Development of The Welfare Index at Sub-district Level in West Java 2020: A Small Area Estimation Approach

N E Monadiyan[1] **, F Haris**[2]

[1] BPS-Statistics of Cianjur Regency, Perintis Kemerdekaan St. No.3 Sirnagalih, Cilaku, Cianjur Regency, West Java, Indonesia 43285
[2] BPS-Statistics of Depok Municipality, Boulevard Grand Depok City Street, Kalimulya, Sukmajaya, Depok Municipality, West Java, Indonesia 16413

*Corresponding author's e-mail: nimasezra@gmail.com, faisalhaaaaaris@gmail.com

Abstract. Welfare as an alternative measure for poverty, is an important indicator to measure. But there is no composite indicator that specifically measures the prosperity of a population yet. Considering the increase of data needed and the lack of available data at the smaller level, this study develops the Welfare Index at sub-district level for West Java in 2020 using a small area estimation approach to explain the condition of welfare of the population. The indicators in sub-district level formed in this study were created from two kinds of data. The first type of indicators were formed from SUSENAS using Small Area Estimation and the other type of indicators were formed from PODES aggregation. All the indicators were then processed with factor analysis to form the Welfare Index. The Welfare Index formed shows the range of 22.86 to 83.76 and generally higher in the northern part of West Java. This index has a correlation of 0.798 with the Human Development Index because of the components that defined both indexes. The existence of this correlation shows that the Welfare Index formed is able to explain the conditions/phenomena being measured.

1. Introduction

Good health and well-being is the third goals of Sustainable Development Goals (SDGs). Public welfare is also the main goal of the establishment of the Republic of Indonesia. Welfare itself can be defined in many ways. According to the Oxford dictionary, welfare is the general health, happiness and safety of a person, an animal or a group. Indonesia Macroeconomic Outlook by University of Indonesia also stated that a prosperous society is a society that can fully enjoy prosperity, is not poor, does not suffer from hunger, enjoys education, is able to implement gender equality, and enjoys health facilities equally [1]. A prosperous society can be achieved if the population is able to participate in national development. Thus, the strategy for national development must aim to improve the welfare of the population.

In order to evaluate national development, usually single indicators are used, such as poverty rate, school participation rate, unemployment rate, infant mortality rate, etc. There are also some composite indexes, such as the Human Development Index. The Human Development Index (HDI) is a summary measure of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable and having a decent standard of living.

There are so many indicators that are being used to portray a prosperous society, but there are no composite indicators that specifically measure the prosperity of a population. World Bank in World Development Report states that poverty is a lack of prosperity [2]. So, welfare cannot be separated from poverty and a measurement of welfare can give us an alternative measure for poverty. Considering the importance of welfare, this study aims to develop a composite index to explain the condition of welfare.

Researches that formulate the welfare index have been carried out previously both in Indonesia and in other countries. In 2016, Miko Armiento formulated The Sustainable Welfare Index (SWI) for Italy. This index is formed from 17 components adopted from the Index of Sustainable Economic Welfare (ISEW) and Genuine Progress Indicator (GPI). The 17 indicators are private consumption expenditure, welfare losses due to income inequality, services provided by non-paid domestic work, services provided by durable goods, expenditure for durable goods, public expenditure in health and education, research and development expenditure, cost of vehicle accidents, cost of commuting, cost of noise pollution, cost of water pollution, cost of air pollution, cost of urbanization, depletion of non-renewable resources, social cost of annual carbon dioxide emissions, net fixed capital formation, and variation of the net international investment position [3].

In the same year, Benjamin Held et al formulated The National and Regional Welfare Index (NWI/RWI) in Germany. Similar to the SWI formed by Miko Armiento, the NWI/RWI is formed from 20 indicators adopted from the ISEW and GPI. The indicators that make up this index are Index of Income Inequality, private consumption weighted by Index of Income Inequality, value of household work, value of voluntary work, non-defensive public expenditure on health care and education, net value of the costs and benefits of consumer durables, costs of commuting, costs of traffic accidents, costs of crime, costs of alcohol tobacco and drug abuse, costs of avoidance and repair, damage costs of water pollution, damage costs of soil degradation, damage costs of air pollution, damage costs of noise, value of increase/loss of ecosystems, value of increase/loss of agricultural areas, replacement costs due to the consumption of non-renewable energy resources, damage costs of yearly GHG emissions, and costs of the use of nuclear energy [4].

In Indonesia itself, the welfare of population was studied by Eko Sugiharto in 2007. Adopting indicators from the Statistics Indonesia, the welfare of population is measured using 8 indicators, namely income, household consumption or expenditure, housing conditions, housing facilities, health of household members, ease of obtaining health services, ease of entering children into education level, and ease of obtaining transportation facilities [5].

Considering the condition of population and the availability of data in Indonesia, this study adopts the indicators for constructing the welfare index from Statistics Indonesia. Since the main data source for welfare indicators in Indonesia is the National Socio-Economic Survey (SUSENAS) in March, several indicators must be adjusted to the availability of SUSENAS data. SUSENAS in March are designed to estimate variables in District/Municipality Level. In order to prevent welfare problems and to improve the level of welfare of the population, a welfare indicator at a smaller level is needed. With the estimation at a smaller level, the effort to tackle the welfare problem will be more effective, more focused and consume fewer resources. Hence, overcoming welfare problems that occur in the community will also be right on target and reach the wider community. In conclusion, the development of the welfare index at the sub-district level is urgently needed right now.

