Regional income inequality model based on theil index decomposition and weighted variance coefficient

To cite this article: H R Sitepu et al 2018 J. Phys.: Conf. Ser. 983 012111

View the article online for updates and enhancements.
Regional income inequality model based on theil index decomposition and weighted variance coefficient

H R Sitepu*, O Darnius¹ and W N Tambunan¹

¹ Mathematics Department, University of Sumatera Utara, Indonesia

* Corresponding author: h.ranysitepu@gmail.com

Abstract. Regional income inequality is an important issue in the study on economic development of a certain region. Rapid economic development may not in accordance with people's per capita income. The method of measuring the regional income inequality has been suggested by many experts. This research used Theil index and weighted variance coefficient in order to measure the regional income inequality. Regional income decomposition which becomes the productivity of work force and their participation in regional income inequality, based on Theil index, can be presented in linear relation. When the economic assumption in j sector, sectoral income value, and the rate of work force are used, the work force productivity imbalance can be decomposed to become the component in sectors and in intra-sectors. Next, weighted variation coefficient is defined in the revenue and productivity of the work force. From the quadrate of the weighted variation coefficient result, it was found that decomposition of regional revenue imbalance could be analyzed by finding out how far each component contribute to regional imbalance which, in this research, was analyzed in nine sectors of economic business.

1. Introduction

Rapid economic development may not be in accordance with per capita income development. When it occurs in some regions, there will be regional income inequality. Regional income inequality today becomes an important issue in the study on economic development of a country, especially Indonesia which economic development leads to equal and fair distribution of prosperity. The interest in how big the regional revenue imbalance has caused the experts in development economy to develop some methods of measuring the imbalance. Some variations of statistics applications have been developed by measuring regional revenue imbalance such as Gini coefficient, variation coefficient, weighted variation coefficient, income logarithmic variants, and Theil index.

Williamson developed weighted variation coefficient to measure regional income inequality [1]. Theil developed theil index to measure regional inequality[2]. Some researchers attempted to collaborate with new methods to access and measure regional inequality. Akita T. and Kataoka M who measured regional income inequality in Japan in the period of the post-war Japan used weighted variation coefficient and theil index[1]. Sharrocks and Wan[3] and Marquez and Lasarte [2] developed spatial decomposition by using theil index. Anselin and Arribas-Bel developed statistics application in spatial data to measure regional inequality [4], while Rey and Smith [5] developed spatial decomposition by using gini coefficient. The difference in per capita income may partly be caused by the difference in productivity of work force. Carlino found out that many calculations on regional inequality in the United States could be explained by regional differences in the ratio of work force with population [6]. The difference in the participation of work force or unemployment can cause the
difference in per capita income among the regions even though the average productivity per worker is similar. Duro and Esteban analyzed the factors of these differences in productivity by using their index decomposition in which part of the population aggregate was used as the weight [7].Goerlich Gisbert widened the decomposition from Duro and Esteban to theil index of the other inequalities in which the part of the income aggregate was used as the weight [8]. Theil can calculate the inequality in a certain region or among the regions all at once so that the scope of the analysis will be widened. Secondly, theil index can calculate the contribution of each region to regional inequality as a whole so that it can make an important policy.

PDRB (Gross Regional Domestic Product) is one of the important indicators for recognizing the economic condition of a certain region in a certain period. A domestic concept is used to calculate regional income. It indicates that the whole value-added caused by various sectors or businesses which are operating in a certain region is calculated and included without paying attention to ownership and production factors. Therefore, aggregative PDRB indicates the capacity of a certain region to yield income or compensation to the factors of production in the process of production in that region.

Regional income inequality can partly be caused by the factor of work force productivity and their participation. Therefore, the criterion of regional revenue imbalance stated in PDRB will be decomposed in order that it can explain in detail some factors which influence regional inequality which is related to per capita income and work force productivity from each business and economic sector.

