LOCAL INEQUALITY ANALYSIS IN THE US: EVIDENCE FROM SOME METROPOLITAN STATISTICAL AREAS

Margherita Gerolimetto, Stefano Magrini

/

1. Introduction

Understanding the impact of policy changes on the distribution of income first requires a good representation of the distribution. There are various ways to do this that range from simple approaches, as the calculus of inequality measures, to more sophisticated approaches, as the kernel estimation of the distribution and, if possible, the observation of its evolution over time. All these methods can be jointly employed to have a clearer view of the concentration of the income, to compare different distributions and to understand the impact of difference policy actions.

In this work, we are interested in the analysis of income inequalities in the US between 2010 and 2018. In particular, we concentrate on a local perspective, by studying the inequality recorded in seven Metropolitan Statistical Areas, which have been chosen in order to cover geographically the US territory. It is important to emphasize that evaluating the degree of inequality at a local level, such a city or a metropolitan statistical area, is as important as at the national level. The connection, for instance, between inequality and crime is as strong within urban areas as it is across countries. Moreover, urban inequality seems as likely to generate political uprisings as inequality across large geographic units (Glaeser *et al*, 2009).

The aim of this work is two-fold. On the one hand, we study inequality, measured via some of the most well-known inequality indices. On the other hand, we complete the analysis via the kernel estimate of the distribution and some distribution dynamics analysis. The idea is that to have a completely informative inequality analysis, different perspectives should be considered.

The structure of the paper is as follows. In section 2, we present an overview of the main inequality measures. In section 3, we recall two tests to compare distributions. In section 4, we present our empirical analysis on 7 Metropolitan Statistical Areas across the USA. In section 5, we present some conclusions.

2. Inequality measures: an overview

Measures of inequality are widely used to study income and welfare. They are often a function that assigns a value to a specific distribution of income so that direct and objective comparisons across different distributions are possible: i) dynamic comparison (i.e. comparing inequality measures across time) and ii) comparisons for policy analysis (i.e. comparing the redistributive effects of current tax policy).

To do this, inequality measures should have certain properties and behave in a certain way, given certain events. No single measure satisfies all the properties, so the best approach is to look at more than one measure, trying to capture all the different perspectives into which the phenomenon is observed. In this overview, we focus on indices and ratios.

The Gini index (Gini, 1912) is the most widely cited measure of inequality; it measures the extent to which the income distribution within an economy deviates from a perfectly equal distribution. In its most intuitive definition, the Gini index is calculated as the ratio of the area between the Lorenz curve and the 45-degree line to the area underneath the 45-degree line. The larger is the index the higher the level of inequality. Being scale invariant, the Gini index allows for direct comparison between two populations regardless of their size. Among its limitations, one is that it is not easily decomposable or additive and it does not respond in the same way to income transfers between people in opposite tails of the distribution as it does to transfers between people in the middle of the distribution. Moreover, very different distribution can be characterized by the same value of the Gini index.

The Atkinson's inequality measure (Atkinson, 1970) is known for being a welfare-based measure of inequality. It presents the percentage of total income that a given society would have to forego in order to have more equal shares of income between its citizens. The index depends on the degree of risk aversion to inequality that characterizes a society. An important feature of the index is in that it can be decomposed into within-group and between-group inequality. Furthermore, it can provide welfare implications of alternative policy options, thus allowing the researcher to possibly include some normative content in the analysis.

The Theil (Theil, 1967) index is a special case of the General Entropy index. It ranges between zero (perfect equality) to one, if normalized. A key feature of these measures is that they are fully decomposable, i.e. inequality can be broken down by population groups or income sources or using other dimensions, which is very useful for policy makers. Another peculiarity of this index is that its mathematical expression depends on the value of a parameter, α, that represents a weight to distances between incomes in different parts of the income distribution. For low values of α, the index is more sensitive to changes in the lower tail of the distribution while for higher values it becomes more sensitive to changes in the upper tail of the distribution. The most commonly adopted values for α are 0, 1, 2. When α=0 the index is called "Theil's L", when α=1 the index is called "Theil's T", or more simply Theil index, when α=2, the index is called "coefficient of variation". Similarly, to the Gini index, when income redistribution occurs, change in the indices depends on the level of individual incomes involved in the redistribution and the population size.

