2014-4

On the Mathematical Formulation of the Cohort Cumulative Social Increase Ratio

Takashi Inoue

May 2014

Working Paper Series, Institute of Economic Research
Aoyama Gakuin University

On the Mathematical Formulation of the Cohort Cumulative Social Increase Ratio

Takashi Inoue

Department of Public and Regional Economics

College of Economics

Aoyama Gakuin University

4-4-25 Shibuya, Shibuya-ku

Tokyo, 150-8366 JAPAN

Abstract

The purpose of this paper is to mathematically formulate a new measure, the Cohort Cumulative Social Increase Ratio (CoCSIR), developed by Inoue (2002), and to consider its mathematical properties. The CoCSIR is a theoretical measure to estimate the long-term influence of the social increase on an area, and shows what proportion of the native place population of the area conforms to the social increase accumulated for the long term.

This paper presents the basic and practical formulas of the measure. The discussion in this paper suggests the versatility and availability of it. For example, the measure can be calculated for any size of area, and if using one of the practical formulas, requires only census data by age class for calculation. Therefore the measure has an advantage especially for analyzing the population of developing countries that are not providing small area statistics and/or exact vital statistics.

1. Introduction

As is well known, cohort analysis can clearly explain long-term patterns of regional population in many cases. For example, the effect of cohort size accounts to some extent for the so-called migration turnaround observed in major developed countries (Ishikawa 2001). Kawabe (1985) presented an idea to accumulate net migration rates after age 10-14 by cohort [1], and attempted to explain such long-term patterns in Japan. By improving the idea, Inoue (2002) developed a new measure, which was referred to as Cohort Cumulative Social Increase Ratio [2], and attempted to consider the internal migration between metropolitan and nonmetropolitan areas in Japan. With regard to the measure, Shimizu (2006, 2009) examined its availability by applied it to population data of Japan by Prefecture or municipality. However, there is no study that sufficiently considers mathematical formulation of the measure.

Thus the purpose of this paper is to mathematically formulate the Cohort Cumulative Social Increase Ratio (Hereinafter referred to as the CoCSIR) and to consider its mathematical properties. On such consideration, we need to realize that the social increase is quite the same as the amount of net migration.

---

[1] Regrettably values calculated using the idea have an undesirable property that the influence of net migration rates is overestimated in the case of small denominators of these rates.

[2] At first, it had been referred to as Cohort Cumulative Net Migration Ratio.

As mentioned later, the largest advantage of the CoCSIR is that all required for calculation is census data by age class[3]. It also is an advantage that the CoCSIR can be calculated for any size of area (e.g. country, region, Prefecture, Province, State, municipality and other smaller areas). Hence the CoCSIR can be applied also for developing countries that are not providing small area statistics and/or exact vital statistics (mortality statistics as well as migration statistics).

First, this paper derives the basic mathematical formula of the CoCSIR and considers its mathematical properties in Chapter 2, secondly, shows practical mathematical formulas of the CoCSIR in Chapter 3 by comparing four methods of calculating the social increase, and finally presents conclusion in Chapter 4.

---

[3] This paper assumes that we can obtain population data by 5-year age class every 5 years from census. If census is conducted every 10 years, we need to estimate intercensal population. This advantage is confined to one of the practical formulas mentioned later.

2. Basic Formula and Mathematical Properties of the CoCSIR

This chapter derives the basic mathematical formula of the CoCSIR from its definition in Section 1, and considers its mathematical properties in Section 2.

2-1. Basic Mathematical Formula

It is obvious that the size of a cohort in a certain area changes chiefly owing to the social increase, that is, migration from/to the area until the cohort reaches its old age. Although the size changes largely due to ancillary migration before age 10-14, it becomes most stable when the cohort reaches age 10-14, and thereafter changes basically due to voluntary migration. Therefore, it is much significant to consider such change of the cohort size from the viewpoint of the social increase using the size at age 10-14 as a criterion. Thus the CoCSIR is defined for a cohort in a certain area and obtained by dividing the cumulative social increase of the cohort after age 10-14 by the size at age 10-14.

The cohort size rapidly changes especially when changing from age 10-14 to age 15-19, since many members of the cohort leave their homes at that time. For this reason, most members of the cohort ought to recognize the area where they lived at age 10-14 as their native place. Thus this paper defines a member who lived in a certain area at age 10-14 as a "native" of the area [4], and refers the number of the natives as the "native place population" of the area. If using this word, the CoCSIR can

[4] Please note that the natives differ from the members born in the area.

be expressed also as a ratio of the cumulative social increase to the native place population.

