Almost stochastic dominance for poverty level in Central Java Province

Isnandar Slamet¹, Aryanto Agus Wibowo², and Mania Roswitha³

¹Department of Statistics, Sebelas Maret University, Surakarta, Indonesia 57126
²³Department of Mathematics, Sebelas Maret University, Surakarta, Indonesia 57126

E-mail: isnandar06@yahoo.com

Abstract. The criteria for the domination of the distribution function has been used in the investment issues, momentum, agricultural production, and so on. One criteria of domination is stochastic dominance (SD). When this criteria is applied to the dominating area that has smaller value than the dominated area, then almost stochastic dominance (ASD) can be used. In this research, we apply the ASD criteria on data of expenditure per capita based on districts/cities in Central Java. Furthermore, we determine which year the expenditure per capita in the period 2009-2013 is the most dominating to know the level of poverty in Central Java. From the discussion, it can be concluded that the expenditure per capita in Central Java in 2013 dominates expenditure per capita in Central Java in 2009-2012. In other words, the level of poverty in Central Java in 2013 is lower than in 2009-2012.

1. Introduction

Poverty is one of the major problems faced by developing countries, including Indonesia. The changes of the level of poverty in Indonesia is one of the big concern that is observed every year. According to the Indonesia Central Statistics Agency [1], the poor are the people who have an average per capita expenditure below the poverty line. Therefore, the poverty line is used to determine whether a person is poor or not.

Generally, there are three indexes that are used to measure poverty levels. These are head count index, poverty depth index, and poverty severity index. The third index is a family of poverty indices F-G-T (Foster-Greer-Thorbecke) which is often used to determine the changes in poverty over time and regions. According to Madden and Smith [6], the weaknesses of the three indexes in knowing the changes in poverty is sensitive to the selection of the poverty line as well as the size of the poverty level. To overcome this problem, it is required a more robust approach to the poverty line as well as the size of the poverty level. Madden and Smith [6] explained that the stochastic dominance (SD) method is a powerful method in the selection of the poverty line. As noted by Ravallion [7], the poverty line used in the stochastic dominance interval is the region interval which produces same conclusion when the specific poverty line is used.

The SD method is the method which can be used to compare two distribution functions at which certain distribution function is more dominant than other (Heyer [3]). In discussion of SD, there are three criteria which we call first degree stochastic dominance (FSD), second degree stochastic dominance (SSD), and third degree stochastic dominance (TSD). Based on the definition of SD, a random variable Y dominates a random variable X when the condition of...
the expectation of utility price $Y$ is higher or equal than the expectation of the utility price $X$. However, if the criteria $SD$ is not satisfied, then the alternative decision-making almost stochastic dominance ($ASD$) can be used (Leshno and Levy [4]). In $ASD$, there are almost first stochastic dominance ($AFSD$) and almost second stochastic dominance ($ASSD$). The values of $AFSD$ can be calculated from the areas that violate the criteria of stochastic dominance divided by the total absolute value of the area between the cumulative distributions of $X$ and the cumulative distribution of $Y$ (Leshno and Levy [4]). The value of $ASSD$ can be obtained from the quotient between the areas that violate the criteria stochastic dominance divided by the average area dominated.

In this paper, we apply the method $ASD$ (Levy [5]) for analysing of poverty level based on the expenditure per capita per month in Central Java for period 2009 to 2013.

2. Literature Review

Method $SD$ is a method to determine the criteria for decision-making when at least one criterion $SD$ is negative (Levy [5]). Heyer [3] stated that $SD$ is a term that refers to a relationship between two investment distribution functions $F$ and $G$, where one investment dominates the other investment. If there are two investment options $F$ and $G$, then $G$ investment will be favored over $F$ if and only if the expectation of the utility $G$ is higher or equal to the expectation of the utility $F$. Systematically, it can be written as

$$E_F[u(x)] \leq E_G[u(x)], \quad (1)$$

where $E_F[u(x)]$ is the expectation of the utility $F$ and $E_G[u(x)]$ is the expectation of the utility $G$. According to Bain and Engelhardt [2], if $X$ is a continuous random variable, equation (1) can be written as

$$\int_{-\infty}^{\infty} u(x)f(x)dx \leq \int_{-\infty}^{\infty} u(x)g(x)dx, \quad (2)$$

with $f(x)$ and $g(x)$ each of which is the probability distribution function of the random variable $X$ and $u(x)$ is defined as the utility value of the random variable $X$.

