

A Computational Approach to Economic Inequality, Happiness and Human Development

Irina GEORGESCU¹, Jani KINNUNEN², Armenia ANDRONICEANU³,
Ane-Mari ANDRONICEANU⁴

^{1 3 4} Bucharest University of Economics, Bucharest, Romania

² Åbo Akademi, Turku, Finland

¹irina.georgescu@csie.ase.ro, ²jkinnune@abo.fi,

³armenia.androniceanu@man.ase.ro, ⁴ane.androniceanu.drd@gmail.com

In this paper, we study the connections of categorical levels of Human Development Index (HDI), GDP per capita, World Happiness Index, the Gini indexes of Income and Wealth inequalities together with poverty rate for 98 world countries. By clustering analysis we identify four groups of countries with similar features. K-means clustering algorithm is applied to obtain four clusters of sizes 21-26 countries by explaining 68.3% of the total variation in data. The analysis reveals significant differences between the clusters, while also factors with largest differences within the clusters. Secondly, multinomial logistic regression (MLR) is applied in predicting the HDI categories of the full sample of 98 world countries for year 2018. The MLR model can capture also nonlinear relationship. The logistic regression model achieved 91.8% overall accuracy. The results of our research together from earlier literature is followed by suggestions for the future research.

Keywords: Human Development Index, Economic inequality, Happiness Index, Poverty rate.

DOI: 10.24818/issn14531305/24.4.2020.02

1 Introduction

Economic inequalities are universal and inevitable. The income that is created in a year is a certain amount and any favouring of certain categories of population means a minus for another segment of the society in a zero-sum game, whereas in fair markets, the value added through transactions is not away from others, but rather a win-win situation even when inequalities exist. Income inequalities can be cyclical or permanent, random or systematic, normal or abnormal. The criteria for separating the types of economic inequality may be statistical (form of income distribution, European average and comparison between countries, etc.) or may include the social consequences of inequalities on the functioning of society (social cohesion, social conflicts, the community problems it creates, etc.) and the quality of life. Among the collected indicators, Human Development Index and its determinants, we have included the World Happiness Index 2019. The Happiness Index is annually computed by the United Nations starting from 2012, as a composite indicator

based on GDP per capita, social support, health life expectancy, freedom to make life choices, generosity and perceptions of corruption.

We intend to study the impact and interconnectedness of the economic equalities and the happiness index on HDI.

The rest of the paper is structured as follows. Section 2 reviews the relevant literature. Section 3, firstly, presents the data (Section 3.1), secondly, conducts clustering analysis (Section 3.2) and, thirdly, studies by multinomial logistic regression, the relationship between Human Development Index and economic inequalities, GDP per capita, and Happiness Index. Finally, Section 4 concludes the paper and presents some future research opportunities.

2 Literature Review

Social development with its economic, technological and political aspects has large-scale changes due to globalization. The impact of these changes on the development and competitiveness of economies [1]. Economic

inequalities and the mechanism of social stratification are a vast field of research, with a long tradition in sociological research. Explaining the social mechanism of establishing and perpetuating economic inequalities, their control and legitimation would contribute to understanding the differences in people's life chances. Where is the "key" to fulfilling the life that people want and cherish? To whom should we attribute responsibility success or failure in achieving the standard of living that people aspire to? If someone has a strong motivation and well-defined life goals, various resources (financial, educational, social, health, etc.) and a good knowledge of the rules social game, will it reach the level of material well-being it has set itself? Social welfare is directly influenced by poverty and economic inequality [2], [3]. Economic discrepancies how income, wealth, goods and wages are distributed in a population [4]. Economic inequalities and poverty have effects not only on material well-being and the ability to have the standard of living to which people aspire, but also on relations between social groups, generating alienation and tensions between them. The social welfare has in center the equal opportunity policy for all citizens and the essence of income distribution relies on the principles of global economic distribution [5], [6], [7]. Several instruments can be applied to reduce the level of economic inequality [8], [9] (1) redistribution - from high incomes to low incomes; (2) the widely available opportunities and (3) social responsibility. The first instrument, redistribution, means supplementing the lower income of people from the higher income of people [10]. The second instrument, the opportunity grid is a policy that consists in offering various opportunities to the vulnerable, in difficulty or helpless people. The third instrument, the social responsibility belongs to organizations [11]. Social responsibility means that economic agents have responsible social initiatives to the expectations of employees, customers, investors and of a company overall, focused on high quality services, respecting the rights of the employees and solving social problems [12]. If these three instruments are applied by governments,

economic discrepancies would be reduced, and the economic productivity, the labour market and the investment incentives will not be affected.

Economic discrepancies have as main causes [13] wage inequalities, including gender wage inequalities, globalization, technological changes, education, fiscal policy, wealth growth, political reforms, certain market segments monopolies. We underline the focus on increases in R&D expenditures, with significant influences on the development of the global competitiveness [14]. Some conceptual approaches to economic discrepancies are presented below. Papers [15], [16], [17] consider that economic discrepancies are necessary for economic growth. Papers [18], [19] underline the importance of inequalities components to understand how they influence the social welfare. People's chances of living better depend on individual characteristics (options and motivation, financial, educational, social and capacity resources, representations of the rules of the social game), as well as the social structure, the opportunities and facilities offered by the social environment, the rules and practices of access to the world consumption of goods and services. On the long term, economic inequalities can contribute to economic growth and well-being. However, on the short-term, rising income inequality is leading to deepening poverty. Any improvement in the financial situation of a social group necessarily leads to the deterioration of the welfare of other social categories. Problems in developed countries do not arise because society is not rich enough (or even because it is too rich), but because the scale of material differences between people within each society is too large. We use data for 2018-2019 and we study the correlation between well-being and social inequality for 98 world countries. This paper extends the results of [20], [21], [22], [23], [24] and [25] where the authors used data mining techniques such as cluster analysis, logistic regression, neural networks, multiple discriminant analysis and canonical correlation analysis to study the relationship between social inequality, competitiveness and digitalization, respectively.

