A New Generalized Variance Approach for Measuring Multidimensional Inequality and Poverty

Ottó Hajdu¹

Accepted: 15 May 2021 / Published online: 4 June 2021
© The Author(s) 2021

Abstract

The paper suggests a new generalized variance concept for measuring multidimensional inequality of a stratified society, based on multivariate statistical methods, where the members of society form a cloud in the oblique space of dimensions of inequality, such as income, expenditure and property. The cloud presents the multidimensional inequality capsulized in the cloud. The goal is to condense all the inequality information embodied by the cloud into a composite compact metric characterizing both the shape and the inner structure of the cloud. Contrary to the conventional literature that considers multidimensionality as a unidimensional weighted combination of the dimensions, our new composite index measures the inequality of the configuration of the points in the cloud. Our aim is twofold. First, we introduce the Inequality Covariance Matrix (ICM) assigned to the cloud, with elements measuring the correlations among dimensions. Having ICM, we propose the Generalized Variance (GV) of ICM to measure the composite Generalized Variance Inequality (GVI) level. Second, to evaluate the stratum-specific structure of the overall inequality, we suggest a new two-stage procedure. In the first stage, we divide the total GVI into between-groups and within-groups effects. Then, in the second stage the contributions of the strata to the within-groups inequality and, the contributions of the dimensions to the between-groups inequality are calculated. This GVI approach is sensitive to the correlation system, decomposable into stratum effects and, the number of dimensions is not limited. Moreover, including the log-dimensions in the analysis, GVI yields an Entropy Covariance Matrix giving a new Generalized Variance Entropy index. Finally, the GVI of censored poverty indicators means multidimensional poverty measurement. This special complex task is not yet solved in the traditional literature so far.

Keywords Generalized entropy · Generalized variance · Multidimensional inequality · Poverty · Theil-indices · Wilks · Lambda

✉ Ottó Hajdu
hajdu@gti.elte.hu

¹ Department of Comparative Economics, Institute of Business Economics, Eötvös Loránd University, Egyetem tér 1-3, 1053 Budapest, Hungary

1 Introduction

Measuring multidimensional inequality of a stratified society, first, needs defining a composite multivariate index, which is decomposable into subgroup effects generated by the stratification. The individuals of the society (households) form a *multidimensional population cloud* in the *oblique space* of the inequality axes. Such socio-economic inequality dimensions are for instance income, consumption, expenditure, property, etc. Obviously, given the stratification, the overall population cloud embodies the overall inequality present in the society, surrounded by stratum clouds representing their own inequalities. The paper distinguishes groups from clouds and, dimensions from variables.

The fundamental aim of this article is to elaborate a methodological framework for analysing the structure of inequalities in a stratified system of skewed clouds, with special regard to the between-strata and within-strata decomposition of the overall inequality. Once the inequality measurement is defined, the measurement of poverty is also, based on the theory of censored distributions.

1.1 Literary Aspects

Considering the key fields of measuring economic inequality and poverty, the fundamental contributions of the traditional literature so far, without wishing to be exhaustive, are as follows.

First, in terms of dimensionality, dimensions are latent factors for which manifest indicators are observed to define them. The inequality measuring method can be based on a one-dimensional approach on the one side, or a multi-dimensional approach on the other side. However, the one-dimensional approach essentially multivariate if the inequality-poverty indicators used are condensed into a single dimension with appropriate weights as a linear combination. But a true, pure multidimensional approach is both multidimensional and multivariate in the sense that it defines several factors (dimensions) and defines them by corresponding manifest indicators.

The next central task is to define a concise, composite measure—inequality-poverty index—that meets pre-established reasonable requirements, so-called axioms. These axioms define how a specific index required to respond to structural changes happened in the income distribution. A further key problem is to rank the % decomposition of the composite index into subgroup relative (percentage) contributions, provided a stratified population. Finally, ordering importance of factors and indicators remains essential task. Given the importance of factors, the task is twofold. On the one hand, the order of importance of the factors in relation to internal inequality needs to be established. On the other hand, the importance of indicators should be ranked within a given factor.

Below, we highlight the studies that have fundamentally influenced the relevant literature. We do this so that the contribution of the present study can be placed in the present literature. The fundamental studies highlighted for the purposes of this article are as follows.

1. For transfer sensitivity axioms see Shorrocks and Foster (1987) and, for the unit-consistent decomposability see Zheng (2005).
2. For a class of decomposable inequality measures see Foster et al. (1984) and Foster and Shneyerov (1999).

3. The most widely used unidimensional inequality indices are, in fact, the fundamental Gini index, the Sen-Shorrocks-Thon index (Shorrocks, 1995) and, the Atkinson index-family (Atkinson, 1987).
4. The use of the axiomatic approach based on a truncated population has been introduced by Sen (1976) and Cowell and Kuga (1981a, b).
5. The Hybrid Multidimensional Index is suggested in Araar (2009).
6. In the field of multidimensional extensions of inequality, concerning comparisons, relative welfare and dominance conditions are available in Duclos et al. (2006) while, comparing multidimensional indices of inequality is given in Lugo (2005).
7. For the decomposition of the Gini and the Generalized Entropy Indices, see Mussard et al. (2003), Bourguignon (1979), Aristondo et al. (2010) and Dagum (1997).
8. For application and discussion of the generalized entropy measures, see Tsui (1999), Mussard et al. (2003).
9. For a study of Multidimensional Poverty Measures from an Information Theory Perspective see Lugo and Maasoumi (2008).
10. For summarizing the sample information of poverty indicators based on a so-called data matrix see for instance Alkire and Foster (2011).

1.2 Goals and Contributions

The fundamental aim of this paper is to introduce a new concept suitable to analyse the structure of inequality considering a stratified society with outliers, forming a system of skewed clouds in the oblique inequality space defined by correlated, asymmetrical dimensions as axes. As a contrast to the traditional approaches, our concept works as a step-by-step procedure rather than a single, composite, decomposable index. The proposed Generalized Variance Inequality (*GVI*) procedure is based mainly on fundamental statistical tools of multivariate discriminant analysis.

The *GVI* approach is a two-stage procedure. First, the total population inequality is measured and divided into a within-clouds + between-clouds structure (e.g. 80% + 20%). Then, secondly, we allocate the 80% within-clouds inequality to strata and the 20% between-clouds inequality to dimensions.

To measure the composite overall degree of the population inequality, we define the *Inequality Covariance Matrix* (ICM) which consists of all the pairwise covariances among inequality dimensions. This ICM essentially a pairwise metric of the multivariate dispersion. The reason for defining ICM is twofold. On the one hand, its determinant measures the multivariate *Generalized Variance* dispersion in the cloud. On the other hand, provided a stratified society, the overall population covariance matrix is additively decomposed according to the stratification.

Based on these two properties, as a basic result, the paper proposes the *GV* generalized variance metric to measure the composite, multivariate degree of inequality (dispersion) of any multidimensional cloud. Hence, in the case of a stratified population, we have an overall *GV* measure for the population on the one side and, separated single *GVs* for the sub-clouds on the other side. As mentioned, the final task is to allocate the within-clouds inequality among strata and, the between-clouds inequality among dimensions. Apparently, the number of dimensions in this task is unlimited. In this way both the dimensions and the strata can be ranked according to their importance in explaining inequality based on two types of statistical tests: testing the equality of the stratum-specific generalized variances and, testing the equality of the stratum-specific centroids.

