Socio-Economic Planning Sciences 44 (2010) 212–219

Contents lists available at ScienceDirect

Socio-Economic Planning Sciences

journal homepage: www.elsevier.com/locate/seps

On the suitability of income inequality measures for regional analysis: Some evidence from simulation analysis and bootstrapping tests

Boris A. Portnov [a,*], Daniel Felsenstein [b]

[a] Department of Natural Resources and Environmental Management, Graduate School of Management, University of Haifa, Israel
[b] Department of Geography, Hebrew University of Jerusalem, Mount Scopus, Jerusalem, Israel

ARTICLE INFO

Article history:
Available online 13 April 2010

Keywords:
Regions
Inequality measures
Bootstrapping
Random permutation tests

ABSTRACT

The paper looks at the sensitivity of commonly used income inequality measures to changes in the ranking, size and number of regions into which a country is divided. During the analysis, several test distributions of populations and incomes are compared with a 'reference' distribution, characterized by an even distribution of population across regional subdivisions. Random permutation tests are also run to determine whether inequality measures commonly used in regional analysis produce meaningful estimates when applied to regions of different population size. The results show that only the population weighted coefficient of variation (Williamson's index) and population-weighted Gini coefficient may be considered sufficiently reliable inequality measures, when applied to countries with a small number of regions and with varying population sizes.

1. Introduction

The study of inequality across regions is rather different to the study of inequality between individuals. This derives from the fact that regions are groups formed by individuals. This is not as obvious as it may sound. For example, a tradition exists in the regional income convergence literature that treats regions as individual observations regardless of the size of the former (cf. e.g., [1]). As such, large and small regions are assumed to carry equal weight, just as fat and thin people are treated equally when looking at inequality between them.

The computational issues associated with multi-group comparisons of income inequality were noticed (apparently for the first time) by the American economist Max Lorenz. In his seminal paper published in 1905, Lorenz highlighted several drawbacks associated with the comparison of wealth concentrations between fixed groups of individuals. In particular, he found that while an increase in the percentage of the middle class is supposed to show the diffusion of wealth, a simple comparison of percent shares of persons in each income group may often lead to the opposite conclusion. For instance, while the upper income group in a particular period may constitute a smaller proportion of the total population, the overall wealth of this group may be far larger compared to another time period under study ([2]: 210–211). The remedy he suggested was to represent the actual inter-group income distribution as a line, plotting 'along one axis cumulated percents of the population from poorest to richest, and along the other the percent of the total wealth held by these percents of the populations' (ibid. p. 217).

In an essay published in 1912, the Italian statistician Corrado Gini moved Lorenz's ideas a step further, suggesting a simple and easy comprehendible measure of inequality known as the Gini coefficient. Graphically, the calculation of this coefficient can be interpreted as follows:

$$\text{Gini coefficient} = \frac{\text{Area between Lorenz curve and the diagonal}}{\text{Total area under the diagonal}}$$

Mathematically, the Gini coefficient is calculated as the arithmetic average of the absolute value of differences between all pairs of incomes, divided by the average income (see Table 1).[1] The coefficient takes on values between 0 and 1, with zero interpreted as perfect equality [3].

A few years later, Dalton [4] carried out the first systematic attempt to compare the performance of different inequality measures against 'real world' data. As he noted, many inequality measures, though having intuitive or mathematical appeal, react to changes in income distribution in an unexpected fashion. For

* Corresponding author.
E-mail address: portnov@nrem.haifa.ac.il (B.A. Portnov).

[1] The computation includes the cases where a given income level is compared with itself.

0038-0121/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.seps.2010.04.002

Table 1
Commonly used measurements of regional inequality.

Coefficient of variation (CV) (unweighted)	Population weighted coefficient of variation (Williamson index (WI))
$CV = \frac{1}{y}[\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2]^{1/2}$	$WI = 1\frac{}{\bar{y}}[\sum_{i=1}^{n}(y_i - \bar{y})^2 \frac{A_i}{A_{tot}}]^{1/2}$
Theil index (TE(0))	Atkinson index (AT)
$TE(0) = \frac{1}{n}\sum_{u=1}^{n} \log\frac{\bar{y}}{y_i}$	$AT = 1 - [\frac{1}{n}\sum_{i=1}^{n}[y_i \frac{}{\bar{y}}]^{1-\varepsilon}]^{1/(1-\varepsilon)}$
Hoover coefficient (HC)	Coulter coefficient (CC)
$HC = \frac{1}{2}\sum_{i=1}^{n} \lvert\frac{A_i}{A_{tot}} y_i \frac{}{\bar{y}} - \frac{A_i}{A_{tot}}\rvert$	$CC = [\frac{1}{2}\sum_{i=1}^{n}(\frac{A_i}{A_{tot}} y_i \frac{}{\bar{y}} - \frac{A_i}{A_{tot}})^2]^{1/2}$
Gini (U) (unweighted)	Gini (W) (population weighted)
$Gini = \frac{1}{2\,n^2\bar{y}}\sum_{i=1}^{n}\sum_{j=1}^{n}\lvert y_i - y_j\rvert$	$Gini = \frac{1}{2\,\bar{y}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{A_i}{A_{tot}}\frac{A_j}{A_{tot}}\lvert y_i - y_j\rvert$

Note: A_i and A_j = number of individuals in regions i and j respectively (regional populations), A_{tot} = the national population; y_i and y_j = development parameters observed respectively in region i and region j (e.g., per capita income); $\bar{y}$ is the national average (e.g., per capita national income); n = overall number of regions; ε is an inequality aversion parameter, $0 < \varepsilon < \infty$ (the higher the value of ε, the more society is concerned about inequality).

instance, if all the incomes are simply doubled, the variance quadruples the estimates of income inequality. Dalton's second observation was that some inequality measures do not comply with a basic principle of population welfare set forward by Arthur Pigou and formulated as follows: 'if there are only two income-receivers, and a transfer of income takes place from the richer to the poorer, inequality is diminished' (ibid. p. 351). After applying the 'principle of transfers' to various inequality measures, Dalton found that most measures of deviation (e.g., the mean standard deviation from the arithmetic mean, and the coefficient of variation) are perfectly sensitive to transfers and pass the 'test with distinction' (ibid. p. 352). The Gini index, commonly used in empirical studies, was also found by Dalton sufficiently sensitive to income transfers. He also found that the standard deviation is sensitive to transfers among the rich, while the standard deviation of logarithms is less sensitive to transfers among the rich than to transfers among the poor but still changes when a transfer among the rich takes place.

