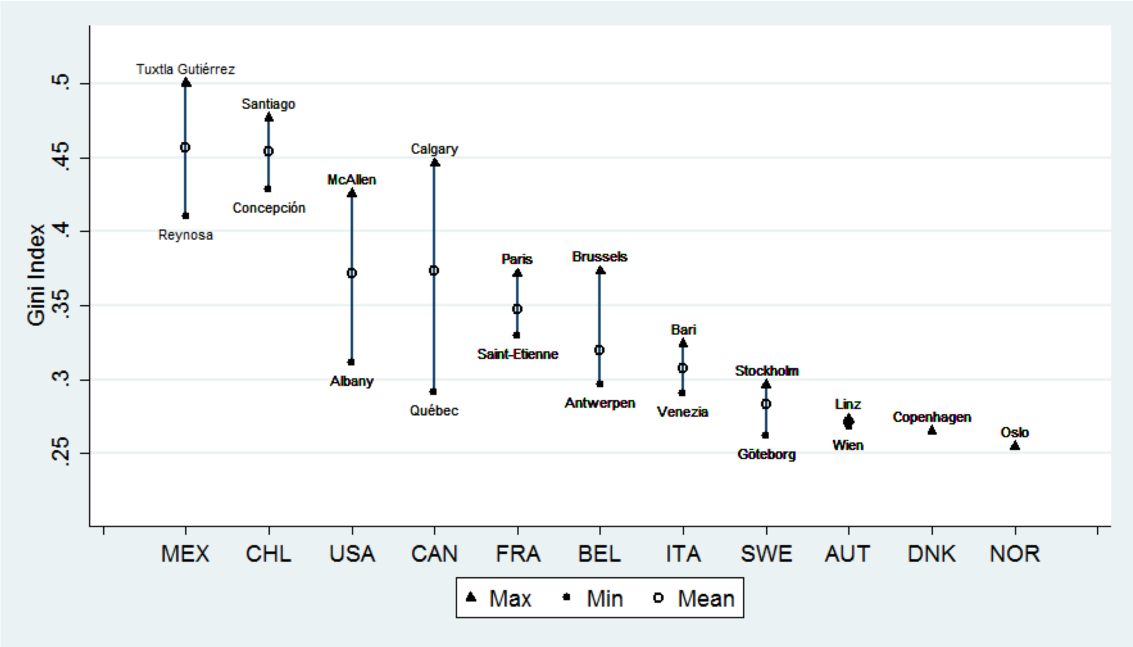
The levels of inequality (using the Gini index) for the 153 different cities in the 11 countries in our sample are displayed in Figure 1, while Table 1 presents the top and bottom cities in the sample according to the Gini index.[13] Clearly, Latin American cities are the ones with the highest inequality indices (as the top 10 unequal cities were always Mexican, the list includes other non-Mexican cities), while Nordic and Austrian cities are the ones with the lowest values. The largest differences among cities of the same country are found in Canada. Large diversity is also present among US cities. In any case, what can be seen in Figure 1 is that the range in inequality levels is much larger when one uses an international dataset like ours, rather than using a sample for an individual country, as previously done in the literature.

[11] We approximated the industrial structure at FUA level with that of the region in which the FUA is located for about half of the FUAs in our sample. For European countries, we considered the NUTS-3 geographical breakdown.

[12] Table OSM.1 in the Online Supplementary Material displays the descriptive statistics for the 11 countries in the sample.

[13] Figure OSM.1 in the Online Supplementary Material displays the distribution of the several inequality measures by country.

Figure 1. Inequality measures – geographical distribution



Note: The Gini index of every city corresponds to the average for all available periods.

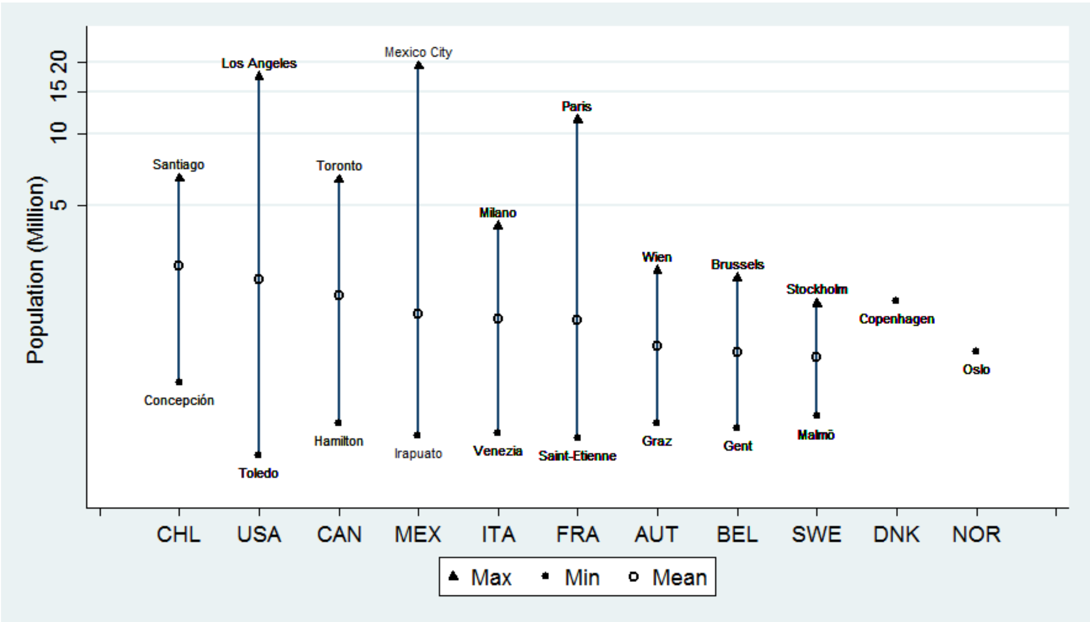
Table 1. Top and bottom inequality cities

	FUA	Year	Gini	Population	GDP pc
1	Tuxtla Gutiérrez (MEX)	2010	0.500	738,261	6,368
2	Oaxaca de Juárez (MEX)	2010	0.490	729,315	6,848
3	Querétaro (MEX)	2010	0.480	1,119,642	17,779
4	Celaya (MEX)	2010	0.480	602,045	11,478
5	Pachuca de Soto (MEX)	2010	0.480	546,513	10,031
11	Santiago (CHL)	2013	0.478	6,604,166	26,322
18	Valparaíso (CHL)	2009	0.458	941,725	14,926
19	Calgary (CAN)	2006	0.457	1,141,328	66,338
25	Concepción (CHL)	2009	0.442	880,061	12,002
35	Toronto (CAN)	2006	0.432	5,965,105	42,170
144	Québec (CAN)	2013	0.287	855,184	34,546
143	Catania (ITA)*	2013	0.285	628,886	24,443
142	Stockholm (SWE)	2000	0.277	1,838,377	49,628
141	Graz (AUT	2004	0.272	574,149	40,950
140	Wien (AUT)	2004	0.266	2,527,748	46,148
139	Linz (AUT)	2012	0.261	611,622	46,469
138	Malmö (SWE)	2000	0.260	609,424	32,272
137	Göteborg (SWE)	2000	0.241	826,126	35,035
136	Copenhagen (DNK)	2000	0.241	1,907,401	45,703
153	Oslo (NOR)*	2000	0.232	1,058,009	57,880

Note: * GDP pc is referred to 2008

In addition to the Gini index, Table 1 also displays the population and the GDP per capita of the reported cities. Figure 2 plots the distribution of population within countries.[14] We find large differences across the cities in our sample. The three largest cities are all in American countries: Mexico City, New York, and Los Angeles. With Paris, these four FUAs have each more than 10 million inhabitants. Above 5 million, we find Chicago, San Francisco, Toronto, Santiago de Chile, Houston, Miami, and Washington. Clearly, most large cities are in the American continent. In fact, the average size of cities in this continent is higher than the one in European cities, except for Copenhagen, which is the only Danish city in the sample.