To develop welfare index in sub-district level, small area estimation is used to estimate welfare index constituent. Small area estimation is a method to estimate with small sample size. In this case, small sample size because SUSENAS data are designed to only provide estimation up to district/municipality level but we want to use it to develop welfare index in sub-district level. In this study we use Empirical Best Linear Unbiased Predictor model to estimate an area with sample and we use Cluster Information for an area with no sample.

Small Area Estimation tend to make a better result with a large number of areas and West Java is the most populated province in Indonesia with almost 50 million population in 2020 census and it has 627 sub-districts. Because of that, it has a relatively big sample size and can be used as pilot study to estimate welfare indicator in sub-district level. If we can develop a welfare index in sub-district level in West Java, we can expand this study to develop the welfare index in another province with smaller sample size and smaller number of sub-districts.

2. Methodology

2.1. Theoretical Framework

According to Statistics Indonesia there are eight criterias to determine welfare, those are yearly income, household consumption or expenditure, housing, ownership of housing facilities, household member health, access to health services, access to education, and access to transportation services [5].

Housing is defined by material of roof, material of wall, ownership status of the dwelling, material of floor and floor area. Ownership of housing facilities defined by house yard, ownership of air conditioners, source of lighting, ownership of vehicles, type of fuel mainly used for cooking, main source of water, drinking water facility, main source of drinking water, toilet facility, distance of waste disposal to the dwelling. Access to health services defined by distance of nearest hospital, distance of drug store, drug handling, drug prices, and contraception. Access to education defined by education fees, distance of school, and admission process. Access to transportation services defined by transportation fare, transportation facilities, and vehicle ownership status.

2.2. Small Area Estimation

According to Rao, Small Area Estimation (SAE) is a method to estimate the parameters of subpopulation with small sample size. In this case, small area means an area that cannot be directly estimated because it can produce a very large standard error [6].

The most common SAE method is Empirical Best Linear Unbiased Prediction (EBLUP). EBLUP was first applied by Fay and Herriot and is a weighted average of the direct estimation and a regression estimation that is obtained by fitting linear regression into the data. The main idea of EBLUP is to explain the diversity of target variables that can be explained by auxiliary variables combined with a specific area random effect. Therefore, the Fay-Herriot model is also commonly known as the linear mixed model [7].

Because SAE usually deals with small sample size, usually there are several areas that do not have samples for estimating. Because of that, sometimes the direct estimation of that area cannot be estimated. Recently there are two ways to estimate an area with zero sample, there are synthetic estimation and utilizing the cluster information. Gonzales stated that synthetic estimation is an indirect estimation that uses variable characteristics of a large sample area to estimate the variable of small sample size [8]. It assumes that an area with small sample size has a similar characteristic with an area that has a large sample size. However, synthetic estimation does not take into account the effect of random areas, so there will be bias in the estimation.

To overcome the bias in synthetic estimation, Rahma Annisa uses cluster information of sampled area with the same characteristic as non-sampled area to improve the precision of the estimation [9]. Cluster information can be obtained by clustering an area with a similar auxiliary variable and it can be added to non-sampled area estimation. One of the proposed models by Rahma Annisa is to add the average of random area effect of the sampled area to the non-sampled area synthetic estimation with the same cluster. By using this technique, the estimation has a smaller MSE and bias compared to synthetic estimation.

2.3. RRMSE

Root Relative Means Squared Error (RRMSE) is the result of dividing the root of Mean Squared Error (MSE) by the average true value in an area [6]. RRMSE is used to calculate the accuracy of an estimation method and usually expressed in percent. BPS - Statistics Indonesia used a maximum estimated RRMSE of 25% as the standard for reliability of estimates. According to international standards, Estimates with RRMSE between 25% and 50% are considered to be high and should be used with caution. On the other hand, estimates with RSE higher than 50% are considered unreliable for general use and used only when the lack of reliability is understood. To calculate RRMSE we use:

$$RRMSE\left(\hat{\theta}_i^{SAE}\right) = \frac{\sqrt{MSE\left(\hat{\theta}_i^{SAE}\right)}}{\hat{\theta}_i^{SAE}} \tag{1}$$

2.4. Factor Analysis

Factor analysis is an interdependence technique to define the underlying structure among the variables in the analysis. Factor analysis provides the tools for analyzing the structure of the interrelationships among a large number of variables by defining sets of variables that are highly interrelated [10]. The simpler method used, the better the index. But, in determining the right aggregation for an index, a lot of supporting theory and previous research are needed. Therefore, as a first step in creating a new index in a study, factor analysis is often used to help researchers grouping the indicators into appropriate dimensions and perform the index aggregation.

The assumptions in factor analysis are more conceptual than statistical. In factor analysis, the overriding concern center is on the character and composition of the variables included in the analysis. From a statistical standpoint, several necessary requirements are: (1) statistically significant Barlett's test of sphericity, that indicates existing sufficient correlations among variables; (2) measure of sampling adequacy (MSA) values must exceed 0.50 for both overall test and each individual variable.

