Measurement of Income Transition and Test of Income Mobility

(Preliminary version)

August, 2013

MYUNG JAE SUNG[*]

School of Economics, Hongik University, Seoul, Korea

ABSTRACT: This paper proposes a new measure of income mobility under the continuous transition, which can be summarized into a single parameter. This measure is also easily testable for its structural changes over time. It is independent of choice of intervals or scale of income, unlike previously developed measures. The income mobility is measured by the variance of the income transition variable defined on a continuous distribution stemming from the income transition rule. It is defined based on the changes in relative income induced by the changes in the household income ranks between two periods (years). The underlying assumptions are twofold: firstly, the probability distribution of household income remains in the same distributional family over time and secondly the expected income rank in the following period equals the realized rank in the previous period within each age group. The proposed measure of income mobility is invariant to the selection of partitions of income intervals adopted in the discrete type income mobility measures using transition matrix as proposed by Shorrocks (1978) and subsequent studies. The continuous measure from Fields & Ok (1996, 1999) is interval-invariant. However, it is scale-dependent, unlike the Shorrocks' type measures. To the contrary, the proposed measure is also scale invariant. Another advantage is the simplicity of calculation and hypothesis testing of structural changes. An additional parametric assumption, such as the log-normality of household income distribution, generally eases the complexity of the measure. It suffices to estimate the variance of income transition variable to infer all characteristics of income mobility since all necessary distributional information is summarized into a unique parameter, the variance of income transition variable, under the log-normality. Thus, its dimension of parameter space of the income mobility shrinks to unity. However, a drawback is misspecification error when the log-normality breaks down. The proposed measure is applied to the case of Korea using a panel data set from the Korea Labor and Income Panel Study (KLIPS). The income mobility expressed in the form of variance was estimated to be 0.183 in 1999, which decreased to 0.106 in 2008. Statistical hypothesis testing shows that this change is statistically significant: the variance estimates are differenced 9 times between the two consecutive years between 1999 and 2008, and the differences turn out to be statistically significant in 7 out of 9 times, implying that they declined significantly over the last decade. The income transition rule can be applied to estimating the poverty inflow/outflow probabilities (PIP/POP), depending on the sizes of households and their income ranks in the previous period. The PIP/POP are analyzed mathematically and their estimates are illustrated graphically. The POP is positively correlated with income mobility. This strongly implies that the POP has decreased significantly over the last decade in Korea. There exist many factors which may reduce income mobility and the POP: key candidates are (rapid) population aging, decreasing labor market flexibility, skill-biased technical development, regulations and so forth. These need to be scrutinized in the future to understand the recent decreasing trend of income mobility/POP.

Key Words: income mobility, lognormal distribution, panel analysis, testing, poverty
JEL Codes: D31, I31

Measurement of Income Transition and Test of Income Mobility

1. Introduction

Life-time income inequality is generally different from snapshot income inequality (e.g., annual income inequality) mainly due to income mobility between periods. A conventional way of measuring income mobility is to estimate transition matrices measuring discrete income transitions with the partition of income domain usually consisting of a finite number of (sub)sets, as proposed by Shorrocks (1978) and subsequently used by Gottschalk (1997), Geweke et al. (1986), Bager-Sjögren & Jenkins (1998), Jarvis & Jenkins (1998), Trede (1998), Gardiner & Hills (1999), Cho & Kim (2007), Auten & Gee (2009), Chen (2009), Tsiu (2009), Shi & Xin (2010), amongst others. These measures can powerfully describe how income changes among individuals and households and trace its longitudinal changes. Unfortunately, they are not invariant to the choice of partitions of income intervals. An interval-invariant measure was developed by Fields & Ok (1996, 1999) to measure absolute mobility. They developed a continuous income mobility measure by summing up the absolute differences of logged incomes of individuals and households between two periods, which was not dependent on the choice of income intervals. The Shorrocks' type discrete measures do not depend on the scale of income or income growth rates between periods. However, the Fields & Ok's continuous type measure is not scale-invariant, since it usually measures the sum of absolute differences (usually logged differences). In both of the above cases, it is not easy to test longitudinal structural changes in income mobility.

This paper proposes a new measure of income mobility under the continuous transition, which can be summarized into a single parameter. The income mobility is measured by the variance of the income transition variable defined on a continuous distribution stemming from the income transition rule, which will be defined and introduced in Section 2.2.2. This is defined based on the changes in relative income induced by the changes in the household income ranks between two periods (years). There are two underlying assumptions: the probability distribution of household income remains in the same distributional family over time and the expected income rank in the following period equals the realized rank in the previous period within each age group, that is the assumptions are the log-normality of household income distribution and the persistence of log-normality over time. An additional parametric assumption such as the log-normality of household income distribution generally eases the complexity of the measure. However, a drawback to this is the potential for misspecification error when the log-normality breaks down. The proposed measure is both interval- and scale-invariant. Furthermore, it is simple to calculate and easy to test a longitudinal structural change in income mobility because, by estimating the variance of the income transition variable, one can infer all characteristics of income mobility since all necessary distributional information is summarized into that single unique parameter under the log-normality. Thus, its dimension of parameter space of the income mobility shrinks to unity, whereas in the Shorrocks' type measures it is as large as the product of numbers of income intervals of the two periods. The proposed method is thus simpler and more powerful to apply since one dimensional

parameter can summarize whole distributional information. In addition to proposing an income mobility measure, this paper also applies this by analyzing recent change in income mobility in Korea using a panel data set from the Korea Labor and Income Panel Study (KLIPS).

Income mobility usually stems from differences in households characteristics as well as sometimes from uncertainty. Household characteristics include age, gender, education level, occupation, health, and many more. These factors are usually persistent over time. For example, a person with a higher education level or someone in good health tends to earn more than those who are less educated or unhealthy. Alongside this, climate conditions, unexpected changes in regulations, or other unexpected and unpredictable occasions caused by uncertainty can affect income changes or mobility.

The income mobility adopted in this paper measures changes in relative income ranks or changes in real income between periods.[1] It is natural that the relative income rank within a certain group changes over time. If it does not, income inequality measured at a certain period is almost the same as that measured for longer periods.

A part of income inequality observed in a certain period consists of life-cycle effects, that is, inequality caused by age differences between generations. Age differences among generations are in fact one of the major sources of income inequality which is observed every period because each generation is at a different point in the life-cycle. The income mobility of young generations is generally higher than that of older generations, implying that population aging may cause a decline in income mobility across the economy as a whole.

Income inequality and income mobility are closely related. Even though cross-sectional information about income inequality is useful when considering short-term inequality, it cannot generally allow for inferences about life-time income inequality, because it does not provide necessary information as to income mobility between periods. To infer life-time income inequality correctly, it is necessary to combine snapshot information about short-term income inequality with the income mobility between periods. Income mobility is therefore another measure of income inequality to the extent that life-time income inequality is involved. Fields (2010) similarly discussed the issue of the relation between income mobility and longer-term income inequality by developing an equalization measure (see also Shi & Xin (2010)).

This can be reviewed in the following example of two contrasting cases. Suppose that, in two imaginary economies, the annual income inequalities are similar to each other and that both inequalities remain constant every year at some value. In one economy, income would not change over time among individuals in terms of its relative rank (low income mobility), while, in the other economy, income ranks change very frequently among individuals over time. In the former case, income ranks hardly change and thus the life-time income rank of an individual would remain almost unchanged in the near neighbor

---

[1] This does not mean the relative ratios of percentage changes in nominal incomes between two periods within each household. There exists no change in the mobility of real income if income increases at a certain rate for all households. Even in that case, the mobility will seem to have nonzero values, if the income mobility is measured in terms of growth rates of nominal incomes over time; however, this does not necessary imply any change in the income ranking of a household over time.

observed in the initial period. In this case, life-time income inequality tends to be very similar to that of annual income. Furthermore, the poverty problem would become more persistent: a poor individual or household would tend to remain in poverty without prospect of escape. However, in the latter case with high income mobility, even though there exist huge differences in income levels among individuals in every period, their life-time incomes would be convergent to their population mean level. Higher income mobility generally implies less correlation, or sometimes negative correlation, of income between periods. An extreme case is the random realizations of income, wherein the realization of income ranks and levels depends purely on luck. In this case, a time-series of income for an individual or a household tends to become more independently distributed. If the income is perfectly mobile and if the correlation of income between periods is zero, then the average annual incomes of individuals or households converge to the population average annual income value. Furthermore, the variance of life-long income converges to zero as the time span increases to infinity. Thus, life-time income inequality shrinks over time. The poverty problem lessens because there is increased chance of an individual or household escaping poverty over time due to the high mobility of income. The smaller the income mobility becomes, the severer the socio-economic problems induced by the rigidity of income flows among individuals/households becomes in the long run, and vice versa. It is almost certain that the real economy lies in-between.

There are two main purposes to this paper: firstly to propose a new estimator of income mobility under continuous income transition between periods, which is both scale and interval invariant, and simpler than its precedent measures and which is also easily testable. Secondly, the aim is to estimate income mobility over the last decade and to test its structural changes in Korea. Following studies may wish to address issues such as the determinants of income mobility and its life-cycle effects. This paper estimates income mobility between annual income for the last decade in Korea using the panel data sets of KLIPS, which were compiled for the period between 1998 and 2008, and tests the hypothesis of structural changes in income mobility.

In Section 2, income mobility is first reviewed and is then defined based on the main findings of previous studies in which the statistical property of household income distribution was investigated in Korea, that is the log-normality. Based on this, the income transition rule between two consecutive years is derived and the income mobility is defined as the variance of the income transition variable. The income transition between two periods is usually defined in most empirical studies as discrete with transition matrices. In this paper, however, the income transition is not defined as discrete but continuous. In Section 3, the income mobility is estimated in the form of the variance of income transition variable and its structural changes are tested. Furthermore, the probabilities of poverty exit/entry (or, equivalently, outflow/inflow) are simulated based both on the log-normality of household income and on the estimates of income mobility. The effects of changes in income mobility on this are also discussed with simulation results. Section 4 concludes the paper and offers suggestions for future research.