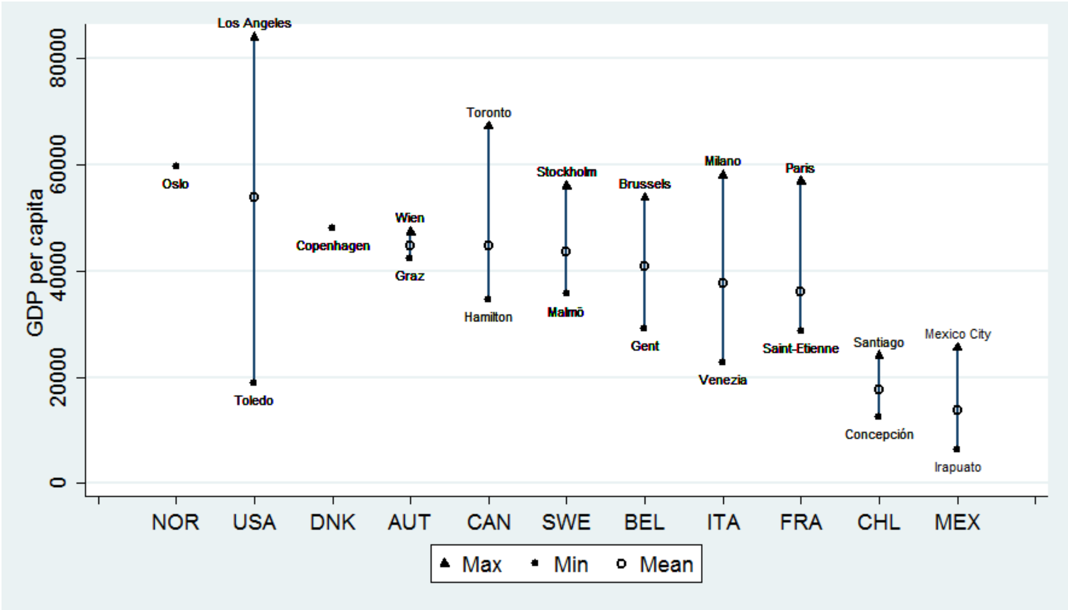
Figure 2. FUAs Population – geographical distribution



Note: FUAs population in the last available year in the sample. See table A1 in the appendix for corresponding year for every country. The vertical axis reports the logarithmic scale. The middle circle reports the country's average of city size.

[14] Figure OSM.2 in the Online Supplementary Material represents the histogram of the overall distribution of population (in logs) across our cities.

Figure 3. FUAs GDP per capita – geographical distribution



Note: FUAs' GDP per capita in the last available year in the sample. See Table A1 in the Appendix for corresponding year for every country. The middle circle reports the country's average of city income.

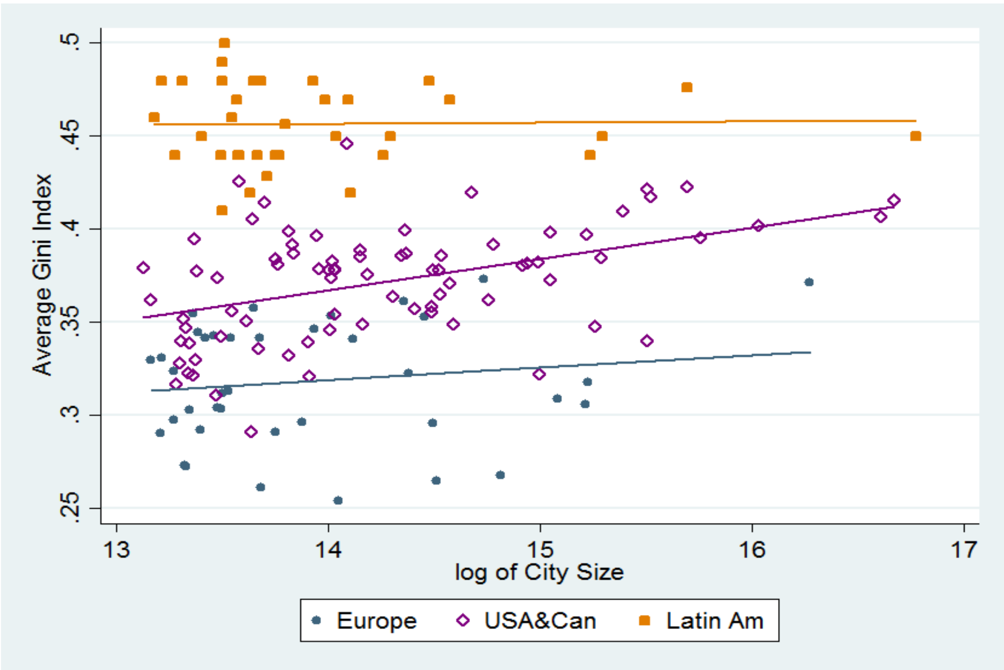
In terms of income, we also find important differences between countries. As expected, cities in Chile and Mexico are among those in the sample with the lowest GDP per capita, significantly lower than in most cities in other countries in the sample.[15] As can be seen in Figure 3, the average income of US cities is only behind Oslo, being the rest of European cities in middle positions, together with Canada.

Finally, we look at the correlation between city size and inequality for the 153 FUAs in our sample. Our sample is mostly composed of relatively large cities in rich countries. According to our theoretical framework (see Section 2), larger cities in richer countries are expected to display a positive association between city size and income inequality. Figure 4 displays the scatterplots between population size (in logs) and the Gini index, showing a positive association between inequality and city size. We have distinguished the association by geographical areas to show the different intensities of such association. Such disparities

[15] The average of the GDP per capita of the considered European countries plus US and Canada more than doubles the average income of Mexico and Chile.

are also displayed in Table 2, which includes our other inequality measures (i.e., the 90/10, 50/10 and 90/50 percentile ratios) and shows a significant and positive correlation in most cases. The strongest association is found for US and Canadian cities: as we have seen above, American cities are, on average, larger and more unequal than European cities. In fact, when we divide the sample into three similar groups of cities according to their size, we see that the correlation is positive and only robustly significant for those cities above 1.5 million inhabitants.[16] Interestingly, medium sized-cities also display a significant association with inequality at the bottom of the distribution (i.e., the 50/10 percentile ratio.)

Figure 4. Association between city size and inequality measures.



Note: the scatterplot represents the average city size and Gini index over all available periods for every city.

[16] Figure OSM.3 in the Online Supplementary Material displays the scatterplots between population size (in logs) and the Gini index, distinguishing by initial city size.

Table 2. Correlations between population size and inequality measures

	Gini Index	p90/10	p50/10	p90/50
Overall sample	**0.289**	**0.337**	**0.336**	**0.320**
Europe	**0.138**	**0.123**	*0.090*	**0.134**
USA & Canada	**0.461**	**0.350**	**0.387**	**0.337**
Chile & Mexico*	0.025	**0.878**	0.115	**0.833**
Small cities (below 0.75 M inh.) (43 cities)	0.039	-0.028	-0.005	-0.044
Medium-sized cities (between 0.75 & 1.5 M inh.) (56 cities)	0.039	0.098	**0.139**	0.070
Large cities (above 1.5 M inh. (54 cities)	**0.340**	**0.327**	**0.295**	**0.341**

Note: Mexico only has data for the Gini index. Europe accounts 38 FUAs, USA and Canada 79, while Chile and Mexico have 36. There is no data available for the income ratios of the 33 Mexican cities. Bolded figures correspond to correlations significant at 5%, while cursive numbers represent significance at 10%.