Finally, we present some ratios that represent a basic inequality measure. In particular, we focus on decile dispersion ratios, which express the income of the richest as a multiple of the income of the poorest. They are simple, direct and easy to understand. At the same time, however, they do not provide as much information as the indexes listed before. A very commonly reported decile ratio is the D9/D1: the ratio of the income of the 10 percent richest part of the population to the income of the 10 per cent poorest. Another frequently adopted ratio is the 20/20 ratio that compares the ratio of the average income of the richest 20 per cent of the population to the average income of the poorest 20 per cent of the population.

3. Inference for comparing distributions

In order to focus on the entire distribution, we also resort on the representation of the empirical distribution, via the kernel density estimation (see for example Silverman (1986) for a very good presentation). In particular, we concentrate on the adaptive kernel density estimation based on the nearest neighbors' approach. As typical feature of the adaptive kernel, the smoothing parameter (bandwidth) employed in the estimate is not constant but instead varies according to the degree of clustering of the data. This allows a less biased estimate, while reaching a smoother graphical representation.

Moreover, in order not to confine ourselves to the pure graphical inspection, we consider to two well-known tests to compare distributions: the Kolmogorov-Smirnov test and Kramér-Von Mises test. The Kolmogorov-Smirnov (Kolmogorov, 1933; Smirnov, 1948) test is a nonparametric goodness-of-fit test to assess whether one random sample obtained from a population can be generated by a certain distribution function, that must be specific and known. The Kolmogorov-Smirnoff (KS hereafter) test may also be used to test whether two underlying one-dimensional probability distributions differ, as in the analysis carried out in this paper. The KS test uses the maximal absolute difference between these curves as its test statistic, that for space reasons we do not present here[1]. An attractive feature of this test is that the distribution of the KS test statistic itself does not depend on the underlying cumulative distribution function being tested.

[1] Interested readers may refer to the original papers instead.

Several goodness-of-fit tests, such as Kramér-Von Mises test (Kramér, 1928; von Mises, 1928), KvM hereafter, are refinements of the KS test. As these refined tests are generally considered to be more powerful than the original KS test, many analysts prefer them. In addition, the independence of the KS critical values of the underlying distribution is not as much of an advantage as it first appears. This is due to the fact that the distribution parameters are typically not known *a priori* and have to be estimated from the data. So, in practice the critical values for the KS test have to be determined by simulation just as for the Kramér-Von Mises (and related) tests.

4. Empirical analysis

We now present our empirical analysis. Data come from the IPUMS-USA database (Ruggles *et al*., 2020) and consists of per capita income net of transfers, relatively to year 2010 and year 2018 for the following 7 Metropolitan Statistical Areas (MSAs): Chicago-Naperville-Elgin, IL-IN-WI; Dallas-Fort Worth-Arlington, TX; Los Angeles-Long Beach-Anaheim, CA; Miami-Fort Lauderdale-West Palm Beach, FL; Minneapolis-St. Paul-Bloomington, MN-WI; Phoenix-Mesa-Scottsdale, AZ.

The principle behind the concept of Metropolitan Statistical Area (MSA) is that of a core area containing a substantial population nucleus, together with adjacent communities having a high degree of economic and social integration with that core. So, the area defined by the MSA is typically characterized by significant social and economic interaction. As of September 2018, (OMB Bulletin, 2018) there are 392 regions that meet the requirements to be designated as MSA in the U.S. and Puerto Rico (384 in the United States and 8 in Puerto Rico). Currently delineated metropolitan statistical areas are based on application of 2010 standards (which appeared in the Federal Register on June 28, 2010) to 2010 Census and 2011-2015 American Community Survey data, as well as 2018 Population Estimates Program data. The Bureau of Labor Statistics (BLS) uses MSA data to analyze labor market conditions within a geographical area. Within a metropolitan statistical area, workers can presumably change jobs without having to move to a new location, creating a relatively stable labor force. Consequently, MSAs offer a more representative view of the income variable behavior. At the same time, the disadvantage of adopting these geographical units is that they do not correspond to natural political units which make them awkward units for analyzing or discussing public policy.