Once the variables are specified and the correlation matrix is prepared, the decision on the method of extracting the factors and the number of factors selected to represent the underlying structure in the data must be made. A several stopping criteria to determine the initial number of factors to retain are: (1) factors with eigenvalues greater than 1.0; (2) a predetermined number of factors based on research objectives and/or prior research; (3) enough factors to meet a specified percentage of variance explained, usually 60 percent or higher; (4) factors shown by the scree test to have substantial amounts of common variance; (5) more factors when heterogeneity is present among sample subgroups.

To assist in the process of interpreting a factor structure and selecting the final factor solution, the factor rotation is encountered. No specific rules have been developed to guide in selecting the technique of factor rotation, but most programs have the default rotation of varimax. In interpreting factors, the decision made regarding the factor loadings. The loading must exceed 0.70 for the factor to account for 50 percent of the variance of a variable. But the significance level of the factor loadings are different based on the sample size. The factor loadings 0.30 is significant for sample sizes of 350 or greater.

Once all the significant loadings have been identified, the communalities must be considered to assess whether the variables meet acceptable levels of explanation. The communalities less than 0.50 are not having sufficient explanation. Respecifying the factor model could be done if necessary. Respecification of a factor analysis includes such options: (1) deleting a variable(s); (2) changing rotation methods; (3) increasing or decreasing the number of factors.

2.5. Composite Index Construction

The following steps are used in the construction of a composite indicator [11]. While the workflow of the index construction can be seen in Figure 1.

- Theoretical framework and data selection
  To get a clear understanding and definition of the multidimensional phenomenon to be measured, and structure the various sub-groups of the phenomenon, the theoretical framework should be formed.
- Multivariate analysis
  Factor analysis is used to compose the Welfare Index in this study. The steps taken in factor analysis are data eligibility test, variable reduction, factor formation, factor rotation, up to form the factor scores.

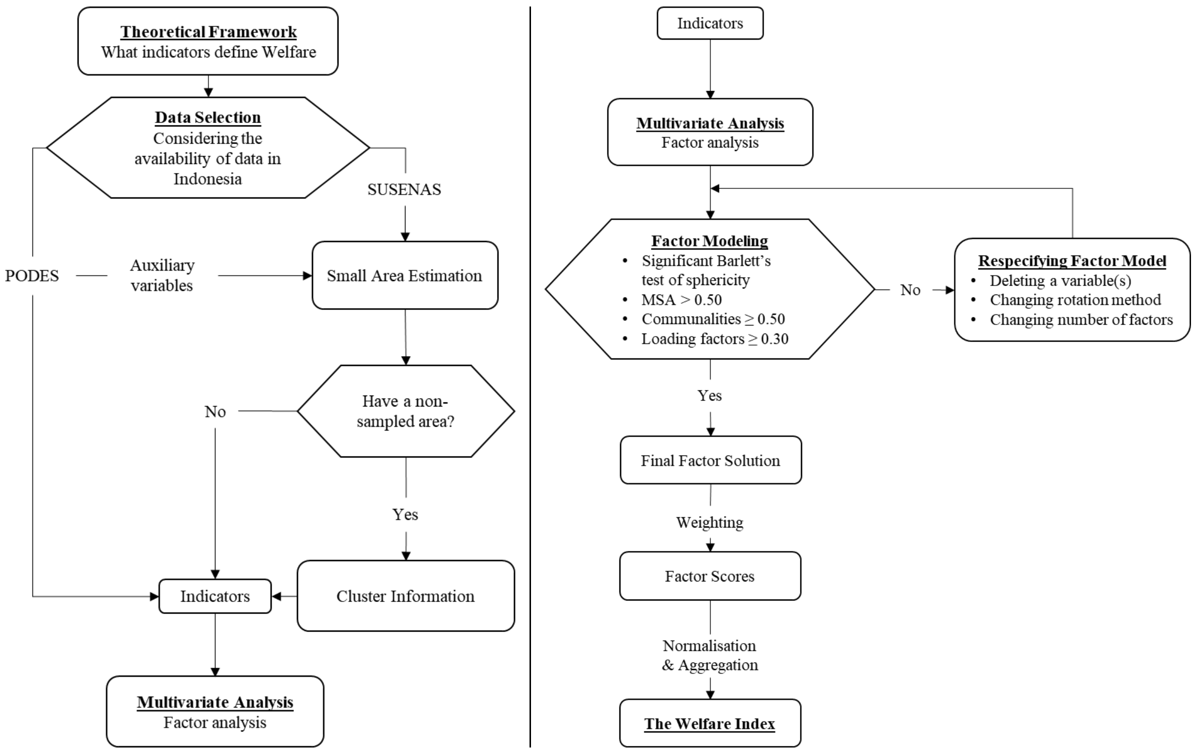

Figure 1. The workflow of the Index Construction

- Normalisation
  Normalisation is required prior to any data aggregation as the indicators in a data set have different measurement units. This study uses min-max normalisation to have an identical range [0, 1].
- Weighting and aggregation
  According to factor analysis, the weighting refers to the variance explained. The weighting for each factor is the variance explained by the factor divided by the total variance explained by all factors.