All in all, our descriptive analysis highlights two main facts. First, there are large differences in size, income, and income inequality between the cities in our sample. Although a large share of these differences is observed between countries, there are also important differences within countries; in fact, differences between countries account for just the 50% of the total variation in urban GDP per capita in our sample. For instance, while the GDP per capita in Seattle in 2013 was higher than 86 thousand US dollars, in another city in the USA, like McAllen, it was just around 18 thousand dollars, and in the Mexican city of Tuxla Gutiérrez it was just above 6 thousand US dollars. And second, there is a positive association between city size and income inequality at the city level. Larger cities have, on average, a higher Gini index. This association is stronger among North American cities, among larger cities (above 1.5 million inhabitants), and for inequality in the top part of the income distribution. Our next step is to investigate this association further, testing whether it holds after controlling for other factors, looking across different geographical areas, and using different econometric techniques.

4. INEQUALITY AND CITY SIZE: AN EMPIRICAL ANALYSIS

As described in Section 2, the relationship between city size and income inequality at city level is not straightforward and therefore requires empirical research. Moreover, heterogeneities based on several characteristics of cities, like their initial size, income levels, and other contextual factors, may play a relevant role in the size-inequality relationship. In this section, we empirically explore the relationship between city size and city-level inequality, using our sample of 153 comparable FUAs and several estimation techniques.

Econometric Analysis for OECD FUAs

To econometrically test the relationship between city size and income inequality we run some reduced-form regressions, like those in equation (1), and in line with previous papers studying this relationship for single countries:

$$inequality_{it} = \alpha_1 income_{it-1} + \beta Pop_{it-1} + \psi X_{i\,t-1} + \varepsilon_{it} \qquad (1)$$

Where $inequality_{it}$ is income inequality in city i in time t, $income_{it-1}$ is income per capita (in logs) in $t-1$, $X_{i\,t-1}$ is a vector including time–variant factors potentially influencing income inequality beyond income per capita (following the literature, we consider labour participation rates, the share of population over 65 years old, and the share of population between 25 and 64 years old with tertiary education). Finally, ε_{it} is a city-time specific shock. Our variable of interest is Pop_{it-1}, measured as the total population (in logs) for each considered FUA-year observation.

Equation (1) is estimated using our (unbalanced) panel, considering as many cities as possible (up to 153 in main estimations) and the longest time span for every city (from 2000 to 2014). Right-hand-side variables are included one period before to reduce problems of reverse causality. Time fixed effects are included to control for global shocks. As our dataset includes cities for 11 different countries, we are also able to include country-specific fixed effects to control for time-invariant country-specific characteristics. This gives our estimations a clear advantage over previous papers relying on data for one single country. Finally, in some estimations, we also include city-specific random effects. All our panel estimations report robust standard errors clustered at the city level.

Table 3 presents our main results. Results show a positive and highly significant association between city size (Pop) and income inequality ($Gini$). The significance of this association holds when we control for income at the city level and year fixed effects (column 2). Interestingly, the significance of the coefficient for city size also holds when we introduce country fixed effects (column 3), suggesting that the positive association with inequality cannot be fully explained by differences across countries. The coefficient for city size remains significant even when we further introduce city-specific random effects (column 4).[17]

Table 3. Main results

	(1)	(2)	(3)	(4)	(5)
Dependent variable: *Inequality* (Gini Coefficient)					
Log(pop)	0.0155***	0.0166***	0.0131***	0.0114***	-0.0855***
	(0.0044)	(0.0029)	(0.0027)	(0.0028)	(0.0299)
Log*(income)*		-0.0540***	0.0005	-0.0062	-0.1434***
		(0.0041)	(0.0222)	(0.0183)	(0.0483)
*Log(pop)*Log(income)*					0.0095***
					(0.0030)
Year FE	NO	YES	YES	YES	YES
Country FE	NO	NO	YES	YES	YES
City (Random) Effects	NO	NO	NO	YES	YES
Observations	301	300	300	300	300
No. of cities	153	153	153	153	153
R-Square	0.052	0.688	0.801	0.807	0.812

Note: The time span goes from 2000 to 2014. Robust standard errors (clustered by city) in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

[17] We also performed more demanding estimations with city-specific fixed effects (see Table OSM.2). Fixed effects allow us to control for city-specific characteristics that are time-invariant or that change over time but very slowly, like the industrial structure of cities. The coefficient for city size remains positive and significant. However, a fixed-effects estimation relies on variation over time, which in our dataset is very limited, and therefore lowers the possibility of more flexible specifications (i.e., more controls, non-linear specifications, etc.). Our dataset covers only a short time span for every city: for inequality, the between-cities variance in our dataset represents on average 8 times the within-city variance. For population, the between/within variance ratio is over 30.

Our main results suggest a statistically significant association between city size and income inequality at the city level. The magnitude of the coefficients suggests that the association is also economically significant (although these results should be taken with caution). According to our coefficients, doubling city size is associated with an increase of around 1 percentage point in income inequality (measured by the Gini coefficient). This magnitude is similar to that found in previous papers for single countries (see for instance Lee et al., 2016, who report an elasticity between 0.7% and 1.2% for 60 cities in the UK, or Florida and Mellander, 2016, who report an elasticity of 1.4% looking at metropolitan areas in the USA).

Our estimates also suggest that, in our sample, the inequality-increasing dynamics of city size dominate over inequality-reducing ones. However, this may be because our sample is dominated by relatively rich cities. In the study for British cities of Lee et al. (2016), when controlling for median wage, inequality loses significance, suggesting that "higher median wages in large cities makes them more unequal – the presence of more skilled and better-paid workers drives inequality" (Lee et al 2016, p. 1723). As already mentioned, in low-income cities the size-inequality association may be negative. To capture the role of income levels, in column 5 of Table 3, we interact city size with city income. While the coefficients for size and income are both negative and significant, it is the interaction between the two that yield a positive and significant coefficient. This suggests that it is only for richer cities (above an income per capita of around 20 thousand 2010-PPP US dollars) that the size-inequality relationship is positive and significant.[18] We further explore this in next section by looking at heterogeneity across world regions and initial city size.

[18] Figure A.1 in the Appendix shows the marginal effect of city size for different values of initial income per capita.

Results by world region and by initial city size

A key advantage of our data set is that it considers metropolitan areas from several countries, which provides a source of variation in income levels and in several other city and country characteristics. Consequently, it is only natural to explore whether the city size-inequality relationship changes for different groups of countries. In column 1 of Table 4, we let the coefficient for size to vary across three clearly distinct geographical areas: i) North America (i.e., USA and Canada), ii) Europe, and iii) Latin America (i.e., Chile and Mexico). The coefficient for size is only positive and significant for cities in USA and Canada and for cities in Europe. By contrast, for cities in Chile and Mexico, where income levels are significantly lower (below the threshold found in column 5 of Table 3), the coefficient is non-significant. This is in line with, and in fact explains, the positive and significant coefficient for our interaction between income and size in column 5 of Table 3. Being the size-inequality clearly different in cities in Chile and Mexico, in column 2 of Table 4 we exclude them from the analysis. As expected, our coefficient for size is positive and significant. A positive coefficient for Europe in column 1 also means that our results in Table 3 are not driven by the USA. This is interesting, as most papers to date have focused on the USA. However, the coefficient for city size is higher in magnitude when we consider only cities in the USA (column 3).

In a similar spirit of that of the analysis by world region, it seems natural to explore heterogeneity of our results distinguishing between small, medium-sized, and large cities. Increasing city size is expected to lead to higher levels of income inequality when cities are already large, but not necessarily when cities are small. In this line, in column 4 of Table 4, we let the coefficient for size to vary for cities of different initial city size. We distinguish between cities with less than 750 thousand inhabitants (44 cities of 153 in our sample), cities with a population between 750 thousand and 1.5 million (56 cities), and cities with more than 1.5 million (53 cities). We find that the coefficient for city size is non-significant for cities below 1.5 million inhabitants, and positive and significant for cities above that threshold. In other words, and as expected, increasing city size is associated with higher inequality when cities are already large.