2. Income Distribution and Income Mobility

In this section, the statistical properties (that is, log-normality) of household income in Korea are discussed. The log-normality of household income plays a critical role in defining income mobility. Continuous type income mobility is then introduced and defined.

2.1. The Statistical Properties of Household Income: Literature Survey

Household income is earned primarily by supplying labor and capital in the market. Labor/capital supply decisions are often affected by the interactions between individuals or household members through virtual income, cooperation or competition, and sometimes by government intervention. The income flow of a household sometimes reveals regularity in its trend over time. The statistical characteristics of household income flow are diversified by types of households or by types of incomes earned by each household member in terms of the size of the household, the age, gender and occupation of its individual members, industry, and so forth. It has long been debated whether (household) income follows a specific statistical probability distribution such as the log-normal distribution, which has long been used without rigorous tests that can demonstrate its validity. Recently, multiple studies have found that, despite these problems, the logarithm of household income as a whole has a unique distributional pattern of normality in Korea. Kim & Sung (2003were the first to show that household annual income in Korea is distributed log-normally. They constructed a null hypothesis of log-normality and empirically tested the normality of the logarithm of household income with the Jarque-Bera normality test[2], using the Household Income and Expenditure Survey (HIES) dataset for 1982-2002 released by Statistics Korea to achieve this. They found that the household annual incomes from HIES followed log-normal distributions excluding for the years 1997 and 1998, during which the Korean economy was in a crisis that coincided with huge increases in the foreign exchange rate and unemployment rate, sharp drops in income and imports, and so forth. During these years, the income distribution deviated significantly from a log-normal distribution. They concluded that household annual income generally follows a log-normal distribution, and that the deviation from a normal distribution during the two years was temporarily induced by the economic crisis. They also tested the log-normality of annual incomes of 41 age subgroups from age 25 to 65 between 1982 and 2002. They found that the null hypothesis of log-normality was rejected for 44 age subgroups and accepted in 817 subgroups out of a total of 861[3] at the 5% significance level. Over the nineteen year period studied, excluding 1997 and 1998 for which the log-normality was rejected for the whole sample

---

[2] Besides this, there exist several other statistical methods of testing the (log-)normality. For samples of size three or less, the W-statistic from Shapiro-Wilk (1965) is used; the statistical probability distribution of W has been verified up to the size of three. For the samples of size three to six, the test statistic proposed comes from Royston (1982), which converts the test statistic which is equivalent to the W-statistic of size three. For the samples larger than 2,000, the test statistic $Z_n = ((1-W_n)\gamma - \mu)/\sigma$ is used where Z_n is a standard normal variable, and γ, μ, and σ are functions of n by simulations, also proposed by Royston (1982). If Z_n is large enough, the null hypothesis of a normal distribution is rejected. For the samples of size 2,000 or larger, the Kolmogorov D statistic is used (Stephens 1974). This paper adopts the Jarque-Bera test proposed by Jarque & Bera (1980), which is the linear combination of squared skewness and kurtosis. It is known that the power of test is relatively weak. However, this method is the most widely used in economic studies.

[3] The number of age subgroups comprises 41 per year. There exists therefore 861 subgroups between 1982 and 2002.

due to the economic crisis, there were 779 age subgroups and the null hypothesis of log-normality was rejected in only 13 of these at the 5% significance level. This amounts to only 1.7% of all subgroups, which is far smaller than the significance level.

Sung (2005) tested the log-normality of household annual income using a different survey data set from KLIPS and concluded that the Chi square test statistic is large and that the null is rejected. However, the empirical distribution function was estimated nonparametrically using a kernel method, which showed that the discrepancies between the densities of a normal distribution and the nonparametrically estimated densities of log-income are very small, even though its normality is rejected when using the Jarque-Bera method. The discrepancies are slightly notable only in the narrow region around the income level of eight to ten million Korean Won (KRW)[4] which represents the absolute poverty lines, that is the minimum subsistence levels for a four-person household, as set by the government in the early 2000s, and negligible elsewhere. The estimated densities were slightly but still observably higher than the normal densities below the poverty lines and slightly but obviously lower right above the poverty lines. The study concluded that, even though the log-normality was rejected using formal statistical tests, the benefits from assuming the log-normality of household income would be much larger than the costs incurred by the biases and, therefore, that it would not be at all problematic to assume log-normality.[5] Sung (2008) and Weon & Sung (2007) confirmed this conclusion.

2.2. Income Mobility

2.2.1. Theoretical Introduction

Income mobility is often measured by changes in relative income ratios or ranks between periods. Measures of income mobility are classified as either discrete or continuous. Discrete type measures generally count relative frequencies or empirical probabilities of income shifts between income groups. These are often expressed as transition matrices as shown in Shorrocks (1976, 1978). Continuous type measures of income mobility, by contrast, usually measure the correlations of incomes or changes in the relative income ratios/ranks using a continuous mapping between periods.

Fields & Ok (1999) developed two measures of income movement, which were defined as any function $m_n : R_{++}^{2n} \to R_+$ that is continuous where R_+^n is the space of income distributions with population size $n \geq 1$ (see also Mitra & Ok (1998)). Maasoumi & Trede (2001) compared income mobility between Germany and the United States using the relative ratios of short-run and long-run inequality indexes.

However, due to analytic difficulties, the majority of empirical studies examine income mobility with a discrete type measure using income transition matrices of relative frequencies between periods among income deciles/quintiles or certain partitions of sample; examples include Jarvis & Stephen (1998),

---

[4] US $1 was roughly equal to 1,200 KRW in the early 2000s.
[5] These are Sung (2008) and Weon & Sung (2007).

Jenkins (2000), Auten & Gee (2009), Chen (2009), Tsui (2009), Shi & Xin (2010) and many more.

Many studies have also been devoted to applying income mobility. Van Kerm (2004) studied the decomposition of income mobility into two sources: the mobility induced by a change in the shape of income distribution and the mobility induced by reordering of individuals. He found that reranking has been the major force behind income mobility. Lee & Solon (2009) measured intergenerational income mobility using regression analysis and found that the intergenerational mobility did not change much for the cohorts born between 1952 and 1975 using the Panel Study of Income Dynamics.

This paper develops a measure of income mobility defined as a continuous random variable which is named the 'income transition variable', which is derived from the 'income transition rule'. Both of these are defined below. Its structural changes during the last decade in Korea are also tested. The income mobility measure proposed here assumes two basic statistical properties: the log-normality of (household annual) income and its persistency over time. The persistency of log-normality means that the log-normality of household annual income is generally well preserved over time. The log-normality of income is not tested here; the findings of previous studies such as Kim & Sung (2003), Sung (2005, 2008), and Weon & Sung (2007) in which the log-normality of household income was established through empirical tests are accepted as valid.

In order to establish the income transition rule, it is assumed that the expected income rank of a household in the following year (or period) is identical to the income rank realized in the previous year. The underlying assumption is that the income earning power of a household persists over time. In other words, because the demographic characteristics of a household, including gender, age, education level, health, occupation, productivity and so on, are not easily changeable in the short term. Given this fact, the realization of income has in general a stylized pattern for most households. Therefore, it is highly probable that a rich household will remain in one of the high income deciles in the following year, whereas a poor household will tend to remain in one of the poor deciles in the following year. This is illustrated in more in detail in the following section.

2.2.2. Concept/Definition of Income Mobility

In this section, the concept of income mobility adopted in this paper is introduced and the income transition rule and variable are defined accordingly; as stated above the log-normality of household income distribution and its persistency are assumed, based on earlier research. Furthermore, whether innate or acquired, demographic characteristics are considered to be important determinants of income, which will play a crucial role in defining continuous income transition between periods.

The KLIPS data set is partitioned into 41 age groups from age 25 to 65; age group 25 consists of households whose heads are 25 years old or under and age group 65 consists of the households whose heads are 65 or above. Households within the same age group are further diversified in terms of their other characteristics such as education levels, work experience, occupations, industries, gender, health condition, and the location of their residences or workplaces. These are significant factors that affect the

determination of their market income as earned by supplying labor/capital in the market. For simplicity of discussion, let these factors be entitled 'income characteristics' in this paper. Certain assumptions can be made here, for example, an individual with higher education/experience tends to earn higher income on average, ceteris paribus.

While age and gender are neither changeable nor manageable, most other income characteristics are more or less changeable, but are not usually flexible in the short term. Their chances depend on an individual's endowments, endeavors, and desire or willingness to invest in human capital over a sustained period. Therefore, the relative income ranks within an age group would change with low income mobility especially in the short or medium run. At least in the short run, relative income ranks within an age group would tend to persist and, so, their income expectations would not change much. That is, the income of an individual whose income was realized at the p-th percentile income within age group (A-1) in the previous year (t-1) would be realized with high probability near the p-th percentile income within age group A at t.[6] More formally, this can be specified so that the expectation of income of an individual (or household) whose income in the previous year was realized at the p-th percentile within an age group (for instance, age group (A-1) at (t-1)) is equal to the p-th percentile income within the same age group (age group A at t) in the following year.

