



IARIW-Bank of Korea Conference “Beyond GDP: Experiences and Challenges in the Measurement of Economic Well-being,” Seoul, Korea, April 26-28, 2017

Inequality of Opportunity in South Korea

Woojin Lee (Korea University) and Jinsoo Cho (Korea University)

Paper prepared for the IARIW-Bank of Korea Conference

Seoul, Korea, April 26-28, 2017

Session 5B: Theory and Measurement of Well-being: The Equality of Opportunity and the Capability Approaches

Time: Thursday, April 27, 2017 [Afternoon]

Inequality of Opportunity in South Korea

Woojin Lee

Department of Economics,
Korea University,
Anam-dong, Seoungbuk-gu,
Seoul 136-701, Korea.

wojinlee@korea.ac.kr

Jinsoon Cho

Department of Economics,
Korea University,
Anam-dong, Seoungbuk-gu,
Seoul 136-701, Korea.

ursong@korea.ac.kr

Abstract

We decompose inequality of individual achievement, measured by income, education, and health, into the part due to unequal circumstance and due to differential exercise of individual effort. We use individual data of people in the Korea Labor and Income Panel Studies who were born between 1960 and 1980. Unlike Ko and Lee (2013), who used only one circumstance variable (father's level of education), we use five circumstance variables: father's education, gender, birth year, grown-up region until 14 years old, and the number of siblings. We chose individual education and individual income as the variables for individual achievement. Regarding inequality of individual education, the circumstance is found to account for 47% of the inequality, and the effort is found to account for 53%. Among the circumstance variables, father's education contributes the most (about 31% of inequality) and the gender the second largest (7%). Birth year accounts for 4%, the number of siblings accounts for 3%, and grown-up region accounts for 2%. Regarding inequality of individual income, circumstances account for 52% of the inequality, individual effort accounts for 48% of the total inequality. Gender is the most important circumstance variable (40%), and father's education (11%) is the next. Birth year and grown-up region exhibit little effect (less than 1%). The number of siblings shows almost no effect. Regarding inequality of individual health, the circumstance is found to account for 38% of the inequality. Gender contributes the most (23%) of the inequality, and age(13%) is the next.

Key Words: Inequality of opportunity, circumstances, individual effort, Shapley value decomposition

JEL Code: D6, H3, J3

1. Introduction

There is a consistent discussion about inequality, as inequality is getting worse. Traditionally, a discussion about inequality was mainly about inequality of outcome, but recently, an active discussion about inequality of opportunity is in progress.

The research about inequality of opportunity gives a great significance in a way that it gives policy alternatives by understanding the cause of inequality such as an effort and circumstances, not just analyzing the degree of inequality. An empirical study of inequality has great significance about Korea's latest 'dirt spoon, and gold spoon' issue.

Despite the fact that inequality of opportunity is quite a familiar term to us, strict definition about the term needs to be preceded in order to do an empirical research on the inequality of opportunity. In this paper, we used the concept of the inequality of opportunity defined by Roemer (1993, 1998). Roemer first separates the inequality in a society into the inequality that is attributed to circumstances and to effort, and then call the inequality that is attributed to circumstance inequality of opportunity. In this context, circumstances are the vector of factors that is beyond individuals' control, but that have great effect on their achievements. To be specific, parent's education, parent's income/ fortune, parent's job, race, gender, birth place or growing region, number of siblings, birth year and genetically determined physical/cognitive ability are the factors that individuals can't choose, but these have influence on their achievements greatly. We decompose the inequality by the part that is attributed to difference in circumstances and that in effort and then compute the contribution of each circumstance factors using the Shapley-value decomposition.

Among various data that tracked individuals' information, we chose KLIPS which is proper for the inequality of opportunity research. We used 5 years (2010 to 2014) of KLIPS

data. Our analysis sample is individuals who are born 1960 (54 years old at 2014) and 1980 (34 years old at 2014). These individuals are the individuals that are doing the most active economic activities compared to other ages in the same year. Then, we used five circumstance variables: father's education, gender, number of siblings, grown-up region, and birth year. We used inequality in labor income, individual's education and BMI as an inequality of individual achievement: dependent variables.

In Korea, Ko and Lee (2013) is the first research that decomposed the inequality with the part with circumstances and that with effort. However, this paper used only one circumstance variable: father's level of education. On the other hand, this paper used five circumstance variables: father's education, gender, number of siblings, grown-up region, and birth year. Thus, a key difference between this paper and Ko and Lee (2013) is that this research not only shows how much circumstances contribute to inequality, but also shows how much each five circumstance variable contributes to inequality of opportunity.

Our paper extends to the work of Erikson et al. (2015), who estimated inequality of opportunity in long-run income in Sweden. As circumstance variables, they used parental income, family structure, number of siblings, IQ, and non-cognitive ability. Unlike our paper, Erikson et al. (2015) controlled gender and compared the inequality of opportunity by gender.

We used age when analyzing the effect of birth year. There seems to be no problem with using age instead of birth year, as age follows birth year. However, there must be some attention when treating age entirely a circumstance variable. First, we can definitely treat individual's age as a circumstance variable in a sense that it reflects the one's birth year. However, at the same time, the individual accumulate experience and improve skills as they grow old. To be specific, proper interpretation about the fact that income of individuals increase as one grows older would be the result of accumulation of one's skills and

experience, not the result of factors that is beyond individuals' control. On the contrary, in South Korea, education level of individuals decrease as one grows older, which can be explained by the lack of opportunity in education for those who were born in a long time ago. Therefore, as age contains the factor as a circumstance and the factor as one's effort at the same time, the interpretation should be done with precaution. In Section 3, we present the method to decompose a circumstance factor among age.