We begin our empirical analysis by the calculus of the inequality measures we presented in the second section. These results are reported in Table 1. For all 7 MSAs we calculate the 4 inequality indexes and 2 ratios, for 2010 and 2018. We also computed the percentage variation of the indexes. As expected, all indexes and ratios are very well aligned. Apart from a couple of exceptions, all the considered inequality measures increase over the 8-year time span under examination. For two MSAs, in particular, this increase appears to be very severe and is emphasized in bold in the Table: they are Dallas-Fort Worth-Arlington, TX and Miami-Fort Lauderdale-West Palm Beach, FL. A third case captured our attention: it is Minneapolis-St. Paul-Bloomington, MN-WI. Contrary to the previous two mentioned MSAs, this area instead shows an increase in the inequality measures that is not so severe. Interpreting this as a virtuous behavior, we recall that Minneapolis-St. Paul-Bloomington is an area characterized by a special form of local governance, where there is a high degree of overlap between the administrative delimitation of the area and its economic delimitation.

Table 1 – *Inequality indices and ratios in 2010, 2018 and percentage variation (in italic)*

MSA name	Year	Gini	Theil L	Theil T	Atkinson	S80S20	P90P10
	2010	0.483	0.491	0.432	0.388	16.450	12.014
Chicago	2018	0.497	0.516	0.467	0.403	17.236	13.368
	Δ%	*2.903*	*5.115*	*8.069*	*3.913*	*4.778*	*11.268*
	2010	0.475	0.466	0.414	0.373	15.262	10.341
Dallas	2018	0.501	0.517	0.473	0.404	16.175	11.818
	Δ%	**5.327**	**10.946**	**14.120**	**8.376**	**5.980**	**14.285**
	2010	0.502	0.512	0.464	0.401	16.565	11.889
Los Angeles	2018	0.518	0.539	0.510	0.417	18.090	13.000
	Δ%	*3.188*	*5.342*	*9.977*	*4.036*	*9.202*	*9.346*
	2010	0.490	0.473	0.454	0.377	14.462	10.021
Miami	2018	0.517	0.521	0.523	0.406	16.566	11.538
	Δ%	**5.636**	**10.186**	**15.258**	**7.777**	**14.552**	**15.142**
	2010	0.450	0.437	0.374	0.354	14.156	10.407
Minneapolis	2018	0.460	0.449	0.402	0.362	14.043	10.909
	Δ%	**2.202**	**2.753**	**7.588**	**2.182**	**-0.799**	**4.823**
	2010	0.507	0.528	0.489	0.410	18.194	12.900
New York	2018	0.516	0.547	0.508	0.421	19.466	12.939
	Δ%	*1.728*	*3.513*	*3.914*	*2.643*	*6.989*	*0.304*
	2010	0.457	0.433	0.380	0.351	13.593	10.323
Phoenix	2018	0.466	0.444	0.404	0.358	13.014	10.934
	Δ%	*1.925*	*2.617*	*6.240*	*2.080*	*-4.255*	*5.927*

The graphical representation of the distributions of the logarithm of income and the results of the KS and KvM tests are reported in Figures 1-7. The estimates of the distributions are obtained with a nearest neighbors Gaussian kernel, where the percentage of neighbors is set equal to 25%. Although the 2010 and 2018 distribution representations seem to overlap, in line with the increase in inequality indexes documented in Table 1 both tests lead always to a rejection of the null hypothesis that the distribution is the same. Consistently with those results, we note that the

MSA for which the rejection is less strong is Minneapolis-St. Paul-Bloomington, for which, as previously seen, the increase in inequality measures is generally smaller.