Each age group has different mean and variance of income. This implies that the income ranks of the specific amount of income differ from age group to age group with probability one.[7] Of course, income ranks within an age group are generally different from those in the whole sample with probability one. Let $Y_{tA} \in R^1$ be the random variable representing the income of a household within age group A at time t. Define $X_{tA}^p \in R^1$ as the value of income, Y_{tA}, corresponding to the p-th percentile income within age group A at t for some $p \in [0,1]$. Let $X_{(t+1)(A+1)}^{p'}$ be the realized income corresponding to the p'-th percentile at (t+1) within age group (A+1) at (t+1) for some $p' \in [0,1]$. Also, let $X_{(t+1)(A+1)}^{*p}$ be the income value corresponding to the p-th percentile within the same age group (A+1) at (t+1). The transition of income between the two consecutive years within an age group can then be formulated for a scalar γ as follows:

$$X_{(t+1)(A+1)}^{p'} = (1 + \gamma_{(t+1)(A+1)}^{p}) \cdot X_{(t+1)(A+1)}^{*p} \quad \text{where } \gamma \in R^1, \gamma > -1. \tag{1}$$

Observe from equation (1) that the income determination process at (t+1) does not involve any information at t other than the income percentile realized at t. The variabilities of X and γ at (t+1) do not

---

[6] Note that age groups A at t and (A+1) at (t+1) are identical age groups for all t and A. Of course, it is quite possible that an age group becomes slightly different in its household composition over time due to the attrition/addition caused by demises, births, or social migrations. For simplicity of discussion, these factors are not considered here, because the differences are negligible between adjacent periods in general.

[7] It is possible to have identical income percentiles in more than two age groups. However, its probability is zero because income distribution is continuous.

depend on those at t. This implies that the degrees of dispersions of X or γ at t do not explicitly determine or affect those at (t+1).

In equation (1), X^* is not observable. However, it is not difficult to infer its value from the assumed or empirically accepted log-normal distribution of X with its estimated values of mean and variance for all $p \in [0,1]$. It is also nonparametrically estimable using sample data. $(1+\gamma)$ is the scalar multiple representing the relative ratio of X (realization) to X^* (conditional mean value). Taking logarithm on both sides of Equation (1) yields the following equation:

$$Z_{(t+1)(A+1)}^{\ p'} = \Gamma_{(t+1)(A+1)}^{\ p} + Z_{(t+1)(A+1)}^{*\ p} \tag{2}$$

where $Z = \ln\left(X_{(t+1)(A+1)}^{\ p'}\right)$, $Z^* = \ln\left(X_{(t+1)(A+1)}^{*\ p}\right)$ and $\Gamma = \ln\left(1+\gamma_{(t+1)(A+1)}^{\ p}\right)$.

Equation (2) applies for all $p \in [0,1]$. Note that Z, Z^* and Γ are random variables. In this paper, for simplicity of discussion, an assumption that Γ is independent of A and p is imposed. Of course, it is possible that Γ depends on either of these. However, even if this were true, the basic result would be qualitatively the same except for the more complicated calculation processes. Under independence, the subscripts and superscripts can be dropped from equation (2) without loss of generality.

$$Z = \Gamma + Z^* \tag{3}$$

Z is the logarithm of household income to be realized at (t+1). It is known to be normally distributed by the basic assumption of log-normality of household income. Z^* is the imaginary value of the logarithm of household income of the same income ranks within the same age group realized at time t. Therefore, Z^* is also distributed normal. Γ is the random variable which reshuffles the order of household incomes Z^*, resulting in Z. As assumed earlier, Z and Z^* are normal variables with identical means and variances. Since Z^* is distributed normal, the normality of Γ follows from the statement that Z is distributed normal. Conversely, based on the normality of Z, the normality of Γ follows from the normality of Z^*. Therefore, the normality of Γ is the necessary and sufficient condition for the normalities of Z and Z^*. Equation (3) is named 'the income transition rule.' Likewise, Γ is named as the 'income transition variable.'

Consider the following mean and variance of Z:

$$E(Z) = E(\Gamma) + E(Z^*) \tag{4}$$

$$Var(Z) = Var(\Gamma) + Var(Z^*) + 2Cov(\Gamma, Z^*) \tag{5}$$

Note that the probability distribution functions of Z and Z^* are identical. Therefore, $E(\Gamma)=0$ or, equivalently, $E(\gamma X^*)=0$. Especially, the latter equation implies that γ is orthogonal to X^*. However, this does not necessarily mean either $E(\gamma|X^*)=0$ nor $E(\gamma)=0$. Note, nonetheless, that the expectation of γ needs

to be nonzero. This is because equation (4) would not hold, if $E(\gamma)=0$. From this observation of $E(Z)=E(Z^*)$ and $Var(Z)=Var(Z^*)$, the following relations hold:

$$\mu_\Gamma = E(\Gamma) = 0 \tag{6}$$

$$\sigma_{\Gamma Z^*} = Cov(\Gamma, Z^*) = -\frac{\sigma_\Gamma^2}{2} \tag{7}$$

The covariance of Γ and Z^*, $Cov(\Gamma, Z^*)$, depends only on the size of variance of Γ. Since the variances of Z and Z^* are identical, the variance of Γ which is located on the right hand side of equation is nullified exactly by $2\,Cov(\Gamma, Z^*)$ as shown in equation (7). Otherwise, the variance of Z located on the left hand side will either increase with time to infinity, or shrink to zero.[8] In either case, Z cannot have a legitimate probability distribution, because it degenerates or explodes in the future. In the real economy, household income follows a log-normal distribution and its longitudinal characteristic of log-normality is stably preserved over time. In this sense, equation (7) can be entitled the 'stability condition' of normally distributed household incomes.

Based on the fact that Z (or Z^*) follows a normal distribution, estimation of the variance of Γ implies estimation of the income transition rule. Furthermore, this also implies that the dimension of parameter space regarding the income transition rule is just 1 under the condition of log-normality of household income. As such, estimation of the variance of Γ is sufficient for estimating the income mobility between periods.

(Γ, Z^*) is a pair of normally distributed random variables whose correlation is not zero. A simple example is the joint bivariate normal distribution. The joint bivariate normality of (Γ, Z^*) is stronger than necessary to satisfy equation (3). However, this assumption that is slightly stronger than necessary is often convenient in applied economic analyses. For simplicity of discussion, I assume that (Γ, Z^*) follows a joint bivariate normal distribution:

$$(\Gamma, Z^*) \sim \text{BVN}(0, \mu_Z, \ \sigma_\Gamma^2, \sigma_Z^2, \rho) \quad \text{where} \quad \rho = \frac{\sigma_{\Gamma Z^*}}{\sigma_\Gamma \cdot \sigma_{Z^*}}.\text{[9]}$$

Thus, the conditional probability distribution of Γ given Z^* follows a normal distribution with mean and variance which are expressed as functions of Z^*; its location parameters (mean and variance) depend

---

[8] If σ_Γ^2 is smaller than the absolute value of $2Cov(\Gamma, Z^*)$, σ_Z^2 will explode to infinity as time goes to infinity. To the contrary, if the reverse condition holds, it will degenerate to zero. The latter case particularly implies that Γ is always zero and, thus, that income ranks will never been altered. In real economic activities, this phenomenon is never observed.

[9] Z^* and Z have identical distribution functions. Therefore, $\rho = -\dfrac{\sigma_\Gamma}{2\sigma_{Z^*}} = -\dfrac{\sigma_\Gamma}{2\sigma_Z}$.

on the value of Z^*. It is well known that the conditional expectation of a variate from a bivariate normal distribution is linear in the conditioning variable. Note that

$$\mu_\Gamma = 0$$

$$E(\Gamma|Z^*) = \mu_\Gamma + \frac{\sigma_{\Gamma Z^*}}{\sigma_{Z^*}^2}(Z^* - \mu_{Z^*}) = \frac{\sigma_{\Gamma Z}}{\sigma_Z^2}(Z - \mu_Z)$$

$$\text{Var}(\Gamma|Z^*) = \sigma_\Gamma^2(1 - \rho^2).$$

Therefore, the conditional normal distribution is expressed as following:

$$\Gamma|Z^* \quad \sim \quad N\left(\frac{\sigma_{\Gamma Z^*}}{\sigma_{Z^*}^2}(Z^* - \mu_{Z^*}), \sigma_\Gamma^2(1 - \rho^2)\right) \quad = \quad N\left(\frac{\sigma_{\Gamma Z}}{\sigma_Z^2}(Z - \mu_Z), \sigma_\Gamma^2(1 - \rho^2)\right). \tag{8}$$

In what follows, a way of estimating the variance of the income transition variable, Γ, is proposed. Since it is not directly observed from the data, Γ needs to be estimated under the constraint of normal distribution conditional on Z^*. In order to estimate σ_Γ^2, information about income transition patterns taking place between years is required for all households, soothe use of a panel data set is essential.

In this paper, σ_Γ^2 is estimated using the KLIPS data sets. The estimation process consists of three steps. In the first step, income percentiles together with cumulative relative frequencies are calculated by each age group in both years. In the second step, the relative incomes are calculated by taking the ratios of the income values realized in the following year to those of the same income percentiles at which income was realized in the previous year. In the final step, σ_Γ^2 is estimated/calculated from the relative income ratios which are taken by natural logarithm.