As a result, the mathematical basic formula of the CoCSIR is derived from the above definition as follows: let CoCSIR of area *i*, of cohort of birth year *b* to *b*+4, and at age *a* to *a*+4 be ${}_iCoCSIR_{b,b+4}(a,\ a+4)$,

$$ {}_iCoCSIR_{b,b+4}(a, a+4) = \frac{\sum_{k=3}^{a/5} {}_i s_{b,b+4}(5k, 5k+4)}{{}_i p_{b,b+4}(10,14)} \quad \text{for } a = 15, 20, 25, \cdots \qquad (1) $$

Here ${}_is_{b,b+4}(5k,\ 5k+4)$ is the social increase of area *i*, of cohort of birth year *b* to *b*+4, and between age 5*k*−5 to 5*k*−1 and age 5*k* to 5*k*+4; ${}_ip_{b,b+4}(10,\ 14)$ is the population of area *i*, of cohort of birth year *b* to *b*+4, and at age 10 to 14. As mentioned above, the size of area *i* in equation (1) is optional, namely, the CoCSIR can be calculated for any size of area.

We should note that neither the numerator nor the denominator of the CoCSIR means the actual population surviving at age *a* to *a*+4. This is because both the population of the numerator and the denominator decrease[5] by death by age *a* to *a*+4. The CoCSIR is a theoretical measure to estimate the long-term influence of the social increase on a

[5] The decrease rate of the denominator is slightly higher than that of the numerator, because the population of the denominator is exposed to death risk for longer duration than that of the numerator. The two rates, however, are almost the same until old age. Therefore, the CoCSIR is almost equivalent to the ratio of the two actual population surviving at age *a* to *a*+4 if *a* is small.

certain area, and shows what proportion of the native place population of the area conforms to the social increase accumulated for the long term.

2-2. Mathematical Properties

In order to consider mathematical properties of the CoCSIR, we introduce the following four conditions by extremely simplifying rural-to-urban migration patterns typical in developed countries.

1) The whole area is divided into two areas: an urban area (i=1) and a rural area (i=2).
2) All the natives of area 1 move only within the area.
3) A part of the natives of area 2 move into area 1 from age 10-14 until age 20-24, and thereafter partly return to area 1 gradually until old age.
4) The mortality level is ignorable compared to that of mobility in the two areas.

Since the sum of the social increase equals to zero, under the above conditions, the following important equation holds for any a:

$$ {}_1CoCSIR_{b,b+4}(a,a+4)\cdot{}_1p_{b,b+4}(10,14)+{}_2CoCSIR_{b,b+4}(a,a+4)\cdot{}_2p_{b,b+4}(10,14)=0 $$

Obviously the following inequalities hold for any a:

$$ {}_1CoCSIR_{b,b+4}(a,a+4)>0, \quad {}_2CoCSIR_{b,b+4}(a,a+4)<0 $$

$$ {}_1CoCSIR_{b,b+4}(20,24)\geq {}_1CoCSIR_{b,b+4}(a,a+4) $$

$$ {}_2CoCSIR_{b,b+4}(20,24)\leq {}_2CoCSIR_{b,b+4}(a,a+4) $$

At that time, $-{}_2CoCSIR_{b,b+4}(a,a+4)$ shows what proportion of natives of area 2 live in the outside (i.e., area 1) at age a to a+4, and ${}_1CoCSIR_{b,b+4}(a,a+4)/\{1+{}_1CoCIR_{b,b+4}(a,a+4)\}$ shows [6] what proportion of residents in area 1 are the natives of area 2 at age a to a+4.

Moreover, if the CoCSIR is determined only by a and i and the native place population is determined only by i, namely, if ${}_iCoCSIR_{b,b+4}(a,a+4)={}_iCoCSIR_{b+5,b+9}(a,a+4)$ and ${}_i\,p_{b,b+4}(10,14)={}_ip_{b+5,b+9}(10,14)$ for any b, the migration system between areas 1 and 2 reaches a stationary state. In other words, the migration between areas 1 and 2 becomes always constant.