Further, having the log-dimensions included among dimensions leads to the *interpretation of the covariance* between the so-called *relative income* and its *logarithm*. This covariance is (shown by this paper) the *sum* of the *Theil-Redundancy Index (TRI)* and the *Mean Logarithmic Deviation (MLD)* index, connecting the *GVI* approach with *Entropy Theory*. Consequently, the *Inequality Covariance Matrix* contains the covariance and the variances of dimension “income” and its logarithm therefore termed *Entropy Covariance Matrix* denoted by ECM.

2 Highlighted Properties and Contributions of the GVI Procedure

Provided a stratified population of multidimensional clouds, *GVI* gives the *between-clouds + within-clouds* decomposition of the *population inequality*. In addition, the stratum-specific contributions to the *average within-clouds inequality* and, the dimension-specific contributions to the *between-clouds inequality* are also computed. The *GVI* is purely multidimensional because it measures clouds, rather than arbitrary, univariate, weighted sums of dimensions. The proportion of the total population inequality that is not explained by the stratification is reported by the standard Wilks’ Lambda ratio. Based on the group-specific contributions of the strata to the *average within-clouds inequality* (computed via the standard Box-M statistic) it is possible to test the equality of stratum-specific covariance matrices. Both the dimensions and the strata are ranked according to their importance in explaining inequality. The *GVI* method is sensitive to the skewness characteristics of a cloud simultaneously: the *correlations* among the dimensions and, the distortions due to distributional *asymmetry* with outliers. This sensitivity is necessary when a symmetrical distribution is required for statistical inference.

For the sake of clear computations and interpretations, the paper provides computational results based on artificial data and, besides, empirical results based on Hungarian Households’ expenditure data surveyed by the Hungarian Central Statistical Office in the year of 2003. The computations and plots are carried out by the means of the SPSS and Statistica, and R packages.

The structure of the article is as follows. Section 2 introduces the new Generalized Variance Inequality *GVI* concept, yielding the *GVI* inequality and *GVP* poverty indices based on the idea of the *Inequality Clouds* and, the *Inequality Covariance Matrix* ICM defined in the section. Section 3 gives the Multivariate Discriminant Analysis of the Generalized Variance Inequality, with special regard to the between-groups and within-groups sub-group decomposition, for the case of a stratified population. The hypothesis of homogeneous variances and equal centroids are tested. A three-dimensional empirical case study is presented. Section 4 introduces the Generalized Variance Entropy concept *GVE*, based on the *Entropy Covariance Matrix* ECM. The purpose of this section is to connect the *GVI* concept with the standard Entropy Concept of Information Theory to develop the *GVE* concept. Section 5 suggests—defines—the within-groups *Censored Head-Count Ratio* poverty application based on the *GVI* procedure. Section 6 is about the limitations of the method proposed. Finally, Sect. 7 concludes.

Fig. 1 Generalized Variance. Source: Own elaboration (based on Sharma, 1996)

3 The Generalized Variance Inequality Concept: GVI

Following the fundamental principle that inequality is essentially a special aspect of dispersion (Tsui, 1999), this paper suggests the *generalized variance* (GV) multivariate metric to measure the degree of inequality, introducing the *GVI* procedure. The generalized variance means the *determinant* of the covariance matrix **C** of the multivariate data applied. Depending on the features of the data set, the *GVI* methodology results in several inequality-poverty methods. The main features of the *GV* metric concerning the *GVI* approach are as follows.

3.1 Generalized Variance

For simplicity, let us consider a society considering only the household income (*Y*) and expenditure (*X*). These dimensions are obviously correlated with one another and the individuals of society belong to an area, bounded by a parallelogram presented in Fig. 1, plotting *X* and *Y* as points in the space spanned by the individual members. If *X* and *Y* are collinear (they coincide), then the alpha angle is zero. This case indicates perfect redundancy of *X* and *Y*. In the other extreme case *X* and *Y* are orthogonal, with zero redundancy in the data. Therefore, the area of the parallelogram counts for redundancy in the data set. The squared area is used to measure the generalized variance *GV*.

The angle α indicates the intensity of correlation between *X* and *Y* because $\cos(\alpha) = \text{Corr}(X, Y)$. The generalized variance equals the *determinant* of the covariance matrix:

$$C_{(2,2)} = \begin{matrix} & \text{Dimension} & X & Y \\ \begin{matrix} X \\ Y \end{matrix} & & \begin{matrix} \sigma_X^2 = \text{Var}_X \\ C_{Y,X} \end{matrix} & \begin{matrix} C_{Y,X} \\ \sigma_Y^2 = \text{Var}_Y \end{matrix} \end{matrix} \quad (1)$$

The *GV* squared area is computed as follows:

$$GV = \det(C_{(2,2)}) = \text{Var}_X \cdot \text{Var}_Y - C_{X,Y}^2 \quad (2)$$

In general, the *GV* determinant is bounded, because the determinant of any positive definite covariance matrix is bounded. Its lower and upper bounds are based on the *squared canonical correlation coefficient*

$$0 \leq R^2 = \frac{C_{X,Y}^2}{\text{Var}_X \cdot \text{Var}_Y} \leq 1 \quad (3)$$

from which

$$0 \leq GV = 1 - R_{X,Y}^2 \leq \text{Var}_X \cdot \text{Var}_Y \quad (4)$$

follows.

Clearly, an increase in the population inequality increases the $\text{Var}_X \cdot \text{Var}_Y$ upper bound but the squared covariance reduces it to the extent of redundancy due to the correlation. In this context the GV metric yields the so-called *generalized variance inequality GVI* index.

Moreover, the GV serves also to measure poverty. As mentioned, *poverty* can be measured using *censored Y^c* distributions where the individual incomes, greater than the poverty line, are replaced with the line level (see Hamada & Takayama, 1978):

$$Y_{ji}^c = \min\{Y_{ji}, l_j\}, \quad i = 1, 2, \dots, n \quad (5)$$

where y_{ji} is the *level* of individual “ i ” in the “ j ” poverty dimension with l_j poverty line.

The use of censored distribution ensures that any income movements among those living above the poverty line leave the poverty level unchanged if the set of the poor remains unchanged. Based on censored distribution, all poverty information is retained even the population size as well. So, it is reasonable to apply the GV index to a censored population to compute the level of poverty. As a result, in the case of using *censored distributions* as inequality dimensions, GV counts for the *multivariate degree of multidimensional poverty* automatically. In this context GV yields the generalized variance poverty index: GVP .

3.2 Inequality Clouds

Unlike groups, a cloud is not only a set of points in the space of interest, but also, it presents the whole configuration of the members constituting the population, with nearest and farthest neighbours, similar clusters, extreme outliers, etc. The reason for using clouds while measuring inequality is to map both the surface topography and the internal structure of the society.

Let us expand the set of dimensions with the households’ properties: *income*, *expenditure*, and *property*. These correlated latent inequality axes exhibit asymmetrical densities forming *skewed clouds*, containing distorting, extreme, outlier observations as well. Such a cloud termed skewed, presents the multidimensional inequality capsulized in the cloud.

Using factor analysis terminology, dimensions are factors, measured by observable, manifest, proxy indicators having strong loading coefficients, such as annual per capita income, monthly average expenditure, the price of apartment owned, the number of durables, cars, etc.