We mainly used income data which has a value greater than zero. It was to prevent overestimating the effect of those who has no income at all such as housewives to the results. However, the results of using the whole sample and the results of using male sample were generally similar.

We find that regarding inequality of individual education, the circumstance is found to account for 47% of the inequality, and the effort is found to account for 53%. Among the circumstance variables, father's education contributes about 31% of inequality and the gender about 7%. Regarding inequality of individual income, circumstances account for 52% of the inequality, individual effort accounts for 48% of the total inequality. Gender is the most important circumstance variable, which accounts for 40% of the inequality, and father's education about 11%. When it comes to the health inequality, circumstance attributes 38% of the inequality and individual effort accounts for 62% of the inequality. Gender contributes the most in inequality as 22%, and age about 14%.

The rest of the paper is structured as follows. In Section 2, we provide short literature review of papers inside and outside of Korea. In Section 3, we present the model and method and we describe the data in Section 4. In Section 5, we report the results. Lastly, we summarize the results and provide a further discussion.

2. Literature Review

The first research about inequality of opportunity in Korea is Kim and Lee (2008). This paper computes optimal tax and redistribution policy to reduce the level of inequality. Also, they find out that Korean tax-benefit policies have played almost no role in correcting unequal opportunities for income acquisition among people.

Another way of inequality research is decomposing inequality into the part due to unequal circumstance and due to differential exercise of individual effort. In Korea, there is Ko and Lee (2011) that used a circumstance variable in the analysis. In this paper, they consider father's education as a circumstance variable and analyzed how this circumstance explains inequality of son's achievement. The result shows that father's education account 16~25% of son's education inequality and 2~12% of son's income inequality.

Kim et al. (2016) analyzes the inequality of circumstance using three circumstance variables : education, father's education, and gender. The result shows that, the circumstance variables account for 88.3% increase in inequality of opportunity of male who is in their thirties for 10 years. As following policy implications, this paper suggests the policy that can improve the inequality, improve the quality of employment, and to expand budget on public education.

For research in overseas, there are Erikson et al. (2015). This paper uses 5 circumstance variables: parental income, parental education, number of siblings, family structure, IQ, and non-cognitive ability. This paper controlled gender and compared inequality of opportunity by gender using other circumstance variables. Their result is that, the circumstance variable accounts 31% of male's inequality and 25% of female's inequality. The result shows that inequality of opportunity is greater to male than female.

Peichl et al. (2015) uses various circumstance variables including height of individuals to analyze the part of inequality of opportunity that is explained by the spouse in Germany. According to this paper, the effect of income of the spouse is decreasing due to assortative mating, and less responsibility of the individual on the spouse variable, higher the level of inequality of opportunity.

Ferreira et al. (2008) analyzes inequality of opportunity in Latin America using various circumstance variables including a race. In this paper, they explore inequality of opportunity of expenditure and income. The result shows that the inequality in opportunity is higher in income than expenditure in any cases.

Checchi, Peragine (2010) compares inequality of opportunity of South Italy and North Italy using several circumstance variables. As a result, the paper shows that inequality of opportunity in labor market is greater in South Italy, which is less developed region. This paper analyzed the reason of this result is can be linked to internal migration flows such as brain drain.

Nilsson (2005) analyzes inequality of opportunity in Sweden, using many circumstance variables about family background. The result shows that the inequality of opportunity in labor income, stable family relationship (nurturing parent matches with biological parent) accounts the most.

Bourguignon et al. (2003) explores inequality of opportunities about income in Brazil. This paper limited their sample on city dwellers. As a result, among circumstance variables, parental education accounts 55% of inequality of income, and adding father's occupation, two circumstance variables accounts 80% of inequality.

3. Methods

We now explain the methods we used in this paper in three ways. First, we show the conceptual structure of the model and the regression specifications we used in analysis. Second, we present the method of decomposing the inequality that is due to difference in circumstance and due to difference in effort using Shapley-value decomposition. Lastly, we outline the method of decomposing the estimated value of effort to type-specific effort and individual effort.

1. The conceptual framework of the model and regression specification

The approach we use is originally based on Erikson et al. (2015). We measure the inequality by two measures: Gini coefficient (Gini) and coefficient of variation (CV). Gini is sensitive to a variation in inequality of middle class, but not sensitive to variation in inequality of upper class. On the other hand, CV is sensitive to variation in upper class than in middle class.

We intend to explore which part is attributable to circumstance and which part is attributable to effort. Circumstances are the factors that are beyond individual's control, but affects individual achievements. There are various circumstance variables beside out circumstance variables: father's education, gender, age, grown-up region, and number of siblings, such as height, appearance, family background, and IQ. In this paper, we used only 5 circumstance variables due to a matter of measurement. Individual effort is the part that affects individual by diligence and will when everyone is under same circumstances. The problem is that it is hard to measure the effort. Therefore, we substituted the parts that measurable circumstances account for from the variation of a dependent variable, and set the

rest residual as an effort. Thus, the part of our result which is regarded as the part due to an effort could actually be due to circumstances.

We use the following regression for the analysis.

$$Y = \mu + \sum_{j=1}^J X_j \beta_j + \epsilon. \quad (1)$$

In this regression, Y is a variable that presents individual achievements such as individual education and income, and (X_1, \dots, X_J) is a vector of circumstances with J being the number of circumstances. ϵ is a residual that we treat the part as individual effort for convenience, despite the fact that part of it reflects the effect of circumstances. The aim of our analysis is to measure how much is due to variation of (X_1, \dots, X_J) among variation of Y.