Two other fundamental requirements for a 'robust measure' of inequality, set forward by Dalton, are the principle of *proportional addition to incomes*, and the principle of *proportional increase in population*. According to the former, a proportional rise in all incomes diminishes inequality, while the proportional drop in all incomes increases it. According to the latter principle, termed by Dalton the 'principle of proportional additions to persons,' a robust inequality measure should be invariant to proportional increase in the population sizes of individual income groups. Dalton's calculations showed that most commonly used measures of inequality comply with these basic principles. Only the most 'simple' measures, such as absolute mean deviation, absolute standard deviations and absolute mean difference, fail to indicate any change, when proportional additions to the numbers of persons in individual income groups are applied (ibid. pp. 355–357, see also [5], pp. 87–112).[2]

Yitzhaki and Lerman [6]noted another deficiency inherent to most inequality measures, viz., insensitivity to the position which a specific population subgroup occupies within an overall distribution. Their Gini decomposition technique (see below) takes group-specific positions into account. In particular, they suggested weighting subgroups by the average rank of their members in the distribution. This is in contrast to the weighting system used more conventionally according to which between-group inequality is weighted by the rank of the average [7,8]. This latter system results in a large residual when inequality is decomposed into within and between groups. In contrast, the Yitzhaki approach results in a more concise decomposition with no residual [9].

Other empirical studies proposed and used a variety of additional inequality measurements, such as the population weighted coefficient of variation (Williamson's index), Theil index, Atkinson index, Hoover and Coulter coefficients [3,6,10–15]. However as the Gini measure ranges between 0 and 1 and is unaffected by change of scale (the population principle), it has become probably the most attractive measure for inequality in regional analysis.

While there have been numerous attempts to test the conformity of commonly used inequality measures with basic inequality criteria – e.g., principles of transfer, proportional addition to incomes, and proportional addition to population – (see *inter alia* [4,5,11]), there appears to be no systematic attempt to verify whether all of these measures are equally suitable for regional analysis, in which individual countries may be represented both by a different numbers regions and by regions of different population sizes. The lack of interest to this aspect of inequality measurement may have a simple explanation. Since commonly used inequality indices (some of which appear in Table 1) are abstract mathematical formulas, one can assume that they can be applied to both large and small countries alike and provide fully comparable results. However, it is well known that the use of different measurement indices in regional analysis gives rise to highly variable results. For example, the notion of optimal regional convergence (i.e., that point where regional convergence also reduces overall nation-level inequality) has been shown to be highly dependent on the type of inequality index used [16] as is the measurement of regional price convergence [17].

The present paper attempts to determine whether commonly used inequality measures are sufficiently sensitive to changes in the ranking, size and number of regions into which a country is divided. The paper is organized as follows. First, we look at the specificity of measuring regional inequality. Given the fact that the unit of observation (i.e., a region) is a group measure, it presumably needs some weighting as regions of a country come in different sizes. We then proceed to discuss the general principles that should govern in our view, the selection of robust inequality measures. Then we move to testing the compliance of different commonly used inequality measures against the set of criteria that should characterize a robust inequality measure. The tests are run in two phases. First, we use a number of pre-designed distributions, to verify whether a particular inequality measure meets our intuitive expectations concerning inequality estimates. Then, in the second stage of the analysis, we run more formal permutation tests to verify whether different inequality measurements respond sensibly to changes in the population distribution across the space.

2. Sizes and shapes of regions

General economic theory does not suggest *a priori* that the size and number of regions in a country should affect the distribution of inequality. Beenstock [18] investigated this issue, testing whether regional amalgamation (decreasing the number of regions) impacts on inequality between them. His analysis shows that unifying any two regions will increase the earnings of each and reduce inequality between them but the same cannot be said for the level of overall inequality between all regions. In terms of regional size, a similar conclusion is drawn from economic theory. Beenstock's work shows that regional size *per se* has little impact on the determination of regional inequality and regional social and

[2] Dalton ([4]: 352) distinguishes between measures of relative dispersion and measures of absolute dispersion. Whereas the former measures are dimensionless, the measures of absolute dispersion are estimated in units of income. The latter measures are easily transformed in the former by normalization.

economic heterogeneity is likely to be a more important conditioning factor for regional gaps. However, factor mobility is expected to be greater in small regions reducing the impacts of regional differences (ibid).

In a seminal paper, Robinson [19] distinguished between two types of correlation – *ecological* and *individual*. The former is obtained for a group of people, while the latter is estimated for indivisible units, such as individuals. According to Robinson, ecological and individual correlations tend to be dissimilar. As a result, any assumption about an individual based on average data obtained for a group to which the individual belongs may result in an assessment error, known as 'ecological fallacy' [20].