Figure 2 illustrates the cloud of Households considering socio-economic dimensions.

Our primary aim is to condense all the inequality information embodied by the cloud into a composite compact metric characterizing both the shape and the inner structure of the cloud. Contrary to the conventional literature that considers multidimensional inequality as a unidimensional weighted combination of the dimensions, our new composite index measures the inequality of the configuration of the points in the cloud. This cloud-based approach ensures that the skewness is not smoothed out from the inequality tendencies.

Fig. 2 Population inequality cloud of Households. *Source:* Own elaboration, using the Statistica package

This is not the case when a multivariate case is simply reduced to a weighted unidimensional approach.

The covariance matrix C is termed *Inequality Covariance Matrix* and denoted by ICM. The structure of ICM in this 3-dimensional case takes the form:

$$C_{(3,3)} = \text{ICM}_{(3,3)} = \begin{array}{ccccc} & \text{Dimension} & X & Y & P \\ \begin{array}{c} X \\ Y \\ P \end{array} & & C_{X,X} & C_{Y,X} & C_{P,X} \\ & & & C_{Y,Y} & C_{P,Y} \\ & & & & C_{P,P} \end{array} \quad (6)$$

where P stands for *property*. The geometric meaning of the determinant is the volume of the cloud. Apparently, the number of dimensions in ICM is not limited, it can be expanded at will. The *GVI* determinant, in general, is computed by the product of the λ eigenvalues of C .

4 Discriminant Analysis of the Generalized Variance Inequality¹

Let us stratify the population into $g = 1, 2, \dots, G$ strata based on socio-economic categorical variables in order to decompose the total population inequality into stratum effects, where the stratum-specific sub-clouds surround the central population cloud.

¹ As a comparison, for an overview of the decomposition properties of the Gini and the Generalized Entropy indices see Mussard et al. (2003).

Fig. 3 Stratum-specific clouds surrounding the central population cloud. *Source:* Own elaboration, using the Statistica package

Returning to the Households example, Fig. 3 shows such a cloud of clouds, using the settlement type as stratification variable with four outcomes: Capital, Towns, Cities and Villages. The question is, what factors and to what extent contribute to this stratified inequality structure that is, in other words, how to interpret and measure the dispersion of clouds.

It is apparent from Fig. 3 that the overall population inequality (centred in the Figure) is divided into two components of the between-strata and the within-strata inequalities. From socio-economic point of view, our focus is on the ratio and the stratum-ordered distribution of the average, i.e. within-strata inequality. Next, the explanatory dimension-ordered explanatory contributions to the between-strata inequality are of interest.

4.1 The Multivariate Within-Groups and Between-Groups Decomposition

The decomposition of the overall inequality of a stratified population into between-groups and within-groups components, is based on the additive decomposability of any covariance matrix as follows:

$$\mathbf{C}_{Total} = \mathbf{C}_{Between} + \mathbf{C}_{Within} \quad (7)$$

where the *within-groups* covariance matrix is the weighted average of the sub-group covariance matrices:

$$\mathbf{C}_{Within} = \sum_{g=1}^G \frac{n_g}{n} \mathbf{C}_g \quad (8)$$

where n_g stands for the size of group g and $\mathbf{C}_g$ stands for the covariance matrix of the group. The variables of a *stratified* covariance matrix termed *discriminator* variables.

4.1.1 Homogeneity Analysis

Given a socio-economic stratification, the basic goal to use $\mathbf{C}_{within}$ is to test the *homogeneity* of the group-specific covariance matrices. The H_0 null hypothesis of the equality of the covariance matrices is

$$H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_G \quad (9)$$

where Σ_g denotes the group-specific hypothetical covariance matrix in the population. H_0 equivalently states, that the hypothetical Generalized Variances, hence, the group specific *GVI* values are also equal in the population. Acceptance of the null hypothesis concludes that the inequalities per group are the same. In this case, there is no need to analyse any within-groups structure.

The standard method to test H_0 is based on the *Box-M likelihood-ratio* statistic²:

$$M = \sum_{g=1}^G (n_g - 1) [\text{Indet} \mathbf{C}_{Within} - \text{Indet} \mathbf{C}_g] \quad (10)$$

The Box-M is a weighted sum of -2LogLikelihood differences. Based on additivity of the structure, the paper suggests the distribution of the categories in M for computing the percentage group-specific contributions to the within-groups inequality.

4.1.2 Analysis of Variance

Let us continue with testing the between-groups component. Given the stratification, the null hypothesis tests the equality of the group-specific centroids by the means of the Multivariate Analysis of Variance (MANOVA) method:

² For the statistical hypothesis behind the M statistic and the procedure how to test it see e.g. Sharma (1996).

Table 1 Descriptive statistics of the proxy variables. *Source:* Own elaboration from HCSO data, using the SPSS package

Statistic	Income	Expenditure	Property
Mean	855,02	720,50	4,86
Median	752,61	609,02	3,74
Skewness	3,968	3,161	3,722
Kurtosis	35,670	20,057	29,102
<i>Percentiles</i>			
25	594,62	450,35	2,03
50	752,61	609,02	3,74
75	990,41	851,79	6,06

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_G \quad (11)$$

where μ_g denotes the group-specific hypothetical centroid in the population.

The test is based on the proportion of the total inequality unexplained by the groups is measured by the standard Wilks' Λ ratio computed as the within-groups generalized variance divided by the total generalized variance:

$$\text{Wilks}' \Lambda = \frac{\det C_{\text{Within}}}{\det C_{\text{Total}}} \quad (12)$$

Because the additive decomposition for the determinant of the total covariance matrix is not held,³ the proportion of inequality *not explained* by the stratification variable is: $1 - \text{Wilks}' \Lambda$.

Therefore, H_0 states that Wilks' $\Lambda = 1$ while H_1 concludes that Wilks' $\Lambda < 1$ significantly.

Acceptance of the null hypothesis concludes that the group centroids are the same, that is the only source of inequalities is the within-groups effect. Two basic factors reduce the value of Lambda. Increasing the number of groups on the one hand or increasing the number of discriminant dimensions for a given number of groups. However, the dimensions defining the covariance matrix have different discriminating power in the formation of the Lambda. The priority ranking is given by a stepwise procedure, in which the first step is to find the discriminator that reduces Lambda the most, then the next most, and so on, until the procedure stops. The decrease can be statistically tested step by step.

So, practically, we divide the total population inequality into a between-strata and within-strata decomposition, to have the Wilks' Lambda ratio of the total population inequality. Next, we present the percentage distribution of the strata and, the importance ranks of the dimensions in the Wilks' Lambda ratio inequality.

4.2 A 3-Dimensional Empirical Application⁴

Consider the stratified population of Hungarian Households ($n = 3,138,330$, representing the total population in 2003) with their *per capita annual income* (HUF, thousands), *annual per capita expenditure* (HUF, thousands) and the *per capita current property*

³ $\det C \neq \det C_{\text{Between}} + \det C_{\text{Within}}$.

⁴ The source of the data: Hungarian Central Statistical Office, 2003.