$$W_i = \frac{Variance\ explained_i}{Total\ variance\ explained}, \text{ where } \sum_i W_i = 1 \tag{2}$$

Meanwhile, to form the factor scores, the standardized values of indicators in each factor need to be weighted by the square factor loadings divided by the total of square factor loadings.

$$w_{ij} = \frac{l^2_{xij}}{\sum_j l^2_{xij}}, \text{ where } \sum_j w_{ij} = 1 \tag{3}$$

$$f_i = \sum_j x_{ij} \times w_{ij}, \text{ where } x_{ij} \text{ is the standardized value of } j \text{ indicator in } i \text{ factor} \tag{4}$$

Once the weighting and the factor scores are formed, the following linear aggregation is used.

$$Social\ Welfare\ Index = \sum_i W_i \times f_i', \text{ where } f_i' \text{ is the normalized factor scores} \tag{5}$$

- Decomposing the indicator
  Decomposing a composite indicator is needed to identify if the results are dominated by a few indicators and to explain the relative importance of the sub-components of the composite indicator.
- Link to other variables
  This step should be made to correlate the composite indicator with other relevant measures, taking into consideration the results of sensitivity analysis.

3. Result

3.1. Data Description

To estimate The Welfare Index at sub-district level in West Java 2020 we use two types of data, The National Socio-Economic Survey (SUSENAS) in March 2020 and The Village Potential Data Collection (PODES) in 2020. In Small Area Estimation, SUSENAS data used as direct estimator and PODES 2020 data used as auxiliary variables. SUSENAS data is designed to estimate up to district/municipality level. Therefore, the direct estimator for sub-district level has a relatively high RSE and some sub-district have no samples at all. PODES data are also used in composite index construction along with SUSENAS data.

We use SUSENAS data to estimate expenditure per capita (y_1), household food expenditure (y_2), household non-food expenditure (y_3), percentage of household with home ownership (y_4), percentage of household with decent drinking water (y_5), percentage of population with health complaints (y_6), percentage of smoker (y_7), percentage of population with health insurance (y_8), mean years of schooling (y_9), and percentage of population with vehicle (y_{10}). From PODES, we use a percentage of villages with decent public bathing, washing, and toilet facilities (y_{11}), educational facilities to population ratio (y_{12}), medical facilities to population ratio (y_{13}), percentage of villages with decent roads (y_{14}) and percentage of villages with public transportation (y_{15}).

3.2. Small Area Estimation

The direct estimation from SUSENAS data to estimate indicators in sub-district level have a really high RSE and there are some indicators with RSE value 0% and have no sample at all. This is because some sub-district are not selected as SUSENAS samples and some sub-districts have a uniform sample that causes the MSE of the estimated sub-district to be 0. This makes the estimated indicators to be unreliable and cannot be used in general.

Overall, if we make a composite index from this data we will get 214 sub-district with unreliable indexes and 190 sub-district with indexes that can be used with caution. So, there are only 223 sub-district or 35.57% of the data that is reliable and can be used to map the level of welfare in West Java. so, to make our estimated index reliable we need to use Small Area Estimation.

To estimate all indicators we use PODES 2020 data as auxiliary variables. We use PODES variables that have been aggregated to sub-district level and their derivative variables, such as coverage, ratio and others. To select suitable auxiliary variables we use stepwise regression and we use the selected auxiliary variable to find an optimum number of clusters to calculate cluster information. Once we get the best suitable auxiliary variable and the most optimum number of cluster, we estimate the sample area using EBLUP and add cluster information to the non-sampled area.

After that we estimate the sample area using EBLUP and add cluster information to the non-sampled area. To see how efficient our SAE compares to direct estimation we compare the average of MSE and Relative efficiency between SAE and direct estimator. In table 1 we can see that the optimum cluster of auxiliary variables are between 2 and 8. Also, the average of mean squared error in all small area estimation variables is less than direct estimation. The relative efficiency of small area estimation variables are between 0.293 and 0.734. Thus we can say that small area estimation are more efficient than direct estimation.

Table 1. Optimum number of cluster, Average MSE comparison and Relative Efficiency of the model

Variable	Optimum Number Of Cluster	Average Mean Squared Error		Relative Efficiency
		Direct Estimation	Small Area Estimation	
y_1	2	35297649331.16	15622650623.11	0.443
y_2	8	52530220817.41	33380496004.16	0.635
y_3	5	279324857670.04	81887294599.28	0.293
y_4	2	0.003808	0.002279	0.598
y_5	2	0.008495	0.006237	0.734
y_6	3	0.004073	0.002899	0.712
y_7	2	0.000910	0.000530	0.582
y_8	3	0.005338	0.003539	0.663
y_9	3	0.322073	0.184585	0.573
y_{10}	2	0.004460	0.002663	0.597

Based on Table 2, we can see that out of 10 indicators that use RRMSE , there are 6 indicators that do not have an RRMSE above 25%. There are 3 indicators with a sub-district that have RSE between 25% and 50%, we have to be careful to use the estimated indicators in these sub-district. And lastly, there is still one indicators that has a sub-district with an RRMSE above 50, namely percentage of population with health complaints.