Papers [22], [23] use the same dataset and classify world countries by means of multinomial logistic regression and multiple discriminant analysis.

We concluded that the multinomial logistic regression classifier was somewhat more efficient, with an accuracy 92.9% compared with the discriminant analysis accuracy of 83.7%.

3 Data, Methodology and Analysis

In this section we will, firstly, describe the data in section 3.1 and, secondly, apply the K-means algorithm [25], one of the most commonly used clustering algorithms to group the 98 world countries with respect to the measures of well-being and economic inequalities presented in 3.1. Thirdly, section 3.3 presents the logistic regression analysis, where human development categories represent the dependent variable.

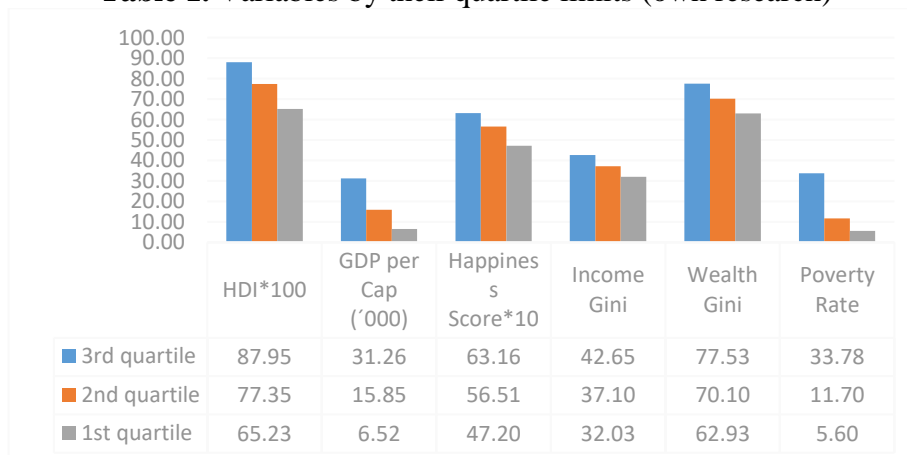
3.1 Data

The data variables used in this study include: the *well-being measures* of Human Development Index, HDI from United Nations Development Programme [26], GDP per capita (in 2011, purchasing power parity, PPP, terms also from [26]) and Happiness score from the United Nations Sustainable Development Network [27], together with the *measures of economic inequality*, i.e. Income Gini, Wealth Gini, and Poverty rate reported by World Economic Forum's Inclusive Development Index

[28]. The variables' quartile limits are shown in Table 1. Specifically, the HDI's limits determine the four categories used as the dependent variable in the logistic regression analysis.

The categories of HDI are determined as follows: if a country's HDI*100 (we have scaled HDI by 100 in Table 1 for visual purposes) value > 87.95 (3rd quartile limit), then it belongs to the top category "Very High"; if the HDI value (*100) is between 77.35 (2nd quartile limit) and 87.95, then the country belongs to "High" HDI category; if the country's HDI value (*100) is between 65.23 (1st quartile limit) and 77.35, then it belongs to "Medium" category; and if the HDI value (*100) < 65.23 , then the country falls to the bottom category of "Low" HDI. For GDP per capita the limits in Table 1 are USD 31.26 thousands (3rd quartile), USD 15.85 thousands (2nd quartile) and USD 6.52 thousands in 2011 money (PPP). Thus, the economic well-being difference is very large between the top-25% "Very High" GDP category and the bottom-25% "Low" GDP category, around minimum of 5 times (6.62 versus 31.26), while around minimum of 2 times difference (15.85 and 31.26) between also between "Very High" and "Medium" GDP categories. Similarly, Poverty rate shows great differences between the categories, and the other variables in Table 1 are viewed accordingly.

Table 1. Variables by their quartile limits (own research)



We take a look at the multiple dependences between the variables of the models and highlight the most correlated variables. Figure 1

presents the correlation plot. The scale of blue represents positive correlations and the scale of red represents negative correlations. We

note: HDI is mostly positively correlated with GDP per capita (0.83) and the Happiness score (0.77), and negatively with Poverty rate (-0.83); GDP per capita is, further, strongly positively correlated with Happiness score (0.76) and negatively with Income Gini (-0.55) and Poverty rate (-0.53); Income Gini is somewhat positively correlated with Wealth Gini (0.36) and Poverty rate (0.26), while

negatively with other variables; Wealth Gini does not show other significant correlations. Poverty rate shows strong negative correlation with Happiness score (-0.61). Wealth Gini is not significantly correlated with happiness, but income inequalities and poverty lowers it, while HDI (and its underlying pillars of education, health and income) and GDP per capita drive the world happiness.



Fig. 1. Correlation plot (own research)

After these suggestive relations, next, we conduct a clustering analysis and discuss the relatedness of the variables in each constructed cluster of countries.