Now, we present the regression specification by dividing Y as individual education(Edu), individual income(Wage) and individual health(BMI).

(1) A regression specification to estimate education equation

We estimate education equation by a regression of individual education on circumstance variables.

$$Edu_i = \alpha_0 + \alpha_1 Fdeu_i + \alpha_2 Male_i + \alpha_3 Age_i + \alpha_4 Growreg_i + \alpha_5 Nsib_i + \epsilon_i, \quad (2)$$

with *Fedu* as father's education, *Male* as a dummy variable that presents male, *Age* as an

individual's age according to birth year, *Growreg* as a dummy variable that presents whether the grown-up region is above metropolitan city and *Nsib* as a number of siblings.

As we can see from a following result, an individual education has a strong negative relation with an individual birth year. This reflects that for those who were born in earlier times, the opportunity of education was very limited in Korea. As Korea went through a rapid economic growth, the opportunity of education varied considerably in the time of a birth. Therefore, the effect of birth year in education equation is completely considered as the part that is due to circumstance factors.

(2) A regression specification to estimate income equation

In the same context with (1), we gain a following regression of wage on circumstance variables.

$$Wage_i = \beta_0 + \beta_1 Fdeu_i + \beta_2 Male_i + \beta_3 Age_i + \beta_4 Growreg_i + \beta_5 Nsib_i + \epsilon_i. \quad (3)$$

However, there is a critical error in this equation, as individual education is treated as a residual, omitted from the equation. Thus, a more accurate equation is

$$Wage_i = \beta_0 + \beta_1 Fdeu_i + \beta_2 Male_i + \beta_3 Age_i + \beta_4 Growreg_i + \beta_5 Nsib_i + \beta_6 Edu_i + \epsilon_i. \quad (4)$$

If we compare equations (3) and (4), we can see that $\epsilon_i = \beta_6 Edu_i + \varphi_i$. However, it is hard to distinguish the part due to circumstances from individual education, if we insert education

directly to the equation like (4). Therefore, in order to decompose the part accountable to circumstances among education, we substitute (2) to education variable in (4), and gain the equation

$$\begin{aligned}
 Wage_i &= \gamma_0 + \gamma_1 Fdeu_i + \gamma_2 Male_i + \gamma_3 Age_i + \gamma_4 Growreg_i + \gamma_5 Nsib_i + \mu_i. \\
 \gamma_j &= \beta_j + \widehat{\alpha}_j \beta_6, j = \{0, \dots, 5\}, \\
 \mu_i &= \varphi_i + \beta_6 \epsilon_i.
 \end{aligned} \tag{5}$$

However, we can't instantly analyze the circumstances and effort with estimated coefficients and variables as there is a problem in birth year (Age) variable. We can only observe birth year as itself, not the birth year distinguished with two parts, one as circumstance part, and the other as effort part. The birth year is definitely a circumstance variable in a way that it is beyond individual's control. On the other hand, individuals earn higher wages as they grow old, which is due to accumulation in their experience and skills. According to (5), age variable is

$$\gamma_6 Age_i = (\beta_3 + \widehat{\alpha}_3 \beta_6) Age_i. \tag{6}$$

In equation (6), $\widehat{\alpha}_3 \beta_6 Age_i$ present the effect of a birth year on income through education. Therefore, we can say that this is a part of birth year as a circumstance variable. On the other hand, $\beta_3 Age_i$ is the part where the effect as a circumstance variable and as an effort are mixed. Thus, for the sake of precision in our analysis, we omit this part from individual income and use the following equation.

$$Wage_i^{Adj} = \delta_0 + \delta_1 Fdeu_i + \delta_2 Male_i + \delta_3 Age_i + \delta_4 Growreg_i + \delta_5 Nsib_i + \rho_i. \quad (7)$$

$$Wage_i - \widehat{\beta}_3 Age_i = Wage_i^{Adj} \quad (8)$$

Now, every circumstance variables in (7) including Age are the variables that have the characteristic only as circumstance variables. We use this regression to find out how much each circumstance variables contributes income, and how much estimated residual contributes an individual effort.

(3) A regression specification to estimate health equation

We estimate health equation by a regression of individual's health indicator on circumstance variables.

$$BMI_i = \varepsilon_0 + \varepsilon_1 Fdeu_i + \varepsilon_2 Male_i + \varepsilon_3 Age_i + \varepsilon_4 Age_i^2 + \varepsilon_5 Growreg_i + \varepsilon_6 Nsib_i + \varepsilon_7 Workout_i + \tau_i, \quad (9)$$

with Workout as a dummy variable that presents if individual workout regularly. Also, we added Age_i^2 to observe if there is a turning point in age.

2. Decomposition of inequality using Shapley-value decomposition

Shapley-value decomposition is equilibrium allocation that in a coalition, players allocate a payoff by each player's marginal contribution. In this paper we use the approach originally based on Lee and Lee (2016).

Shapley-value decomposition is useful when decomposing inequality measures by a

marginal contribution of each source. In this paper, we decompose individual income to the sources such as the part by father's income, the part by gender, etc. In this context, we can measure the marginal contribution of each source by using the sources of individual income are players of a game, and various combinations of income sources and the resulting inequality measures.