Table 4. Results by world region and initial city size

	(1)	(2)	(3)	(4)
Dependent variable: *Inequality* (Gini Coefficient)				
Log(pop)		0.0151***	0.0200***	
*USA-CANDA	0.0164***			
*EUROPE	0.0097***			
*LATAM	-0.0005			
*Log(pop)**				
pop<750				-0.0106
750<pop<1.5M				0.0143
pop>1.5				0.0127***
Year FE	YES	YES	YES	YES
Country FE	YES	YES	YES	YES
City (Random) Effects	YES	YES	YES	YES
Observations	300	261	140	300
No. of cities	153	117	70	153
R-Square	0.813	0.684	0.301	0.807

Note: The time span goes from 2000 to 2014. Log(*income*) is included as control. Column 2 exclude Mexican and Chilean cities. Column 3 considers only cities in the USA. Column 4 includes all cities, plus dummies for small- and medium-sized cities. Robust standard errors (clustered by city) not shown for simplicity. *** p<0.01, ** p<0.05, * p<0.1

Other inequality measures

We now look at percentile ratios. This allows us to study the size-inequality relationship at different segments of the income distribution, with the 90/50 ratio capturing inequality among the rich and the 50/10 capturing inequality among the low-income residents. Results are presented in Table 5. According the estimates in columns 1 to 3, the positive association between city size and income inequality is significant along the whole income distribution. Following findings in Table 4, in columns 4 to 6 of Table 5, we let the coefficient for size to vary for initially small, medium-sized, and large cities. For small cities, we find non-significant coefficients. For medium-sized cities, we find positive and significant coefficients for the log of population when we use either the 90/10 or the 50/10 ratio as dependent variable, but not when we use the 90/50 ratio. By contrast, for large cities, we also find a positive and significant coefficient when we use the 90/50 ratio, with a parameter actually higher than the one for the 50/10 ratio. This suggests that as medium-sized cities grow, inequality increases mainly in the bottom part of the income distribution, or, in other words, among those with lower incomes. By contrast, as larger cities grow, inequality increases both in the bottom but mainly in the top part of the income distribution: among the relatively rich. This is in line with the idea in the urban economics literature describing how agglomeration economies in large cities disproportionally benefit high-skill, high-income urban residents, which in turns increases the sorting of high-skilled workers from smaller to larger cities (see for instance Behrens and Robert-Nicoud 2014).[19]

Table 5. Other inequality measures

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	p90/10	p50/10	p90/50	p90/10	p50/10	p90/50
Log(pop)	0.4652***	0.0998***	0.0845***			
pop<750				-0.8025	-0.2127	-0.0459
750<pop<1.5M				1.0174**	0.3010**	0.0060
pop>1.5				0.4483***	0.0659**	0.1240***
Year FE	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	YES	YES
City (Random) Effects	YES	YES	YES	YES	YES	YES
Observations	267	267	267	267	267	267
No. of cities	120	120	120	120	120	120
R-Square	0.764	0.732	0.803	0.767	0.735	0.801

Note: Log(*income*) is included as control. The time span goes from 2000 to 2014. Columns 4 to 6 include dummies for small- and medium-sized cities. Robust standard errors (clustered by city) not shown for simplicity.
*** p<0.01, ** p<0.05, * p<0.1

Confounding factors and different lag structures

In Table 6 and Table OSM.3 in the Online Supplementary Material, we explore the robustness of our results. First, we do so by introducing more time-variant city-specific variables that may help us control for omitted time-variant characteristics of cities. In column 1 of Table 6, we start by controlling for potential confounding factors that the literature has suggested as relevant to explain inequality in cities. We consider labour participation rates

[19] In columns 4 to 6 of Table 5, we include dummies for medium-sized and large cities. These dummies show that controlling for time and country fixed effects, medium-sized cities have on average lower levels of inequality than smaller or larger counterparts.

and the share of population over 65 years. Higher labour participation rates are significantly associated with lower inequality, as expected. We also consider the share of people between 25 and 64 years old with tertiary education, but at the expense of losing observations, and the industry composition of cities, looking at the share to GDP of industry and construction, public services and private services (see Table OSM.3 in the Online Supplementary Material). Larger cities have, on average, a higher share of highly educated people and a higher share of private services. Nevertheless, even controlling for all these time-varying city-specific factors, our key coefficient for city size remains positive and significant. Following a comment from a referee, we believe it is worth to signal the parameter for education. When regressing inequality on the share of highly educated people we find a positive and significant coefficient, as expected from the literature. However, when we control for city size and income levels, the coefficient for high education becomes negative, what reinforces our hypothesis of the role of city size in explaining inequality at the city level. Second, in column 2 of Table 6, we test our results to lagging our Right-Hand-Side (RHS) variables one more period. Although lagging RHS variables comes at the expense of reducing our sample size, our coefficient for size remains positive and significant. Third, in column 3 of Table 6, we consider a potential quadratic effect of income. Interestingly, we find a positive coefficient for income per capita and a negative for its square, being both highly significant, suggesting an inverted-U relationship. This is in line with the traditional Kuznets' hypothesis, stating that as income per capita increases, inequality first increases and then declines. This inverted-U relationship between income and inequality has been traditionally reported for a global panel of countries as well as for a panel of European regions (Castells-Quintana et al. 2015). To the best of our knowledge, this is the first time that is reported for a panel of cities.

Instrumental Variables (IV) estimations

Our results so far highlight an interesting and relevant association between city size and income inequality. This association, though, does not necessarily imply a causal effect of city size on inequality. Our estimates may suffer from reverse causality and omitted variables explaining at the same time city size and income inequality.[20] To ameliorate these endogeneity concerns, in some of our estimations we have included country-specific fixed effects and controlled for several time-variant city-specific controls like income levels, age structure, labour participation rates, and share of highly educated people. We have also tested for different lag structures. Finally, to further reduce endogeneity problems, we perform Instrumental Variables (IV) estimations. Finding valid time-variant instruments for city size is not an easy task. For instruments to be valid they should not only be relevant (that is, explain city size) but also exogenous and affect inequality only through city size (the exclusion restriction). One potential option according to recent papers is to rely on a cross-sectional specification and use historical population data as instrument.[21] Using a cross-section specification can also give a different insight to our analysis, as inequality is fairly stable over time. Column 4 in Table 6 presents cross-section OLS results, showing a positive and significant coefficient for city size.[22] Columns 5 presents our IV estimates instrumenting city-population size in recent years with population size *circa* 1870, using data from Mitchell (2013). According to first stage results, population size *circa* 1870 is relevant to explain current population of cities in our sample.[23] IV estimates reinforce the idea of a positive (and

[20] In any case, our main goal is not to show a causal effect of city size on income inequality; the association in its own is interesting and relevant for policy debate, beyond causality concerns.

[21] See for instance Duranton (2015), Castells-Quintana (2018) and Chen et al. (2018). "Migrants tend to move to large cities for jobs and higher incomes. The historical population of each city (…) is, to a great extent, exogenous to current inequality" (Chen et al., 2018, p. 50).

[22] For city size, we use the first available value in our data set. For inequality, we use the average Gini coefficient considering all periods in our data set.