Under the presumption that Γ follows a normal distribution, the methods of estimating σ_Γ^2 can be direct or indirect. The direct method is to derive the estimate by using the ordinary variance formula such that

$$\hat{\sigma}_{\Gamma_t}^2 = \sum_i w_{it}(\Gamma_{it} - \bar{\Gamma}_t)^2 \quad \text{where} \quad \sum_{i=1}^N w_{it} = 1 \quad \text{and} \quad \bar{\Gamma}_t = \sum_{i=1}^N w_{it}\Gamma_{it} \quad \text{for } t=1,\ldots,T. \tag{9}$$

Since $\mu_{\Gamma_t} = E_t(\Gamma) = 0$, $\bar{\Gamma}$ can be removed from equation (9). By analogy principle, equation (9) can be transformed into the following equation:

$$\hat{\sigma}_{\Gamma_t}^2 = \sum_i w_{it}\Gamma_{it}^2. \tag{10}$$

11

The indirect method is as follows. Note that Γ follows a normal distribution with zero mean. The indirect estimate of σ_Γ^2 can thus be acquired by minimizing the mean squared error between the nonparametrically estimated densities and the normal densities of Γ among all normal distributions with zero means. In other words, the mean squared error minimizing value for the variance is the indirect estimate of σ_Γ^2. This method is in general less sensitive to outliers and therefore more robust. In the sample, it is common to observe unrealistically extreme values of Γ. These phenomena are usually or most frequently observed in low income groups. It is claimed and empirically verified by Kim & Sung (2003) that low income groups in Korea tend to underreport their income as being below the poverty line to tax authorities and statistical offices. The tendencies are exacerbated by a government welfare program, the National Basic Livelihood Security System (NBLSS). Many households whose incomes are around the minimum subsistence level tend to underreport their incomes to remain eligible for NBLSS support. For these households, the reported values of income levels are very small and thus, a small change in income in the following period generally yields extremely high values of Γ. As a result, a direct estimate of σ_Γ^2 frequently yields exceptionally high values of σ_Γ^2 due to the existence of outliers. A direct estimate of σ_Γ^2 is often unrealistically large and sometimes volatile. Therefore, the indirect estimate is more robust and preferable to the direct estimate because it is less sensitive to outliers[10], meaning that the distortive effects can be minimized. In what follows, indirect estimates of σ_Γ^2 are pursued in estimating income mobility parameters. Equations (3) and (8) are used for estimating income mobility.[11]

2.2.3. Discussion

Shorrocks (1978) develops a measure of income mobility by estimating transition probabilities from discrete income transitions among income intervals between two distinct periods. This measures only interval shifts and therefore does not depend on absolute or percentage changes of income. In other words, it is scale-invariant. However, the income transitions between two periods depend on the choice of partitions of income domain, depending on the widths of intervals as well as on the number of intervals. Due to this, a large change in income within an interval sometimes does not have any impact on transition probabilities, while a small change in income between intervals affects the measure. The dimension of the parameter space to be estimated is the product of the number of income intervals of the two periods. In this case, it is not at all easy to summarize the scale of income mobility with a couple of statistics.

A continuous mobility measure which is interval-invariant has been proposed by Fields & Ok (1996, 1999). The following is a well-known Fields & Ok type measure of income mobility:

---

[10] Γ follows a probability normal distribution of mean zero. This does not necessary imply that its sample mean ($\overline{\Gamma}$) is also zero. In other words, the probability that $\overline{\Gamma}=0$ is zero. According to the Law of Large Numbers, $\overline{\Gamma}$ converges to 0, as the sample size increases to infinity.

[11] A change in income over time can be distinguished as income mobility versus income risk. Income mobility is usually defined as the change in income induced from the structural changes in income distribution. Income risk is usually defined as the temporary change in income.

$$m(x, y) = \frac{1}{N} \sum_{i=1}^{N} \left| \ln(y_i) - \ln(x_i) \right| \quad \text{for i=1,2, ... ,N,}$$

where *x* and *y* denote the base and final year incomes, respectively, for a population of size N. m is, in fact, a measure of per capita dollar movement. This can be interpreted as the mean percentage income changes between the two years (Chen, 2009). m does not depend on the number or widths of income intervals. However, its scale depends on income growth rates between periods. Thus, m is scale-dependent. Often, m is decomposed into two parts: total mobility induced by economic growth and total mobility induced by transfer of income between losers and winners.

Using any of the above measures, it is difficult, albeit not impossible, to test the structural changes in income mobility over time.

The income mobility measure proposed in this paper is the variance of income transition variable Γ, which is a continuous real-valued variable. Therefore, Γ is interval-invariant. Furthermore, Γ is scale-invariant as well.

Note that, in equation (1), the information at t explicitly considered in determining the income at (t+1) is only the income rank at t. In other words, there exists no explicit link between the two periods except for the income percentile realized in the previous period. Also, note that the degree of dispersion of (1+γ) does not necessarily depend on that of X for all t. This can be easily calculated such that $1+\gamma = \frac{X^{p'}}{X^{*p}} = \frac{c \cdot X^{p'}}{c \cdot X^{*p}}$ for any nonzero constant c: (1+γ) is independent of the degree of distributional dispersion of income X. In this sense, the logged (1+γ), Γ, is invariant to the scale of X (equivalently, Z=ln(X)).

Note that $\frac{(N-1)s_\Gamma^2}{\sigma_\Gamma^2}$ follows a chi-square distribution with the degree of freedom (N-1), because Γ is distributed normally. Given this statistical property, the longitudinal change in the variance of Γ as an index of income mobility can be tested with the asymptotic theory by the central limit theorem. This is discussed in more detail in Section 2.3 below.

2.3. Testing Methods and Data

The core task of this paper is to estimate variance of the income transition variable as an index for income mobility. Based on equation (3) representing the 'income transition rule', the income transition variable, Γ, is a normal variable with zero mean. In a sense, its function is to reshuffle the orders of household incomes within each age group without affecting means and variances of the normal random variables of Z and Z^* on both sides of equation (3). Hence, the larger the value of the variance of Γ, the more the orders of household incomes become reshuffled, while the smaller the variance, the less the income orders are reshuffled. The variance of income transition variable can as such be interpreted as an

index of income mobility. In this method, the estimation of income mobility is equivalent to estimating the variance of income transition variable.

The income transition variable follows a normal distribution. This implies that the variate, which is defined as the product of the variance estimate and the sample size divided by its corresponding population parameter, $\frac{(N-1)s_\Gamma^2}{\sigma_\Gamma^2}$, follows a χ^2-distribution with the degree of freedom (N-1).[12] The mean and variance of a χ^2-random variable are the degree of freedom and twice the degree of freedom, respectively. Therefore, the mean and variance of s_Γ^2 becomes σ_Γ^2 and $\frac{2\sigma_\Gamma^4}{N-1}$, respectively. Construct a null hypothesis that the variances are the same between two periods, 1 and 2, such that $H_0 : s_1^2 = s_2^2$ or $H_0 : s_1^2 - s_2^2 = 0$. For simplicity of discussion, assume that the two samples of periods 1 and 2 are independent each other. Note that both of the two variance estimates are scalar multiples of Chi-square variables and also, that the difference has an asymptotic normal distribution by the law of large numbers. Its asymptotic mean is zero. Its asymptotic variance has the form of $\frac{2\sigma_1^4}{N_1-1} + \frac{2\sigma_2^4}{N_2-1}$ when the two samples are drawn independently.

In this paper, I use the KLIPS data set compiled for the period 1998 to 2008. Its basic sample characteristics are shown in Tables A-1 and A-2 in the Appendix. The KLIPS is a panel data set and thus the samples are serially correlated. Therefore, the asymptotic variance of the difference of sample variances, ($s_1^2 - s_2^2$), ranges from zero when they are perfectly negatively correlated with each other to two times as large as $\left(\frac{2\sigma_1^2}{N_1-1} + \frac{2\sigma_2^2}{N_2-1} \right)$ when they are perfectly positively correlated. In the real economy, they are in-between. Even though Γ is distributed normal and, so, the sample variance of Γ multiplied by some scalar is distributed χ^2, the difference of sample variances are distributed asymptotically normal by the law of large numbers. Based on this, the null hypothesis of identical variances of the income transition variables for two different periods can be tested using a t-test.

3. Results

3.1. Income Mobility and Testing Hypothesis

As defined above, the income mobility in this paper is measured by the size of the variance of income transition variable.[13] In this paper, the underlying assumption is modified for simplicity of discussion such that the distribution of income transition variable is invariant to age and income percentile (or,

---

[12] Note that Γ follows a probability normal distribution of mean zero. Therefore, its distribution exactly follows a Chi-square distribution.

[13] The variance of income transition variable may depend on some demographic, economic, or other characteristics such as time, age, and income percentile et cetera. This is likely to be analyzed in future research.

equivalently, rank). However, it is assumed that it is changeable over time. The raw data of KLIPS are released for 11 years and, thus, the variance of the income transition variable can be estimated ten times. The estimation results for the variance of income transition variable for the period between 1998 and 2008 are shown in Figure 1 and Table 1.

Two methods of estimating the variance of the income transition variable are adopted: a direct method versus an indirect method. The direct method uses a numerical formula of variance definition where the estimate is derived by taking (weighted) average of squared deviations from the mean value (which is zero). The indirect method takes as the variance estimate the value of variance of a normal variable with zero mean which minimizes the mean squared errors of the difference between its distribution function and that of the (nonparametrically) estimated distribution function.[14] The direct estimate is heavily affected by outliers in general, while the indirect estimate is more robust to outliers. As such, the indirect estimate seems to be preferable to the direct estimate. However, the indirect estimate is also burdensome due to its technical complexity and calculation burden, while the direct method is simple to calculate.

Estimates of variances are discussed mainly focusing on the indirect method. The indirect estimate of the variance of Γ was 0.18288 in 1999 and 0.10598 in 2008, respectively, which was 42.0 percent smaller than the 1999 value. The variance of Γ dwindled over time with a couple of small humps over the last decade and its locus was roughly close to linear. The direct variance estimate of Γ was 0.20763 in 1999 and 0.16256 in 2008, respectively, which was 21.7 percent smaller than the 1999 value. Except for the absolute values, the two types of variance estimates are quite similar in their decreasing trends over time.