When decomposing inequality measures by Shapley-value decomposition, we should pay attention about how to deal with sources that are not included in S which is a universal set comprised of various combinations of income sources. In this paper we used 'equalized shapley-value method'.

Originally designed by Cantreuil and Trannoy (1999), 'equalized shapley-value method' is the the method that measure a inequality of S by adding the mean of sources that are not included in S , assuming that sources that are not included in S gives its mean to all individuals. An individual is presented as i , and an income source as j , then x_i^j is j^{th} income source of an individual i . Now we can present a income distribution of j^{th} income source as a n dimension vector $x^j = \{x_1^j, \dots, x_n^j\}$. When a subset of all income sources is $S \subset J$, there is total the number of 2^k of subset S that can be made by income sources with the number of k . J is a universal set of income sources $J = \{1, \dots, j, \dots k\}$.

An income made from income sources in set S is a sum of each income sources in set S , and it's income distribution can be presented as a n dimension vector as follows. Then, with a principle of equalized shapley-value decomposition, by adding a mean of income sources that are not included in S to income sources in S , we get

$$y^e(S) = [\sum_{j \in S} x_1^j + \sum_{j \notin S} \mu(x^j), \dots, \sum_{j \in S} x_n^j + \sum_{j \notin S} \mu(x^j)], \quad (10)$$

with $\mu(x^j)$ as a mean of j^{th} income source.

An equalized shapley-value of j^{th} income source is

$$\text{Sh}_j^\epsilon(I, J, G) = \sum_{\substack{S \subset J \\ j \in S}} \frac{(s-1)!(k-s)!}{k!} [G(y^\epsilon(S)) - G(y^\epsilon(S - \{j\}))]. \quad (11)$$

Therefore, we can compute marginal contributions of each circumstance variable with sums with various combinations of circumstances and resulting inequality measures.

3. Decomposing the residual to type-specific effort and individual effort

We divide each circumstance variables in to groups, and then call the combination of the groups as types. Then we categorize every individual to these types. We will introduce standards of categorizing in Section 4. As mentioned above, we measure an effort by the residual of a regression. Likewise, we are trying to explore what part of inequality of income Y , which has a distribution of F_Y , is due to circumstance or effort. Type t is a combination of father's education as 3groups, gender, birth year, grown-up region and number of sibling as 2groups. Therefore there exist 48 distinctive types. The key idea is that an individual should not hold attributable for income difference according to type.

Denote each of the J circumstances as X_j , which have the value of K_j . Each type t consists of specific collection of $t \in T$. The set T consists of elements $X^t = \{X_1 = x_1^t, X_2 = x_2^t, X_3 = x_3^t, \dots, X_J = x_J^t\}$. The type of specific individual's income in the sample is Y_i^t . Then we measure the individual's effort by the deviation of Y_i^t from the expected income of an

individuals from type t , $E[Y|X^t]$. We use the equation (11).

$$Y_i^t = \mu + \sum_{j=1}^J X_{ji}^t \beta_j + \epsilon_i^t \quad (12)$$

However, there also exists type-specific part in residual. Therefore, we will decompose the residual of the regression to type-specific effort and individual effort.

4. Data

1. Samples and Sources

KLIPS (Korean Labor and Income Panel Study) is a Korean census which conducts once a year on city-dwelling 5000 households and all members of the households. It's first census started at 1998, and now it's in progress in 2014, 17th census. KLIPS has both household data and household member data. We used household member data. Our sample is household members who are born between 1960 and 1980, registered in KLIPS.

(1) Dependent Variables (Variables for Individual Achievement)

We used three dependent variables, individual education, individual labor income and individual health in order to measure individual achievements. First, we used the responds about individual's final school and whether a individual had completed the education to construct a individual education indicator. We combined an individual's final school data and

respond about whether an individual had completed the education to compute an individual's total years of education and measure this as an individual education. A range of the individual education indicator is 0 year (no education) to 23 years (completed PhD). Second, in order to construct individual income indicator, we used 13th census (2010) to 17th census (2014) to get individual pre-tax labor income data from 2009 to 2013. Then, with CPI of each year, we converted the data to pre-tax real labor income data (with CPI of 2013).

Second, we used an average of 5 years of real labor income as an individual real labor income indicator.

Finally, we as an indicator of individual's health, we used the body mass index (BMI).¹ We used the respond about individual's weight and height to construct a BMI indicator. The two variables are an average of 3 years of data.

(2) Explanatory Variable

We used five circumstance variables: father's education, gender, birth year, grown-up region, and number of siblings.

We constructed a father's education indicator in the same way as an individual education indicator from above. We combined the respond about father's last school and whether he completed an education, and computed total years of father's education. A range of the father's education indicator is 0 year (no education) to 18 years (completed bachelor's degree). In case of a gender indicator, 1 denotes male, 0 denotes female. As a measure of a birth year indicator, we count an individual's age in full.

¹ We define BMI as follows: $BMI = \frac{\text{mass}_{\text{kg}}}{\text{height}_{\text{m}}^2}$

We constructed a grown-up region indicator by responds about a region an individual grew up until 14 years old, which 0 denotes cities up to metropolitan cities, and 1 denotes otherwise. However, as city-dwellers are the census subjects of KLIPS, it is insufficient to interpret effect of a grown-up region as a difference of accessibility of infrastructures between rural and urban areas.

In case of a number of sibling indicator, 1 denotes an individual who has siblings more than a median, 0 denotes an individual who has siblings less than a median. However, we found out that birth rate is keep decreasing rapidly. Therefore, we used median of 1960s as a standard for individuals born in 1960s and the same for 1970s.