3.2 K-means Clustering Analysis

K-means clustering is, firstly applied using all the six variables presented above. We obtain four clusters of sizes 26, 26, 21, and 25 countries (clusters 1-4, respectively). The four clusters are well delimited and relatively

compact, as shown in Figure 2. Within cluster sum of squares for each cluster are 53.589, 46.202, 39.052, and 45.648 (for CL1-CL4, respectively). The between_SS / total_SS = 68.3 % gives the total variance in the data explained by the k-means clustering. This suggests a relatively good explanatory power of our clustering model, while keeping the number of clusters still relatively low.

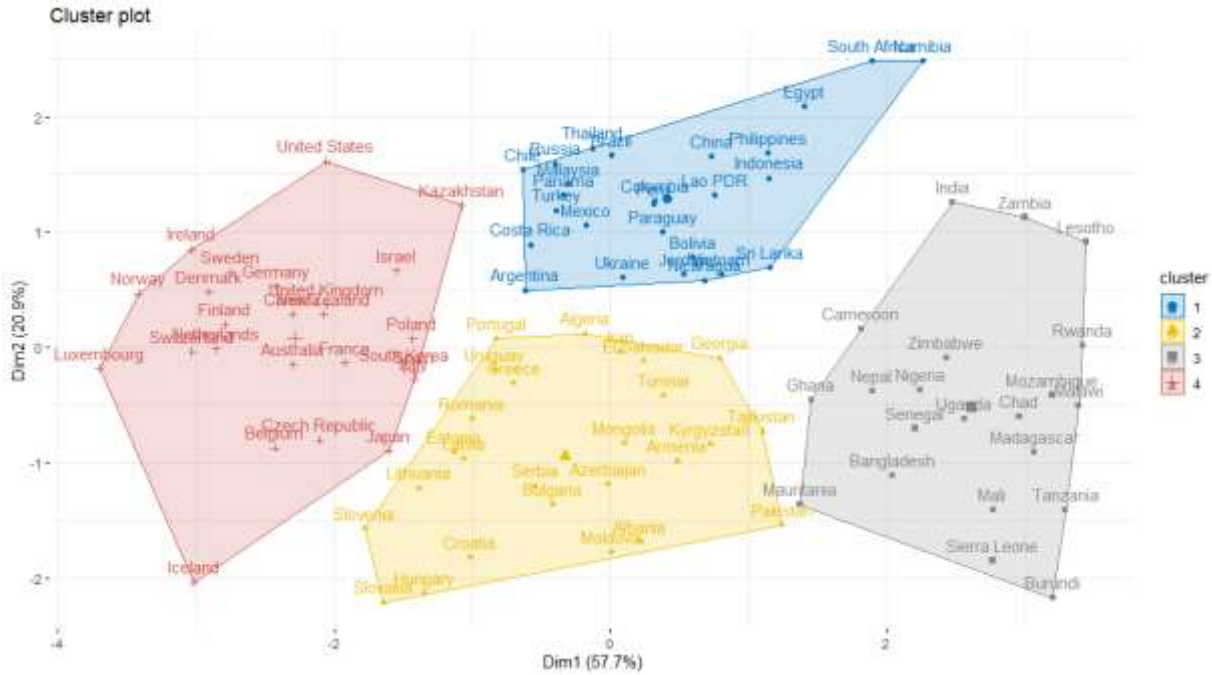


Fig. 2. Cluster plot (own research)

The k-means clustering method uses the standardized cluster means presented in Table 2.

Table 2. Standardized cluster means (own research)

| Cluster | HDI | GDP per Cap | Happiness Score | Income Gini | Wealth Gini | Poverty Rate |
|---------|--------|-------------|-----------------|-------------|-------------|--------------|
| 1 | -0.041 | -0.365 | 0.019 | 1.050 | 0.865 | -0.337 |
| 2 | 0.223 | -0.152 | -0.141 | -0.312 | -1.032 | -0.527 |
| 3 | -1.530 | -0.987 | -1.218 | 0.260 | -0.083 | 1.672 |
| 4 | 1.096 | 1.367 | 1.150 | -0.985 | 0.242 | -0.506 |

The absolute cluster mean values and the standard deviations for each cluster as well as for the total sample of 98 countries are seen in Table 3. The values are neither scaled in any

way. Thus, the HDI ranges from 0-1 (not 0-100 as in Table 1), GDP is in US dollars (not thousands), etc.

Table 3. Clusters data (own research)

| | HDI | GDP per Cap | Happiness Score | Income Gini | Wealth Gini | Poverty Rate |
|--------------|------|-------------|-----------------|-------------|-------------|--------------|
| CL1_avg | 0.74 | 14128 | 5.65 | 44.94 | 79.45 | 14.26 |
| CL1_Std.dev. | 0.06 | 6455 | 0.87 | 5.83 | 6.70 | 15.13 |
| CL2_avg | 0.78 | 18013 | 5.47 | 34.79 | 57.80 | 9.18 |
| CL2_Std.dev. | 0.08 | 9572 | 0.58 | 4.25 | 8.61 | 9.39 |
| CL3_avg | 0.52 | 2813 | 4.24 | 39.05 | 68.63 | 68.12 |
| CL3_Std.dev. | 0.07 | 1509 | 0.58 | 5.43 | 7.75 | 18.91 |
| CL4_avg | 0.91 | 45663 | 6.94 | 29.78 | 72.34 | 9.73 |
| CL4_Std.dev. | 0.03 | 14647 | 0.58 | 3.75 | 9.38 | 4.65 |

| | | | | | | |
|----------------|------|-------|------|-------|-------|-------|
| Total_avg | 0.75 | 20779 | 5.63 | 37.12 | 69.58 | 23.30 |
| Total_Std.dev. | 0.15 | 18020 | 1.13 | 7.38 | 11.29 | 26.51 |

Figure 3 presents the box plots of the four clusters by each variable and we will interpret it together with Table 3 for each cluster, but in

order from richest to poorest clusters: CL4 > CL2 > CL1 > CL3.