Decreases in the variance estimates of the income transition variable, Γ, imply that the income mobility decreased and therefore that the changes in income ranks between years became narrower. This observation can be tested statistically with the null hypothesis of identical variances of Γ between two periods. The simplest case is that the samples are independent of each other. The other two extreme cases are either perfect negative correlation or perfect positive correlation. In the former case, any nonzero value of test statistic guarantees rejection of the null hypothesis because the variance of the difference of variance is zero. In the latter case, a large value t-statistic which exceeds the critical value results in rejection of the null hypothesis. Suppose, for the second case, that the two samples are perfectly positively correlated. In this case, the asymptotic variance of the t-test statistic is twice as large as the variance of the case where samples are independently drawn. This implies that the critical value of the former needs to be $\sqrt{2}$ times larger than the latter case to reject the hull hypothesis. The t-statistics shown in Table 1 are calculated based on independent cases. Therefore, if the t-statistics in Table 1 are $\sqrt{2}$ times larger than the ordinary critical value, then the null hypothesis is rejected in any case at a given significance level.

Figure 1. Variance Estimates of Γ as an Income Mobility Index

---

[14] A nonparametric kernel density estimation method is adopted.
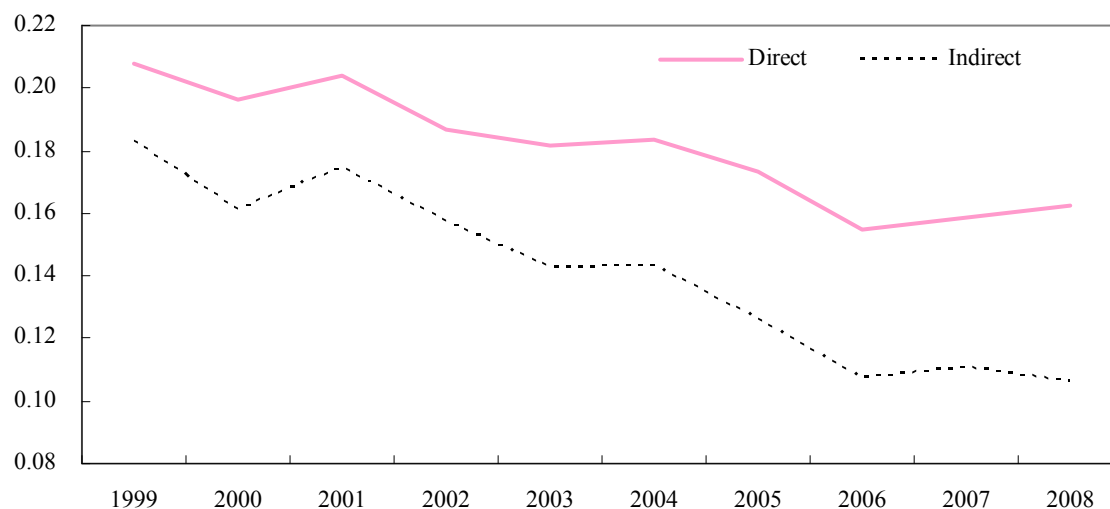
In what follows, the null hypothesis that the variances of income transition variable are the same between the two (consecutive) periods is tested. The test results show that the null hypothesis is rejected in seven out of nine cases tested at the 5 percent significance level, based on the indirect method; the null was rejected three times, based on the direct method.

Note that the income transition variable, Γ, is an index measuring the magnitude of change in income mobility[15] between two periods. Even though the variance of Γ rebounded slightly up several times during the last decade, this did not outweigh its decreasing trend. This implies that income mobility became smaller over time and its change is structural. Conversely, it also implies that the rigidity of income changes became stronger over time.

Identification of the factors involved is beyond the main scope of this paper and is not explicitly analyzed here. However, one can posit several plausible factors which reduced the income mobility over time at a statistically significant level: these are the decline of labor market flexibility, population aging, skill-biased technical development, restructuring of financial sectors, information asymmetry and so forth. These need to be scrutinized in the future.

Of these, it has been argued that population aging is the strongest and most significant factor that affected income mobility during the last decade in Korea. Sung & Park (2009) found that population aging is the largest and most significant factor that has widened the income inequality since the mid 1990s and that the inequality is expected to grow even in the distant future due to population aging. They predicted that the income inequality would be 27.5 percent larger in 2050 compared to 2008 in Korea solely due to population aging. Likewise, it seems reasonable to suggest that population aging will also play a role in restricting income mobility in the future. This is because population aging implies that the older generations, who are mostly retirees and are unlikely to succeed in reentering the labor market, will claim a greater population share. This increased share of retirees is an important factor in the lowering of income mobility.

---

[15] The magnitude includes both frequencies and widths.

Table 1. Results of Hypothesis Testing of Structural Change in Variance of Income Transition Variable

	Avg.	Variance (Direct)	Variance (Indirect)	Log-normality χ^2- Test Stat .	Sample Size	Std. Dev. for Difference between Variances(Direct)	Difference between Variances (Direct)	t-Value (Direct)	Std. Dev. for Difference between Variances(Indirect)	Difference between Variances (Indirect)	t-Value (Indirect)
1999	-0.00912	0.20763	0.18288	0.331	4,375						
2000	-0.01634	0.19609	0.16146	9.410	3,989	0.00591	0.03709	6.277***	0.00477	0.06854	14.370***
2001	0.03328	0.20383	0.17436	1.115	3,785	0.00648	-0.01154	-1.782*	0.00553	-0.02142	-3.875***
2002	-0.05755	0.18666	0.15735	6.668	3,903	0.00645	0.00774	1.200	0.00542	0.0129	2.381**
2003	0.01301	0.18168	0.14234	14.228	3,967	0.00623	-0.01717	-2.754**	0.00530	-0.01701	-3.211***
2004	-0.01274	0.18330	0.14352	17.470	4,287	0.00574	-0.00498	-0.867	0.00468	-0.01501	-3.205***
2005	-0.03737	0.17335	0.12604	50.242	4,455	0.00552	0.00162	0.293	0.00432	0.00118	0.273
2006	-0.02543	0.15511	0.10721	95.548	4,575	0.00531	-0.00995	-1.873*	0.00402	-0.01748	-4.344***
2007	-0.01200	0.15845	0.11066	64.859	4,743	0.00483	-0.01824	-3.780***	0.00343	-0.01883	-5.483***
2008	-0.01580	0.16256	0.10598	91.323	4,777	0.00455	0.00334	0.735	0.00316	0.00345	1.092

Notes: 1. 1998 was the year when the KLIPS was first compiled. Estimation of variance requires information of income ranks in the previous year. Therefore, it is not possible to estimate the variance of Γ for the year 1998.

2. In χ^2-tests, critical values at significance levels of 1%, 5%, and 10% are 9.210, 5.991, 4.605, respectively.

3. *, **, *** indicate that the null hypothesis is rejected at significance levels of 1%, 5%, or 10%.

4. $\frac{(N-1)s^2}{\sigma^2} \sim \chi^2(N-1)$ under the assumption that the variate follows a normal distribution. The variance of a χ^2-variable with the degree of freedom of (N-1) is 2(N-1); so, $\text{Var}(s^2) = \frac{2\sigma^4}{N-1}$. Therefore, the difference between the two variances of the adjacent years (1 and 2, for instance) is equal to $\frac{2\sigma_1^4}{N_1-1} + \frac{2\sigma_2^4}{N_2-1}$ under the assumption of mutual independence between them.

3.2. Application: Poverty Outflow/Inflow Probabilities

In this section, poverty inflow/outflow (or, exit/entry) probabilities are estimated using the distributional characteristics of the income transition variable, Γ, which is known to be distributed normally. Poverty inflow probability (PIP) is defined as the probability that income at (t+1) is larger than the poverty line for a given income percentile realized at t. The PIP depends on the percentile (or rank) of income at t. If a household belonged to the poor income group in the previous period, then the PIP can be renamed as the poverty continuation probability. Poverty outflow probability (POP) is simply the reverse of the PIP, so that their sum always makes unity. By the same token, the POP is defined as the probability that the income realized in the following year will exceed the poverty threshold level. This can be renamed as the non-poverty continuation probability, as it relates to those who were not poor in the previous year.

The higher the income percentile for a household in the previous year is, the higher the POP is. By contrast, the POP becomes smaller as the income percentile in the previous year becomes lower. The minimum subsistence levels set by the government differ by household size. Therefore, the POP also depends on the household size. Based on the estimates of Γ and its variance, the POPs are calculated for all income percentiles. The left-hand-side of Figure 2 shows the loci of POPs by household size. Likewise, the loci of PIPs for various types of households are shown in the right-hand-side of Figure 2.

Figure 2. POPs and PIPs by Sizes of Households

POPs   PIPs

(unit: %)
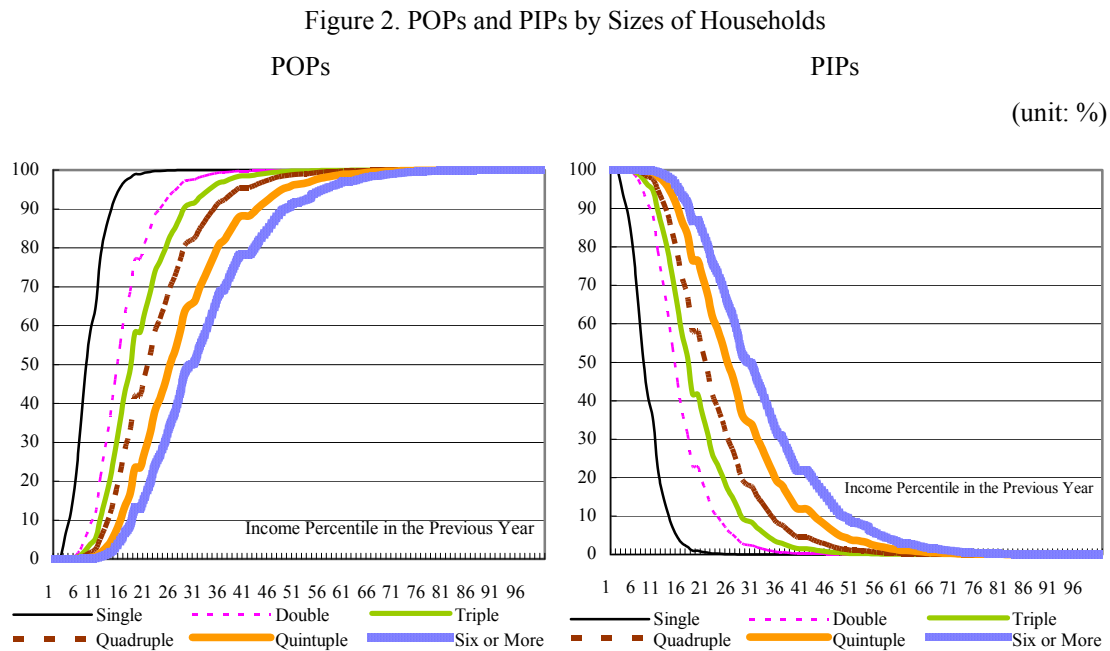
Note: The POPs and PIPs are estimated based on the indirect variance estimate of Γ (0.10598) using the 2008 KLIPS data set.