For the extended work to see the effect of the circumstances on individual's health, we added the dummy variable that indicates whether the individual works out regularly or not as an explanatory variable. We used an average of 3 years of data.

(3) Type classification

We classified each circumstance variables to groups, and call combinations of groups as types. Then, classify every individual by the types. The table below provides a standard for a type classification.

[Table 1 about here]

Combinations of the groups make 48 types, and descriptive statistics for the 48 types are given in Table 2.

[Table 2 about here]

2. Descriptive Statistics

In Figure 1, we show the cumulative distribution of income with a groups of each circumstance variables. The result shows that a difference between groups is the most distinctive in father's education and gender, as those are the circumstance variables that affect income inequality the most.

[Figure 1 about here]

In Table3, we show the correlation of each circumstance variables which is considered to affect the most on individual's education with individual's education years. The result shows that both father's education and gender has fairly strong positive relationship with individual's education. Among individuals who have more than 13 years of education, only 22% of them were those whose father has 0 to 6 years of education. On the other hand, those whose father has over 13 years of education were 78% of total. Also, among female, only 33% of them had more than 13 years of education, but male were 47%. On the other hand, 12% of female had 0 to 9 years of education, but in case of male, it was less than 9%.

[Table 3 about here]

Table 4 shows the regression result of each dependent variables that indicates individual's achievement on circumstance variables.

[Table 4 about here]

According to the result, all circumstance variables, father's education, gender, grown-up region, the number of siblings and the birth year have statistically significant effect on an individual's education. The higher a father's education is, the higher an individual's education. Male have higher education than female. Individuals who grew up in metropolitan cities have higher level of education than those who are not. The more siblings an individual have, the lower an individual's education. Finally, when people gets older, their education level gets lower.

According to the regression result of income, except grown-up region and number of siblings, circumstance variables have statistically significant effect on an individual's income. An individual has higher income if the individual's father has higher level of education and if the individual is a male. As we explained in Section 3, education in the regression of income is used as an explanatory variable to omit the effort accountable part of birth year variable which has both effort and circumstance accountable parts.

According to the result, BMI is higher if the individual is male and old. Also, the quadratic term of age indicates that among the samples of age 31 to 51, BMI increases until the age of 46.5 and decreases after.

5. Result

In section 4, we identified basic information and relationship with an individual achievement by descriptive statistics of circumstance variables. Now, we turn to the decomposition of inequality of achievement into the part due to circumstances and the part due to an effort based on the model we outlined in Section 3. Especially, we explore how much each circumstance variable accounts the inequality. First, we focus on decomposition of education inequality, then that of income inequality and lastly that of health inequality.

1. Decomposition of education inequality

The estimated Gini for education inequality is 0.096, and the estimated CV is 0.188, which is fairly low level of inequality compared to income inequality. We decomposed this individual inequality into father's education, gender, grown-up region, the number of siblings, birth year and residual as an effort.

[Table 5 about here]

According to table 5, 47% of education inequality is accounted for by circumstances and 53% is accounted for by an effort. To be specific, father's education shows the biggest contribution, 31%, gender as 7%, birth year as 4% and the number of siblings as 3%. Grown-up region accounts 2% of education inequality. A birth year indicator which is treated as a circumstance variable in education inequality reflects the difference in level of accessibility in education as economic development is in progress in Korea.

Among estimated CV of education inequality, circumstance accounts 42% and effort

accounts 58%, which indicates that inequality of circumstance is smaller in the inequality estimated by CV. Further, among circumstance variables, father's education and gender and the number of siblings and birth year and grown-up region are important contributors to inequality, accounting for 29%, 7%, 3%, 3%, and 2 percent of education inequality.

2. Decomposition of Income Inequality

Prior to discussing the decomposition of income inequality, we found out that 95% of individuals who have 0 incomes in total sample is female. We can assume that considerable amounts of respondents are housewives from the fact that KLIPS is the census that is based on a massive amount of survey questions from each individual. However, in Korea, as most housewives became housewives by their own choices, it might be meaningless to analyze what factors affect the most to housewives' 0 incomes. Of course, we can't ignore some cases that they were trying to get a job but due to circumstance factors, when their opportunity costs exceed income, they give up their job and become housewives. However, most cases are not. On this account, if we include the samples which have 0 incomes, among circumstance variables, effect of gender might be overestimated. Therefore, we analyzed by omitting the individuals with 0 income in order to avoid overestimation.

Table 6 displays the result when we decompose income inequality of individuals who have income bigger than 0. The estimated Gini for income inequality is 0.348. Among that, circumstances shows 52% and effort shows 48% of contribution. Gender accounts the largest proportion of inequality with 40%, then, father's education with 7%. Grown-up region and birth year barely have an effect, which was less than 1%. Finally, the number of sibling had no effect at all.

The income inequality measured by CV was 0.682, which was considerably high number. Circumstances attributed 47% and effort attributed 53%. Each circumstance variable's attribution was, gender with 38%, father's education with 8%, and rest were insignificant. The number of sibling also had no effect at all.

[Table 6 about here]

Although our main analysis is about sample of individuals who have income bigger than 0 in order to avoid overestimation, additionally, we analyzed the inequality of total sample, and a sample with male and income bigger than 0. The following result is on Table 7.