[23] Table A.4 in the Appendix show results from first stage of column 5-Table 6. Under-identification tests are reported in Table 6, reinforcing the validity of historical population data as instrument for current city size.

significant) association between city size and inequality in relatively rich and large (OECD) metropolitan areas.[24]

Table 6. Robustness checks

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Gini	Gini	Gini	Average Gini	Average Gini
Log(*pop*)	0.0113***	0.0135***	0.0164***	0.0103**	0.0090**
	(0.0028)	(0.003)	(0.0028)	(0.0044)	(0.0039)
Log(*income*)	0.0078	0.0447	2.3965***	0.0485	0.0703**
	(0.0209)	(0.0274)	(0.7839)	(0.0349)	(0.0277)
Log*(income)*^{2}			-0.1135***		
			(0.0379)		
Year FE	YES	YES	YES	NO	NO
Country FE	YES	YES	YES	YES	YES
City (Random) Effects	YES	YES	YES	NO	NO
Controls	YES	YES	YES	YES	YES
Observations	299	143	143	150	94
No. of cities	153	117	117	150	94
R-Square	0.809	0.707	0.729	0.83	0.77
K-P LM stat					13.99***

Note: Controls include *labour participation rate* and the *share of population over 65*. Columns 1 to 3 are estimated using panel data; the time span goes from 2000 to 2014. In columns 2 and 3 right-hand-side variables are lagged one period and exclude cities in Mexico and Chile. Columns 4 and 5 are estimated using cross-section data using, for each city, the average Gini for all periods in our dataset as dependent variable, and the earliest available period for population. Column 4 is estimated by OLS, while column 5 presents an instrumental variables estimation using historical data as instrument for city-population. Robust standard errors (clustered by city in columns 1 to 3) in parentheses. *** p<0.01, ** p<0.05, * p<0.1

5. DISCUSSION, CONCLUSIONS AND POLICY IMPLICATIONS

In this paper, we have studied the relationship between city size and income inequality at city level. To do so, we have used novel panel data at the Functional Urban Areas (FUA) level for more than 150 cities in 11 OECD countries. We have implemented several estimation techniques (from pooled-OLS to IV estimations), controlled for several relevant factors in the size-inequality relationship, and studied the heterogeneity of the results across different initial characteristics of cities.

[24] We also performed IV estimations splitting the sample by initial city size (not reported here). While we find a positive and significant association for cities above 750 thousand inhabitants, we found a non-significant coefficient for cities below that threshold, in line with our results in Table 4.

Our results are in line with previous empirical evidence for individual countries, suggesting that inequality tends to increase with city size: according to most estimates, as a city doubles in size, its Gini index is expected to grow by around 1 percentage point. Our results are robust to different estimation techniques, different measures and the inclusion of several controls, including average income, education levels, demographic factors and industry composition. As our data includes cities from different countries, we are also able to find that the positive association between city size and income inequality is significant only in relatively rich countries, and in large cities. In particular, we find that larger city size is associated with higher inequality in cities with an income per capita level above around 20 thousand US dollars (2010-PPP) and an initial population of 1.5 million inhabitants. We can connect these findings with the theoretical insights in the literature, reviewed in Section 2: large-rich cities tend to have a widening of the distribution of skills, reinforce monopoly rents, and allow for higher returns to the high-skilled. Looking at the size-inequality association at different parts of the income distribution, we find that while it is mainly inequality among those with lower incomes which increases as medium-sized cities grow, for larger cities the positive association is mainly driven by inequality among the relatively rich. This is in line with previous findings suggesting that larger cities tend to disproportionally concentrate high-income earners.

Our results, together with previous findings in the literature, suggest that while city growth may be desirable when cities are small, as it allows for better economic performance, the increasing growth that some large cities experience may come at some cost. Excessive city size not only may lead to congestion diseconomies, which reduce economic performance (see for instance Frick and Rodriguez-Pose 2018), but also to higher inequalities and the risk of less cohesive societies. As high and increasing inequalities continue to be a major challenge for large cities, better understanding of inequality dynamics within cities can be of great value. In this regard, further research could investigate in more detail the role of structural change and sectoral composition (for instance following the literature on job polarization - Autor and Dorn 2013), or take a deeper look at how different city characteristics, like for instance the fragmentation of governance or the quality of urban infrastructure, can alleviate the rise of inequality that large cities experience.

REFERENCES

Ahrend R, Farchy E, Kaplanis I, Lembcke A (2017) What makes cities more productive? Evidence from five OECD countries on the role of urban governance. *Journal of Regional Science* 57(3): 385-410.

Alperovich G (1995) The relationship between income inequality and city size: a general equilibrium model of an open system of cities approach. *Urban Studies* 32(6): 853-862.

Autor D, Dorn D (2013) The Growth of Low-Skill Service Jobs and the Polarization of the U.S. Labor Market. *American Economic Review* 103(5): 1553–97.

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

Baum-Snow N, Pavan R (2013) Inequality and city size. *The Review of Economics and Statistics* 95(5): 1535-1548.

Behrens K, Robert-Nicoud F (2014) Survival of the fittest in cities: Urbanisation and inequality. *The Economic Journal* 124(581): 1371-1400.

Bettencourt L, Lobo J, Helbing D, Kühnert C, West G (2007) Growth, innovation, scaling, and the pace of life in cities. *PNAS* 104(17): 7301-7306.

Bettencourt L, Lobo J, Strumsky D, West G (2010) Urban scaling and its deviation: Revealing the structure of wealth, innovation and crime across cities. *PLOS ONE* 5(11): e13541.

Boulant J, Brezzi M, Veneri P (2016) Income Levels and Inequality in Metropolitan Areas: A Comparative Approach in OECD Countries, OECD Regional Development Working Papers, 2016/06, OECD Publishing, Paris. http://dx.doi.org/10.1787/5jlwj02zz4mr-en

Castells-Quintana D (2018) Beyond Kuznets: Inequality and the size and distribution of cities. *Journal of Regional Science* 58(3): 564-580.

Castells-Quintana D, Royuela V (2014) Agglomeration, inequality and economic growth. *Annals of Regional Science* 52(2): 343-366.

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

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

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

Duncan O, Reiss A (1956) *Social Characteristics of Urban and Rural Communities, 1950*. John Wiley and Sons, New York.

Duranton G (2015) Agglomeration effects in Colombia. *Journal of Regional Science* 56(2): 210-238.

Duranton G, Puga D (2004) Micro-foundations of urban agglomeration economies. In Henderson, V. and J. Thisse (eds.) *Handbook of Regional and Urban Economics*, vol. 4. Amsterdam: North-Holland. pp. 2063-2117.

Duranton G, Puga D (2013) The growth of cities. In Henderson, V. and W. Strange (eds.) *Handbook of Economic Growth*, vol. 2A. Amsterdam: North-Holland. pp. 781-853.

Farbman M (1975) The size distribution of family income in U.S. SMSAs. *The Review of Income and Wealth* 2: 217-237.

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

Frech H, Burns L (1971) Metropolitan interpersonal income inequality: A comment. *Land Economics* 47(Feb. 1971): 104-106.

Frick S, Rodríguez-Pose A (2018) Big or small cities? On city size and economic growth. *Growth and Change* 49(1): 4-32.

Gabaix X (1999) Zipf's law for cities: an explanation. *Quarterly Journal of Economics* 114(3): 739-767.

Glaeser E, Resseger M, Tobio K (2015) Inequality in cities. *Journal of Regional Science* 49(4): 617-646.

Haworth C, Long J, Rasmussen D (1978) Income distribution, city size, and urban growth. *Urban Studies* 15(1): 1-7.

Henderson V (2010) Cities and Development. *Journal of Regional Science* 50(1): 515-540.

Holmes N, Berube A (2016) City and metropolitan inequality on the rise, driven by declining incomes. In Brookings website: https://www.brookings.edu/research/city-and-metropolitan-inequality-on-the-rise-driven-by-declining-incomes/

Hortas-Rico M, Ríos V (2019) The Drivers of Local Income Inequality: A Spatial Bayesian Model Averaging Approach. *Regional Studies* (forthcoming).

Kuznets S (1955) Economic Growth and Income Inequality. *American Economic Review* 45: 1–28.

Lee N, Sissons P, Jones K (2016) The Geography of Wage Inequality in British Cities. *Regional Studies* 50(10): 1714-1727.

Levernier W, Partridge M, Rickman D (1998) Differences in metropolitan and non-metropolitan U.S. family income inequality: A cross-county comparison. *Journal of Urban Economics* 44(2): 272–290.