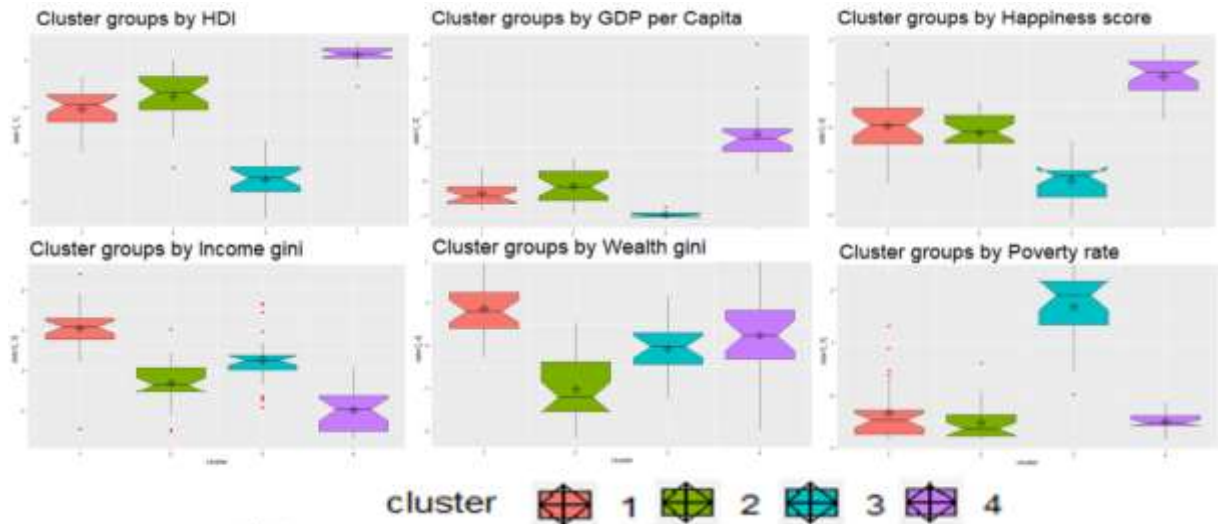


Fig. 3. Cluster box-plots by variables (own research)

Cluster 4 of rich and developed countries: On the top of Figure 3, well-being variables are presented: starting from the top-left, cluster 4 shows the highest HDI (0.91 on average with small standard deviation of 0.03 as shown in Table 2), GDP per capita (USD 45663 on average), as well as Happiness score (values cluster 3 shows the lowest HDI (0.52 mean in Table 2). Cluster 4 is represented by 25 countries including the richest Western states, such as Luxembourg, Ireland, Norway, Switzerland, USA, Canada, UK, Australia, for instance. Also, not surprisingly, Japan, South Korea and New Zealand belong to cluster 4, while rather surprisingly also Kazakhstan and Poland, which by their GDP per capita would be in the middle of the second, not the top quartile and are close to cluster 1 (Kazakhstan) and cluster 2 (Poland). On the bottom of Figure 3, Income Gini shows the lowest value for cluster 4 (but with large variance), while only cluster 1 shows the higher Wealth Gini values than cluster 4. By Poverty rate, the differences between cluster 4 (9.73% with small variance within the cluster) and cluster 2 (9.18%) are small, while cluster 1 is somewhat higher (14.26%), but in cluster 3,

poverty explodes (68.12% with large variability).

Cluster 2 of countries-in-transition: Figure 3 and Table 3 show that 26 countries of CL2 are well-being by HDI (0.78) and GDP per capita (USD 18K) on average, but with great variability within the cluster, which together with the lowest wealth Gini and the second lowest income Gini only after cluster 4 and even the lowest poverty rates, which, however, are not reflected well in the happiness of these countries (only cluster 3 shows lower happiness on average). The cluster consists of relatively rich countries like Estonia, Slovenia, Slovakia, Lithuania, Portugal and Hungary (all with GDP per capita > USD 28K) and include, e.g. Romania and Bulgaria (with GDP per capita of ca. USD 25 and 19K, respectively), but the large variability is seen as poor countries like Kyrgyzstan and Pakistan with GDP per capita of only ca. USD 3.5K and 4.9K and HDI of 0.67 and 0.56, respectively, belong to the same cluster as their economic inequalities are very close to the cluster average. Many former Soviet countries belong to this group together with some MENA countries like Algeria, Tunisia and Iran, as well as, with Latin-

American countries like El Salvador and Uruguay. Lowest happiness scores of this cluster are detected in Tunisia, Georgia, Iran, Armenia, and Albania, which show some of the highest income, but also wealth inequalities within the cluster.