5. Concluding remarks

In this paper, we study the evolution of per capita personal income inequalities within selected urban areas of the USA between 2010 and 2018. In particular, we adopt the Metropolitan Statistical Area as the basic spatial unit of analysis as this is an urban region characterized by a significant degree of social and economic interaction.

We first calculate several well-known income inequality indexes for 7 large MSA distributed around the US territory and document a significant increase in income disparities over the 2010-2018 period. Then, for each MSA we produce kernel density estimates of the distributions and performed Kolmogorov-Smirnoff and Kramér-Von Mises tests to evaluate whether the distributions are the same. In all cases, we reject the null hypothesis and conclude that the 2010-2018 period is characterized by a significant increase in income inequalities.

Figure 1 – *Kernel Representation of Income Distributions*

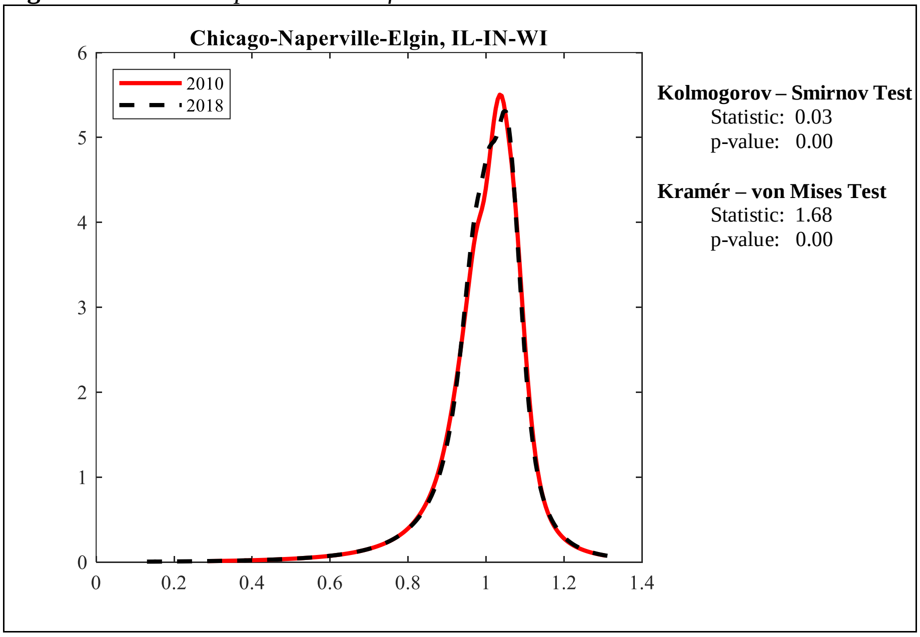
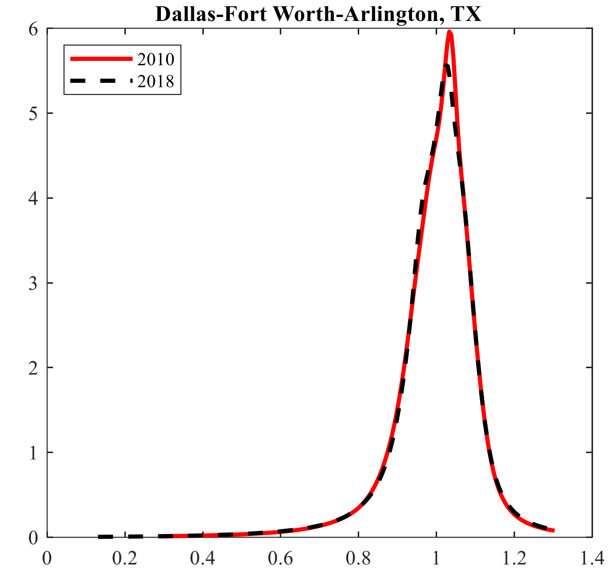