Overall, if all indicators from SAE are included in the composite index there will be 4 sub-district with unreliable index, 175 sub-district with index that can be used with caution and there are 448 sub-district or 71.45% data that is reliable and can be used in general.

Table 2. Percentage of Sub-district based on the RSE of Direct Estimation and RRMSE of SAE Results per Indicator.

Variable		Direct Estimation				SAE		
		0-25	25-50	>50	0 and NA	0-25	25-50	>50
y1	Expenditure Per Capita	92.82	4.78	0.00	2.39	100.00	0.00	0.00
y2	Household Expenditure for Food	96.01	1.75	0.00	2.23	100.00	0.00	0.00
y3	Household Expenditure for Non-Food	83.73	13.08	0.96	2.23	93.14	6.86	0.00
y4	Percentage of Households with Home Ownership	81.18	2.07	0.16	16.59	100.00	0.00	0.00
y5	Percentage of Household with Decent Drinking Water	69.54	11.80	4.94	13.72	91.39	8.61	0.00
y6	Percentage of Population with Health Complaints	75.12	19.30	3.03	2.55	82.93	16.43	0.64
y7	Percentage of Smoker	93.78	3.83	0.32	2.07	100.00	0.00	0.00

Variable		Direct Estimation				SAE		
		0-25	25-50	>50	0 and NA	0-25	25-50	>50
y8	Percentage of Population with Health Insurance	87.40	8.29	0.64	3.67	98.72	1.28	0.00
y9	Mean Years of Schooling	97.45	0.00	0.00	2.55	100.00	0.00	0.00
y10	Percentage of Population with Vehicle	92.66	3.03	0.32	3.99	100.00	0.00	0.00

3.3. Index Construction

The first step in creating the index is to synchronize the direction of each variable. In this study, the value of all indicators will be made in a positive way, so that greater index value states higher welfare of an area. In addition, considering the different units of each indicator, standardization is carried out for each variable.

Table 3. The Indicators of The Welfare Index

Variable		Direction	Used in the model 1 – Yes 0 – No	Additional Information
y_1	Expenditure Per Capita	+	1	
y_2	Household Expenditure for Food	+	1	
y_3	Household Expenditure for Non-Food	+	1	
y_4	Percentage of Households with Home Ownership	+	1	
y_5	Percentage of Household with Decent Drinking Water	+	1	
y_6	Percentage of Population with Health Complaints	-	0	Doesn't meet the requirement of RRMSE
y_7	Percentage of Smoker	-	1	
y_8	Percentage of Population with Health Insurance	+	1	
y_9	Mean Years of Schooling	+	1	
y_{10}	Percentage of Population with Vehicle	+	1	
y_{11}	Percentage of Villages with Decent Public Bathing, Washing, and Toilet Facilities	+	0	Not representative, high in almost all of sub-district
y_{12}	Educational Facilities to Population Ratio	+	1	
y_{13}	Medical Facilities to Population Ratio	+	1	

	Variable	Direction	Used in the model 1 – Yes 0 – No	Additional Information
y_{14}	Percentage of Villages with Decent Road	+	0	Not representative, high in almost all of sub-district
y_{15}	Percentage of Villages with Public Transportation	+	0	Not representative, high in almost all of sub-district

After synchronizing the direction and standardizing the variables, eligibility test is carried out on the data used. Considering the correlation matrix and referring to the Kaiser-Mayer-Olkin (KMO), Barlett's Test of Sphericity, Anti-image Correlation (Measure of Sampling Adequacy/MSA), and the values of communalities, 11 variables are selected to define the welfare index.

Those 11 variables were then grouped into 6 factors. The factor selection considering the percentage of variance explained, prior theory or research, and the scree plot. With six factors formed, the percentage of variance explained is up to 87.25 percent which is sufficient to explain the variance of the data. Then to determine the variable in a factor, the eigenvalues are considered. Based on the sample size, the eigenvalues 0.3 is acceptable. The summary of the selection of indicators in each factor can be seen in Table 4.

Table 4. The Indicators and the Factors Formed for the Welfare Index

No.	Factor	Variable	Notation	Eigenvalues	% of variance
1	Expenditure and mean years of schooling	Expenditure Per Capita	y_1	0.921	29.65
		Household Expenditure for Food	y_2	0.849	
		Household Expenditure for Non-Food	y_3	0.882	
		Mean years of schooling	y_9	0.667	
2	Housing	Percentage of households with home ownership	y_4	0.646	13.51
		Percentage of household with decent drinking water	y_5	0.804	
3	Access to educational and medical facilities	Educational facilities to population ratio	y_{12}	0.922	13.27
		Medical facilities to population ratio	y_{13}	0.621	
4	Access to transportation	Percentage of population with vehicle	y_{10}	0.857	10.64

No.	Factor	Variable	Notation	Eigenvalues	% of variance
5	Access to health insurance	Percentage of population with health insurance	y_8	0.907	10.51
6	Household member health	Percentage of smoker	y_7	0.958	9.65

Using the eigenvalues of each variable in a factor, the factor scores are formed. In order to perform the aggregation, the factor scores must have an identical range, thus the normalisation is used. The normalisation used in this study is min-max normalisation so the factor scores will have an identical range [0, 1] and the Welfare Index will have a value of 0 to 100.