Although Robinson's article in American Sociological Review [19] became a real eye-opener for many social scientists, more than a decade earlier, Gehlke and Biehl [21], reported similar variation of correlation coefficients in line with data aggregation. Follow up studies (see *inter alia* [20,22,23]) shed additional light on Gehlke–Biehl–Robinson's findings, showing that the size of correlation coefficients tends, in general, to increase with data aggregation into areal units of larger size. Openshaw [22] termed this phenomenon the 'modifiable areal unit problem' or MAUP. Although the recent popularity of geographic information systems (GIS) technology ([20,24–26]), has helped to simplify the establishment of such data linkages, formal verification of the correspondence between results obtained from individual data and those obtained from areal aggregates is largely still in its infancy.

Recent work has added empirical evidence to the way in which regional size and shape impact measures of regional inequality and agglomeration. Experimenting with different zones for French regions, Briant Combes and Lafourcade [27] conclude that conclude that interregional inequalities are not particularly sensitive to regional size and shape when compared with issues of specification. The case for this sensitivity lies in the MAUP thesis and the broad claim that spatial aggregation can distort the measurement of social and economic relationships. This distortion can arise through changing both the size (scale effect) and shape (zoning effect) of areas. It comes about when trying to model continuous spatial processes using discrete zones and by assuming spatial stationarity in relationships that are spatially dependent [28].

The evidence on the severity of the MAUP phenomenon is mixed. Fotheringham and Wong [29] show that in a multivariate setting, correlation and regression coefficients are not robust to different areal aggregations. Others have shown that what is sometimes taken as MAUP-type aggregation effect is really the result of mis-specification [20,30]. Menon [31] utilizes the existence of MAUP effects in data on regional industrial clustering in order to 'engineer' an ideal zonal system that serves as a counterfactual for understanding industrial concentration and its labor market determinants. In this way, control of the data generating process is exploited to examine causal relationships.

3. Regions as groups

Related to the issue of regional size is the issue of scale. Table 1 outlines many of the common measures used for measuring regional inequality. While most of them satisfy the 'transfer' principle, many do not comply with the 'population' principle transfer, i.e., they are sensitive to scale. Weighting by population gets around this, giving more weight to more populous regions and assuming that not all groups are of equal size. Another issue still outstanding relates to interpreting these inequality measures. Many, like the use of variance or the Theil and Atkinson indices, do not yield measures in the range of 0–1 and are not readily interpretable. The population-weighted Gini (Gini (W)), which handles scale and yields an index in the 0–1 range, is one of the most popular inequality measures in recent regional studies dealing with income inequality [32].

The question then arises as to the exact form of Gini (W) to use. A generic definition of Gini formulated by Schechtman and Yitzhaki [33], that allows for incorporating different weighting schemes, is as follows:

$$G_t = \frac{2\mathrm{cov}(Y_{it}, R_{it})}{\overline{Y}_t} \tag{1}$$

where G_t is the Gini coefficient measuring inequality in Y between n regions at time t. Pyatt [7] suggests weighting Gini by population shares. In this case the numerator of eq. (1) is the weighted covariance, and the denominator is the weighted average:

$$G_t = \frac{w2\mathrm{cov}(Y_{it}, R_{it})}{w\overline{Y}_t} \tag{2}$$

Yitzhaki [9] suggests that between-group inequality can be measured by replacing R in eq. (1) by the average rank in the region rather than the rank of the average:

$$G_t = \frac{2\mathrm{cov}(\overline{R}_{it}, Y_{it})}{\overline{Y}_t} \tag{3}$$

The difference between these two forms of weighted Gini is that the Pyatt version weights by the ranks of the averages and the Yitzhaki version weights by the average of the ranks. Thus the Pyatt Gini ranks regions (region i's rank out of total I regions) while the Yitzhaki Gini uses the average rank out of N individuals in region i. If the average of the ranks is the same across regions, the Yitzhaki between-group Gini is equal to zero even if average income is not the same. The Yitzhaki version of Gini is attractive as it accommodates an exact decomposition of Gini. This means that the regional Gini's will sum to the national Gini. Denoting Gini for individuals by G^*, Yitzhaki [9] has shown that $G = G^* - \sum_{i} G_i V_i O_i$, where G_i denotes Gini for individuals in region i, V_i denotes the share of earnings in region i, and O denotes the coefficient of stratification or overlapping (i.e., the extent to which the richest in region j overlaps with the poorest in region i). This would intuitively seem to be an insightful approach to measuring regional (group) inequality.

Recent empirical work on regional income inequality in Israel has tested these Ginis for regional inequality [32]. Three different measures of regional Gini were estimated using average regional earnings for 1991 and 2001(see Table 2); an unweighted Gini, Pyatt's version of Gini for groups and the Yitzhaki formulation for groups. As can be seen, the unweighted and Pyatt version yield similar estimates. However, the Yitzhaki Gini yields an unrealistically low estimate (0.00035) implying almost complete regional equality. The reason for this is that the means of the ranks vary within a narrow margin, between 0.47 and 0.56. Had the regions been of equal size, the Pyatt Gini and the unweighted Gini would have been the same. As can be seen from Table 2, in 2001 the results seem to be similar. The unweighted and Pyatt Ginis are close in magnitude and greater than in 1991 thus suggesting an increase in regional inequality over the period. In contrast, the Yitzhaki formulation does not suggest that inequality increased because in 2001 the means of the ranks vary within a narrow margin.

Table 2
Empirical Testing of Regional Inequality using Different Gini Formulations for Israeli Regions, 1991 and 2001.

Year	Unweighted Gini	Pyatt Gini	Yitzhaki Gini
1991	0.05620	0.05522	0.00035
2001	0.0750	0.0785	−0.0033

Source: Beenstock and Felsenstein (2007).

In sum, treating regions as groups means taking scale into account. This calls for a population weighted inequality indices (such as Gini (W)) to be used. Empirical testing of two different Gini formulations shows that while the Yitzhaki version may be methodologically more elegant, in practice it yields measures that do not reflect reality. This is probably because average ranks do not change very much compared to the ranks of the averages.