Table 2 Pooled within-groups matrices. *Source:* Own elaboration from HCSO data, using the SPSS package

Matrix/variables	Income	Expenditure	Property
<i>Covariance</i>			
Income	186,246.114	126,225.145	584.547
Expenditure	126,225.145	176,853.678	665.065
Property	584.547	665.065	17.394
<i>Correlation</i>			
Income	1.000	0.695	0.325
Expenditure	0.695	1.000	0.379
Property	0.325	0.379	1.000

(*HUF, millions*). The stratification is the settlement type, distinguishing four categories, with the following relative frequency distribution:

$$\{\text{Budapest (17.4\%), Large towns (22.7\%), Other cities (25.8\%), Villages (34.1\%)}\}. \quad (13)$$

Table 1 presents the descriptive statistics of the proxy variables and Table 2 shows the pooled, that is the average within-groups covariance and correlation matrices.

It is apparent, that the pooled correlations are all significant but, as it is expected, the correlation between *Income* and *Expenditure* is the strongest: 0.695. Interestingly, *Property* correlates slightly more strongly with *Expenditure* (0.379) than *Income* (0.325). The magnitude of the covariances is not of interest to the study, their role is of a methodological nature in the measurement of multivariate variance.

4.2.1 Measuring Inequality

The generalized variance inequality concept for this 3-D study yields the 0.823 Wilks' Λ ratio.⁵ Thus, the proportion of the total inequality explained by the four categories is 17.7% and the intensity of the correlation between the categories and the dimensions is $Rho = \sqrt{0.177} = 0.421$.

The hypothesis of the equality of the population stratum covariance matrices is

$$H_0 : \Sigma_{\text{Budapest}} = \Sigma_{\text{Towns}} = \Sigma_{\text{Cities}} = \Sigma_{\text{Villages}} \quad (14)$$

with $\ln(\det(\mathbf{C}_{\text{Within}})) = 26.249$ and $\text{Box-M} = 1,295,244.2$ with practically zero (0.000) p -value, indicating *heterogeneity* (H_1) of the covariance matrices. The stratum-specific log-determinants are as follows:

$$\ln \det(\mathbf{C}_{\text{Budapest}}) = -1.827 \rightarrow 26.249 - \ln \det(\mathbf{C}_{\text{Budapest}}) = -2.382$$

$$\ln \det(\mathbf{C}_{\text{Towns}}) = -3.304 \rightarrow 26.249 - \ln \det(\mathbf{C}_{\text{Towns}}) = 0.427$$

$$\ln \det(\mathbf{C}_{\text{Cities}}) = -3.801 \rightarrow 26.249 - \ln \det(\mathbf{C}_{\text{Cities}}) = 0.496$$

$$\ln \det(\mathbf{C}_{\text{Villages}}) = -4.460 \rightarrow 26.249 - \ln \det(\mathbf{C}_{\text{Villages}}) = 1.768$$

⁵ The magnitude of GV is irrelevant from the group-specific contributions point of view.

Table 3 Importance of dimensions in the between-groups inequality. *Source:* Own elaboration from HCSO data, using the SPSS package

Step	Variables entered	Wilks' Lambda				Exact F			
		Statistic	df1	df2	df3	Statistic	df1	df2	Sig
1	Property	0,855	1	3	3,138,326,000	177,908,616	3	3,138,326,000	0,000
2	Expenditure	0,826	2	3	3,138,326,000	104,974,209	6	6,276,650,000	0,000
3	Income	0,823	3	3	3,138,326,000				

Considering the importance of the settlement types in the within-groups inequality, “*Budapest*” has the smallest score -2.382 while the “*Villages*” the largest score 1.768. Let us **convert** the settlement-effects into a **distance preserving scale with positive scores** only, choosing “*Budapest*” as baseline category. The paper suggests the following scheme to make the category-distance preserving scale⁶:

$$\begin{aligned} Budapest &= 2.382, Towns = 5.191, Cities = 5.260, Villages = 6.533 \\ \text{where, for instance : } 6.533 &= 1.768 + 2 \times \text{abs}(-2.382). \end{aligned}$$

As a result, the decomposition of the *total* inequality is as follows:

$$100\% = 17.7\%_{\text{Between-groups inequality}} + 82.3\%_{\text{Within-groups inequality}} \quad (15)$$

where

$$82.3\%_{\text{Within-groups inequality}} = 8\%_{\text{Budapest}} + 22.8\%_{\text{Towns}} + 26.2\%_{\text{Cities}} + 43\%_{\text{Villages}}.$$

While the contribution of the within-groups inequality to the total inequality is 82.3 percent, 17.7 percent remains for the between-groups inequality.

The key question is the contributions of the categories to the 82.3 percent inequality. The weighted proportional linear contributions of the categories to the within-groups effect, using the weights from Eq. (13) are:

$$M_{\text{Budapest}} = \frac{0.174 \cdot 2.382}{0.174 \cdot 2.382 + 0.227 \cdot 5.191 + 0.258 \cdot 5.260 + 0.341 \cdot 6.533} = 8\% \quad (16)$$

$$M_{\text{Towns}} = \frac{0.227 \cdot 5.191}{0.174 \cdot 2.382 + 0.227 \cdot 5.191 + 0.258 \cdot 5.260 + 0.341 \cdot 6.533} = 22.8\% \quad (17)$$

$$M_{\text{Cities}} = \frac{0.258 \cdot 5.260}{0.174 \cdot 2.382 + 0.227 \cdot 5.191 + 0.258 \cdot 5.260 + 0.341 \cdot 6.533} = 26.2\% \quad (18)$$

$$M_{\text{Villages}} = \frac{0.341 \cdot 4.176}{0.174 \cdot 2.382 + 0.227 \cdot 5.191 + 0.258 \cdot 5.260 + 0.341 \cdot 6.533} = 43\% \quad (19)$$

⁶ The reason for this rule is that the distance of the base-line category from the zero and the distances between adjacent categories remain unchanged.

Table 4 Pooled within-groups censored matrices. *Source:* Own elaboration from HCSO data, using the SPSS package

	cPcIncome	cPcExpenditure	cPcProperty
<i>Covariance</i>			
cPcIncome	786,698	322,970	3,084
cPcExpenditure	322,970	781,792	4,325
cPcProperty	3,084	4,325	0,250
<i>Correlation</i>			
cPcIncome	1,000	0,412	0,220
cPcExpenditure	0,412	1,000	0,309
cPcProperty	0,220	0,309	1,000

*cPc stands for “censored Per capita”

Thus, *Budapest* gives the smallest 8% contribution to the average within-strata inequality, while the “*Villages*” generate the largest proportion 43%.

The other key problem is to order dimensions according to their contributions to the development of the between-strata inequality. The solution is carried out by the means of a stepwise algorithm testing successive reductions in the *Wilks’* Λ due to the gradual entry of each dimension into the evolution of inequality. The results of each step are shown in Table 3.

In this stepwise manner, at each step, the variable that minimizes the overall *Wilks’ Lambda* is entered. It is apparent, that the strata are scattered mostly in terms of *Property*, followed by household *Expenditure* and the order is closed by the *Income*. Besides, based on the Sig = 0.000 values of the F-test, the decrease in *Wilks-Lambda* was significant in both steps.

4.2.2 Measuring Poverty: The Generalized Variance of Censored Dimensions

As mentioned earlier, economic *poverty* can be measured using *censored* Takayama-type Y^c distribution where “the individual Y incomes, greater than the poverty line, are replaced with the line level”. As a result, *GVI* yields the degree of *multidimensional poverty* as follows.