3.2.1. Methods of Estimating PIPs and POPs using Conditional Probability Distribution

Conditional on the log-normality of household income and on the assumption that income realization in the following year depends on the location (that is, the rank) of income within each age group in the previous year, the probability of income realization on a certain interval can be calculated using the conditional probability distribution of a random variable which is distributed normal. As discussed above, household income is distributed log-normal and its log-normality has been preserved for the period since 1982 when the HIES was first released to the public, except for the two years of economic crisis between 1997 and 1998. Equation (3) denotes the income transition rule. The assumption of bivariate normality of logarithm of household income is used to derive the conditional probabilities of PIPs or POPs, based on the conditional distribution function of a normal variable of Γ given Z^*.

Z_t^* is the random variable denoting household income distribution in the following year, even though it is not actually observed. The value of Z_t^* denotes the logarithm of household income value at the p-th percentile within the age group A at t, when the income was realized at the p-th percentile for the household within the age group (A-1) at (t-1)[17] for $p \in [0,1]$; Z_t^* denotes, in fact, the expected value of logarithm of income realization in the following year (t) for the household whose income percentile was p in the previous year (t-1).

The joint random pair (Γ_t, Z_t^*) follows a bivariate normal distribution and their sum constitutes a normal variable, Z, which represents the logarithm of realized income in the following year. A household is classified as poor when its logarithm of income is smaller than the logarithm of the corresponding poverty line. Note that the realization of income (Z_t) in the following year differs from its expected value (Z_t^*) with probability one for a continuous random variable; it is affected by the income transition variable Γ. In other words, the realization of income is not its expected value (Z_t^*) but the sum of the expected value and its transition variable, according to the income transition rule.

Suppose that the income of a household was realized in the previous year at the p-th percentile within the age group (A-1) at (t-1). The probabilities of realization of income on a certain interval in the following year can be easily defined and calculable; the probabilities that the realized income in the following year is greater or smaller than certain values within the age group A at t which is identical to the age group (A-1) at (t-1) are the conditional probabilities given the p-th percentile income variable Z_t^*. Since (Γ_t, Z_t^*) jointly follows a bivariate normal distribution, (Z_t, Z_t^*) also follows a joint bivariate normal distribution. From the well-known statistical properties of a bivariate normal distribution, Γ (or, equivalently, Z_t) conditional on Z_t^* is distributed normal with mean and variance which are linear in Z_t^*.

Special examples of these probabilities are PIPs and POPs, if the threshold points are logarithms of poverty lines.[18] PIPs and POPs depend on the conditional distribution of Z given Z^*, or, equivalently, that of Γ given Z^*, where the joint pair (Γ, Z^*) follows a bivariate normal distribution. Let ρ be the correlation

coefficient of (Γ_t, Z_t^*): $\rho = \dfrac{\sigma_{\Gamma Z^*}}{\sigma_\Gamma \sigma_{Z^*}}$. The joint density of (Γ_t, Z_t^*), the marginal density of Z^*, and the

---

[17] Note that the age group A at t is identical to the age group (A-1) at (t-1).

[18] Note that the sum of PIPs and POPs is always 1. This follows from the definitions of PIPs and POPs.

conditional density of Γ given Z^* are given, respectively, as follows:

$$f\left(\Gamma, Z^*\right) = \frac{1}{2\pi\sigma_\Gamma\sigma_{Z^*}\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{(\Gamma-\mu_\Gamma)^2}{\sigma_\Gamma^2} - \frac{2\rho(\Gamma-\mu_\Gamma)(Z^*-\mu_{Z^*})}{\sigma_\Gamma\sigma_{Z^*}} + \frac{(Z^*-\mu_{Z^*})^2}{\sigma_{Z^*}^2}\right\}\right]$$

$$f_Z\left(Z^*\right) = \frac{1}{\sqrt{2\pi}\sigma_{Z^*}}\exp\left[-\frac{(Z^*-\mu_{Z^*})^2}{2\sigma_{Z^*}^2}\right]$$

$$f\left(\Gamma|Z^*\right) = \frac{f(\Gamma, Z^*)}{f_Z(Z^*)}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_\Gamma\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\times\right.$$

$$\left.\left\{\frac{(\Gamma-\mu_\Gamma)^2}{\sigma_\Gamma^2} - \frac{2\rho(\Gamma-\mu_\Gamma)(Z^*-\mu_{Z^*})}{\sigma_\Gamma\sigma_{Z^*}} + \frac{(Z^*-\mu_{Z^*})^2}{\sigma_{Z^*}^2} - \frac{(Z^*-\mu_{Z^*})^2}{\sigma_{Z^*}^2}(1-\rho^2)\right\}\right]$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_\Gamma^2(1-\rho^2)}}\exp\left[-\frac{1}{2\sigma_\Gamma^2(1-\rho^2)}\left(\Gamma-\mu_\Gamma-\frac{\sigma_{\Gamma Z^*}}{\sigma_{Z^*}^2}(Z^*-\mu_{Z^*})\right)^2\right]$$

Therefore, $\Gamma|Z^*$ follows a normal distribution with mean $\mu_\Gamma + \frac{\sigma_{\Gamma Z^*}}{\sigma_{Z^*}^2}(Z^*-\mu_{Z_*})$ and variance $\sigma_\Gamma^2(1-\rho^2)$.

Since $\mu_\Gamma=0$, its conditional mean shrinks to $\frac{\sigma_{\Gamma Z^*}}{\sigma_{Z^*}^2}(Z^*-\mu_{Z_*})$.

$$\Gamma\big|Z^* \sim N\left(\frac{\sigma_{\Gamma Z^*}}{\sigma_{Z^*}^2}(Z^*-\mu_{Z_*}), \sigma_\Gamma^2(1-\rho^2)\right) \tag{11}$$

Equation (11) suggests that the expectation of the conditioned variable (Γ) is a linear function of the conditioning variable (Z^*). As the correlation increases to unity in absolute value, the conditional distribution degenerates.

Suppose that the income percentile was realized at p within the age group (A-1) in the previous year, meaning that Z^* corresponds to the logarithm of the p-th percentile income in the following year within the age group A such that $Z=z_0$. Suppose also that the poverty line is exogenously set at P_0 by the government. The POPs can then be derived as follows:

$$\Pr\left[Z > \ln(P_0)\big|Z^* = z_0\right] = \Pr\left[\Gamma + z_0 > \ln(P_0)\big|Z^*\right] = \Pr\left[\Gamma > \ln(P_0) - z_0\big|Z^*\right] \tag{12}$$

$$= \int_{\ln(P_0)-z_0}^{\infty} f\left(\Gamma\big|Z^* = z_0\right)d\Gamma$$

The above probability, POP, is the probability of poverty exit in the following year for a household that

was poor in the previous year. However, for a household whose income was above the poverty line in the previous year, it is the non-poverty continuation probability that is calculated for the following year. PIPs are derived by subtracting the above probability from one, and thus, PIP represents the probability of poverty entry in the following year for a household that was not poor in the previous year, or the poverty continuation probability in the following year for a household that was also poor in the previous period.

Using the cumulative density function of a standard normal variable, $F(\cdot)$, the POP expression shown in Equation (12) can be rewritten as:

$$\Pr\left[Z > \ln(P_0) \big| Z^* = z_0\right] = \Pr\left[\Gamma > \ln(P_0) - z_0 \big| Z^*\right] \tag{13}$$

$$= \Pr\left[\frac{\Gamma - \frac{\sigma_{\Gamma Z^*}}{\sigma_{Z^*}^2}(z_0 - \mu_{Z^*})}{\sqrt{\sigma_\Gamma^2(1-\rho^2)}} > \frac{\ln(P_0) - z_0 - \frac{\sigma_{\Gamma Z^*}}{\sigma_{Z^*}^2}(z_0 - \mu_{Z^*})}{\sqrt{\sigma_\Gamma^2(1-\rho^2)}} \Bigg| Z^* = z_0\right]$$

$$= F\left[\frac{-\ln(P_0) + z_0 + \frac{\sigma_{\Gamma Z^*}}{\sigma_{Z^*}^2}(z_0 - \mu_{Z^*})}{\sqrt{\sigma_\Gamma^2(1-\rho^2)}} \Bigg| Z^* = z_0\right].$$

3.2.2. Estimation Results of PIPs and POPs

Absolute poverty lines are set by the Korean government, depending upon the size of a household.[19] The POP is obtainable by substituting for the values of mean and variance of (logarithm of) disposable income into the above equation (13) given P_0, the income rank (p) in the previous year, and the estimates of covariance, variances, and correlation coefficient, whose values depend on household size. This probability becomes a function of the income rank in the previous year. Note that the expected value of income rank in the following year equals the income rank realized in the previous year. Thus, the probability of remaining poor in the following year also tends to be high for a household that was poor in the previous year. However, the PIP tends to be low for a household belonging to a high income decile. The POP in the following year increases with income ranks realized in the previous year. Therefore, the POP curve is upward sloping to the right. As shown in Figure 2, the POP curve is concave downward (or, equivalently, convex upward) and thus, its probability increases with income percentile, but at a decelerated rate.