Once the normalized factor scores are formed, the aggregation runs using the weighting that comes from the percentage of variance explained.

$$\begin{aligned} &Social\ Welfare\ Index \\ &\quad = (0.3399\ Factor_1 + 0.1549\ Factor_2 + 0.1521\ Factor_3 \\ &\quad + 0.1220\ Factor_4 + \ 0.1205\ Factor_5 + 0.1107\ Factor_6) \times 100 \end{aligned} \tag{6}$$

3.4. The Welfare Index

The Welfare Index formed has a value of 0 to 100. The greater index value states higher welfare of an area. The Welfare Index in West Java 2020 shows the range of 22.86 to 83.76, with the average being at 44.46. The lowest index was found in Tanjungsari (Bogor Regency) while the highest index was found in Panyileukan (Bandung City).

As The Welfare Index is a new establishment index, there is no previous literature as a basis for classification, therefore the index was then grouped into five categories based on the quantiles. The quantiles classification method was chosen to distribute the index into groups that contain an equal number of areas/subdistricts, in order to compare the ranking between regions. Quantile classification divides classes so that the total number of features in each class is approximately the same. This type of classification is useful for showing rankings and ordinal data. The distribution of the classification results can be seen in Figure 2, where the areas with darker colors indicate higher welfare index. In general, it can be seen that the areas in the northern part of West Java have higher welfare than the southern part. Meanwhile, the district/city with high welfare in all sub-districts are Bogor City, Depok City, Bekasi City, Bekasi Regency, Cimahi City, Sukabumi City, Cirebon City, Banjar City, and Pangandaran Regency.

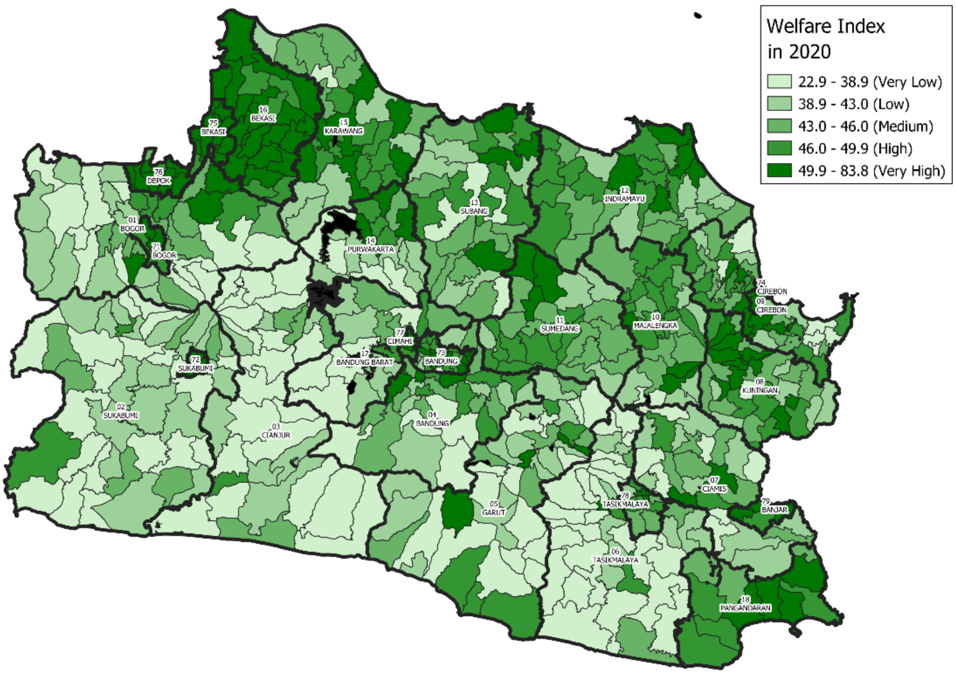

Figure 2. The Welfare Index at Sub-district Level, West Java, 2020

This may happen because these cities are metropolitan areas. The rampant development in these cities certainly spurs the economic growth. High urbanization also increases the welfare of population. Zara Hadijah, in her research in 2020, found that the level of urbanization has a positive relationship to per capita income [12]. World Bank also found that 1 percent urbanization growth in Indonesia increases the value of GDP per capita by 4 percents [13]. These findings indicate that there is a relationship of urbanization and development in a city on the welfare of the population.

The general description of this index is similar to the Human Development Index (HDI) in 2020. The higher welfare index reflects the higher HDI in an area. There is a correlation of 0.798 between the average welfare index in district/city and the 2020 HDI. This correlation exists because some components that define both indexes are quite similar. The welfare index consists of expenditure, housing, education, health, and transportation. While the HDI consists of expenditure, education, and health. The existence of this correlation shows that the welfare index formed is able to explain the conditions/phenomena that are being measured.