Cluster 1 of unequal and poor, but relatively happy countries: the 26 countries of cluster 1 show the highest income and wealth inequalities (the Gini values of 44.94 and 79.45, respectively compared to the full sample's averages of 37.12 and 69.58) and the second highest poverty rates of 14.62% on average (cf. Figure 3 and Table 3). The average GDP per capita is ca. USD 6.5K (in 2011 PPP) and HDI is 0.74. However, the average happiness score is 5.65, which is slightly above the world average of our 98 countries (5.63), while only the richest cluster 4 had a higher score (6.94). Cluster 1 is strongly dominated by Latin-American countries including Argentina, Bolivia, Brazil, Chile, Colombia, Nicaragua, Paraguay, Peru and happy (top-quartile) Chile, Costa Rica, Mexico, and Panama. Bottom-quartile happiness is seen in Egypt, Namibia, Sri Lanka, and Ukraine. Among with some Latin-American relatively well-being countries (specifically, Argentina, Chile, Costa Rica, and Panama) high HDI values in cluster 1 are seen in Malaysia, Russia, Sri Lanka, and Turkey.

Cluster 3 of low development of African and Indian-Bangladeshi countries: the 21 countries of cluster 3 show the lowest GDP per capita of ca. USD 2.8K on average with also the smallest variability over all clusters, while the richest individual country, Ghana produced ca. USD 4.2K. Similarly, the lowest values are seen in HDI (0.52) and happiness score (4.24), while the poverty rate of 68.12% suggest the key problem is a lack of general

economic activity and markets, when people could find employment. The lowest poverty rate is in Mauritania (23.60%), but the second lowest Ghana already shows over third of the people living in poverty (34.90%), while Burundi, Madagascar, Malawi, Mozambique, Rwanda and Sierra Leone have poverty rates between 81.30-91.10% and many other countries are close to such levels. Income inequalities are the second highest only after cluster 1 and wealth inequality is close to the third highest after cluster 1 and close to cluster 4 levels. All countries of cluster 3 belong the bottom HDI quartile of in our sample of 98 world countries and "medium" happiness is on average seen only in Cameroon, Ghana, Nepal, Nigeria, and Senegal. Only Bangladesh, India and Nepal are non-African representatives of cluster 3 countries.

Next, we will study the dependency of human development level (HDI categories) on the other variables of well-being (GDP per capita and Happiness score) and economic inequalities by logistic regression analysis over the whole sample before concluding the analysis.

3.3 Logistic Regression Analysis

In Table 4 the null model (only the intercept) and the final model (the predictors and the intercept) are compared by means of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The final model is chosen since $68,3 < 277,673$ and $114,83 < 285,428$. The lower value 32,3 of the -2 Log Likelihood also indicates that the final model is preferred. The final model is significant $\chi^2(12) = 239,372$, $p < 0,01$, therefore, the final model is a better predictor than the null model.

Table 4. Model fitting information (own research)

| Model | Model Fitting Criteria | | | Likelihood Ratio Tests | | |
|----------------|------------------------|---------|-------------------|------------------------|----|------|
| | AIC | BIC | -2 Log Likelihood | Chi-Square | df | Sig. |
| Intercept Only | 277,673 | 285,428 | 271,673 | | | |
| Final | 68,300 | 114,830 | 32,300 | 239,372 | 15 | ,000 |

In Table 5 the goodness-of-fit model by Pearson (Chi-square 33,506) and Deviance (Chi-square=32,3) statistics with the p-values greater than 0.05 indicate a good fit. In Table

6, pseudo- R^2 s confirm the good fit, with values $> 0,90$; the best is Nagelkerke pseudo R^2 with the highest value of 0,974.

Table 5. Goodness of fit (own research)

| | Chi-Square | df | Sig. |
|----------|------------|-----|-------|
| Pearson | 33,506 | 276 | 1,000 |
| Deviance | 32,300 | 276 | 1,000 |

Table 6. Pseudo R-square (own research)

| | |
|---------------|------|
| Cox and Snell | ,913 |
| Nagelkerke | ,974 |
| McFadden | ,881 |

In Table 7, likelihood ratio tests show that *the intercept, GDP per capita, Poverty rate and Wealth Gini have p-values less than 0,10*; thus, they have a significant contribution to the full effect. Of these statistically significant components, AIC and BIC for the reduced

model shows the lowest values for Wealth Gini and Poverty rates together with Intercept (although lower values are detected for statistically insignificant Income Gini and Happiness score).

Table 7. Likelihood ratio tests (own research)

| Effect | Model Fitting Criteria | | | Likelihood Ratio Tests | | |
|---------------------|------------------------|----------------------|------------------------------------|------------------------|----|-------------|
| | AIC of Reduced Model | BIC of Reduced Model | -2 Log Likelihood of Reduced Model | Chi-Square | df | Sig. |
| Intercept | 70,577 | 109,352 | 40,577 | 8,277 | 3 | ,041 |
| GDP per Cap | 115,393 | 154,168 | 85,393 | 53,093 | 3 | ,000 |
| Income Gini | 62,475 | 101,250 | 32,475 ^a | ,175 | 3 | ,982 |
| Wealth Gini | 68,601 | 107,376 | 38,601 ^a | 6,301 | 3 | ,098 |
| Poverty Rate | 75,390 | 114,164 | 45,390 ^a | 13,090 | 3 | ,004 |
| Happiness Score | 62,745 | 101,519 | 32,745 ^a | ,444 | 3 | ,931 |

In Table 8, the change in logit function for a unit change in the predictor is given by B. GDP per capita and Wealth Gini have the smallest values of the B coefficient, meaning

that these predictor least influence HDI prediction. The odds ratio of each predictor is $Exp(B)$.