4. Basic requirements for regional income inequality measures

We hypothesize that the following three basic requirements need to be met in order for a regional income inequality measure to be fully applicable to countries with different numbers of regional subdivisions and different population size, – the subdivision principle, tolerance to size difference, and rank-order insensitivity:[3]

- *Subdivision principle:* No matter into how many regions (subdivisions) a country is split, inequality estimates should not change, unless the parameter distribution alters. This requirement is basically in line with Dalton's principle of population, according to which neither replication of population nor merging identical distributions should alter inequality.
- *Tolerance to size differences:* A robust inequality measure should produce identical estimates for both geographically even and geographically skewed population distributions, providing that the parameter distribution (e.g., distribution of incomes) remains unchanged. For instance, most residents of a country may be concentrated in a single region or population may be dispersed evenly across 10 districts into which the country is split. As long as the income distribution stays the same, regional inequality should not alter.
- *Rank-order insensitivity*: The inequality estimate should not alter as a result of a change in the sequence in which regions are introduced into the calculation, e.g., ranked either by population size or in alphabetical order. Since regions in a small country may be a subject to rapid changes, both in terms of their population sizes and parameter distributions, the compliance with this principle will secure that inequality estimates do not alter simply as a result of changing the position of a region in the rank-order hierarchy.

In order to verify the compliance of commonly used measures of regional inequality with the above requirements, the analysis will be carried out in two stages: pre-designed sensitivity tests and random permutation tests.

5. Pre-designed sensitivity tests

Eight commonly used inequality measures (see Table 1) are tested here. The tests are designed as follows. First, we introduce the 'reference' distribution (Table 3: 'Reference distribution'). As Table 3 shows, this distribution has 16 internal divisions (regions). The average per capita income in its four central regions doubles that in the 12 peripheral regions – 20,000 and 10,000 Income Units (IUs), respectively. Let us call the former group of regions 'H[igh-income]-regions,' while 12 other regions will conditionally be termed 'L[ow-income]-regions.'

[3] Basic inequality criteria (such as the principles of transfer, proportional addition to incomes, and proportional addition to population, etc), outlined in the previous discussion, fall outside the scope of the present analysis. They could be a legitimate topic for follow up studies.

Table 3
The reference and test distributions.

Reference distribution				Test 1 (Number of regions)			
Average income				Average income			
10,000	10,000	10,000	10,000	10,000	10,000		
10,000	20,000	20,000	10,000	10,000	20,000		
10,000	20,000	20,000	10,000	10,000	20,000		
10,000	10,000	10,000	10,000	10,000	10,000		
Population size				Population size			
10,000	10,000	10,000	10,000	10,000	10,000		
10,000	10,000	10,000	10,000	10,000	10,000		
10,000	10,000	10,000	10,000	10,000	10,000		
10,000	10,000	10,000	10,000	10,000	10,000		
Test 2 (Population distribution)				**Test 3 (District ranking)**			
Average income				Average income			
10,000	10,000	10,000	10,000	20,000	10,000	10,000	20,000
10,000	20,000	20,000	10,000	10,000	10,000	10,000	10,000
10,000	20,000	20,000	10,000	10,000	10,000	10,000	10,000
10,000	10,000	10,000	10,000	20,000	10,000	10,000	20,000
Population size				Population size			
5,000	5,000	5,000	5,000	10,000	10,000	10,000	10,000
5,000	25,000	25,000	5,000	10,000	10,000	10,000	10,000
5,000	25,000	25,000	5,000	10,000	10,000	10,000	10,000
5,000	5,000	5,000	5,000	10,000	10,000	10,000	10,000

As the table shows, in the reference distribution, the population is distributed evenly: there are 10,000 residents in each regional cell. The total population of the reference system is 160,000 residents and the average income is 12,500 IUs per capita.

5.1. *Test 1 – small number of regions*

During this test, we should check whether the overall number of regions matters. To this end, we reduce the overall number of regions to eight, from sixteen in the reference distribution. Total population for this distribution is 80,000 residents, while the average income remains the same and being equal to 12,500 IUs. Since there are no cardinal changes in income or population distribution, robust inequality indices should indicate the *same level of inequality* for both the reference and Test 1 distributions (see Table 3).

5.2. *Test 2 – uneven population distribution*

This test is designed to trace the response of different inequality measures to regional distribution of population: evenly spread population in the reference distribution vs. unevenly spread population in the Test 2 distribution. Compared to the reference distribution, there are no changes in per capita incomes; only the pattern of population distribution is altered. In particular, the populations of the four central (H-regions) increased to 100,000 ($4 \times 25{,}000$) residents, while the populations of surrounding L-regions shrunk to 60,000 (5000×12) residents (see Table 3). The total population in this distribution is 160,000 residents and the average income is 16,250 IUs. Since the percent share of population concentrated in four H-regions increased to 62.5 percent [$100{,}000 \times 100/160{,}000$ (total population) $= 62.5\%$] from 25 percent in the reference distribution [$40{,}000 \times 100/160{,}000 = 25\%$; see Table 3], the regional inequality of per capita incomes *should expectedly decline.*

5.3. *Test 3 – rank-order change*

Our last test is designed to verify whether the sequence in which regions are introduced in the calculation matters. Compared to the reference distribution, there is no change in either the total number of residents (160,000) or in the average per capita income (12,500 IUs). The only change is the location of H-regions: if in the reference distribution these regions are located in the centre of the grid (6, 7, 10 and 11 sequence numbers), in the Test 3 distribution, they are moved to the corners of the grid (1, 4, 13 and 16 sequence numbers – see Table 3). Since the percent share of population concentrated in the H-regions has not changed [40,000 × 100/160,000 = 25%], *no change in inequality should occur.*

6. Sensitivity test results

The results of the tests are reported in Table 4 and discussed below.