In case of total sample, the estimated Gini of income equality is 0.35, which is higher than that of our main analysis. Now circumstances account 58% of income inequality and an effort accounts 42%, which is fairly high number of contribution of circumstances, comparing that of our main analysis was 52%. Also, gender attributes 49% of total income inequality, which is 23% higher attribution than the result from our main analysis.

In case of the sample with male and income bigger than 0, the estimated Gini of income inequality is 0.27. Now, gender is excluded from circumstance variables, and father's education shows the largest attribution, which is 15%. Other circumstance variable shows almost no effect, which is less than 1%.

[Table 7 about here]

3. Decomposition of health inequality

According to the decomposition of health inequality, the estimated Gini for health inequality is 0.063, and the estimated CV is 0.113.

[Table 8 about here]

According to table 8, gender shows the biggest contribution, 29%, age as 6%, father's education as 1% and other variables accounts health inequality less than 1%.

6. Conclusion

In this paper, we analyzed the parts attributable of circumstances and parts attributable of effort from inequality of achievement. We used the sample of individuals who were born between 1960 and 1980 from KLIPS data, and measure an individual's achievement by education years and 5 years (2009-2013) of labor income of individuals. Among measurable circumstance variables, we constructed father's education, gender, birth year, grown-up region, number of siblings as circumstance variables, which enable us to observe not only the contribution of circumstance, but also the contribution of each circumstance variables for inequality.

First, the level of inequality of education inequality is lower than that of income inequality. However, 47% of education inequality was accountable with circumstances. Especially, father's education showed biggest contribution with 31%, and also gender with 7%. Second, in income inequality, despite the omission of sample with 0 income,

circumstance accounted 52% of inequality. Especially, gender attributed the most with 40%, and then father's education with 11%. Gender's 40% attribution is considerably large number, considering that 95% of the omitted sample was female. Lastly, in health inequality, the degree of inequality is low comparing to education inequality and income inequality. Also, the effect of circumstance in inequality is relatively low. Gender showed the most contribution with 23%, age with 13%, and father's education 1%. Therefore, we can conclude that father's education is the most important contributor to education inequality, and gender is the most important contributor to income inequality and health inequality.

One of many limitations in our analysis is that there are numerous factors in circumstances besides our circumstance variables. Particularly, IQ, appearance, height, etc are significant factors that affect our achievement but beyond our control. Therefore, our result considerably underestimates the contribution of circumstances.

References

Ko, J. and W. Lee (2011), "Father's education and son's achievement", *Jaejunghak Yeongu (Public Finance Study)* 4(2), pp.47~87. (written in Korean)

Kim, W and W. Lee, *A Study on the Extent to Which the Korean Tax Regime Equalizes Opportunities for Income Acquisition among Citizens*, Korean Institute of Public Finance Research Report #08-03, 2008. (written in Korean)

Kim, J, H. Seo, and W. Shin (2016), "An analysis on the contribution of the inequality of opportunity to income inequality : Evidence from Korean in 2000s", *Ordo Economics Journal* 19(1) pp.41~67. (written in Korean)

Lee, S and W. Lee (2016), "Analysis of Contribution of Inequality of Income and Assest Inequality Using Shapley-value", *Panel for Korean Economic Analysis, Scheduled*. (written in Korean)

Björklund, A., M. Jäntti, and J.E. Roemer (2012). "Equality of opportunity and the distribution of long-run income in Sweden," *Social Choice and Welfare* 39 (2-3), pp. 675–696.

Bourguignon, F., F.H.G. Ferreira, and M. Menéndez (2003). "Inequality of outcomes and inequality of opportunities in Brazil", World Bank Policy Research Working Paper No. 3174.

Checchi, D. and V. Peragine (2010). "Inequality of opportunity in Italy", *Journal of Economic Inequality* 8 (4), pp. 429–450.

Eriksson, K.H, M. Jäntti, L. Lindahl (2015). "Gender and inequality of opportunity in Sweden", mimeo

Ferreira, F. H., and J. Gignoux (2011). "The measurement of inequality of opportunity: Theory and an application to Latin America", *Review of Income and Wealth*, 57(4), pp. 622-

657.

Gunatilaka, R., and D. Chotikapanich (2009). "Accounting for Sri Lanka's expenditure inequality 1980-2002: Regression based decomposition approaches", *Review of Income and Wealth*, 55(4), pp. 882-906.

Lee, W. (2015). "A Field Guide to Estimating Income Distributions", unpublished lecture note.

Nilsson, W. (2005). "Equality of opportunity, heterogeneity and poverty", PhD thesis. Umeå University.

Peichl, A. and M. Ungerer (2015). "Accounting for the spouse when measuring inequality of opportunity", *ZEW Discussion Papers* No.15 (034). Centre for European Economic Research.

Table 1: Summary of circumstance variables used

Circumstance Variables	Group	Classification Criteria	Obs.
Father's Education	Group 1	0-6 years	1689
	Group 2	7-12 years	1464
	Group 3	Over 13 years	281
Gender	Group 1	Female	1824
	Group 2	Male	1610
Birth Year	Group 1	1960-1970	1903
	Group 2	1971-1980	1531
Grown-Up Region	Group 1	Other cities	2117
	Group 2	Metropolitan cities	1317
Number of Siblings	Group 1	Less than Median	1352
	Group 2	More than Median	2081

Source: Korea Labor and Income Panel Study (Korea)