Equation (2) is a formula $SD$ generally with at least there is an inequality $u(x) \leq 0$. From equation (2) with an investment of $G$ dominate investment of $F$, criteria of $FSD$ is

$$F(t) \geq G(t). \quad (3)$$

When equation (3) is not fulfilled, then the criteria $FSD$ can be used in the decision-making.

According to Levy [5], $S_1$ is defined as the total value when the rule criteria $FSD$ is not fulfilled $S_1(F,G) = t : F(t) < G(t)$. Areas that are not fulfilled the criteria $FSD$ with continuous random variables $X$ and $Y$ can be written follow.

$$\int_{S_1} [F(t) - G(t)]dt. \quad (4)$$

When $X$ and $Y$ are discrete random variables, equation (4) can be written as

$$\frac{1}{n} \sum_{i:y_i > x_i} (y_i - x_i) \quad (5)$$

where $y_i$ is the data that is assumed to dominate and $x_i$ is the data that is assumed to be dominated. Then it can be written as

$$\int_{S_1} [F(t) - G(t)]dt = A_1.$$
For continuous random variables $X$ and $Y$, the total area of $FSD$ between these two variables can be written as

$$\int_{S} [F(t) - G(t)] dt.$$  \hspace{1cm} (6)

For discrete random variables, equation (6) can be written as

$$\frac{1}{n} \sum_{i=1}^{n} |y_i - x_i|$$  \hspace{1cm} (7)

or

$$\int_{S} |F(t) - G(t)| dt = A_1 + A_2.$$  

The results for the areas that do not meet the criteria $FSD$ and the total area between random variables $X$ and $Y$ is continuous or symbolized by $\varepsilon$ and formulated with

$$\varepsilon_{AFSD} = \frac{\int_{S} |F(t) - G(t)| dt}{\int_{S} |F(t) - G(t)| dt}.$$  \hspace{1cm} (8)

If the random variables $X$ and $Y$ are discrete, the equation (8) can be solved using the equation (5) dan (7), and we obtain

$$\varepsilon_{AFSD} = \frac{\sum_{i:y_i > x_i}^n (y_i - x_i)}{\sum_{i=1}^{n} |y_i - x_i|}.$$  \hspace{1cm} (9)

Equation (9) can be written as

$$\varepsilon_{AFSD} = \frac{A_1}{A_1 + A_2}.$$  

If $0 < \varepsilon < 0.5$, equation (9) meets the criteria $AFSD$ with $G$ dominating $F$ with $\varepsilon_{AFSD}$. Thus, $Y$ is said to dominate the $X$ with $AFSD$ (for $\varepsilon$ specific, or $\varepsilon_{AFSD}$) and is denoted by $G \geq \varepsilon_{AFSD} F$ if and only if:

$$\int_{S} [F(t) - G(t)] dt \leq \varepsilon \int_{S} |F(t) - G(t)|.$$  \hspace{1cm} (10)

Using a discrete equation (10) can be written

$$\sum_{i:y_i > x_i}^n (y_i - x_i) \leq \varepsilon \sum_{i=1}^{n} |y_i - x_i|$$

provided $0 < \varepsilon < 0.5$.

According to Levy [5], $G$ dominate the $F$ in $SSD$ if and only if

$$\int_{-\infty}^{t} F(X) dx \geq \int_{-\infty}^{t} G(X) dx$$  \hspace{1cm} (11)

for all $t \in S$. It is assumed that this inequality is true for most of the range of $S$. Suppose that $S_2$ defined as the value of the area ranges $SSD$ violated and formulated with

$$S_2 = \{ t : F(t) < G(t) ; \int_{-\infty}^{t} G(x) dx \geq \int_{-\infty}^{t} F(x) dx \}$$
and complement the area of S2 is denoted by \( S_2 \) and \( S_2 \cup \overline{S}_2 = S \). Areas that do not meet the criteria SSD with random variables X and Y is continuous may be written by the equation

\[
\int_{S_2} (F(t) - G(t)) dt.
\]

(12)

When the random variable X and Y discrete equation (12) can be written

\[
\frac{1}{n} \sum_{i:y_i > x_i} (y_i - x_i)
\]

(13)

with \( x_i \) is the data that is assumed to dominate and \( y_i \) is the data that is assumed to be dominated. While the area range SSD between random variables X and Y is continuous may be written

\[
\int_{S} (F(t) - G(t)) dt + 2 \int_{S_2} (F(t) - G(t)) dt.
\]

(14)

If the total area of SSD between random variables X and Y is discrete, then the equation (14) can be written

\[
\frac{1}{n} \sum_{i=1}^{n} (x_i) - \sum_{i=1}^{n} (y_i) + 2\left[ \frac{1}{n} \sum_{i:y_i > x_i} (y_i - x_i) \right].
\]