Test 1: Somewhat surprisingly, despite the unchanged distributions of incomes and populations, CC indicates a *rise* in inequality. The use of this index for small countries, with a small number of internal divisions (regions), may thus be misleading, specifically when a comparison with countries of larger sizes is planned.

Test 2: While the five indices (WI, CC, HC, Gini (U) and Gini (W)) indeed indicate a drop in regional inequality compared to the reference distribution, three other measures (CV, TE and AT) indicate an *increase* in income disparity (see Table 4). Characteristically, Gini (W) which is essentially a 'Pyatt' Gini, indicates only a marginal drop in inequality (from 0.075 in the reference distribution to 0.072 in the Test 2 distribution) despite a considerable increase in the population share of H-regions. The use of the CV, TE, and AT measures for small countries (which are often characterized by extremely uneven regional distributions of population) may thus lead to erroneous results.

Test 3: The test indicates no performance problems with any of the indices tested. Numerically, the results of the test appear to be identical to those obtained for the reference distribution (see Table 4).

7. Random permutation tests

For more formal sensitivity testing of inequality measures, we used the statistical technique known as *bootstrapping* or *random permutation test* [34]. Traditional methods of calculating parameters for a given statistic (e.g., a certain measure of inequality) are based upon the assumption that the statistic is asymptotically normally distributed and use known transformations for parameter calculation. In contrast, re-sampling techniques, such as bootstrapping, provide estimates of the standard error, confidence intervals, and distributions for any statistic by testing it directly against a large number of randomly drawn re-samples. At least 1000 re-samples are considered as a minimal number recommended for estimating parameters of a statistic, whereas larger numbers of re-runs increase the accuracy of estimates.

Table 4
Results of sensitivity tests.

Inequality index	Reference distribution	Test 1 (Number of regions)	Test 2 (Population distribution)	Test 3 (District ranking)
CV	0.346	0.346	0.353	0.346
WI	0.346	0.346	0.298	0.346
TE	0.022	0.022	0.136	0.022
AT	0.026	0.026	0.251	0.026
HC	0.150	0.150	0.144	0.150
CC	0.061	0.087	0.059	0.061
Gini (U)	0.075	0.075	0.058	0.075
Gini (W)	0.075	0.075	0.072	0.075

Bootstrapping is widely used in a variety of scientific fields, ranging from ecology and biology [35–41] to regional studies and urban economics [42–44].

In the present analysis, we used bootstrapping to generate different distributions of the total population (160,000) across 16 'reference' regions (see Table 3), without any change in the 'reference' income distribution. Next, we calculated values of different regional inequality coefficients listed in Table 1. According to the main underlying assumption of our bootstrapping experiment, the calculated values of a good inequality measure (that is, a measure which is sufficiently sensitive to changes in the interregional distribution of population) should reflect changes in the population distribution patterns, (i.e. *exhibit an undulating pattern*) when graphed against the sequence of distribution tries. Concurrently, the values of an inequality measure, which is insensitive to changes in the population sizes of regions, should *stay constant* over all the bootstrapping tries.

In particular, we ran two separate tests, as described below:

- *Test 1 (Unrestricted test):* The distribution of income was set identical to the reference distribution (see Table 3) and the average income was kept constant (12,500 IUs). Concurrently, the population was distributed across 16 regional cells at random and was allowed to vary slightly around the average population total, which was not restricted a-priori.
- *Test 2 (Restricted test):* The income distribution, the average income, and the total population of the system were kept constant and identical to the reference distribution (see Table 3).

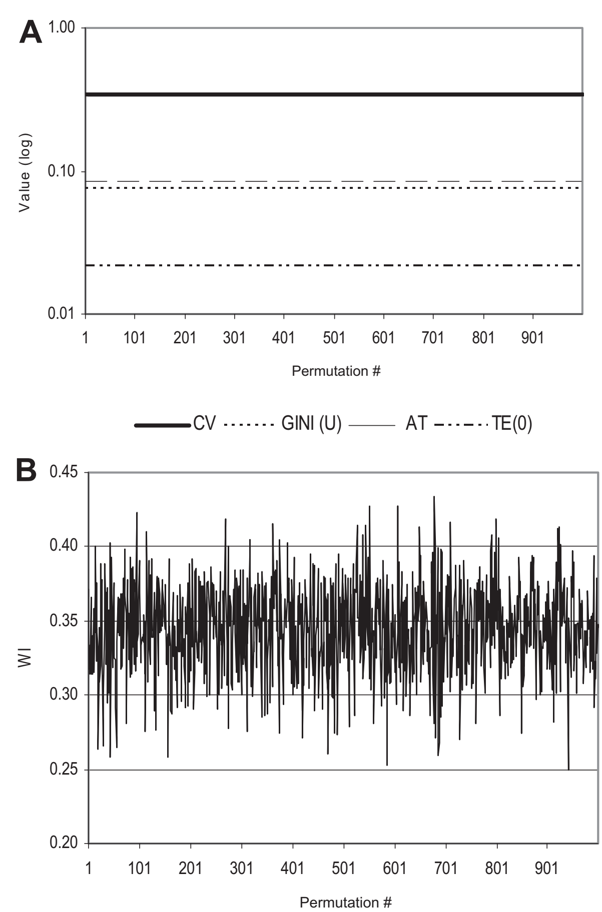

Fig. 1. Results of permutation tests (Test 1: unrestricted test) for selected inequality measures – CV, Gini (U), AT and TE(0) (A) and WI (B) Note: see text for explanations.

In order to comply with these restrictions, the population was redistributed *within* the H-regions and L-regions, without allowing population exchanges *between* these two groups of regions.