Lewis A (1954) Economic development with unlimited supplies of labor. *Manchester School of Economics and Social Studies* 22: 139-191.

Long J, Rasmussen D, Haworth C (1977) Income inequality and city size. *The Review of Economics and Statistics* 59(2): 244-246.

Ma L, Tang Y (2016) A tale of two tails: wage inequality and city size.

Mathur V (1970) Occupational composition and and its determinants: An inter-city class analysis. *Journal of Regional Science* 10: 81-91.

Mitchell B (2013) International Historical Statistics. Edited by Palgrave Macmillan Ltd.

Nord S (1980) An Empirical Analysis of Income inequality and city size. *Southern Economic Journal* 46(3): 863-872.

OECD (2018) *Rethinking urban sprawl. Moving towards sustainable cities.* OECD Publishing, Paris.

OECD (2012) *Redefining "urban". A new way to measure metropolitan areas*. OECD Publishing, Paris.

Perugini C, Martino G (2008) Income inequality within European regions: Determinants and effects on growth. *Review of Income and Wealth* 54(3): 373-406.

Ravallion M, Chen S (2007) China's (uneven) progress against poverty. *Journal of Development Economics,*82: 1-42.

Richardson H (1973) *The Economics of Urban Size*. Saxon House: Westmead.

Rodriguez-Pose A, Tselios V (2009) Education and income inequality in the regions of the European Union. *Journal of Regional Science* 49: 411-437.

Roine J, Vlachos J, Waldenstrom D (2009) The long-run determinants of inequality: what can we learn from top income data? *Journal of Public Economics* 93: 974-988.

Royuela V, Veneri P, Ramos R (2019) The short-run relationship between inequality and growth: evidence from OECD regions during the Great Recession. *Regional Studies* 53(4): 574-586.

Sarkar S (2018) Urban scaling and the geographic concentration of inequalities by city size. *Environment and Planning B: Urban Analytics and City Science* (published online April 2, 2018).

Sarkar S, Phibbs P, Simpson R, Wasnik S (2016) The scaling of income distribution in Australia: possible relationships between urban allometry, city size, and economic inequality. *Urban Studies*. DOI: 10.1177/0265813516676488

Schläpfer M, Bettencourt L, Grauwin S, Raschke M, Claxton R, Smoreda Z, West G, Ratti C (2014) The scaling of human interactions with city size. *Journal of the Royal Society Interface* 11(98): 20130789.

Schmidheiny K, Suedekum J (2015) The pan-European population distribution across consistently defined functional urban areas. *Economic Letters* 133: 10-13.

Stewart F (2005). *Horizontal Inequalities: A Neglected Dimension of Development*. Palgrave Macmillan UK, London, Ch. 5, pp. 101–135.

Stimson R (2016) Some Challenges for Regional Science Research. *Investigaciones Regionales* 36: 11-34.

Todaro M (1969) A model of labor migration and urban unemployment in less developed countries. *American Economic Review* 59: 138-148.

Todaro M (1976) Urban job creation, induced migration and rising unemployment: A formula and simplified empirical test for LDCs. *Journal of Development Economics* 3: 211-226.

Tselios V (2008) Income and educational inequalities in the regions of the European Union: geographical spillovers under welfare state restrictions. *Papers in Regional Science* 87: 403-430.

Tselios V (2014) The Granger-causality between income and educational inequality: a spatial cross-regressive VAR framework. *Annals of Regional Science*, 53: 221-243.

United Nations, Department of Economic and Social Affairs, Population Division (2015) World Urbanisation Prospects: The 2014 Revision. UN DESA Press.

Veneri P (2016) City size distribution across the OECD: Does the definition of cities matter? *Computers, Environment and Urban Systems* 59: 86–94.

Veneri P (2018) Urban spatial structure in OECD cities: Is urban population decentralising or clustering? *Papers in Regional Science* 97(4): 1355-1374.

Weeks J, Hill A, Stow D, Getis A, Fugate D (2006) The impact of neighbourhood structure on health inequality in Accra, Ghana. In *Population Association of America 2006 annual meeting program,* Los Angeles, USA, pp. 1-34.

Williamson J (1965). Regional inequality and the process of national development: A description of the patterns. *Economic Development and Cultural Change* 13 (4, Part 2): 1–84.

World Bank (2006) World Development Report 2006: Equity and development. The World Bank, Washington.

APPENDIX

Table A.1 Definition, sources and building block for measures of income inequalities in FUAs

Country	Source	Link	Original definition of income
Austria	Statistics Austria	Provided by Statistics Austria	Net income = total income including transfer payments – tax paid.
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.
Canada	Statistics Canada	Provided by Statistics Canada	Household disposable income.
Chile	CASEN – Ministry of Social Development	http://www.ministeriodesarrollosocial.gob.cl/basededatoscasen.php	Household disposable income.
Denmark	Statistics Denmark	http://statistikbanken.dk/statbank5a/default.asp?w=1920	Disposable income per fiscal household excluding imputed rent.
France	Insee	http://www.insee.fr/fr/bases-de-donnees/default.asp?page=statistiques-locales/revenu-niveau-vie.htm	Total taxable income from fiscal declarations.
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.
Mexico	CONEVAL	Provided by INEGI	Household total income (monetary and non-monetary income)
Norway	Statistics Norway	https://www.ssb.no/en/statistikkbanken	Ordinary income after special deductions is the equivalent of net income.
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.
US	American Community Survey	http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t#none	Total income.

Source: adapted from Boulant et al. (2016)

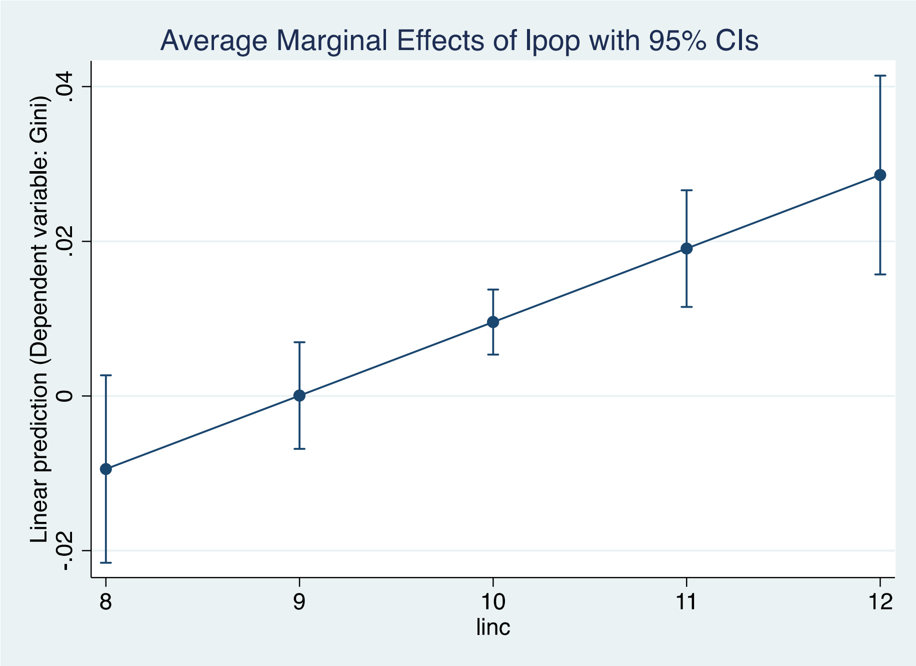
Table A.2. Variable names, definitions and sources