(15)

The results for the areas that do not meet the SSD and the total area of the random variable X and Y is continuous, or can be formulated with \( \varepsilon_{ASSD} \) defined by

\[
\varepsilon_{ASSD} = \frac{\int_{S_2} (F(t) - G(t)) dt}{\int_{S} (F(t) - G(t)) dt + 2 \int_{S_2} (F(t) - G(t)) dt}.
\]

(16)

From equation (16) can be described as follows

\[
\varepsilon_{ASSD} = \frac{\int_{S_2} (F(t) - G(t)) dt}{\int_{S} (F(t) - G(t)) dt + 2 \int_{S_2} (F(t) - G(t)) dt} = \frac{\int_{S_2} (F(t) - G(t)) dt}{\int_{S_2} (F(t) - G(t)) dt + \int_{S_2} (F(t) - G(t)) dt + \int_{S_2} (F(t) - G(t)) dt + \int_{S_2} (F(t) - G(t)) dt} = \frac{\int_{S} (F(t) - G(t)) dt + \int_{S_2} (F(t) - G(t)) dt = \int_{S} (F(t) - G(t)) dt}.
\]

where

\[
\int_{S_2} (F(t) - G(t)) + \int_{S_2} (F(t) - G(t)) dt = \int_{S} (F(t) - G(t)) dt.
\]

Because \( S_2 \cup \overline{S}_2 = S \), we get

\[
-\int_{S_2} (G(t) - F(t)) dt = \int_{S} (F(t) - G(t)) dt.
\]

According to Levy [5]

\[
\int_{S} (F(t) - G(t)) dt = E_F - E_G
\]

(17)

with \( E_F \) is an average value of F and \( E_G \) is an average value of G. Then the area was violated SSD range denoted more simply by \( VA \), where

\[
VA = \int_{S_2} (F(t) - G(t)) dt.
\]

(18)
Equation (17) and equation (18) are substituted in equation (16) in order to obtain
\[ \varepsilon_{ASSD} = \frac{VA}{E_F - E_G + 2VA}. \]

If \( \varepsilon_{ASSD} < 0.5 \) G dominate the F is ASSD. Levy [5] stated that the lower the value \( \varepsilon \) obtained, the higher the degree of dominance. A random variable \( Y \) is said to dominate the \( X \) with \( ASSD \) (for \( \varepsilon_{ASSD} \)) and is denoted by \( G \geq \varepsilon_{ASSD}F \) if and only if:
\[
\int_{S_2} (F(t) - G(t))dt \leq \varepsilon\left[ \int_{S} (F(t) - G(t))dt + 2\int_{S_2} (F(t) - G(t))dt \right]. \tag{19}
\]

For discrete case, equation (19) can be written as
\[
\frac{1}{n} \sum_{i=1}^{n} (y_i - x_i) \leq \varepsilon\left[ \frac{1}{n} \sum_{i=1}^{n} x_i - \frac{1}{n} \sum_{i=1}^{n} y_i + 2\frac{1}{n} \sum_{i: y_i > x_i} (y_i - x_i) \right]
\]
provided \( 0 < \varepsilon < 0.5 \).

### 3. Research Methodology

In this research, we applied ASD on poverty levels in Central Java based on data of expenditure per capita per month in all districts/cities in Central Java for period 2009 until 2013. Suppose there are two investments i.e \( F \) and \( G \). Then, we determine the cumulative distribution functions of \( F \) and \( G \) which we write as \( F(t) \) and \( G(t) \), respectively. Then, we calculate violation area based on SD criteria. Next, we determine the value of \( \varepsilon \) using ASD from the violation area. Finally, we apply the ASD method for data of average expenditure per capita per month to get the conclusion.

### 4. Results

ASD method has been applied to data of average expenditure per capita per month for all districts/cities in Central Java in 2009 until 2013 that were taken from the Central Statistics Agency (BPS) [1]. In this case, the ASD method is used to analyze whether the data of the average expenditure per capita per month for all districts / cities in Central Java in 2013 dominate (higher) than the average data per capita per month spending in 2009, 2010, 2011, and 2012.

#### 4.1. Randomness Test

According to Wackerly [8], randomness test is used to determine whether the data used is random. If the data is random then data is mutually independent and identical. The steps of test is given below.

(i) \( H_0 \) : data of average expenditure per capita per month for all districts/cities in Central Java is random.

(ii) \( H_1 \) : data of average expenditure per capita per month for all districts/cities in Central Java is not random.

(iii) Level of significance \( \alpha = 0.05 \).