For each test, some 1000 permutations (re-samples) of population distribution were run. For the sake of clarity and brevity and to avoid overloading the reader with unnecessary technical details, we discuss below only those results for the tests for inequality indices that appear to exhibit most characteristic trends.

7.1. Unrestricted test

The results of the re-sampling for five inequality indices – CV, Gini (U), AT, TE(0), and WI are reported in Fig. 1. While CV, Gini (U), AT and TE(0) appear to exhibit the response pattern shown in Fig. 1A, the rest of the indices tested (that is, WI, CC, HC and Gini (W)) exhibits the response pattern diagrammed in Fig. 1B. In line with the underlying assumption of our bootstrapping experiment (see Section 7), the conclusion is thus straightforward: the former group of indices is not sensitive at all to the variation in population distribution across regional cells. These indices may thus lead to spurious results when used for small countries, which are often characterized by rapid changes in population patterns, due to (*inter alia*) the impact of immigration.

7.2. Restricted test

When population movements are restricted (i.e., the population is allowed to circulate only within the H-regions and within the L-regions, without direct population exchanges between the two), only the CC index appears to respond to population re-sampling, exhibiting the oscillation response pattern (see Fig. 2B), whereas all other indices tested (i.e., CV, WI, HC, Gini (U), Gini (W), AT and TE (0)) fail to respond to changes in the population distribution across the regional cells (see Fig. 2 A). However, such a situation (in which population movements are geographically restricted) may be considered rather unlikely (specifically for open economies) and thus a failure of an inequality measure to pass this test may be considered only as a minor performance flaw.

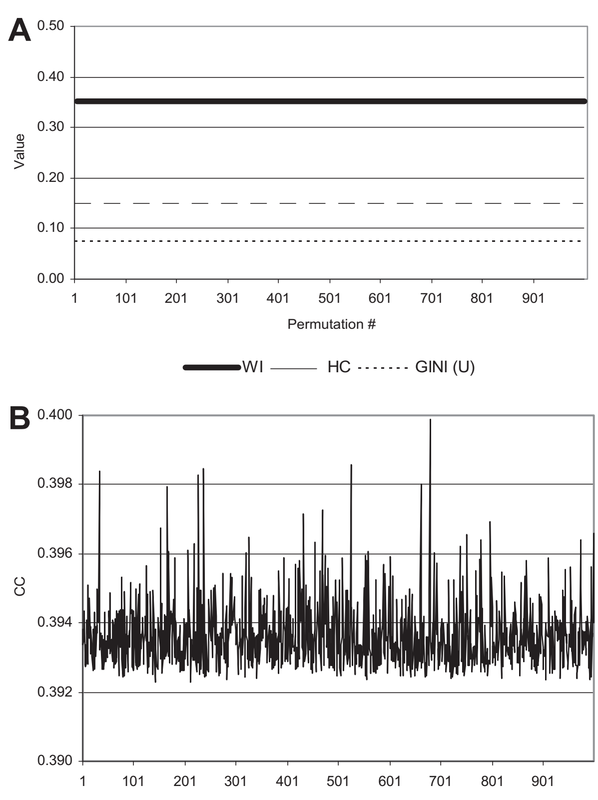

Fig. 2. Results of permutation tests (Test 2: restricted test) for selected inequality measures – WI, HC, and Gini (U) (A) and CC (B) Note: see text for explanations.

8. Conclusions

Though individual studies of regional disparity may deal with separate development measures – population growth, wages, welfare, regional productivity, etc. – the use of an integrated indicator is often essential, particularly if a comparative (cross-country) analysis is required. In order to measure the extent of disparities, various indices of inequality are commonly used. These indices may be classified into two separate groups [14]:

- *Measures of deprivation* (Atkinson index, Theil redundancy index, Demand and Reserve coefficient, Kullback-Leibler redundancy index, Hoover and Coulter coefficients, and the Gini index);
- *Measures of variation,* such as the coefficient of variation and Williamson's index.

In this paper, we did not attempt to assess whether these measurements reflect either the 'true meaning' or 'underlying causes' of regional inequality. Neither did we try to establish whether geographic inequality is a positive socio-economic phenomenon or a negative one. We shall leave these philosophical questions for future studies. Our task was simple: we attempted to determine whether *commonly used inequality measures produce meaningful estimates when applied to countries of different size and with different number of regional subdivisions*, *thus making it possible to directly compare the results of analysis obtained for one particular country with those obtained elsewhere.*

This task is not as abstract as it may sound. There has been a tradition in the regional income convergence literature to treat regions of a country as individual observations, regardless of the numbers of regions into which a country is subdivided and of their population sizes (cf. e.g., [1]). As such, all regions, big and small, are assumed to carry equal weight, an assumption which is hardly justified empirically, considering a large variation in the number of regions and their population sizes that are found in most countries across the globe.

In order to formalize these distinctions, we designed a number of simple empirical tests, in which several hypothetical income and population distributions were compared with the 'reference' distribution in which the population was distributed evenly across regional divisions and assumed to be static.

In the first test, we checked whether the overall number of regions matters. In the second, we checked whether different inequality indices respond to differences in the regional distribution of population, viz., evenly spread population in the reference distribution vs. unevenly spread population in the test distribution. Finally, in the third test, we verified whether different inequality indices were sensitive to the sequence in which regions are introduced into the calculation.

Somewhat surprisingly, none of the indices we tested appeared to pass all the tests, meaning that they may produce (at least theoretically) misleading estimates if used for small countries. However, several indices – CC, WI and Gini (W) – appeared to exhibit only minor flaws and may thus be considered as more or less reliable regional inequality measures, when used empirical studies for comparing the extent of regional inequalities across the time and space, that is, over different time periods and between different countries.