The conventional definition of the poverty line in the literature is the application of 60% of the median level. This study also uses this method. The per capita thresholds applied are:

$$Income = 451.6, Expenditure = 365.4, Property = 2.2.$$

Table 4 presents the Pooled Within-Groups Censored Matrices.

Using the censored variables, the analogue computational results in this censored study are as follows. The *GVI* approach results in a 91.5% *Wilks’* Λ ratio, yielding 8.5 percent proportion of the total poverty explained by the four categories. The homogeneity hypothesis of the censored covariance matrices in the population is

$$H_0 : \Sigma_{Budapest}^c = \Sigma_{Towns}^c = \Sigma_{Cities}^c = \Sigma_{Villages}^c \quad (20)$$

with $\ln(\det(\mathbf{C}_{Within})) = 11.646$. The *Box’s* $M = 1,652,263$ with Sig. = 0.000 significance *P*-value, thus, we reject the null hypothesis that the censored generalized variances are equal.

The stratum-specific log-determinants are:

$$\begin{aligned}\ln \det(\mathbf{C}_{\text{Budapest}}) &= 9.375 \rightarrow 11.646 - \ln \det(\mathbf{C}_{\text{Budapest}}) = -2.271 \\ \ln \det(\mathbf{C}_{\text{Towns}}) &= 8.949 \rightarrow 11.646 - \ln \det(\mathbf{C}_{\text{Towns}}) = -2.697 \\ \ln \det(\mathbf{C}_{\text{Cities}}) &= 11.835 \rightarrow 11.646 - \ln \det(\mathbf{C}_{\text{Cities}}) = 0.189 \\ \ln \det(\mathbf{C}_{\text{Villages}}) &= 12.917 \rightarrow 11.646 - \ln \det(\mathbf{C}_{\text{Villages}}) = 1.271.\end{aligned}$$

The category-distance preserving scale of the settlement-effects *with positive scores*, choosing “Towns” as the baseline category is

$$\text{Budapest} = 3.122, \text{Towns} = 2.697, \text{Cities} = 5.582, \text{Villages} = 6.665.$$

As a result, the decomposition of *total* inequality-poverty is as follows:

$$100\% = 8.5\% \text{Between-groups inequality} + 91.5\% \text{Within-groups inequality}$$

where

$$91.5\% \text{Within-groups inequality} = 11.2\% \text{Budapest} + 12.6\% \text{Towns} + 29.6\% \text{Cities} + 46.7\% \text{Villages}.$$

In contrast to the case of inequality decomposition, *Budapest* contributes the least to the poverty (11.2%) and the *Villages* the most (46.7%).

5 The Generalized Variance Entropy Concept: GVE

The purpose of this section is to connect the *GVI* concept with the standard Entropy Concept of Information Theory to develop the *GVE* concept. As a starting point, let us define the relative $\mathbf{y}$ income distribution of n individuals as

$$y_i = \frac{Y_i}{\bar{Y}} \quad (i = 1, 2, \dots, n) \quad (21)$$

where the individual income Y_i is expressed as a percentage of the average $\bar{Y}$ income. Using these notations, the two well-known *Theil-type* unidimensional income inequality indices are as follows.⁷

First, the *Theil Redundancy Index* is

$$0 \leq TRI = \frac{1}{n} \sum_{i=1}^n y_i \ln(y_i) \leq \ln(n) \quad (22)$$

and, the Mean Logarithmic Deviation is

$$MLD = -\frac{1}{n} \sum_{i=1}^n \log(y_i) \geq 0 \quad (23)$$

The meaning of *TRI* is the standard redundancy measure of the information theory while, *MLD* can be interpreted as the logarithmic *approximation* of the *average* $(y_i - 1)$ gain/

⁷ See Theil (1967).

loss and, its value is clearly non-negative. The simple average $(TRI + MLD)/2$ is commonly used in the literature to measure the bivariate but unidimensional degree of income inequality. The latent dimension here is the Y income level and the two manifest proxy variables are y and $\log(y)$, respectively.

5.1 Entropy Covariance Computations

Since the Generalized Variance Entropy (GVE) index is based on the fundamental covariance meaning of the $(TRI + MLD)$ measure, a brief discussion of this sum is essential. For this reason, first, let us consider the **covariance** between the relative income and its logarithm:

$$C = Cov(y, \log(y)) \quad (24)$$

A fundamental result of this paper is that the covariance C can be decomposed into the sum of TRI and MLD :

$$C = TRI + MLD \quad (25)$$

based on that C can be expressed in the form

$$C = \underbrace{\frac{1}{n} \sum_{i=1}^n y_i \log(y_i)}_{TRI} - \underbrace{mean(\log(y)) \cdot mean(y)}_{MLD} \quad (26)$$

Here we used the fact that the average relative income is 1. Obviously, both TRI and MLD increase due to a *regressive transfer* when a positive amount of income is reallocated from a person to a richer person in the society. Consequently, C also increases in this situation. In addition, TRI and MLD are sensitive to the *size* and *location* of the transfer in the distribution, thus, C inherits this property as well.

Based on that C measures inequality, it is straightforward to extend the covariance measurement to a covariance matrix approach, defining the *entropy covariance matrix* ECM of y and $\log(y)$. The entries of ECM are as follows:

$$ECM = \begin{array}{cc} \begin{array}{c} \text{Variable} \\ y \\ \log(y) \end{array} & \begin{array}{cc} y & \log(y) \\ Var_y & C \\ C & Var_{\log(y)} \end{array} \end{array} \quad (27)$$

All elements of ECM are information theory-based inequality measures: $C = TRI + MLD$, Var_y is the variance of the relative incomes and $Var_{\log(y)}$ is the variance of the logarithms of the relative incomes. Clearly, the composite GV determinant of ECM denoted by GVE is:

$$GVE = Var_y \cdot Var_{\log(y)} - C^2 \quad (28)$$

5.2 Equivalent Entropy Covariance Matrix Structures

The ECM can be re-written in several equivalent alternative forms based on standard inequality indices. Hence, GVE can also be calculated in several ways based on Eq. (28).

First, using the *TRI+MLD* decomposition of C and the *squared coefficient of variation* $V^2_Y = Var_y$ and, equation $Var_{\log(y)} = Var_{\log(Y)}$, the ECM takes the form:

$$ECM = \begin{array}{ccccc} \text{Variable} & y & \log(y) & & \\ y & V_Y^2 & TRI + MLD & & \\ \log(y) & TRI + MLD & Var_{\log(Y)} & & \end{array} \quad (29)$$

Next, based on the *Hirschman-Herfindahl* $HH = \sum_i^n y_i^2$ index, ECM can be written as:

$$ECM = \begin{array}{ccccc} \text{Variable} & y & \log(y) & & \\ y & n \cdot HH - 1 & TRI + MLD & & \\ \log(y) & TRI + MLD & Var_{\log(Y)} & & \end{array} \quad (30)$$

Finally, using the so-called *Generalized Entropy* $GE(\alpha)$ parametric index-family of inequality:

$$ECM = \begin{array}{ccccc} \text{Variable} & y & \log(y) & & \\ y & 2GE(2) & GE(1) + GE(0) & & \\ \log(y) & GE(1) + GE(0) & Var_{\log(Y)} & & \end{array} \quad (31)$$

where

$$GE(\alpha) = \frac{1}{n\alpha(\alpha-1)} \sum_{i=1}^n [(y_i)^\alpha - 1], \quad \underbrace{\alpha \neq 0, \alpha \neq 1}_{L' \text{ Hospital Rule}} \quad (32)$$

For large α , $GE(\alpha)$ is especially sensitive to the existence of large incomes, whereas for small α the index is especially sensitive to the existence of small incomes.