Variable	Description	Years	Method/notes	Source
pop	Total resident population in functional urban areas.	2000-16	Aggregation of demographic data at small geographical level from local administrative sources.	OECD Metropolitan Database http://dx.doi.org/10.1787/data-00531-en
GDP pc	GDP per capita (USD, constant 2010 prices, PPP).	2000-16	Adaptation of regional GDP data from national statistical institutes to FUA boundaries, based on population grids at 1km2 level.	OECD Metropolitan Database http://dx.doi.org/10.1787/data-00531-en
income	Equivalised household disposable income (USD, constant 2010 prices, PPP).	2008-15	Computations based on micro-aggregated data on income. Details in Boulant et al. (2016).	OECD Metropolitan Database http://dx.doi.org/10.1787/data-00531-en
lab_part	Total labour force over working age population (15-64 years old) (%).	2000-16	Adaptation of regional GDP data from national statistical institutes to FUA boundaries, based on population grids at 1km2 level.	OECD Metropolitan Database http://dx.doi.org/10.1787/data-00531-en
Gini	Gini coefficient for equivalised disposable household income.	2008-15	Computations based on simulated income distributions within each FUA, based on available income quintiles and with the hypothesis of a lognormal income distribution. Details in Boulant et al. (2016).	OECD Metropolitan Database http://dx.doi.org/10.1787/data-00531-en
p90/10	90/10 percentile ratio		Calculated using the 90th and 10th percentile in the income distribution.	Constructed using OECD Metropolitan Database
p50/10	90/10 percentile ratio		Calculated using the 50th and 10th percentile in the income distribution.	Constructed using OECD Metropolitan Database
p90/50	90/10 percentile ratio		Calculated using the 90th and 50th percentile in the income distribution.	Constructed using OECD Metropolitan Database
high_edu	Share of pop between 24 and 65 years old with tertiary education.	2000-16	Aggregation of demographic data at small geographical level from local administrative sources	OECD Metropolitan Database
over65	Share of elderly population (over 65 years old) over total resident population.	2000-16	Aggregation of demographic data at small geographical level from local administrative sources.	OECD Metropolitan Database http://dx.doi.org/10.1787/data-00531-en
pop_circa1870	Population in the metropolitan area in 1870 or closest year.		Estimated using historical data for cities and merging those cities belonging today to the same metro area.	Mitchel (2013).

Table A.3. Sectoral composition of FUAs

Country	Spatial scale and available years	Primary source
Austria	NUTS3 (2000-17)	Statistics Austria, Regional Accounts. http://www.statistik.at/web_en/statistics/Economy/national_accounts/regional_accounts/nuts3-regional_gdp_and_main_aggregates/index.html., data accessed the 12th of December 2019.
Belgium	NUTS3 (2000-17)	Eurostat, table nama_10r_3empers (https://appsso.eurostat.ec.europa.eu), data accessed the 30th of July 2019
Canada	Economic areas (2001-18)	Statistics Canada. CANSIM database. Employment by industry, annual, provinces and economic regions - Table 14-10-0092-01 https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410009201, data accessed the 12th of December 2019
Chile	Regions (2013-16)	INE, from National Labour Force Survey, ENE. (Household Survey), data collected in March 2019. http://www.ine.cl/canales/chile_estadistico/mercado_del_trabajo/empleo/series_estadisticas/rama.php
Denmark	NUTS3 (2000-17)	Eurostat, table nama_10r_3empers (https://appsso.eurostat.ec.europa.eu), data accessed the 30th of July 2019
France	NUTS3 (2000-16)	Eurostat, table nama_10r_3empers (https://appsso.eurostat.ec.europa.eu), data accessed the 30th of July 2019
Italy	NUTS3 (2000-16)	Eurostat, table nama_10r_3empers (https://appsso.eurostat.ec.europa.eu), data accessed the 30th of July 2019
Mexico	States (TL2)	INEGI. Encuesta Nacional de Ocupación y Empleo (ENOE), data collected in March 2019.
Norway	NUTS3 (2008-16)	Eurostat, table nama_10r_3empers (https://appsso.eurostat.ec.europa.eu), data accessed the 30th of July 2019
Sweden	TL3 (2000-16)	Eurostat, table nama_10r_3empers (https://appsso.eurostat.ec.europa.eu), data accessed the 30th of July 2019
United States	FUA (2001-18)	Employment by County (https://www.bea.gov/data/employment/employment-county-metro-and-other-areas), data accessed the 12th of December 2019.

Figure A.1. Marginal effect of population size by income levels



Note: marginal effect of log(pop) by levels in log(income), using estimates in column 5 of Table 3.

Table A.4. First-stage results for column 5 of Table 6

	(1)
Dependent variable:	Log(*pop*)
Log(*pop circa 1870*)*USA-Canada	0.1864***
	(0.1121)
Log(*pop circa 1870*)*Europe	0.9211***
	(0.0720)
Log(*pop circa 1870*)*LATAM	1.1208***
	(0.1870)
Country FE	YES
Controls	YES
Observations	94
No. of cities	94
F test of excluded instruments	72.70***

Note: Controls include log(income), labour participation rate, and the share of population over 65. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Supplementary Material

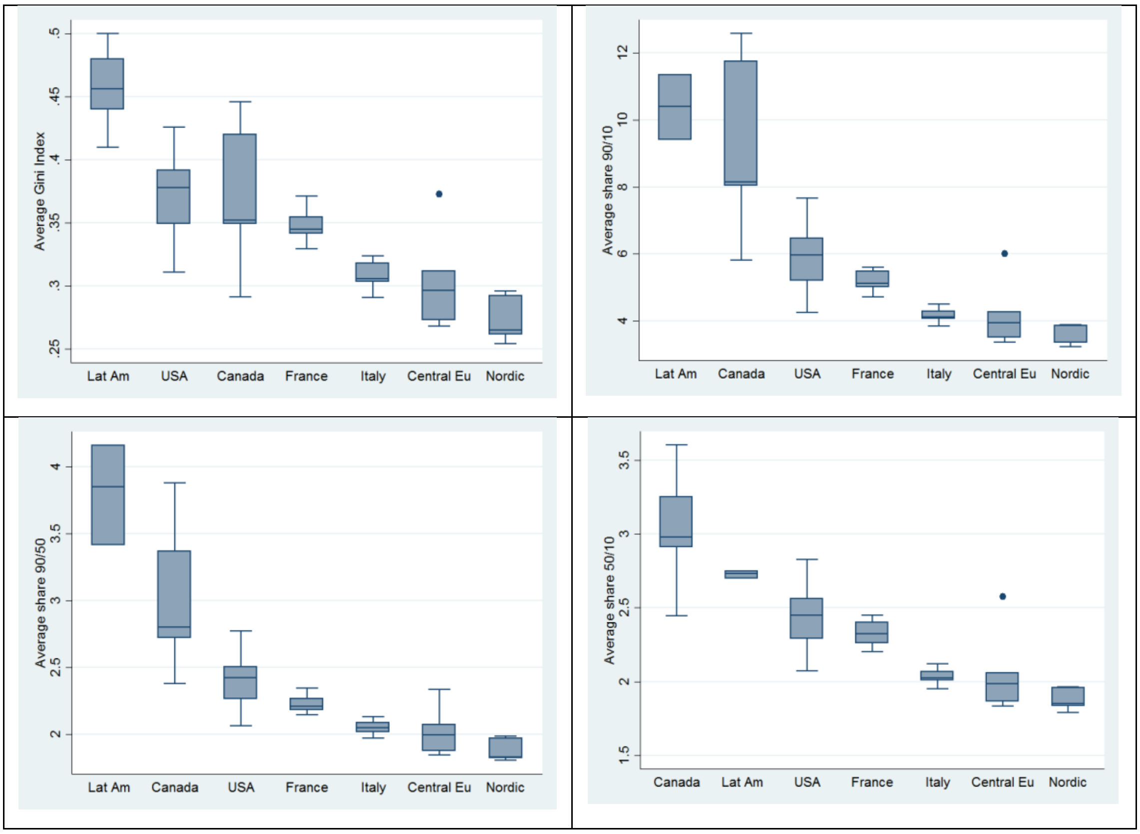
Table OSM.1. Descriptive statistics by country