(iv) Critical area : \( H_0 \) is rejected if \( Z_{Calc} \leq -Z_{0.025} = -1.96 \) or \( Z_{Calc} \geq Z_{0.025} = 1.96 \).

(v) Test of statistic :
\[
Z = \frac{R - E(R)}{\sqrt{V(R)}}
\]
(vi) Conclusion.

Using a significance level $\alpha = 0.05$, we conclude that the data on average expenditure per capita per month for all districts/cities in Central Java for 2009 to 2013 are random.

4.2. Stochastic Dominance

Value development expenditure per capita per month is calculated from the assumed outcome highest expenditure reduced by expenditure will be compared. Value development expenditure per capita per month is calculated from the reduction of expenditure per capita in Central Java in 2013 to the expenditures in 2009, 2010, 2011, and 2012. As an example of the value of the development expenditure in Central Java in 2009 against 2013 in Sragen earned $-1,035$ rupiah. The percentage of development expenditure in Sragen obtained $-0.004$. There are areas that do not meet the criteria $FSD$ by equation (3) and does not meet the criteria $SSD$ by equation (11). Then, we need to calculate the value of $\varepsilon$ through $FSD$ and $SSD$.

4.3. Almost Stochastic Dominance

Criteria $FSD$ and $SSD$ has negative value. Therefore, it can not be used. Furthermore, the criteria $AFSD$ and $ASSD$ are used to find out if the data expenditure per capita per month in Central Java in 2013 dominate the data in 2009, 2010, 2011, and 2012 based on a percentage of its development. The calculation on the values of $\varepsilon_{AFSD}$ and $\varepsilon_{ASSD}$ is presented in Table below.

| Year  | $\varepsilon_{AFSD}$ | $\varepsilon_{ASSD}$ |
|-------|----------------------|----------------------|
| 2009  | 0.00085              | $1.078 \times 10^{-6}$|
| 2010  | 0.014                | $1.749 \times 10^{-6}$|
| 2011  | 0.079                | $1.146 \times 10^{-5}$|
| 2012  | 0.081                | $1.188 \times 10^{-5}$|

Because the value of $\varepsilon_{AFSD}$ and $\varepsilon_{ASSD}$ lies between 0 to 0.5, it means that the criteria $AFSD$ and $ASSD$ can be used. Expenditure per capita per month by district/city in Central Java in 2013 dominates the spending in Central Java in 2009-2012 based on the percentage of development expenditure through the criteria of $AFSD$ and $ASSD$. It explains that in 2013 the lowest poverty rates compared to the 2009-2012 seen from the percentage of development expenditure per capita. Based on the results of $\varepsilon$ obtained, it can be concluded that the value of $\varepsilon_{ASSD}$ is better than the value of $\varepsilon_{AFSD}$ data spending per capita per month in Central Java in 2009 until 2013. This is because the value of $\varepsilon_{ASSD}$ is nearly zero so that the higher degree of dominance.

5. Conclusion

Based on the discussion, it can be concluded that expenditure per capita per month for all districts/cities in Central Java in 2013 dominates expenditure per capita per month in Central Java in 2009-2012 based on the percentage of development expenditure using the criteria of almost stochastic dominance first and second order degree. It explains that in 2013 the lowest poverty rates compared to 2009-2012.

References

[1] Badan Pusat Statistik Jawa Tengah, Politik dan Keamanan Provinsi Jawa Tengah, Badan Pusat Statistik Jawa Tengah, 2011.
[2] Bain, L.J., and Engelhardt, M. Introduction to Probability and Mathematical Statistics, Duxbury Press An Imprint of Wadsworth Publishing Company, Belmont California, 1992.
[3] Heyer D.D., Stochastic Dominance: A tool for Evaluating Reinsurance Alternative, Casualty Actuarial Society (CAS) Form, 2001.

[4] Leshno, M. and H. Levy, Preferred by All and and Preferred by Most Decision Makers: Almost Stochastic Dominance, Journal of Management Science 48 (2002), 1074-1085.

[5] Levy, M., Almost Stochastic Dominance and Efficient Investment Sets, American Journal of Operations Research 2 (2012).

[6] Madden, D. and F. Smith, Poverty in Ireland, 1987-1994: A Stochastic Dominance Approach, The Economic and Social Review 31 (2000), no. 3, 187-214.

[7] Ravallion, M., Poverty Comparisons. A guide to Concepts and Methods, LSMS Working Paper Number 88, The World Bank (1992).

[8] Wackerly, D. D., W. Mandenhall III and R. L. Sceaffer. Mathematical Statistics with Applications, 7ed., Thomson Brooks/Cole, Dukbury, 2008.