As a conclusion, the elements of ECM are functions of the $GE(0)$, $GE(1)$, $GE(2)$ *generalized entropy* indices and, further, the *variance of log-income* which is a function of the *HH-index*.

5.3 The Multivariate Extension of ECM

Let us extend the *number of dimensions* at this stage. The dimensions are per capita *income*, *expenditure*, and *property*. In this 3-dimensional approach the structure of ECM is

$$ECM_{(6,6)} = \begin{array}{ccccccc} \text{Variable} & x & y & p & \log(x) & \log(y) & \log(p) \\ x & C_{x,x} & & & & & \\ y & C_{y,x} & C_{y,y} & & & & \\ p & C_{p,x} & C_{p,y} & C_{p,p} & & & \\ \log(x) & C_{\log(x),x} & C_{\log(x),y} & C_{\log(x),p} & C_{\log(x),\log(x)} & & \\ \log(y) & C_{\log(y),x} & C_{\log(y),y} & C_{\log(y),p} & C_{\log(y),\log(x)} & C_{\log(y),\log(y)} & \\ \log(p) & C_{\log(p),x} & C_{\log(p),y} & C_{\log(p),p} & C_{\log(p),\log(x)} & C_{\log(p),\log(y)} & C_{\log(p),\log(p)} \end{array} \quad (33)$$

where “ x ” stands for relative expenditure “ y ” for relative income, and “ p ” for relative property.

This *multivariate-multidimensional* ECM in addition to the *within-dimensional* covariances also contains the *cross-dimensional* covariances, such as $C_{\log(x),y}$. This covariance between $\log(x)$ and y has a special interpretation. Notice, that both variables express

differences measured on percentage scales. Consider a simple linear regression between the two variables. Then the meaning of the covariance is the *standardized slope coefficient* of this regression, regardless which one is the dependent variable. Further, because of using logs, the problem of distributional frequency asymmetry and the impact of outlier cases are smoothed.

5.4 The ECM measurement of Inequality

Consider now again the stratification of the Hungarian Households: *Budapest, Towns, Cities, Villages*. The ECM matrix of order (6,6) has a 0.779 Wilk's Λ ratio. Hence, the contribution of the within-groups inequality to the total inequality is 77.9 percent, with 22.1 percent remained for the between-groups inequality. The canonical correlation between the categories and the dimensions is $Rho = \sqrt{0.221} = 0.47$.

The hypothesis of the equality of the ECM covariance matrices in the population is

$$H_0 : \Sigma_{Budapest}^{ECM} = \Sigma_{Towns}^{ECM} = \Sigma_{Cities}^{ECM} = \Sigma_{Villages}^{ECM} \quad (34)$$

with $\ln\{\det(ICM_{Within})\} = -11.875$. The *Box-M* equals 3,370,667.9 with practically zero *P*-value. Hence, the group-specific generalized variances (inequalities) significantly differ from each other. The stratum-specific log-determinants are:

$$\ln \det(ICM_{Budapest}) = -8.719 \rightarrow -11.875 - \ln \det(ICM_{Budapest}) = -3.156$$

$$\ln \det(ICM_{Towns}) = -14.053 \rightarrow -11.875 - \ln \det(ICM_{Towns}) = 2.178$$

$$\ln \det(ICM_{Cities}) = -12.829 \rightarrow -11.875 - \ln \det(ICM_{Cities}) = 0.954$$

$$\ln \det(ICM_{Villages}) = -14.464 \rightarrow -11.875 - \ln \det(ICM_{Villages}) = 2.589$$

Considering the importance of the settlement types in the within-groups inequality, apparently, "*Budapest*" has the smallest score -3.156 while, the "*Villages*" the largest score 2.589. The settlement-effects with *positive scores*, choosing *Budapest* as the baseline category is

$$Budapest = 3.156, Towns = 8.49, Cities = 7.266, Villages = 8.901.$$

As a result, the decomposition of the *total* inequality is as follows:

$$100\% = 22.1\%_{Between-groups\ inequality} + 77.9\%_{Within-groups\ inequality}$$

where

$$77.9\%_{Within-groups\ inequality} = 7.4\%_{Budapest} + 26.1\%_{Towns} + 25.4\%_{Cities} + 41.1\%_{Villages}.$$

Thus, *Budapest* has the lowest 7.4% share in the development of the average within-strata inequality while the *Villages* has the highest 41.1 percent.

5.5 The ECM Measurement of Poverty

Consider now the measurement of poverty of Hungarian Households, with the same dimensions (*income, expenditure, property*) and stratification (*Budapest, Towns, Cities, Villages*) applied before. The ECM to be analysed is the $ECM_{(6 \times 6)}^c$ covariance matrix of the

censored per capita data, where the poverty lines are fixed at the 60% of the corresponding median levels for each dimension.

The null hypothesis of this study is

$$H_0 : \Sigma_{Budapest}^{ECM^c} = \Sigma_{Towns}^{ECM^c} = \Sigma_{Cities}^{ECM^c} = \Sigma_{Villages}^{ECM^c} \quad (35)$$

Based on discriminant analysis results, the *Wilk's lambda* ratio is 0.886, with its canonical correlation of $Rho = (1 - 0.886)^{1/2} = 0.338$. According to this figure, the variance *unexplained* by the settlement type in this 3-dimensional poverty case is 88.6 percentage and the correlation between the categories and the dimensions is 0.338. The remaining correlation is due to other socio-economic factors.

The *Box-M* equals 7,224,081 with practically zero *P*-value and $\ln\{\det(ICM_{Within})\} = -38.784$. Hence, the group-specific generalized variances (inequalities) significantly differ from each other. The stratum-specific log-determinants are:

$$\ln \det(ICM_{Budapest}) = -43.600 \rightarrow -38.784 - \ln \det(ICM_{Budapest}) = 4.816$$

$$\ln \det(ICM_{Towns}) = -47.587 \rightarrow -38.784 - \ln \det(ICM_{Towns}) = 8.803$$

$$\ln \det(ICM_{Cities}) = -39.961 \rightarrow -38.784 - \ln \det(ICM_{Cities}) = 1.177$$

$$\ln \det(ICM_{Villages}) = -36.317 \rightarrow -38.784 - \ln \det(ICM_{Villages}) = -2.467$$

Considering the importance of the settlement types in the within-groups inequality, apparently, *Villages* has the smallest score -2.467 while, *Towns* the largest score 8.803. The settlement-effects *with positive scores*, choosing *Villages* as the baseline category is

$$Budapest = 9.75, Towns = 13.736, Cities = 6.111, Villages = 2.467.$$

As a result, the decomposition of the *total* inequality is as follows:

$$100\% = 11.4\%_{Between-groups\ inequality} + 88.6\%_{Within-groups\ inequality}$$

where

$$88.6\%_{Within-groups\ inequality} = 23.5\%_{Budapest} + 43.1\%_{Towns} + 21.8\%_{Cities} + 11.6\%_{Villages}.$$

Thus, *Villages* has the lowest 11.6% share in the development of the average within-strata inequality while the *Towns* has the highest 43.1 percent.