The positive slope of the POP curve can be verified as follows. The correlation of income percentile at (t-1) and expected income value at t (Z^*) is positive. Hence, it suffices to show the correlation of the POP and Z^* for the positive slope. Observe that $Z^*(=z_0)$ denotes a random variable corresponding to an income rank realized in the previous year. Taking a derivative with respect to Z^* evaluated at z_0 on the POP

---

[19] The poverty lines are minimum subsistence levels set by the Korean Ministry of Health and Welfare. See Table A.3. in Appendix.

equation yields the following relation:

$$\frac{\partial F}{\partial z_0} = \frac{1 + \dfrac{\sigma_{\Gamma Z^*}}{\sigma_{Z^*}^2}}{\sqrt{\sigma_\Gamma^2(1-\rho^2)}} F'.$$

$\dfrac{\sigma_{\Gamma Z^*}}{\sigma_{Z^*}^2}$ is negative and its absolute value is less than 1: $\partial F/\partial z_0$ is positive. Therefore, ceteris paribus, the

POP (F) increases with the income rank. The poverty lines are set depending on the household size. The mean and variance of logarithm of disposable income vary with household size. Therefore, the POPs differ by household size. According to Figure 2, the larger the household size is, the higher the POPs are.

The POPs are influenced by the level of the poverty line and the income mean. In what follows, their effects on the POPs are mathematically analyzed. Partial differentiations of Equation (13) with respect to $\ln(P_0)$ and μ_{Z^*} yield the following equations:

$$\frac{\partial F}{\partial \ln(P_0)} = -\frac{1}{\sqrt{\sigma_\Gamma^2(1-\rho^2)}} F' < 0 \tag{14}$$

$$\frac{\partial F}{\partial \mu_{Z^*}} = -\frac{\dfrac{\sigma_{\Gamma Z^*}}{\sigma_{Z^*}^2}}{\sqrt{\sigma_\Gamma^2(1-\rho^2)}} F' > 0 \tag{15}$$

Note that F is decreasing in the (logarithm of) poverty line, $\ln(P_0)$, but increasing in the income rank, which is, more specifically, the conditional expectation of logarithm of income corresponding to the income percentile (μ_{Z^*}) realized in the previous year. The (logarithm of) poverty line, $\ln(P_0)$, is smaller for smaller sized households. Therefore, the POP (F) becomes larger as the household size becomes smaller. However, the mean value, μ_{Z^*}, tends to become smaller for smaller households; thus, the POP (F) becomes smaller for smaller households. Therefore, the direction of change in POP depends on the relative changes in $\ln(P_0)$ and μ_{Z^*} and, also, on the absolute values of Equations (14) and (15). In Korea, the ratios of $\ln(P_0)$ and μ_{Z^*} have conventionally been stable over the last decade. In other words, the government tends to determine the values of P_0 (poverty lines), considering the values of (log-) mean incomes (μ_{Z^*}). Hence, the rates of changes in $\ln(P_0)$ and μ_{Z^*} have been quite similar in Korea. Based

on the estimation results, the absolute value of $\dfrac{\sigma_{\Gamma Z^*}}{\sigma_{Z^*}^2}$ is less than 1. Therefore, ceteris paribus, the POP

tends to be higher for a smaller household and, thus, the POP curve for a smaller household usually lies above the POP curve for a larger household.

According to the income transition rule proposed in the previous section, income realization in the

following year depends on the income percentile realized in the previous year and the value of income transition variable. Income transition variable, Γ, is negatively correlated with income rank or Z^* so that $\sigma_{\Gamma Z^*} = -\dfrac{\sigma_\Gamma^2}{2} < 0$. This implies mean reversion effect. The probability that Γ is positive is larger than 0.5 for households with income ranks lower than the median in the previous year, while the probability that Γ is negative for households with income ranks higher than the median in the previous year. Because a normal distribution is symmetric around its mean, median equals mean. If the income realized in the previous year is higher than the median (log-)income, that is, $Z^*(=z_0) > \mu_{Z^*}$, the conditional variable $\Gamma | Z^*$ has negative mean value and its conditional probability of having a negative value for Γ exceeds 0.5. An exactly symmetric argument applies and, so, the conditional probability of having a positive value for Γ exceeds 0.5 for the case where the income was realized below its median in the previous year.

Note that PIP is defined as the probability that the income realization in the following year is smaller than the poverty line given the condition that its income rank in the previous year was p for $p \in [0,1]$. Thus, PIP = 1-POP. Therefore, PIP and POP are symmetric to each other around the horizontal line at the probability level 0.5.

3.2.3. Sensitivity Analysis

A longitudinal change in income rank is reflected in Γ. Therefore, POP/PIP change through changes in Γ. More specifically, they do depend not only on its variance (σ_Γ^2), but also on the covariance of Γ and Z^* ($\sigma_{\Gamma Z^*}$). Because $\sigma_{\Gamma Z^*} = -\dfrac{\sigma_\Gamma^2}{2}$ under the log-normality of household income over time, POP/PIP depend on the value of σ_Γ^2 given income distributions.

The longitudinal variability of income ranks over time is positively correlated with the size of income mobility, which is measured by the level of σ_Γ^2 in this paper. Ceteris paribus, the higher is the income mobility (σ_Γ^2) for a poor household, the higher is the POP. However, this trend is reversed for a non-poor household. Therefore, the sign of correlation between the POP and the income mobility measured by σ_Γ^2 depends on the income ranks or income percentile in the previous year. For some income percentiles, the POP would not change much, even though σ_Γ^2 varies significantly. Its range is around 25% for households of size 3 and 30% for those of size 4, for instance. Therefore, the POP curves tend to rotate clockwise around certain income ranks depending upon the household size, as σ_Γ^2 increases. These are illustrated in Figure 3 for the cases of households of size 3 and 4, respectively.

Figure 3. POPs for Different Values of Variance (Standard Deviation) of Income Transition Variable (Γ)
(Simulation Results for Households of Size 3)   (Simulation Results for Households of Size 4)

(unit: %)
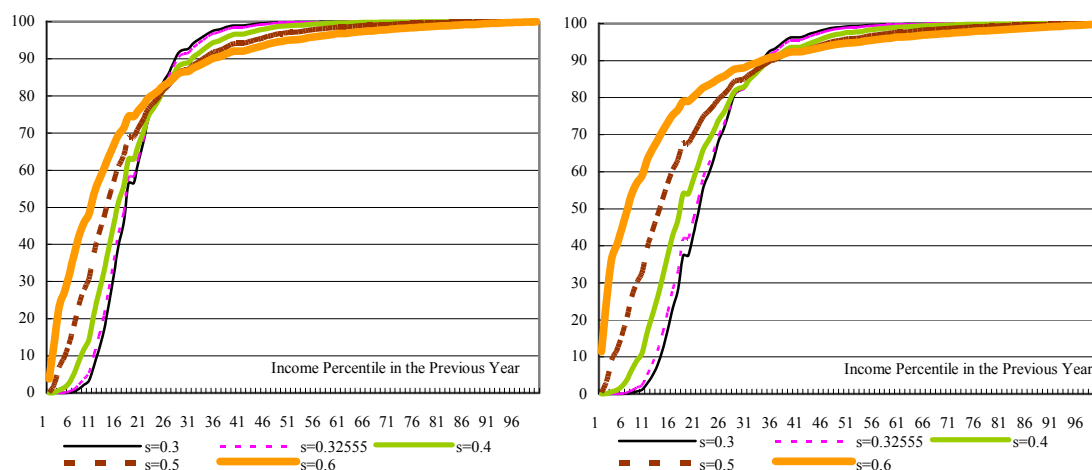
Note: Simulations are conducted using the 2008 KLIPS data set for numerous values of s, which denotes the sample estimate of σ_Γ. 0.32555 is the square root of 2007-2008 indirect variance estimate 0.10598 as shown in Table 1.

4. Concluding Remarks

This paper develops a new definition of income transition between two consecutive periods on a continuous distribution based on the log-normality of household income. It also analyzes the income mobility that can be derived therefrom. Most studies on income mobility and transition between periods have made use of discrete methods such as income transition matrices. A fundamental difference in this paper is the continuity of income mobility using parameterized statistical properties of household income distribution. The income transition rule is derived parametrically based on the assumption that the expectation of following period income is equal to the income level corresponding to the same income percentile that was realized in the previous year. In other words, the income in the following year is realized around its conditional expected value corresponding to the income percentile realized in the previous year. An additional assumption of the log-normality of household income eases calculation of mobility and enables testing its structural changes over time.

The advantages of the proposed measure of income mobility are its interval- and scale-invariances, the simplicity of calculation and hypothesis testing, and the wider applicability of simulation by altering the variance estimates. These are hardly expectable in the mobility measures already established by previous studies such as Shorrocks (1978), Fields & Ok (1996, 1999) and etc. Owing to these advantages, estimation of income mobility becomes easier and more reliable, more simply tested and has wider applicability.

However, the proposed method does have some disadvantages, including misspecification error. The mobility is assumed to be identical across income ranks/percentiles and age for simplicity of discussion. However, the validity of this needs to be easily verified on a case-by-case basis in related future studies.