Table 2: Summary statistics, by type

	Edu (Average)	Income (Average)	Father's edu (Average)	Age	Obs.
Type 1	11.0	878.0	3.9	46.0	53
Type 2	12.4	1274.2	5.1	35.3	73
Type 3	12.1	1171.5	4.8	44.3	32
Type 4	12.1	1168.1	5.0	35.4	46
Type 5	11.0	1199.5	3.8	46.7	445
Type 6	12.3	891.7	4.4	37.4	129
Type 7	11.6	1098.3	4.0	46.2	101
Type 8	12.5	935.8	5.0	36.8	28
Type 9	12.9	4515.5	4.4	45.1	67
Type 10	12.7	3189.1	4.8	35.9	80
Type 11	12.4	2714.8	5.0	44.3	41
Type 12	13.1	3232.3	5.1	36.3	52
Type 13	12.0	3950.6	3.5	46.8	351
Type 14	12.6	3377.9	4.1	37.5	84
Type 15	12.2	3623.8	3.9	46.7	82
Type 16	12.1	2937.5	4.3	37.5	25
Type 17	13.1	1221.6	10.7	43.9	37
Type 18	13.1	1279.6	10.6	35.1	130
Type 19	13.4	1461.0	10.4	44.2	45
Type 20	13.4	1326.9	10.7	35.0	166
Type 21	12.3	1129.9	10.2	46.3	131
Type 22	12.9	1460.8	10.2	36.3	96
Type 23	12.9	1633.7	10.4	45.5	86
Type 24	12.9	1484.7	10.5	36.0	81
Type 25	13.3	3758.3	10.4	44.2	51
Type 26	13.4	3382.0	10.5	34.9	134
Type 27	14.5	4515.1	10.5	44.4	75
Type 28	13.4	3695.9	10.8	35.3	177
Type 29	13.3	3822.2	10.3	45.9	109
Type 30	13.8	3890.9	10.1	36.4	43
Type 31	13.8	4408.5	10.3	46.3	65
Type 32	13.1	3310.4	10.2	36.3	38
Type 33	13.5	1166.3	15.3	43.5	8
Type 34	13.0	1186.8	15.4	34.3	13
Type 35	14.9	1781.2	15.9	44.8	20
Type 36	13.9	1591.5	15.7	34.5	48
Type 37	13.7	1832.4	14.6	47.1	21
Type 38	14.4	1317.2	14.0	38.4	5
Type 39	14.9	1720.4	15.1	44.4	19
Type 40	15.4	1077.7	16.2	37.7	11
Type 41	14.7	4552.4	15.7	44.1	13
Type 42	14.3	4540.2	15.3	34.3	16
Type 43	15.6	4234.8	15.4	44.2	17
Type 44	14.2	3958.3	15.8	35.2	41
Type 45	15.2	5692.3	15.7	45.5	18
Type 46	12.7	4159.5	15.4	36.8	10
Type 47	14.8	5700.7	15.3	45.7	16
Type 48	14.4	6063.5	15.2	38.4	5

Source: Autor's estimation, using Korea Labor and Income Panel Study (Korea)

Table 3: Summary of individual's education, by circumstance variables

	Individual's Education			
Father's Edu	0-9 years	10-12 years	Over 13 years	Total
0-6 years				
7-12 years	0.166	0.610	0.224	1.000
Over 13 years	0.042	0.436	0.522	1.000
Total	0.011	0.213	0.776	1.000
Gender	0-9 years	10-12 years	Over 13 years	Total
Female	0.120	0.551	0.329	1.000
Male	0.077	0.449	0.474	1.000
Total	0.100	0.503	0.397	1.000

Source: Korea Labor and Income Panel Study (Korea)

Table 4: Regression results

	Education	Labor Income	BMI
Fathers Education	0.1834*** (19.65)	23.22* (2.37)	-0.0191 (-1.32)
Male	0.6468*** (8.67)	1918.1*** (25.51)	1.8518*** (15.59)
Grown-up region	0.2157** (8.67)	1.0844 (0.01)	0.0926 (0.74)
Number of Siblings	-0.0592* (-2.27)	32.32 (6.27)	-0.0596 (-0.48)
Age	-0.0237** (-3.27)	45.39*** (6.27)	0.4000*** (2.63)
Age ²			-0.0043** (-2.37)
Education		269.45*** (15.63)	
Workout			-0.4834 (-0.82)
Constant	12.09*** (39.29)	-3737.1** (-10.01)	13.512*** (4.25)
Obs	3427	2851	1627

Source: Autor's estimation, using Korea Labor and Income Panel Study (Korea)

Table 5: Equalized Shapley-value decomposition result of education inequality

	GINI		CV	
	AC	RC	AC	RC
Fedu	0.030	0.309	0.052	0.279
Male	0.007	0.071	0.012	0.066
Growreg	0.002	0.024	0.004	0.021
Nsib	0.003	0.029	0.005	0.024
Age	0.004	0.039	0.006	0.032
Resid	0.051	0.529	0.108	0.578
Total	0.096	1.000	0.188	1.000

Source: Autor's estimation, using Korea Labor and Income Panel Study (Korea)

Note :

1. AC stands for absolute contribution, and RC stands for relative contribution.

**Table 6: Equalized Shapley-value decomposition result of income inequality
(when income>0)**

	GINI		CV	
	AC	RC	AC	RC
Fedu	0.037	0.107	0.052	0.076
Male	0.138	0.396	0.262	0.384
Growreg	0.002	0.007	0.004	0.005
Nsib	0.000	0.000	0.000	0.000
Age	0.003	0.008	0.004	0.006
Resid	0.168	0.482	0.360	0.528
Total	0.348	1.000	0.682	1.000