Table 8. Parameter estimates (own research)

| HDI_4 | B | Std. Error | Wald | df | Sig. | Exp(B) | |
|-------|-----------------|------------|----------|---------|------|--------|------------|
| High | Intercept | 555,324 | 5823,602 | ,009 | 1 | ,924 | |
| | GDP per Cap | -,012 | ,149 | ,007 | 1 | ,933 | ,988 |
| | Income Gini | -1,239 | 61,438 | ,000 | 1 | ,984 | ,290 |
| | Wealth Gini | -,226 | 21,845 | ,000 | 1 | ,992 | ,798 |
| | Poverty Rate | -4,583 | 93,848 | ,002 | 1 | ,961 | ,010 |
| | Happiness Score | -11,664 | ,866 | 181,254 | 1 | ,000 | 8,599E-006 |
| Low | Intercept | 546,209 | 5823,614 | ,009 | 1 | ,925 | |
| | GDP per Cap | -,014 | ,149 | ,009 | 1 | ,925 | ,986 |

| | | | | | | | |
|--------|-----------------|---------|----------|--------|---|------|------------|
| | Income Gini | -1,286 | 61,438 | ,000 | 1 | ,983 | ,276 |
| | Wealth Gini | ,024 | 21,845 | ,000 | 1 | ,999 | 1,024 |
| | Poverty Rate | -4,504 | 93,848 | ,002 | 1 | ,962 | ,011 |
| | Happiness Score | -10,726 | 1,758 | 37,232 | 1 | ,000 | 2,196E-005 |
| Medium | Intercept | 553,926 | 5823,601 | ,009 | 1 | ,924 | |
| | GDP per Cap | -,013 | ,149 | ,008 | 1 | ,930 | ,987 |
| | Income Gini | -1,222 | 61,438 | ,000 | 1 | ,984 | ,295 |
| | Wealth Gini | -,096 | 21,845 | ,000 | 1 | ,996 | ,909 |
| | Poverty Rate | -4,689 | 93,848 | ,002 | 1 | ,960 | ,009 |
| | Happiness Score | -11,157 | ,000 | . | 1 | . | 1,428E-005 |

The three multinomial logistic equations are:

$$\ln\left(\frac{P(HDI = low)}{P(HDI = veryhigh)}\right) = 546,209 - 0,014GDPpercapita - 1,286IncomeGini + 0,024WealthGini - 4,504PovertyRate - 10,726HappinessScore$$

For a one-point increase in GDP per capita, the chances that a country has a low HDI rather than a very high HDI decrease with 1,4% (1-0,986); for a one-point increase in Income Gini, the chances that a country has a low HDI rather than a very high HDI decrease with 72.4% (1-0,276); for a one-point increase in Wealth Gini, the chances that a country has a low HDI rather than a very high HDI increase

with 2,4% (1,024-1); for a one-point increase in Poverty Rate, the chances that a country has a low HDI rather than a very high HDI decrease with 98,9% (1-0,011); for a one-point increase in Happiness Score, the chances that a country has a low HDI rather than a very high HDI decrease with 99,99% (1-2,196E-005).

$$\ln\left(\frac{P(HDI = medium)}{P(HDI = veryhigh)}\right) = 553,926 - 0,013GDPpercapita - 1,222IncomeGini - 0,096WealthGini - 4,689PovertyRate - 11,157HappinessScore$$

For a one-point increase in GDP per capita, the chances that a country has a medium HDI rather than a very high HDI decrease with 1,3%; for a one-point increase in Income Gini, the chances that a country has a medium HDI rather than a very high HDI decrease with 70,5%; for a one-point increase in Wealth Gini, the chances that a country has a medium

HDI rather than a very high HDI decrease with 9,1%; for a one-point increase in Poverty Rate, the chances that a country has a medium HDI rather than a very high HDI decrease with 99,1%; for a one-point increase in Happiness Score, the chances that a country has a low HDI rather than a very high HDI decrease with 99,99%.

$$\ln\left(\frac{P(HDI = high)}{P(HDI = veryhigh)}\right) = 555,324 - 0,012GDPpercapita - 1,239IncomeGini - 0,226WealthGini - 4,583PovertyRate - 11,664HappinessScore$$

For a one-point increase in GDP per capita, the chances that a country has a very high HDI rather than a very high HDI decrease with 1,2%; for a one-point increase in Income Gini, the chances that a country has a high HDI rather than a very high HDI decrease with 71%; for a one-point increase in Wealth Gini, the

chances that a country has a high HDI rather than a very high HDI decrease with 20,2%; for a one-point increase in Poverty Rate, the chances that a country has a high HDI rather than a very high HDI decrease with 99%; or a one-point increase in Happiness Score, the

chances that a country has a low HDI rather than a very high HDI decrease with 99,99%. In Table 9, the overall accuracy is 91.8%, proving that the logistic regression classifier

is well chosen for modeling the relationship between social inequality, well-being and happiness.