In this research, economic sector was assumed to be divided into the main \( j \) sector so that PDRB was used to study the influence of each sector on regional imbalance on PDRB. The main purpose of this research is to generalize regional income inequality model based on PDRB with a decomposition model by collaborating theil index and weighted variation coefficient, such that the model can be used to analyze the sectors that influence the regional income inequality.

### 2. Method

Decomposition technique on theil index and weighted variation coefficient would be used for PDRB by giving some definitions and notations to make easier its decomposition analysis.

#### 2.1. Concept and Definition of Work Force

The concepts and definitions used in the data gathering of work force by the Central Bureau of Statistics were the Work Force Concept suggested by the International Labor Organization (ILO). This concept divides population into two groups: working age population and non-working age population.

- **Working Age Population.** Working age population is the people who are above 15 years old, according to the regulation under Law No. 13/2013 on manpower.
- **Non-Working Age Population.** Non-working age population is the people who are under 15 years old.
- **Work Force.** Work force is working age people who are above 15 years old, either employed or unemployed.
- **Non-Work Force.** Non-work force is working age people that, in the reference period, do not have or do any economic activity at school, household, or others (retired employees, stipend receivers, deposit/interest receivers, the elderly, etc.)
- **Employment.** Employment is an activity of doing something in order to obtain or to help obtain earnings or profit at least in one hour of the previous week. This one hour work has to be done consecutively without being intermittent. It includes those who are working or having jobs do not do their work actively in the previous week because, for example, they were sick or absent. In some countries, the concept of working is based on gainful worker concept. This concept determines that a person works or does not is based on his usual activity, and this concept does not have any limited time.
- **Employment.** There are two definitions of unemployment: standard definition and relaxed definition. Standard definition of unemployment includes those who have no job and are
attempting to find one or preparing for a business. Relaxed definition of unemployment includes those who are not active in searching for a job but ready to work. Since 2001, definition of unemployment which has been used by Sakerns (National Work Force Survey) is relaxed definition so that unemployment includes four criteria: searching for a job, preparing for a business, being hopeless or discouraged worker, and having been hired but not ready to work.

2.2. Assumption and Notation
By using assumption and notation, Duro and Esteban [7] assumed that $X$, $E$, and $N$ respectively were denoting a criterion of regional income like PDRB, the amount of work force, and the population from the $i$ region. It was found that per capita income of the $i$ region could be annotated by

$$x_i = \frac{X_i}{N_i}.$$  

Then, the average per capita income could nationally be annotated by $\bar{x} = \frac{X}{N}$ where $X$ and $N$ consecutively indicate nationally or regionally the income value and the population in a general sense.

In general modeling, if it is assumed that economic or business sector is divided into $m$ sector, region is divided into $n$ regional and $X_{ij}$ indicates regional $i$ PDRB in $j$ sector, where $i = 1, 2, ..., n$ and $j = 1, 2, ..., m$, the total regional $i$ PDRB value will be the total of PDRB value of $m$ sector in regional $i$ which is denoted by

$$X_i = X_{i1} + X_{i2} + ... + X_{im}$$  

(1)

The same is true to the total of work force in regional $i$ which is the total of work force from the $m$ sector in regional $i$ which is denoted by

$$E_i = E_{i1} + E_{i2} + ... + E_{im}$$  

(2)

where $E_{ij}$ indicates the amount of work force in regional $i$ from the $j$ sector.

When work force productivity in regional $i$ is denoted by $y_i$ and defined as the difference in regional PDRB income in regional $i$ with the population in regional $i$, $y_i = \frac{X_i}{N_i}$ and work force participation in regional $i$ are denoted with $e_i$ and defined as the difference between work force in regional $i$ and the population in regional $i$ at $e_i = \frac{E_i}{N_i}$ per capapita income of regional $i$ can be decomposed to two components: work force productivity and work force participation in the following model:

$$x_i = y_i e_i$$  

(3)

Some researchers used decomposition in $i$; for example, Garrido-Yserte and Mancha-Navarro (2010) used it for the economic system in Spain, Doran and Jordan [9] used it for European NUTS 2 regions, and Marquez and Lasarte (2014) used it for provincial cases in Spain.