Mitigating regional development disparities between rapidly developing and economically successful core regions and lagging peripheral areas, has been a prime objective of regional development policy in most developed countries worldwide. In this respect, efficient regional inequality measures may become an indispensable tool for gauging the success of failure of such a policy, enabling policy-makers and regional scientists to compare the extent of development disparities before and after the policy intervention, or in relationship to other countries, characterized by similar development levels. However, this important objective can be achieved only if inequality measures used in the analysis faithfully reflect the extent of interregional disparities but no less importantly, *the number of people facing them*. For instance, if in a country with e.g., 20 million residents, the population size of the least developed region may be 50,000 or, alternatively, 5,000,000 residents. This difference may have different implications for "regional equalization" policies, both in sense of policy tools to be used and their magnitude. However, if regions are assumed to be indivisible units, equal in size, as with many widely used inequality measures, this important distinction may go unnoticed thereby distorting development policy overall. In this respect, the present analysis serves to caution regional analysts against using different inequality measures *indiscriminately* or comparing their values directly, between countries and time periods.

Although an inequality measure may be sufficiently sensitive to differences in the population size across regions this sensitivity may not always be sufficient, unless *intra-regional* disparities are also taken into account. Even when controlling for regional size, treating a region as the prime unit of observation in the analysis of regional inequality may hide considerable internal income disparities among its residents. Therefore future research efforts will need to develop regional inequality indices that account for both disparities between individual regions (in terms of their sizes and development levels), as well as their *internal income heterogeneities*. Further research may also be needed to develop and a set of mathematical tools that can detect potential performance problems of various regional inequality measures, without a need to carry out extensive empirical simulations, such as those undertaken in the present study. Although future work on the performance of different inequality indices may thus be needed to verify the generality of our observations, the present analysis clearly cautions against indiscriminate use of inequality indices for regional analysis and comparison.

Acknowledgement

Our gratitude is due to Jacques Silber, Michal Beenstock and to two anonymous reviewers for their helpful comments on earlier versions of the paper.

References

[1] Barro RJ, Sala-i-Martin X. Convergence. Journal of Political Economy 1992;100(2):223–51.
[2] Lorenz MO. Methods of measuring the concentration of wealth. Journal of the American Statistical Association 1905;9(70):209–19.
[3] Atkinson AB. The economics of inequality. 2nd ed. Oxford: Clarendon Press; 1983.
[4] Dalton H. The measurement of the inequality of incomes. The Economic Journal 1920;30(199):348–61.
[5] Champernowne DG, Cowell FA. Economic inequality and income distribution. Cambridge, UK: Cambridge University Press; 1998.
[6] Yitzhaki S, Lerman RI. Income stratification and income inequality. Review of Income and Wealth 1991;37(3):313–29.
[7] Pyatt G. On the interpretation of disaggregation of the Gini coefficient. Economic Journal 1976;86:243–55.
[8] Silber J. Factor components, population subgroups and the computation of the Gini index of inequality. The Review of Economics and Statistics 1989;71(2):107–15.
[9] Yitzhaki S. Economic distance and overlapping distributions. Journal of Econometrics 1994;61:147–59.
[10] Williamson JG. Regional inequalities and the process of national development: a description of the patterns. 1975 reprint. In: Friedmann J, Alonso W, editors. Regional policy. Cambridge, Massachusetts: The MIT Press; 1965. p. 158–200.
[11] Sen A. On economic inequality (the Radcliffe lectures series). Oxford: Clarendon Press; 1973.
[12] Coulter P. Measuring unintended distributional effects of bureaucratic decision rules. In: Busson T, Coulter P, editors. Policy evaluation for local government. New York: Greenwood Press; 1987.
[13] Sala-i-Martin X. Regional cohesion: evidence and theories of regional growth and convergence. European Economic Review 1996;40:1325–52.
[14] Kluge G. Wealth and people: inequality measures; 1999 [Internet edition].
[15] WBG. Inequality measurement. Washington, D.C.: World Bank Group; 2001 [Internet edition].
[16] Persky JJ, Tam M-YS. The optimal convergence of regional incomes. Journal of Regional Science 1985;25(3):337–51.
[17] Wojan TR, Maung AC. The debate over state-level inequality: transparent method, rules of evidence and empirical power. The Review of Regional Studies 1998;28(1):63–80.
[18] Beenstock M. Country size in regional economics. In: Felsenstein D, Portnov B, editors. Regional disparities in small countries. Berlin: Springer; 2005. p. 25–46.
[19] Robinson WS. Ecological correlations and the behavior of individuals. American Sociological Review 1950;15:351–7.
[20] Portnov BA, Dubnov J, Barchana M. On ecological fallacy, assessment errors stemming from misguided variable selection and the effect of aggregation on the outcome of epidemiological study. Journal of Exposure Science and Environmental Epidemiology 2007;17:106–21 [2007].
[21] Gehlke C, Biehl K. Certain effects of grouping upon the size of the correlation coefficient in census tract material. Journal of American Statistical Association 1934;29:169–70.
[22] Openshaw S. The modifiable areal unit problem. In: Concepts and techniques in modern geography. Monograph series, vol. 38. London: Geo Books; 1984. 41 pp.
[23] Unwin DJ. GIS, spatial analysis and spatial statistics. Progress in Human Geography 1996;20(4):540–51.
[24] Gotway C, Young LJ. Combining incompatible spatial data. Journal of the American Statistical Association. 2002;97(48):632–47.
[25] 1996 reprint. In: Elliott P, Cuzick J, English D, Stern R, editors. Geographical and environmental epidemiology. methods for small area studies. Oxford: Oxford University Press; 1992. 404 pp..
[26] Scoggins A, Kjellstrom T, Fisher G, Connor J, Gimson N. Spatial analysis of annual air pollution exposure and mortality. Science of the Total Environment Apr 5 2004;321(1–3):71–85.
[27] Briant A, Combes PP, Lafourcade M. Dots to boxes: do the size and shape of spatial units Jeopardize economic geography estimates? Journal of Urban Economics 2010;67:287–302.
[28] Fotheringham AS, Brunsdon C, Charlton M. Geographically weighted regression; the analysis of spatially varying relationships. Chichester: Wiley; 2002.
[29] Fotheringham AS, Wong DWS. The modifiable areal unit problem in multivariate statistical analysis. Environment and Planning A 1991;23:1025–44.
[30] Amrhein C. Searching for the elusive aggregation effect: evidence from statistical simulations. Environment and Planning A 1995;27:105–19.
[31] Menon C. The bright side of MAUP: an enquiry on the determinants of industrial agglomeration in the United States. Working Paper. Department of Geography and Environment, London School of Economics; 2009.
[32] Beenstock M, Felsenstein D. Mobility and mean reversion in the dynamics of regional inequality. International Regional Science Review 2007;30(4):335–61.
[33] Schechtman E, Yitzhaki S. A measure of association based on Gini's mean difference. Communications in Statistics Theory and Methods 1987;A16:207–31.
[34] Hesterberg T, Monaghan S, Moore DS, Clipson A, Epstein R. Bootstrap methods and permutation tests. ch18. In: Moore DS, McCabe GP, Duckworth WM, Sclove SL, editors. The practice of business statistics. NY 18: WH Freeman and Co.; 2002. p. 4–25.
[35] Wasserman S, Bockenholt U. Bootstrapping: applications to psychophysiology. Psychophysiology 2007;26(2):208–21.
[36] Cameron CA, Gelbach JB, Miller DL. Bootstrap-based improvements for inference with clustered errors. The Review of Economics and Statistics 2008;90(3):414–27.
[37] Sanderson MJ. Confidence limits on phylogenies: the bootstrap revisited. Cladistics 2008;5(2):113–29.
[38] Sauerbrei W, Schumacher M. A bootstrap resampling procedure for model building: application to the cox regression model. Statistics in Medicine 2007;11(16):2093–109.