Table 9. Classification (own research)

| Observed | Predicted | | | | Percent Correct |
|--------------------|-----------|-------|--------|-----------|-----------------|
| | High | Low | Medium | Very High | |
| High | 21 | 0 | 3 | 0 | 87,5% |
| Low | 0 | 24 | 1 | 0 | 96,0% |
| Medium | 2 | 2 | 20 | 0 | 83,3% |
| Very High | 0 | 0 | 0 | 25 | 100,0% |
| Overall Percentage | 23,5% | 26,5% | 24,5% | 25,5% | 91,8% |

4 Conclusions

We analyzed 98 diverse world countries with respect to their economic inequalities (income and wealth inequalities and poverty rates), level of human development (HDI), gross domestic product per capita and happiness index. This was done by, firstly, running k-means clustering analysis over all the variables and the full sample and, secondly, by conducting logistic regression analysis with HDI as the categorical dependent variable and all others being explanatory variables.

The clustering analysis was pre-set for four clusters, which allowed studying potential direct links between the obtained clusters and the HDI categories and the results from the logistic regressions. The links are clearly non-linear in general and it is challenging to establish, particularly, linear relationships, which hold both in the highly developed countries with advanced markets and relatively low income inequality and low poverty rates (cf. cluster 4) and in the least-developed countries with disastrous poverty rates and top-level economic inequalities (cf. cluster 3), and the country clusters somewhere in-between the two extremes, such as the countries-in-transition (cf. cluster 2 dominated by ex-Soviet and Eastern European countries) and Latin-American relative poor, highly unequal, but relative happy countries (cf. cluster 1). However, with only four clusters, our k-means clustering model explained 68.3% of the total variance over all our data and the key descriptive features were detected.

The logistic regression analysis achieved 91.8% overall prediction accuracy and

implied that the intercept, GDP per capita, Poverty rate and Wealth gini contribute statistically significantly to the full effect from other scale variables to the categorical HDI. Criteria of AIC and BIC for the reduced model supported this but actually showed the lowest values for Income gini and Happiness score. This suggests that with a different sample, either with even a larger set of world countries or by a selection done through some comparative context, also these variables may appear significant, e.g. under (multinomial logistic) regression.

For future research, we find an interesting option to use different clustering methods, with a larger set of world countries so that the number of clusters is optimized based on a broader set of features measuring the functionalities of the economic and political systems and countries’ positioning within their value-creating networks through international trade as well as monetary and know-how flows. Time-series analysis could be used between the features to detect possible causalities. The role of inequalities in different combinations (or clusters) of countries can be more direct (and linear) in some groups, while machine-learning techniques could detect more complex relationships in other groups.

References

[1] H. Fyliuk, I. Honchar and V. Kolosha. “The interrelation between economic growth and national economic competitiveness: the case of Ukraine. *Journal of Competitiveness*, 11(3), 2019, pp. 53–69. <https://doi.org/10.7441/joc.2019.03.04>

- [2] Y. Bilan, H. Mishchuk, N. Samoliuk and H. Yurchyk. "Impact of income distribution on social and economic well-being of the state. Sustainability, 12(1), 2020, pp. 429.
- [3] H. Mishchuk, N. Samoliuk, Y. Bilan and D. Streimikiene. Income inequality and its consequences within the framework of social justice. *Problemy Ekorozwoju*, 13(2), 2018, pp. 131-138.
- [4] S. Stjepanović; D. Tomić and M. Škare. "A new approach to measuring green GDP: a cross-country analysis", *Entrepreneurship and Sustainability Issues* 4(4), 2017, pp. 574-590. [http://dx.doi.org/10.9770/jesi.2017.4.4\(13\)](http://dx.doi.org/10.9770/jesi.2017.4.4(13))
- [5] J. Belas, L. Belas, M. Cepel, Z. Rozsa. "The impact of the public sector on the quality of the business environment in the SME segment", *Administratie si Management Public*, 32, 2019, pp. 18-31. DOI: 10.24818/amp/2019.32-02
- [6] N. Meyer, D.F. Meyer. "The relationship between the creation of an enabling environment and economic development: A comparative analysis of management at local government sphere", *Polish Journal of Management Studies*, 14 (2), 2016, pp. 150-160.
- [7] M. Haseeb, S. Kot, H.I. Hussain and K. Jermsttiparsert. "Impact of economic growth, environmental pollution, and energy consumption on health expenditure and R&D expenditure of ASEAN countries", *Energies*, 12 (19), 2019, art. no. 3598.
- [8] R. Prakash, Garg, P. 2019. Comparative assessment of HDI with Composite Development Index (CDI), *Insights into Regional Development* 1(1): pp. 58-76. [https://doi.org/10.9770/ird.2019.1.1\(5\)](https://doi.org/10.9770/ird.2019.1.1(5))
- [9] J. Baltgailis. "The issues of increasing the effectiveness of teaching comparative economics", *Insights into Regional Development*, 1(3), 2019, pp. 190-199. [https://doi.org/10.9770/ird.2019.1.3\(1\)](https://doi.org/10.9770/ird.2019.1.3(1))
- [10] A. Androniceanu, M. Tvaronavičienė. "Developing a holistic system for social assistance services based on effective and sustainable partnerships", *Administratie si Management Public*, 33, 2019, pp. 103-118, DOI: 10.24818/amp/2019.33-06.
- [11] A. Androniceanu. "Social responsibility, an essential strategic option for a sustainable development in the field of bio-economy", *Amfiteatru Economic*, 21(52), 2019, pp. 347-364.
- [12] G. Çera, M. Meço, E. Çera and S. Maloku, "The effect of institutional constraints and business network on trust in government: an institutional perspective", *Administratie si Management Public*, 33, 2019, pp. 6-19, DOI: 10.24818/amp/2019.33-01.
- [13] L. Smékalová. "Evaluating the cohesion policy: targeting of disadvantaged municipalities", *Administratie si Management Public*, 31, 2018, pp. 143-154, DOI: 10.24818/amp/2018.31-10.
- [14] D. Kiselakova, B. Sofrankova, V. Cabinova, E. Onuferova and J. Soltesova. "The impact of R&D expenditure on the development of global competitiveness within the CEE EU Countries", *Journal of Competitiveness*, 10(3), 2018, pp. 34-50. <https://doi.org/10.7441/joc.2018.03.03>.
- [15] E. Duřová Spiřáková, L. Mura, B. Gontkovičová and Z. Hajduová. "R&D in the context of Europe 2020 in selected countries", *Economic Computation and Economic Cybernetics Studies and Research*, 51(4), 2017, pp. 243 - 261.
- [16] A. G. Raišienė, S. Bilan, V. Smalskys, J. Gečienė, "Emerging changes in attitudes to inter-institutional collaboration: the case of organizations providing social services in communities", *Administratie si Management Public*, 33, 2019, pp. 34-56, DOI: 10.24818/amp/2019.33-03.
- [17] D. Jařková. "Assessment of social development in Slovakia in the context of human resources", *Central European Journal of Labour Law and Personnel Management*, 2 (2), 2019, pp. 21-32. doi: 10.33382/cejllpm.2019.03.02;
- [18] H. Z. Ivancsóné, E. Printz-Markó. "Territorial differences between countries with regard to the wellness lifestyle of