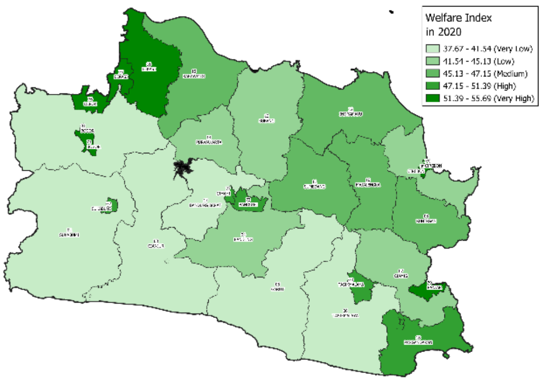

Figure 3. The Average of Welfare Index in District Level, West Java, 2020

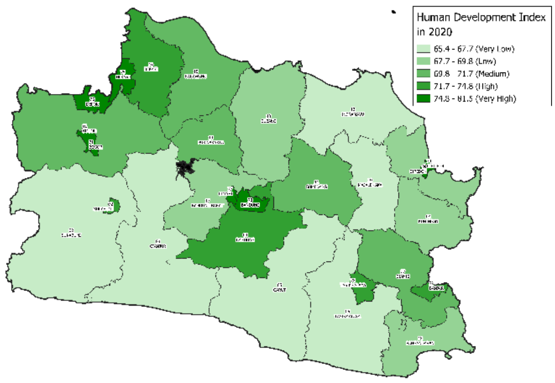

Figure 4. The Human Development Index, West Java, 2020

Factor 1 is the factor with the highest weight in the Welfare Index. This factor explains 29.65 percent variance of the origin data. Expenditure is a key variable to welfare because it indirectly describes the poverty of an area. In Economics, Engel's Law states that higher income will decrease the proportion of food expenditure. This law explains the relation of expenditure and poverty. The poverty of an area itself is a reflection of the welfare of the population living in the area. According to World Bank in the World Development Report, poverty is a lack of welfare [2]. According to the relation of variables defining this factor to welfare, it is reasonable that the factor has the highest variance explained for the Welfare Index. The general description of this factor resembles the Welfare Index and is still related to the high urbanization in certain cities that affects per capita income.

Factor 2 consist of housing indicators. The availability of adequate housing can improve the health status of the population. One indicator of a decent house is the availability of proper drinking water for consumption. Improper drinking water can certainly threaten health because it causes various diseases. Health itself will certainly affect the welfare of population. Similar to factor 2, factor 6 is the household member health which is consist of the percentage of smoker. Although some negative impacts of urbanization on public health were found, especially in the provision of proper water, the general description of factors 2 and 6 shows a high level of public health in big cities as well as the general description of the Welfare Index. This indicates that the high urbanization in certain cities in West Java does not have a negative impact on the public health.

Similar to the Welfare Index formed, the factor scores generated by all the factors tend to be higher in the northern part of West Java, except for factor 3. Generally, the factor scores generated by factor 3 tend to be higher in the southern part of West Java. This situation is possible because of the government development obligation. Hence, even in sub-districts with low populations there must be health and education facilities. Therefore, in sub-districts with low population, the ratio of health facilities to population and the ratio of educational facilities to population are higher.

Factors 3, 4, and 5 are generally explain the accessibility of population to public facilities and government assistance. Factor 4 specifically describes public transportation access. Transportation is a determinant of the affordability of the other facilities such as health and education. The availability of transportation also supports the socio-economic life of the population. In general, all sub-districts in the province of West Java already have public transportation and decent roads so that these indicators are not representative to distinguish welfare between regions. The indicators that available and can explain the differences in transportation access, namely vehicle ownership. Vehicle ownership status certainly influenced by income, therefore the general description of factor 4 also tends to be high in urban areas with more rapid development.

Factor 5 describes access to health insurance. This factor consists of indicators of ownership of health insurance. The value of this factor is high in several cities, namely Depok City, Bekasi City, Bekasi Regency, Bogor City, Sukabumi City, Karawang Regency, Cimahi City, Bandung City, Cirebon City, Banjar City, and Pangandaran Regency. Related to the level of urbanization and development in these cities, access to insurance ownership is also higher because access to transportation tends to be better. In addition, urban communities tend to work more in the formal sector which usually provides health and employment protection.