Source: Autor's estimation, using Korea Labor and Income Panel Study (Korea)

Table 7: Equalized Shapley-value decomposition result of income inequality

	Overall Sample				Male, Income is bigger than 0			
	GINI		CV		GINI		CV	
	AC	RC	AC	RC	AC	RC	AC	RC
Fedu	0.036	0.083	0.048	0.057	0.040	0.150	0.071	0.133
Male	0.212	0.486	0.407	0.484				
Growreg	0.001	0.001	0.001	0.001	0.001	0.005	0.002	0.004
Nsib	0.001	0.001	0.001	0.001	0.002	0.006	0.003	0.005
Age	0.003	0.007	0.004	0.005	0.001	0.004	0.001	0.003
Resid	0.183	0.421	0.38	0.452	0.224	0.836	0.457	0.855
Total	0.435	1.000	0.841	1.000	0.268	1.000	0.535	1.000

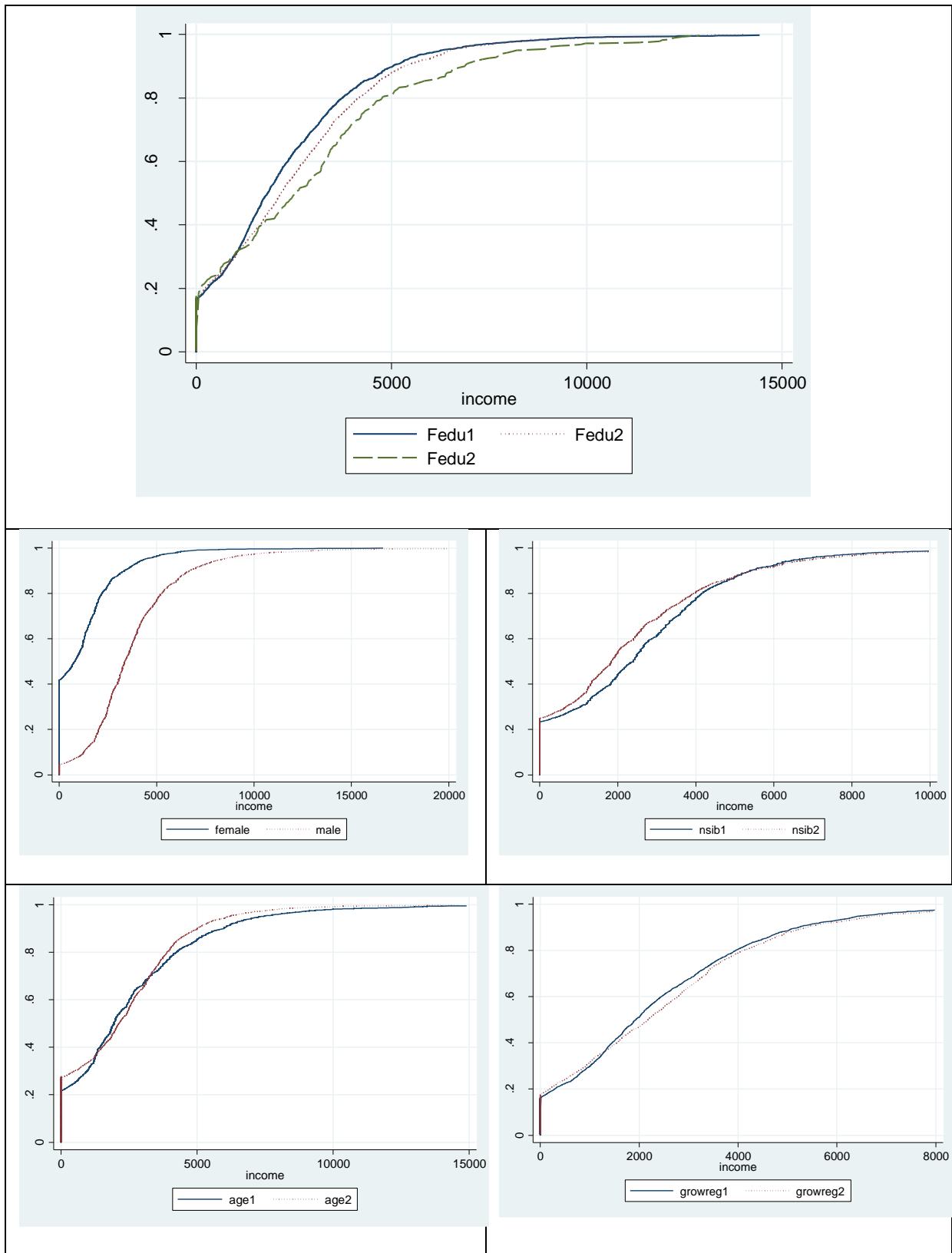
Source: Autor's estimation, using Korea Labor and Income Panel Study (Korea)

Table 8: Equalized Shapley-value decomposition result of health inequality

	GINI		CV	
	AC	RC	AC	RC
Fedu	0.001	0.013	0.001	0.010
Male	0.014	0.221	0.026	0.228
Growreg	0	0.005	0.001	0.004
Nsib	0	0.003	0	0.003
Age	0.007	0.141	0.015	0.133
Workout	0.002	0.001	0	0.003
Resid	0.039	0.615	0.07	0.618
Total	0.063	1	0.113	1

Source: Autor's estimation, using Korea Labor and Income Panel Study (Korea)

Figure 1: Income distribution (CDF) by circumstance variables



Source: Autor's estimation, using Korea Labor and Income Panel Study (Korea)