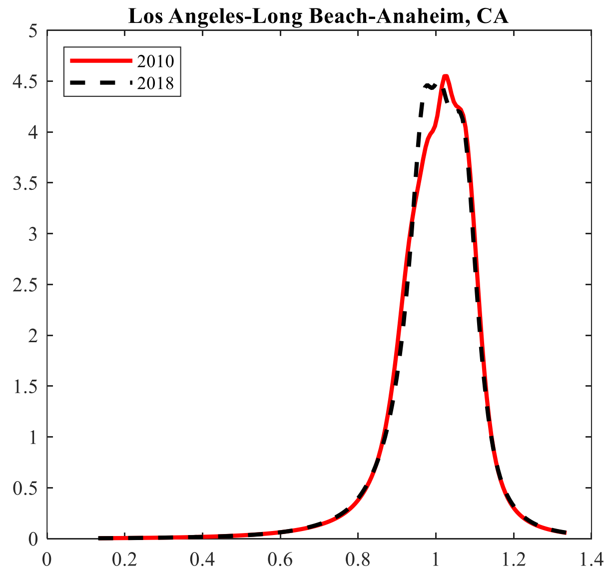
Figure 2 – *Kernel Representation of Income Distributions*



Kolmogorov – Smirnov Test
Statistic: 0.02
p-value: 0.00

Kramér – von Mises Test
Statistic: 1.37
p-value: 0.00

Figure 3 – *Kernel Representation of Income Distributions*



Kolmogorov – Smirnov Test
Statistic: 0.03
p-value: 0.00

Kramér – von Mises Test
Statistic: 3.84
p-value: 0.00

Figure 4 – *Kernel Representation of Income Distributions*

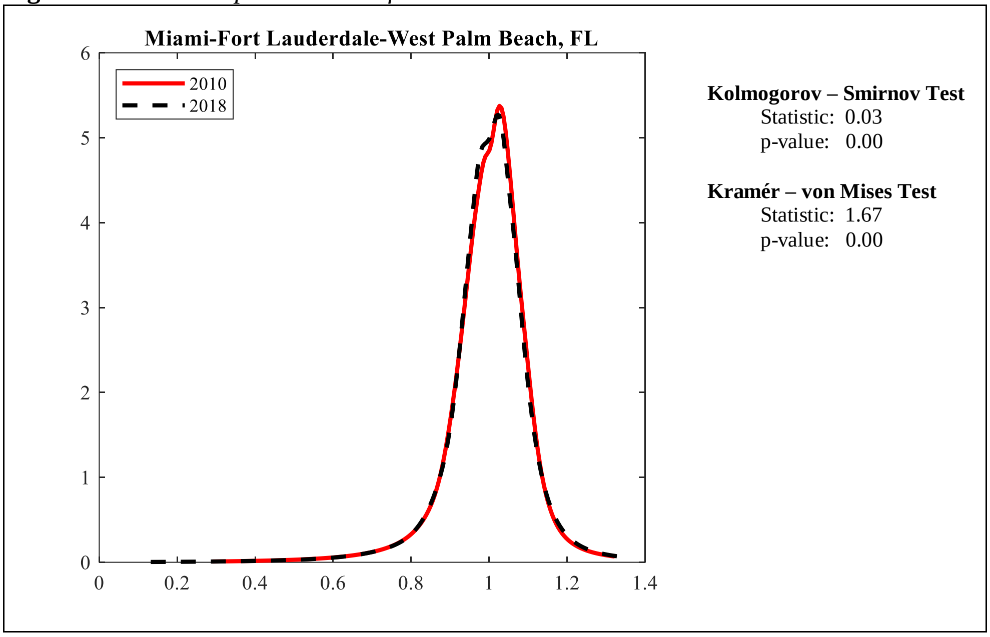


Figure 5 – *Kernel Representation of Income Distributions*

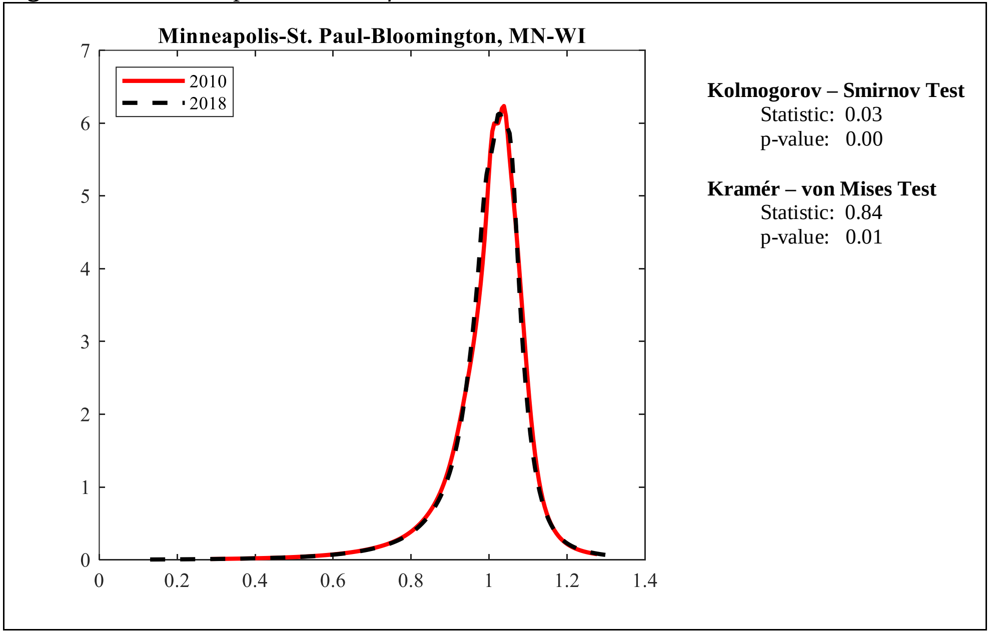


Figure 6 – *Kernel Representation of Income Distributions*

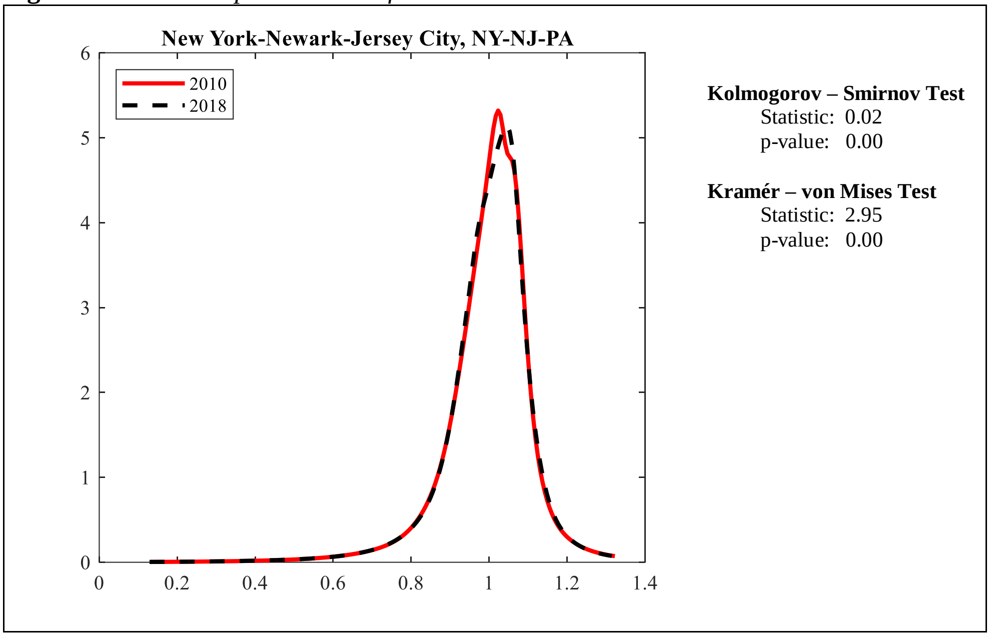