Using the KLIPS data sets for the last decade, the income mobility is defined and measured by the variance (σ_Γ^2) of income transition variable (Γ), which is distributed normal and reshuffles the income

24

orders every period in accordance with the income transition rule. Its structural change is tested by taking differences in the variances of income transition variable between adjacent years. It was found that it decreased significantly for the last decade in Korea.

What remains to be done is to scrutinize the causes of the decrease in income mobility by lowering the variance of income transition variable. Rapid population aging, change in labor market flexibility and rigidity, skill-biased technical development, or regulations are all plausible candidates for explaining lowered mobility. This is an area that needs to be rigorously researched in the future.

References

Auten G. and G. Gee, "Income Mobility in the United States-New Evidence from Income Tax Data," *National Tax Journal*, Vol. 62, No. 2, pp. 301-328, 2009.

Bager-Sjögren, L. and N. A. Klevmarken, "Inequality and Mobility of Wealth in Sweden, 1983/84-1992/93," *Review of Income and Wealth*, Series 44, No. 4, pp. 473-495, 1998.

Chen, W., "Cross National Differences in Income Mobility," *Review of Income and Wealth*, Series 55, No. 1, pp. 75-100, 2009.

Cho, Y. and K. Kim, "A Study on Determinants of Poverty Exit/Entry," A paper presented at the Symposium co-hosted by Allied Korean Economic Associations held in Seoul in 2007, 2007.

Fields, G., "Does Income Mobility Equalize Longer-term Incomes-New Measures of an Old Concept," *Journal of Economic Inequality*, Vol. 8. No. 4, pp. 409-427, 2010.

Fields, G. and E. A. Ok, "The Meaning and Measurement of Income Mobility," *Journal of Economic Theory*, 71, pp. 349-377. 1996.

_____ and _____, "Measuring Movement of Incomes," *Economica*, Vol. 66, No. 264, pp. 455-471, 1999.

Gardiner, K. and J. Hills, "Policy Implications of New data on Income Mobility," *The Economic Journal*, Vol. 109, No. 453, pp. F91-F111, 1999.

Geweke, J., R. C. Marshall, and G. A. Zarkin, "Mobility Indices in Continuous Time Markov Chains," *Econometrica*, Vol. 54, No. 6, pp.1407-1423, 1986.

Gottschalk, P., "Inequality, Income Growth, adn Mobility-The Basic Facts," *Journal of Economic Perspectives*, Vol. 11, No. 2, pp.21-40, 1997.

Jarque, C. M., and A. K. Bera, "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals," *Economic Letters*, Vol. 6, pp. 255-259, 1980.

Jarvis, S. and S. P. Jenkins, "How Much Income Mobility is There in Britain?" *The Economic Journal*, Vol. 108, No. 447, pp. 428-443, 1998.

Kim, J. M. and M. J. Sung, "Estimation of Cohort Lifetime Age-income Profile Using Characteristics of Income Distribution," *Journal of Korean Economic Analysis*, Vol. 9, No. 3, pp. 153-248, 2003.

Lee, C. and G. Solon, "Trends in Intergenerational Income Mobility," *Review of Economics and Statistics,* Vol. 91, No. 4, pp. 766-772, 2009.

Maasoumi, E. and M. Trede, "Comparing Income Mobility in Germany and the United States Using Generalized Entropy Mobility Measures," *The Review of Economics and Statistics*, Vol. 83, No. 3, pp. 551-559, 2001.

Mitra, T. and E. A. Ok, "The Measurement of Income Mobility-A Partial Ordering Approach," *Economic Theory*, Vol. 12, No. 1, pp.77-102, 1998.

Royston, J. P. "An Extension of Shapiro and Wilk's W test for Normality to Large Samples," *Applied Statistics*, 31, pp. 115-124, 1982.

Shapiro, S. S. and M. B. Wilk, "An Analysis of Variance Test for Normality (Complete Samples)," *Biometrika*, 52 (3-4), pp.591-611, 1965.

Shi, X., J. Nuetah, and X. Xin, "Household Income Mobility in Rural China-1989-2006," *Economic Modelling*, Vol. 27, pp.1090-1096, 2010.

Shorrocks, A. F., "Income Mobility and Markov Assumption," *The Economic Journal*, Vol. 86, No. 343, pp.566-578, 1976.

_____, "The Measurement of Mobility," *Econometrica*, Vol. 46, No. 5, pp.1013-1024, 1978.

Stephens, M. A. "EDF Statistics for Goodness of Fit and Some Comparisons," *Journal of the American Statistical Association*, 69, pp. 730-737, 1974.

Sung, M. J., *A Study on Life-time Poverty and Policy Proposals in Korea through Pseudo-panel Analyses*, Research Report 05-01, Korea Institute of Public Finance, 2005.

_____, " Income Mobility and Its Implication on Government Welfare Expenditure, " *Korean Journal of Policy Studies*, Vol. 23, No. 1, pp. 127-152, 2008.

Sung, M. J. and K. Park, "Effects of Demographic Changes on Income Inequality in Korea," *Kyong Je Hak Yon Gu*, Vol. 57, No. 4, pp. 5-37, 2008.

Trede, M., "The Age Profile of Mobility Measures: An Application to Earnings in West Germany," *Journal of Applied Econometrics*, Vol. 13, No. 4, pp. 397-409, 1998.

Tsui, K., "Measurement of Income Mobility: a Re-examination," *Social Choice Welfare*, Vol. 33, pp. 629-645, 2009.

Van Kerm, P., "What Lies behind Income Mobility? Reranking and Distributiuonal Change in Belgium, Western Germany and the USA," *Economica*, Vol. 71, No. 282, pp.223-239, 2004.

Weon, J. and M. J. Sung, *Increasing Income Inequality in Korea and Policy Implications*, Research Report 07-10, Korea Institute of Public Finance, 2007.

Appendix. Miscellaneous Tables

Table A.1. Population Shares by Age of Heads

	Population (A)	Age 0-16 (B)	Age 17-64(C)	Age 65- (D)	B/A (%)	C/A (%)	D/A (%)
1960	25,012,374	11,526,114	12,759,810	726,450	46.08	51.01	2.90
1980	38,123,775	14,708,142	21,959,600	1,456,033	38.58	57.60	3.82
2000	47,008,111	11,299,381	32,313,834	3,394,896	24.04	68.74	7.22

2005	48,138,077	10,493,652	33,277,783	4,366,642	21.80	69.13	9.07
2010	48,874,539	9,285,680	34,232,006	5,356,853	19.00	70.04	10.96
2020	49,325,689	7,010,949	34,613,615	7,701,125	14.21	70.17	15.61
2030	48,634,571	6,303,000	30,520,864	11,810,707	12.96	62.76	24.28
2040	46,343,017	5,493,374	25,808,736	15,040,907	11.85	55.69	32.46
2050	42,342,769	4,360,318	21,826,694	16,155,757	10.30	51.55	38.15

Source: Statistics Korea

Table A.2. Descriptive Statistics for KLIPS of 2008

(unit: 10,000 KRW)

	Average	Std. Dev.	Max.	Min.
Age of Head	50.84	14.33	93	16
Household Size	3.08	1.33	10	1
Household Income (Total)	3,736.1	3,775.81	73,120	0
Wage & Salary Income	3,121.5	3,122.12	73,120	0
Financial Income	53.89	386.16	12,900	0
Rental Income	178.69	1,294.34	30,000	0
Social Security Cash Benefits	94.78	411.24	6,600	0
Private Transfer Income	187.56	924.94	30,000	0
Other Income	99.69	965.15	30,000	0
Consumption Expenditure	2,322.63	1,601.73	22,800	0

Note: The statistics are based on author's own calculation for the 2008 KLIPS data, of which the sample size is 5,116.

Table A.3. Minimum Subsistence Levels by Household Size

(unit: KRW/year, ln(KRW)/year)

KRW \ H. Size	1	2	3	4	5	6
2000	3,888,132	6,439,368	8,856,912	11,140,776	12,667,056	14,293,608
2001	4,004,772	6,632,544	9,122,616	11,475,000	13,047,072	14,722,416
2002	4,144,944	6,864,696	9,441,924	11,876,628	13,503,732	15,237,708
2003	4,269,288	7,070,628	9,725,172	12,232,932	13,908,840	15,694,848
2004	4,418,712	7,318,104	10,065,564	12,661,080	14,395,644	16,244,160
2005	4,817,592	8,022,048	10,895,148	13,635,984	15,635,016	17,733,600
2006	5,019,708	8,405,868	11,271,768	14,045,064	16,238,904	18,508,584
2007	5,231,052	8,812,944	11,674,392	14,466,420	16,864,944	19,315,560
2008	5,556,564	9,411,828	12,319,236	15,190,176	17,854,536	20,546,232
2009	5,890,140	10,029,156	12,974,232	15,919,308	18,864,372	21,809,448
2010	6,052,128	10,304,964	13,331,028	16,357,092	19,383,156	22,409,220
ln(KRW) \ H. Size	1	2	3	4	5	6
2000	15.17	15.68	16.00	16.23	16.35	16.48
2001	15.20	15.71	16.03	16.26	16.38	16.50
2002	15.24	15.74	16.06	16.29	16.42	16.54
2003	15.27	15.77	16.09	16.32	16.45	16.57
2004	15.30	15.81	16.12	16.35	16.48	16.60
2005	15.39	15.90	16.20	16.43	16.57	16.69
2006	15.43	15.94	16.24	16.46	16.60	16.73
2007	15.47	15.99	16.27	16.49	16.64	16.78
2008	15.53	16.06	16.33	16.54	16.70	16.84
2009	15.59	16.12	16.38	16.58	16.75	16.90
2010	15.62	16.15	16.41	16.61	16.78	16.92

Source: Ministry of Health and Welfare, Korea.