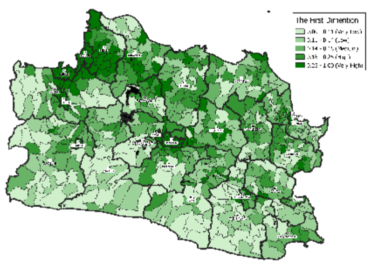

Figure 5. Scores of Factor 1: Expenditure and mean years of schooling

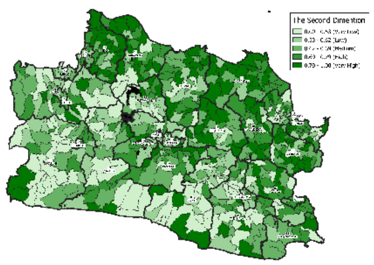

Figure 6. Scores of Factor 2: Housing

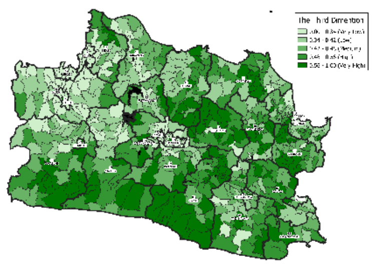

Figure 7. Scores of Factor 3: Access to educational and medical facilities

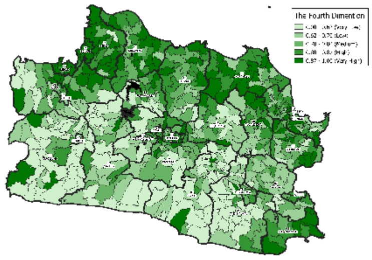

Figure 8. Scores of Factor 4: Access to transportation

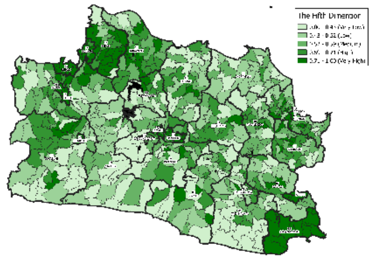

Figure 9. Scores of Factor 5: Access to health insurance

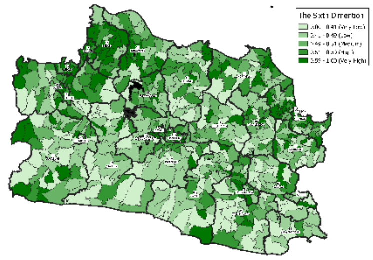

Figure 10. Scores of Factor 6: Household member health

4. Conclusion

Considering the availability of data in Indonesia, the Welfare Index at sub-district level as an indicator to explain the condition of welfare of the population can be generated. For West Java in 2020, the Welfare Index shows higher value at the northern part of West Java with the range of index at 22.86 to 83.76. The distribution of this index shows the same pattern as the Human Development Index and both indexes have a correlation of 0.798. It shows that the Welfare Index formed is able to explain the conditions/phenomena being measured.

Further research must be obtained by considering other variables which currently cannot be estimated to the lowest level yet. Various theories and concepts of welfare also needed to be considered so that the index formed is able to more describe the phenomenon that is being measured.

References

[1] LPEM FEB University of Indonesia 2019 *Indonesia Economic Outlook* (Jakarta: Macroeconomic Analysis Series)

[2] World Bank 2001 *World Development Report 2000/2001: Attacking Poverty*. *World Development Report* (New York: Oxford University Press)

[3] Armiento, Mirko 2018 The Sustainable Welfare Index: Towards a Threshold Effect for Italy. *Ecological Economics* 152(2018) pp 296-309.

[4] Held, Benjamin et al 2018 The National and Regional Welfare Index (NWI/RWI): Redefining Progress in Germany. *Ecological Economics* 145(2018) pp 391-400.

[5] Sugiharto, Eko 2007 The Welfare Level of Fisherman Society of Benua Ilir Village Based on Badan Pusat Statistik Indicator. *Jurnal Perikanan EPP* Vol.4 No.2 2007 pp 32-36.

[6] Rao, J N K & Molina, Isabel 2015 *Small Area Estimation* (New Jersey: John Wiley & sons, Inc).

[7] Fay, R E & Herriot, R A 1979 Estimation of Income from Small Place: An Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association* 74 pp 269-277.

[8] Gonzales, M E 1973 Use and Evaluation of Synthetic Estimates. *Proceedings of The American Statistical Association, Social Statistics Section* pp 33-36.

[9] Annisa, Kurnia, & Indahwati 2014 Cluster Information of Non-Sampled Area in Small Area Estimation. *IOSR Journal of Mathematics* Vol 10 Issue 1 Ver II (February 2014) pp 15-99.

[10] Hair 2014 *Multivariate Data Analysis* (USA: British Library Cataloguing-in-Publication Data).

[11] OECD 2008 *Handbook on Constructing Composite Indicators* (France: OECD Publishing).

[12] Hadijah, Zara and Sadali, M. I. 2020 The Effect of Urbanization on Poverty Reduction in Indonesia. *Jurnal Wilayah dan Lingkungan*, 8(3) pp 290-306

[13] World Bank 2018 *September 2018 Indonesia Economic Quarterly: Urbanization for All* (The World Bank Group)

[14] Haris, Faisal and Ubaidillah, Azka 2020 Mean Squared Error of Non-Sampled Area in Small Area Estimation. *Proceedings of the 1st International Conference on Statistics and Analytics ICSA 2019* (Bogor)

[15] Monadiyan, N E 2020 The Impact of Parental Social Capital and Characteristic of Family to The Subjective Child Well-Being in 2019, Case Study at RW 14 Bidaracina, East Jakarta City. *Prosiding Seminar Nasional Official Statistics 2019,* 2019(1) pp 493-502

[16] Prakarsa 2020 *Indeks Kemiskinan Multidimensi Indonesia* 2015-2018 (South Jakarta: Prakarsa)

[17] Prasad, N G N & Rao, J N K 1990 The Estimation of the Mean Squared Error of Small-Area Estimators. *Journal of the American Statistical Association* 85 pp 163-171

[18] Santoso, K N 2020 Composite Index Analysis of Non Decent Work in Indonesia. *Jurnal Kependudukan Indonesia* Vol.15 No.1 Juni 2020 pp 19-32

Acknowledgments

The authors would like to thank the support of Subdirectorate of Household Statistics (BPS-Statistics Indonesia) and Division of Statistics Integration, Processing, and Disemination BPS-Statistics of Jawa Barat Province.