	GDP pc[1]	Gini Index[2]	Population[3] (thousands)	Number of FUAs	Average city size FUAs[4]
Canada	44,017.59	34	36,708.0	9	2.218.987
Chile	22,767.04	47,7	18,054.7	3	2.834.938
Mexico	17,330.73	43,4	129,163.3	33	1.730.686
USA	54,225.45	41,5	325,719.2	70	2.490.941
Austria	45,436.69	30,5	8,809.2	3	1.323.321
Belgium	42,658.58	27,7	11,372.1	4	1.240.312
Denmark	46,682.51	28,2	5,769.6	1	2.025.171
France	38,605.67	32,7	67,118.6	15	1.693.135
Italy	35,220.08	35,4	60,551.4	11	1.679.620
Norway	64,800.06	27,5	5,282.2	1	1.280.812
Sweden	46,949.28	29,2	10,067.7	3	1.189.879

Notes:

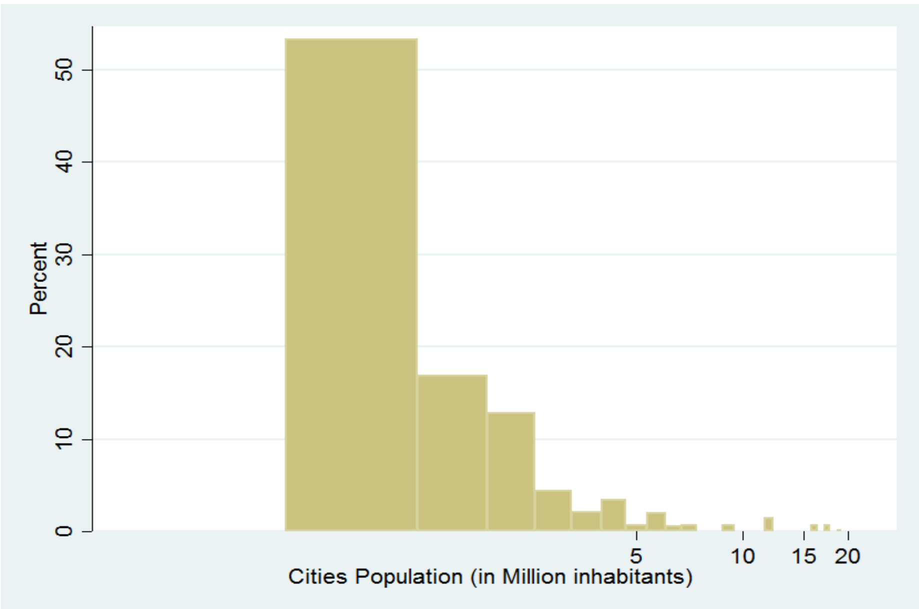
1. 2017 GDP per capita, PPP (constant 2011 international $). Source: World Bank
2. GINI index. 2015 data. 2013 for Canada. Source: World Bank
3. Population, total. 2017, Source: World Bank
4. Last year available. For most countries this corresponds to 2013, with the exceptions of Mexico (2010), France (2011), Austria (2012), and USA and Denmark (2014)

Figure OSM.1. Inequality measures – geographical distribution



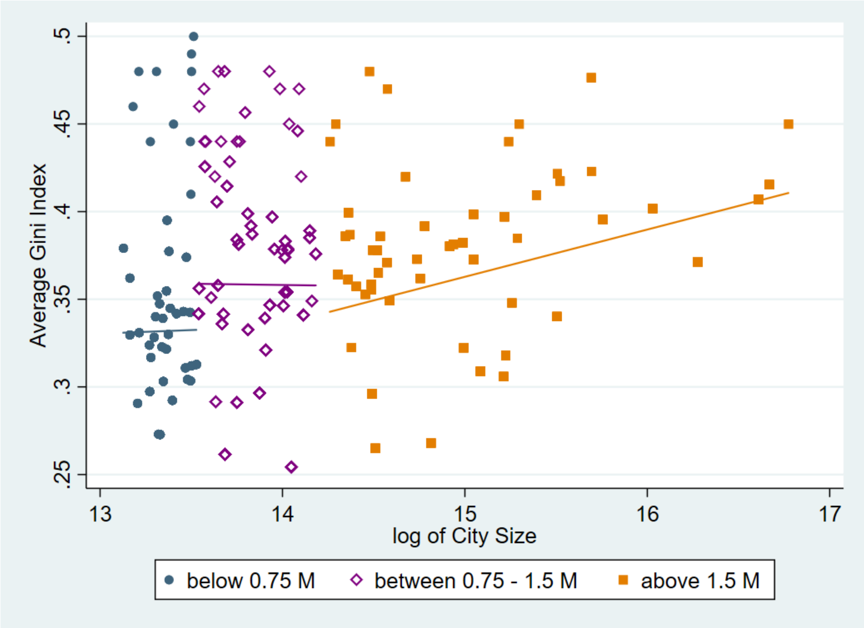
Note: we consider the times series average of every inequality measures for every metropolitan area. The ratio measures are not available for Mexico. Nordic countries include Denmark, Sweden and Norway; Central Europe consider Belgium and Austria, and Latin American countries refer to Chile and Mexico.

Figure OSM.2. FUAs Population – histogram



Note: FUAs population in the last available year in the sample. See table A1 in the appendix for corresponding year for every country. The horizontal axis reports the logarithmic scale.

Figure OSM.3. Association between city size and inequality measures, by initial population size.



Note: the scatterplot represents the average city size and Gini index over all available periods for every city.

Table OSM.2. Fixed-Effects Estimations

	(1)	(2)	(3)
Dependent variable:	Gini	Gini	Gini
Log(*pop*)	0.0608*** (0.0207)	0.1053*** (0.0272)	0.0856** (0.0364)
Year FE	NO	NO	YES
City Fixed Effects	YES	YES	YES
Observations	301	148	148
No. of cities	153	120	120
R-Square	0.09	0.324	0.45

Note: In columns 2 and 3 log(pop) is lagged one period; the time span goes from 2000 to 2014. Robust standard errors (clustered by city) in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table OSM.3. Confounding factors

	(1)	(2)	(3)	(4)
Dependent variable:	Gini	Gini	Gini	Gini
Log(*pop*)	0.0107***	0.0096***	0.0135***	0.0114***
	(0.0029)	(0.0030)	(0.003)	(0.0034)
Log(*income*)	0.0152	0.0252	0.0811***	0.0780***
	(0.0208)	(0.0216)	(0.0242)	(0.0203)
lab_part	-0.0005*	-0.0005**	-0.0004	-0.0003
	(0.0002)	(0.0003)	(0.0003)	(0.0003)
over65	0.0212	-0.0046	-0.0519	-0.0755
	(0.0765)	(0.0756)	(0.0643)	(0.0539)
high_edu	-0.0007	-0.0009*	-0.0024***	-0.0023***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)
ind_share		-0.0015***		-0.0009
		(0.0005)		(0.0011)
public_services		-0.0011*		0.0007
		(0.0006)		(0.0011)
private_services		-0.0005		0.0005
		(0.0005)		(0.0009)
Year FE	YES	YES	YES	YES
Country FE	YES	YES	YES	YES
City (Random) Effects	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Observations	272	271	129	128
No. of cities	143	143	110	109
R-Square	0.785	0.796	0.696	0.683

Note: *lab_part* is the labour participation rate, *over65* is the share of population over 65 years old, *high_edu* is the share of population between 25 and 64 with tertiary education, *ind_share* is the share of industry (including construction) to GDP, *public_services* is the share of public services to GDP, and *private_services* is the share of private services to GDP. Columns 1 to 4 are estimated using panel data; the time span goes from 2000 to 2014. In columns 3 and 4 right-hand-side variables are lagged one period and exclude cities in Mexico and Chile. Robust standard errors (clustered by city) in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$