3. Result and Discussion

3.1. Per Capita Income Decomposition
Supposed $x_{ij}$ was per capita income in regional $j$ which was annotated by $x_{ij} = \frac{X_{ij}}{N_{ij}}$, $x_{ij}$ could be decomposed in multiplication into 3 components

$$x_{ij} = f_{ij} e_i = y_{ij} g_{ij} e_i$$  

(4)

where $\frac{X_{ij}}{E_i}$ was PDRB value from regional $j$ per total of work force in regional $i$, $y_{ij} = \frac{X_{ij}}{E_{ij}}$ of work force productivity from $j$ sector in regional $i$, $g_{ij} = \frac{E_{ij}}{E_i}$ was part of work force in $i$ from $j$ sector.

By using equations (1) and (4), per capita income could be written as follows:
\[ x_i = \left( \frac{X_{i1} + X_{i2} + \cdots + X_{im}}{E_i} \right) e_i \]
\[ = (f_{i1} + f_{i2} + \cdots + f_{im}) e_i \]
\[ = (Y_{i1}g_{i1} + Y_{i2}g_{i2} + \cdots + Y_{im}g_{im}) e_i \]

From equation (5), it was found that regional inequality in per capita income could be related to sectoral work force productivity, part of sectoral work force, and work force participation.

One of the indexes used to measure regional inequality in per capita income was the Theil index. Duran and Esteban made theil index decomposition by using part of population aggregate as the weight which was denoted as follows:

\[ T = \sum_{i=1}^{n} p_i \log \left( \frac{x_i}{X_i} \right) \]

where \( p_i \) indicated regional \( i \) population on the national population which was denoted by

\[ p_i = \frac{X_i}{N} \]

Goerlich-Gisbert (2001) widened theil index decomposition in the equation by using part of income aggregate as the weight which was denoted as follows:

\[ T = \sum_{i=1}^{n} q_i \log \left( \frac{x_i}{X_i} \right) \]

where \( q_i \) indicated the values of income in regional \( i \) nationally which was denoted as \( q_i = \frac{x_i}{X_i} \), and the value of \( \frac{x_i}{X} \) could be denoted as \( \frac{x_i}{X} = \frac{q_i}{p_i} \).

Therefore, theil index in the equation (7) could be written as follows:

\[ T = \sum_{i=1}^{n} \frac{x_i}{X} \log \left( \frac{x_i}{X} \right) \]

(8)

Since the economic sector was included in \( j \) sector where \( j = 1, 2, \ldots, m \), regional imbalance in work force productivity for \( j \) sector could be measured as follows:

\[ T_j = \sum_{i=1}^{n} \frac{X_{ji}}{X_j} \log \left( \frac{X_{ji}/X_j}{E_{ji}/E_j} \right) \]
\[ = \sum_{i=1}^{n} \left( \log (y_{ji}) - \log (\bar{y}_j) \right) \frac{X_{ji}}{X_j} \]

(9)

For \( j = 1, 2, \ldots, m \), where \( \bar{y} = \frac{X_j}{E_j} \) was work force productivity in \( j \) sector nationally.

\( X_j \) dan \( E_j \) was respectively PDRB value and the value of work force in \( j \) sector nationally.