[39] Fewster RM, Buckland St, Siriwardena GM, Baillie SR, Wilson JD. Analysis of population trends for farmland birds using generalized additive models. Ecology 2000;81(7):1970–84.
[40] Plant C, Sorg C, Riedl V, Wohlschläger AM. Reduced regional integration in the default network of patients with very mild Alzheimer's disease detected by bootstrapping Rest-fMRI data. NeuroImage 2009;47(1):S39–41.
[41] Baltanás A, Rincón PA. Application of a cluster-bootstrapping method for identifying the dietary patterns of fish populations. Ecology of Freshwater Fish 2006;1(2):130–8.
[42] Hall MJB, Kenjegalieva K, Simper R. Accounting for environmental factors, bias and negative numbers in efficiency estimation: a bootstrapping application to the Hong Kong banking sector. Loughborough University; 2010 [Discussion Paper Series, #2010_03].
[43] English DBK. Calculating confidence intervals for regional economic impacts of recreation by bootstrapping visitor expenditures. Journal of Regional Science 2002;40(3):523–39.
[44] Enflo K, Hjertstrand P. Relative sources of European regional productivity convergence: a bootstrap frontier approach. Regional Studies 2009;43 (5):643–59.

Boris A. Portnov ("On the Suitability of Income Inequality Measures for Regional Analysis: Some Evidence from Simulation Analysis and Bootstrapping Tests") is an Associate Professor in the Department of Natural Resources and Environmental Management, Graduate School of Management, University of Haifa, Israel. Prof. Portnov is Academic Manager of the MBA Real Estate Valuation and Management Program. He earned an MA in architecture from Poltava Civil Engineering Institute, Ukraine Republic, a Ph.D. from Central Scientific and Project Institute of Town-Building, Moscow, and a D.Sc. (Second Russian Doctoral Degree) from Moscow Architectural Institute. Professor Portnov's current research program covers five interrelated aspects of population geography, urban and regional planning, and environmental studies: (1) environmental factors of real estate appraisal; (2) urban clustering; (3) environmental epidemiology; (4) interregional inequality; and (5) sustainability of urban growth in peripheral areas. Professor Portnov is an Associate Editor of the *International Journal of Sustainable Society*, and serves on the Editorial Boards of *Socio-Economic Planning Sciences; International Journal of Society Systems Science*, and *Open Family Studies Journal*. He has authored or co-authored four books and one textbook, including **Urban Clustering: The Benefits and Drawbacks of Location**, Ashgate, 2001; and **Regional Inequalities in Small Countries,** Springer, 2005. Professor Portnov's research appears in more than 100 refereed articles and book chapters, which are in a variety of journals, including Chonobiology International, Environmental Science and Policy, International Migration, Journal of Regional Science, Regional Studies, Space & Polity, Socio-Economic Planning Sciences, and *Urban Studies*. He holds, or shares, three patents in the former USSR.

Daniel Felsenstein ("On the Suitability of Income Inequality Measures for Regional Analysis: Some Evidence from Simulation Analysis and Bootstrapping Tests") is an Associate Professor in the Department of Geography, Hebrew University of Jerusalem, Israel. He is Director of the Institute of Urban and Regional Studies and Academic Director of the Center for Computational Geography. His research interests are in regional science, economic geography, and spatial econometrics. Current research topics deal with empirically testing alternative models of agglomeration, explaining the dynamics of interregional income disparities, developing a 'job-chains' model for evaluating local economic development initiatives, and the development and application of land-use simulation models. This, and related work, has appeared in five authored or co-edited books and more than 80 articles in international, peer-reviewed journals. In terms of public and professional service, Professor Felsenstein serves on the European Regional Science Association Council and the IGU Commission on the Dynamics of Economic Spaces and is a member of the Israeli National Council of Land Appraisers and Surveyors.