6 The Censored Within-Groups Head-Count Ratio

For a poverty measurement methodological application of the GVI concept, let us consider the artificial data of 100 individuals censored at poverty line of 30:

$$y^c = \{1, 2, \dots, 29, 30 | 30, 30, \dots, 30\}.$$

The structure of the censored ECM^c matrix is as follows⁸:

⁸ For the overview of the classic one-dimensional poverty indices see Foster and Sen (1997), Zheng (1997). The underlining, core unidimensional poverty indices are Watts (1968), Sen (1976), Anand (1977), Hamada and Takayama (1978), Thon (1979), Kakwani (1980), Takayama (1979), Clark et al. (1981), Chakravarty (1983), Blackorby and Donaldson (1984), Foster et al. (1984), Hagenaars (1987), Atkinson (1987), Shorrocks (1995).

$$\text{ECM}^c = \begin{array}{cc} \text{Variable} & y^c \\ y^c & \text{Var}_y^c = 0.10127 \\ \ln(y^c) & C^c = 0.18627 \end{array} \quad \begin{array}{c} \ln(y^c) \\ C^c = 0.18627 \\ \text{Var}_{\ln(y)}^c = 0.38463 \end{array} \quad (36)$$

with determinant

$$GVE^c = \det(\text{ECM}^c) = \text{Var}_y^c \cdot \text{Var}_{\ln(y)}^c - (C^c)^2 = 0.004257 \quad (37)$$

from which—after normalization—the censored relative poverty value is

$$GVI_{\%}^c = 100 \left(1 - \frac{(\text{Cov}^c)^2}{\text{Var}_y^c \cdot \text{Var}_{\ln(y)}^c} \right) = 10.924\% \quad (38)$$

The corresponding canonical correlation is

$$\text{Rho}^c = \sqrt{0.10924} = 0.33051 \quad (39)$$

The interpretation of Rho^c is, that the proportion of variance explained by the poverty line is 10.924 percentage with a 0.33051 canonical correlation intensity.

In addition, the poverty line divides the society into two groups: the poor people on the one side constituting the 30% of the population and the remaining 70% set of the non-poor. Hence, for the censored distribution the within-groups *weighted average* covariance matrix is as follows:

$$\text{ECM}_{\text{Within}}^c = 0.3 \times \text{ECM}_{\text{Poor}} + 0.7 \times 0_{\text{Non-poor}} = \begin{array}{cc} \text{Variable} & y^c & \ln(y^c) \\ y^c & 0.03416 & 0.07793 \\ \ln(y^c) & 0.07793 & 0.20974 \end{array} \quad (40)$$

where

$$\text{ECM}_{\text{Poor}} = \begin{array}{cc} \text{Variable} & y^c & \ln(y^c) \\ y^c & 0.11387 & 0.25976 \\ \ln(y^c) & 0.25976 & 0.69913 \end{array} \quad (41)$$

and the censored covariance matrix of the non-poor is a zero-matrix. Hence, the censored within-groups *GVE* value is

$$GVE_{\text{Within}}^c = \det(\text{ECM}_{\text{Within}}^c) = 0.03416 \cdot 0.20974 - 0.07793^2 = 0.001092 \quad (42)$$

with the normalized version

$$GVE_{\text{Within}\%}^c = 100 \frac{0.001092}{0.03416 \cdot 0.20974} = 15.241\% \quad (43)$$

Recalling now that the censored within-groups variance is computed as the weighted average of the poor's non-null and the non-poor's null covariance matrices, this normalized

value can be interpreted as an adjusted version of the standard *Head-Count Ratio* (H) which simply counts for the proportion of the poor people in the population. The reasons for this interpretation are as follows. Clearly, an increase in $GVE^c_{Within\%}$ is due to an increase in.

- the Head-Count Ratio on the one hand or/and in
- the Generalized Variance measured below the poverty line on the other.

In our example the *adjusted* Head-Count Ratio equals 15.2%, clearly smaller than the $H=30\%$ standard “*Head-Count-Ratio*”.

Let us consider now the *deprivation felt by the poor against the poverty line*, as the distribution sensitive component of poverty. We require this level to be sensitive to the proportion of the non-poor in the population. Let ID^c denote the measure of this deprivation and let us assume a *multiplicative* decomposition of the poverty factors as follows:

$$GVE^c_{\%} = GVE^c_{Within\%} \cdot ID^c_{Between\%} \quad (44)$$

Apparently, $ID^c_{Between} = 10.2/15.9 = 72\%$ is an *implicit level* of the poor’s deprivation (shortfall) measured against (from) the poverty line. As a comparison, the classic *Income Gap Ratio* (the average percentage shortfall of income from the poverty line) is $100(1-15.5/30) = 48.3\%$.

The $GVE^c_{\%}$ metric has all the properties of the original GVE measure. Due to an increase in the poverty line, the censored $Var_y Var_{\log y}$ upper product bound also increases. The poverty line can be defined both for the dimensions separately and, also, for a single weighted combination.

7 Limitations

The questions of this paper are twofold. First, how to convert the settlement-effects into a distance-preserving scale *with positive scores only*, based on a baseline category, for instance “Budapest”. The reason for using a scale of positive values is a mathematical requirement for logarithm calculations. Further, the choice of the starting point and locations of the division points on the scale remains the task of the researcher. Application of equidistant dividing points is not necessary.

The second key question is how to compute the contributions of the categories to the 82.3% within-strata inequality in Eq. (15). Several methods provided to change the linear weighting scheme. As an alternative method, the study suggests the so called “odds” approach of the logistic regression model. Using this model, the proportional contributions of the categories to the within-groups effect are:

$$M_{Budapest} = \frac{\exp(2.382)}{\exp(2.382) + \exp(5.191) + \exp(5.260) + \exp(6.533)} = 0.010 \quad (45)$$

$$M_{Towns} = \frac{\exp(5.191)}{\exp(2.382) + \exp(5.191) + \exp(5.260) + \exp(6.533)} = 0.168 \quad (46)$$

$$M_{Cities} = \frac{\exp(5.260)}{\exp(2.382) + \exp(5.191) + \exp(5.260) + \exp(6.533)} = 0.180 \quad (47)$$

$$M_{Villages} = \frac{\exp(6.533)}{\exp(2.382) + \exp(5.191) + \exp(5.260) + \exp(6.533)} = 0,642 \quad (48)$$

Thus, *Budapest* gives the smallest 1% contribution to the average within-strata inequality, while the “*Villages*” generate the largest proportion 64.2%.

Note that in the above fractions, both the numerator and the denominator use unweighted $\exp(.)$ = “odds” values. The basic reason for this is that the shape of the exponential function automatically involves an implicit weight system. Hence, we avoid the problem of overweight.

Finally, a theoretical problem arises when measuring poverty. Because the measurement of poverty in the present study is based on a censored distribution and includes relative incomes, this results in a decision making situation to censor first and then form relative incomes, or vice versa. This article followed the former approach.