Besides that, by using sectoral PDRB and work force values, regional inequality in work force productivity could be measured by

\[ T = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{X_{ij}}{X} \log \left( \frac{X_{ij}/X}{E_{ij}/E} \right) \]
\[ = \sum_{j=1}^{m} \sum_{i=1}^{n} \left( \log (y_{ij}) - \log (\bar{y}) \right) \frac{X_{ij}}{X} \]

(10)

Equation (10) could be decomposed in multiplication to become 2 components: component in \( (T_w) \) sector and \( (T_B) \) intersectoral component as follows:

\[ T = \sum_{j=1}^{m} \left( \frac{X_j}{X} \right) T_j + \sum_{j=1}^{m} \left( \frac{X_j}{X} \right) \log \left( \frac{X_j/X}{E_j/E} \right) \]
\[ = T_w + T_B \]
was the mean weight of regional inequality in the work force in each sector where \( T_j \) was denoted in equation (9), while \( T_B \) was regional inequality in the work force among the sectors which were denoted by

\[
T_B = \sum_{j=1}^{m} \left( \frac{X_j}{X} \right) \log \left( \frac{X_j / X}{E_j / E} \right)
\]

\[
= \sum_{j=1}^{m} \left( \log(\bar{y}_j) - \log(\bar{y}) \right) \frac{X_j}{X}
\]

For the measurement of inequality using theil index, when the theil index was close to 1 (one), it indicted that it was very imbalanced, but when theil index was close to 0, it indicated that it was equally distributed. Williamson (1965) calculated regional inequality in per capita income by using weighted variation coefficient which was denoted as follows:

\[
V = \frac{1}{x} \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 \frac{N_i}{N}}}
\]

where part of the population was used as weight with \( n \) the regional sequence.

The measurement of regional inequality by using weighted variation coefficient was nearly the same as theil index, when the index was close to 1 which indicated that it was very imbalanced; on the contrary, when theil index was close to 0, it indicated that it was equally distributed.

Finding out and analyzing how far the factors or components influenced regional revenue imbalance could be measured by calculating weighted variation coefficient which was denoted as follows:

\[
W_{jk} = \frac{1}{\bar{x}_j \bar{x}_k} \sum_{i=1}^{n} (x_{ji} - \bar{x}_j) (x_{ki} - \bar{x}_k) \frac{N_i}{N}
\]

where \( W_{jk} \) was weighted co-variation coefficient between \( j \) sector and \( k \) sector (\( j, k = 1, 2, ..., m ; \ j \neq k \)).

4. Result and Discussion

PDRB which is an economic indicator in a region can be decomposed to be work force productivity and work force participation by using theil index decomposition and by assuming regional income which is divided into \( m \) sector. Therefore, theil index decomposition and weighted variation coefficient can also be used to analyze some factors which contribute to regional revenue imbalance. This model can be used to find out the difference in interregional income inequality in Indonesia and to find out what factors which contribute to this inequality.

References

[1] Akita T and Kataoka M 2003 Regional income inequality in the post war Japan 43rd Congress of European Regional Science Association (Finlandia: Jyvaskyla)

[2] Marquez M A and Lasarte E 2014 Decomposition of Regional Income Inequality and Neighborhood Component: a Spatial Theil Index. Working Papers, EC Series

[3] Shorrocks A and Wan G 2005 Journal of Economic Geography 5 59

[4] Anselin L & Arribas-Bel D 2013 Papers in Regional Science 92 3-

[5] Rey S J and Smith R J 2010 Letters in Spatial and Resource Science 6 55

[6] Carlino G A 1992 Are regional per capita earnings diverging? Business Review 3

[7] Duro J A and Esteban J 1998 Economics Letters 60 269

[8] Goerlich-Gisbert F J 2001 Economics Letters 70 303

[9] Doran J and Jordan D 2013 Journal of Economics Studies 40 22

[10] Garrido-Yserte R and Mancha-Navarro T 2010 The Spanish regional puzzle: Convergence, divergence, and structural change. Regional Policy, Economic Growth and Convergence ed JR Cuadrad-Roura, (Berlin: Springer-Verlag)
[11] Shorrocks A F 1982 *Econometrica* 50 193
[12] Akita T 2003 *Annals of Regional Science* 37 55