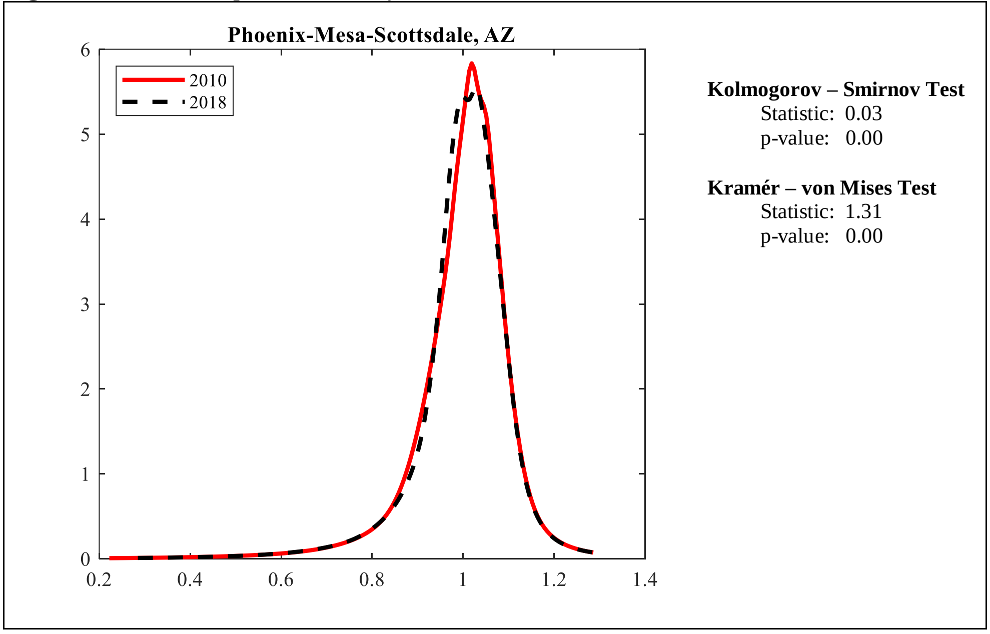


Figure 7 – *Kernel Representation of Income Distributions*



References

ATKINSON A.B. 1970. On the measurement of economic inequality, *Journal of Economic Theory*, Vol. 2, pp. 244–263.

GINI C. 1912. Variabilità e Mutabilità. *Studi Economico-giuridici*, *Università di Cagliari*, Anno III, Parte 2°.

GLAESER E.L., RESSENGER M., TOBIO K. 2009. Inequality in cities, *Journal of Regional Science*, Vol. 49, pp. 617-646.

KOLMOGOROV A. 1933. Sulla determinazione empirica di una legge di distribuzione, *Giornale dell'Istituto Italiano degli Attuari*, Vol. 4, pp. 83–91.

KRAMÉR A. 1928. On the composition of elementary errors II, *Scandinavian Actuarial Journal*, Vol. 1, pp. 13–74.

von MISES R.E. 1928. *Wahrscheinlichkeit, Statistik und Wahrheit*. Berlin: Julius Springer.

RUGGLES S., FLOOD S., GOEKEN R., GROVER J., MEYER E., PACAS J., SOBEK M. *IPUMS USA: Version 10.0 [dataset]*. Minneapolis, MN: IPUMS, 2020. https://doi.org/10.18128/D010.V10.0

SILVERMAN B.W. 1986. *Density estimation for statistics and data analysis*. London: Chapman and Hall.

SMIRNOV N. 1948. Table for estimating the goodness of fit of empirical distributions, *Annals of Mathematical Statistics*, Vol. 19, pp. 279–281.

THEIL H. 1967. *Economics and Information Theory*. Amsterdam: North-Holland Publishing.