8 Conclusions

The article proposes a new concept for measuring economic multidimensional inequality in a stratified population using standard multivariate statistical techniques. Provided a *stratified population* in the space of *multidimensional clouds*, the *GVI* procedure gives the *between-clouds* + *within-clouds* additive decomposition of the *total population inequality*. In addition to the literature, the between-within decomposition is subject to subsequent decompositions: the stratum-specific *contributions* to the *average within-clouds inequality* on the one hand and, the dimension-specific *contributions* to the *between-clouds inequality* on the other hand are computed. In contrast with the literature, the *GVI* concept is multidimensional because it is based on measuring dispersion of clouds with different multidimensional shape, rather than using a unidimensional decomposition of an arbitrary weighted sum of the indicators. Because inequality is dispersion, the *GVI* approach measures inequality using the *Generalized Variance (GV)* metric. Numerically, *GV* is computed as the determinant of the covariance matrix considered. Of course, an increase in *GV* indicates increasing multidimensional dispersion and, consequently, increasing *GV* Inequality as well. The advantage of using *GVI* is twofold. First, given a stratified society the overall *GVI* can be expressed as a function of the separated *GVI*s. Secondly, the proportion of the total inequality not explained by the stratification is reported by the standard Wilk’s Lambda ratio. Based on the group-specific *contributions* of the strata to the *average within-clouds inequality* it is possible to test the equality of stratum-specific covariance matrices. Both the dimensions and the strata are ranked according to their importance in explaining inequality. The *GVI* method considers the correlations among the (*socio-economic*) *dimensions*. Further, the use of logarithmic transformation reduces the bias due to distributional asymmetry (including outlier cases) when a symmetrical distribution is required for statistical inference. The numerical calculations of *GVI* can be carried out by the means of any standard statistical package and the number of dimensions is not limited. The *GVI* approach combines the covariance and information theory and, besides, incorporates classic inequality measures, such as the parametric Generalized Entropy indices. Finally, *GVI* works as a measure of poverty when the poverty indicators are censored at the poverty line. The

poverty application defines a new *Generalized Variance Inequality and Poverty (GVIP)* method.

Funding Open access funding provided by Eötvös Loránd University.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alkire, S., & Foster, J. (2011). Understandings and misunderstandings of multidimensional poverty measurement. *Journal of Economic Inequality*, 9, 289–314.
- Anand, S. (1977). Aspects of poverty in Malaysia. *Review of Income and Wealth*, 23(1), 1–16.
- Araar, A. (2009). The hybrid multidimensional index of inequality. CIRPEE WP 09-45. <https://poseidon01.ssrn.com/delivery.php?ID=145065017119117031106112020078006027008078002074040050125081075031027025111067004099043027042032027032054071127031007003027115040087059020045088082098001066002076018007075006067070086024114123117122079126126031121099028084125097097090117021098008007&EXT=pdf>
- Aristondo, O., Lasso De La Vega, C., & Urrutia, A. (2010). A new multiplicative decomposition for the Foster–Greer–Thorbecke poverty indices. *Bulletin of Economic Research*, 62, 259–267.
- Atkinson, A. B. (1987). On the measurement of poverty. *Econometrica*, 55, 749–764.
- Blackorby, C., & Donaldson, D. (1984). Ethically significant ordinal indexes of relative inequality. *Advances in Econometrics*, 3, 131–147.
- Bourguignon, F. (1979). Decomposable income inequality measures. *Econometrica*, 47, 901–920.
- Chakravarty, S. R. (1983). Ethically flexible measures of poverty. *Canadian Journal of Economics*, 16, 74–85.
- Clark, S., Hemming, R., & Ulph, D. (1981). On indices for the measurement of poverty. *The Economic Journal*, 91, 515–526.
- Cowell, F. A., & Kuga, K. (1981a). Additivity and the entropy concept: An axiomatic approach to inequality measurement. *Journal of Economic Theory*, 25, 131–143.
- Cowell, F. A., & Kuga, K. (1981b). Inequality measurement: An axiomatic approach. *European Economic Review*, 15, 287–305.
- Dagum, C. (1997). A new approach to the decomposition of the Gini income inequality ratio. *Empirical Economics*, 22, 515–531.
- Davis, E. (2012). *Linear algebra and probability for computer science applications*. CRC Press.
- Duclos, J. Y., Sahn, D. E., & Younger, S. D. (2006). Robust multidimensional poverty comparisons. *The Economic Journal*, 116, 943–968.
- Foster, J. E., Greer, J., & Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, 52, 761–766.
- Foster, J. E., & Sen, A. (1997). *On economic inequality: After a quarter century. Annex to the enlarged edition of on economic inequality by Amartya Sen*. Clarendon Press.
- Foster, J. E., & Shneyerov, A. A. (1999). A general class of additively decomposable inequality measures. *Economic Theory*, 14, 89–111.
- Hagenaars, A. (1987). A class of poverty indices. *International Economic Review*, 28(3), 583–607.
- Hamada, K., & Takayama, N. (1978). Censored income distributions and the measurement of poverty. *Bulletin of the International Statistical Institute*, 47, 617–632.
- Kakwani, N. C. (1980). On a class of poverty measures. *Econometrica*, 48(2), 437–446.
- Lugo, M. A. (2005). Comparing multidimensional indices of inequality: Methods and application. ECINEQ WP, 2005–2014. <http://www.ophi.org.uk/wp-content/uploads/ssLugo-2005.pdf>

- Lugo, M. A., Maasoumi, E. (2008). Multidimensional poverty measures from an information theory perspective. ECINEQ WP, 2008–85.
- Maasoumi, E. (1986). The measurement and decomposition of multi-dimensional inequality. *Econometrica*, 54, 991–997.
- Mussard, S., Seyte, F., & Terraza, M. (2003). Decomposition of Gini and the generalized entropy inequality measures. *Economics Bulletin*, 4, 1–6.
- Savaglio, E. (2006). Three approaches to the analysis of multidimensional inequality. In F. Farina & E. Savaglio (Eds.), *Inequality and economic integration*. Routledge.
- Sen, A. (1976). Poverty: An ordinal approach to measurement. *Econometrica*, 44, 219–231.
- Sharma, S. (1996). *Applied multivariate techniques*. Wiley.
- Shorrocks, A. F. (1980). The class of additively decomposable inequality measures. *Econometrica*, 48, 613–625.
- Shorrocks, A. F. (1995). Revisiting the Sen poverty index. *Econometrica*, 63, 1225–1230.
- Shorrocks, A. F., & Foster, J. (1987). Transfer sensitive inequality measures. *Review of Economic Studies*, 54(3), 485–497.
- Takayama, N. (1979). Poverty, income inequality and their measures: Professor Sen's axiomatic approach reconsidered. *Econometrica*, 47, 747–759.
- Theil, H. (1967). *Economics and information theory*. North-Holland Publishing Company.
- Thon, D. (1979). On measuring poverty. *Review of Income and Wealth*, 25(4), 429–440.
- Tsui, K. Y. (1999). Multidimensional inequality and multidimensional generalized entropy measures: An axiomatic derivation. *Social Choice and Welfare*, 16, 145–157.
- Tsui, K. Y. (2002). Multidimensional poverty indices. *Social Choice and Welfare*, 19(1), 69–93.
- Watts, H. W. (1968). An economic definition of poverty. In D. P. Moynihan (Ed.), *On understanding poverty*. Basic Books.
- Zheng, B. (1997). Aggregate poverty measures. *Journal of economic surveys*, 11(2), 123–162.
- Zheng, B. (2005). *Unit-consistent decomposable inequality measures*. University of Colorado at Denver and HSC.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.