- their youth”, *Forum Scientiae Oeconomia*, 6(3), 2018, pp. 101-117.
- [19] A. Ključnikov, M. Sobeková Majková. “Impact of gender in the perception of administrative burdens among young entrepreneurs - evidence from Slovakia. *Journal of Competitiveness*, 8(2), 2016, pp. 17-30. DOI: 10.7441/joc.2016.02.02.
- [20] J. Kinnunen, A. Androniceanu and I. Georgescu, “The role of economic and political features in classification of countries-in-transition by Human Development Index”, *Informatica Economica*, 23(4), 2019, pp. 26-40.
- [21] I. Georgescu and J. Kinnunen, “The digital effectiveness on economic inequality: a computational approach”, to appear in *Business Revolution in A Digital Era*; Dima, A.M., D’Ascenzo, F., Eds.; Springer: Cham, 2020.
- [22] I. Georgescu, J. Kinnunen, A. Androniceanu and A.-M. Androniceanu, “Global well-being and economic inequality” in *Proc. of the 19th International Conference on Informatics in Economy (IE 2020)*, 21-24 May 2020, Bucharest, Romania, pp. 266-273.
- [23] I. Georgescu, A.-M. Androniceanu, J. Kinnunen, “A discriminant analysis to the quantification of Human Development Index under economic inequality” in *Proc. of the 14th International Management Conference*, 5-6 November 2020, Bucharest Romania.
- [24] A.-M. Androniceanu, I. Georgescu, M. Tvaronavičienė and A. Androniceanu. “Canonical correlation analysis and a new composite index on digitalization and labor force in the context of industrial revolution 4.0”. *Sustainability*, 12(17), 2020, 6812.
- [25] I. Georgescu, *Inteligența computațională*, ASE Publishing House, 2015
- [26] J. F. Helliwell, R. Layard, J. D. Sachs, and J.-E. De Neve, “World Happiness Report 2020”. United Nations’ Sustainable Development Solutions Network, 2020. Available at: <https://happiness-report.s3.amazonaws.com/2020/WHR20.pdf>
- [27] UNDP, “United Nations’ Human Development Database”, 2020. Available at: <http://hdr.undp.org/en/data#>
- [28] WEF, “The Inclusive Development Index 2018”, 2018, Available at: www3.weforum.org/docs/WEF_Forum_IncGrwth_2018.pdf



Irina GEORGESCU holds a PhD in Economics from Turku Centre for Computer Science, Turku, Finland. Currently she is a lecturer at the Department of Economic Informatics and Cybernetics, Bucharest Academy of Economic Studies. Her research interests lie in the areas of fuzzy sets and systems, computational intelligence and risk theory. She is the author of about 40 journal papers and 2 books published in Springer Verlag.



Jani KINNUNEN is a researcher at Åbo Akademi, Finland holding a MSc in Economics. His research interests include data science related to economics and political science; application work has focused on freedoms in economic and political systems, mergers and acquisitions and real options. He has about 60 publications from international conferences and journals.



Armenia ANDRONICEANU is Full Professor of The Bucharest University of Economic Studies from Bucharest, Romania and Director of the International Centre for Public Management. Her works were published by prestigious international journals from Romania and abroad. She is member of the Editorial board of several international journals and reviewer of few of them. Research interests: economics; public and private management; strategic management; management of the international organizations; management of change and innovation; management of the human resources, project management and comparative public administration.



Ane-Mari ANDRONICEANU has a degree in business administration and a master's degree in business entrepreneurship. She has published papers in the field of management. She is currently a PhD student at the Doctoral School of Management within the Bucharest University of Economic Studies. Her research interest is in organizational management and economic models in the context of digitalization and internationalization of the business environment.