However, even if ${}_iCoCSIR_{b,b+4}(a,a+4)={}_iCoCSIR_{b+5,b+9}(a,a+4)$ for any b, the migration system becomes unstable if ${}_i\,p_{b,b+4}(10,14)$ changes according b. For example, if a baby boom occurs in year b to b+4, the native place population concerning cohort of birth year b to b+4 becomes much larger than that concerning any other cohort. Therefore, when the baby boomers grow to age 15-19 and age 20-24, rural-to-urban migration dominates, and thereafter return migration prevails. Such change of migration system is thought to be caused by the effect of cohort size.

Meanwhile, also in the case that ${}_i\,p_{b,b+4}(10,14)={}_ip_{b+5,b+9}(10,14)$ for any b and that ${}_iCoCSIR_{b,b+4}(a,a+4)$ changes according b, the migration system becomes unstable. For example, since an economic boom generally promotes migration, especially rural-to-urban migration, if the boom takes place during year b+19 to b+24 (i.e., when cohort of birth year b to

---

[6] The quotient shows a ratio of the social increase to the sum of itself and the native place population.

b+4 reaches age 15-19 and age 20-24), ${}_1CoCSIR_{b,b+4}(15,19)$ and ${}_1CoCSIR_{b,b+4}(20,24)$ are expected to show higher values than any other CoCSIR.

These considerations suggest that the CoCSIR has a potential to detect the influence of social phenomena such as baby and economic booms on the internal migration.

3. Practical Formulas of the CoCSIR

This chapter compares four methods of calculating or estimating the social increase ${}_{i}s_{b,b+4}(5k,\ 5k+4)$, and through the comparison, considers practical mathematical formulas of the CoCSIR.

3-1. Direct Calculation

The simplest and most exact method is to directly calculate the social increase ${}_{i}s_{b,b+4}(5k,\ 5k+4)$ from the numbers of in- and out-migrants. In fact, however, this method is scarcely adopted, since it requires long-term exact migration statistics by area and by age class and such statistics can hardly be provided even in developed countries.

3-2. Residual Method

If we can obtain long-term exact mortality statistics by area and by age class, we can estimate the social increase ${}_{i}s_{b,b+4}(5k,\ 5k+4)$ using the demographic equation. This method is referred to as the residual method, expressed as follows:

$${}_{i}s_{b,b+4}(5k,5k+4)={}_{i}p_{b,b+4}(5k,5k+4)-{}_{i}p_{b,b+4}(5k-5,5k-1)+{}_{i}d_{b,b+4}(5k,5k+4) \quad (2)$$

Here ${}_{i}p_{b,b+4}(5k,\ 5k+4)$ is the population of area i, of cohort of birth year b to b+4, and at age 5k to 5k+4; ${}_{i}d_{b,b+4}(5k,\ 5k+4)$ is the death population of area i, of cohort of year b to b+4, and between age 5k−5 to 5k−1 and age 5k to 5k+4.

This method is less exact than direct calculation because of the problem on inconsistency between census and mortality statistics. Moreover it is difficult to obtain long-term mortality statistics by area and by age class although it is not as difficult as the above migration statistics. For this reason, the social increase ${}_{i}s_{b,b+4}(5k,\ 5k+4)$ is usually estimated using either of the following two methods.

3-3. Life Table Survival Ratio Method

This method literally utilizes the survival ratio of life table data derived from exact mortality statistics. Modifying equation (2), we get equation (3) in which the survival ratio appears.

$$
\begin{aligned}
&{}_{i}s_{b,b+4}(5k,5k+4) \\
&\quad = {}_{i}p_{b,b+4}(5k,5k+4) - {}_{i}p_{b,b+4}(5k-5,\ 5k-1)\cdot\left\{1-\frac{{}_{i}d_{b,b+4}(5k,5k+4)}{{}_{i}p_{b,b+4}(5k-5,5k-1)}\right\}
\end{aligned} \qquad (3)
$$

In equation (3), variable $\left\{1-{}_{i}d_{b,b+4}(5k,5k+4)/{}_{i}p_{b,b+4}(5k-5,5k-1)\right\}$ corresponds to the survival ratio of area *i*. The method assumes that the ratio equals that of the whole area including area *i*. For example, it is assumed that the ratio of every prefecture is the same as that of the whole Japan. According to the assumption, equation (3) is transposed as follows:

$$
\begin{aligned}
&{}_{i}s_{b,b+4}(5k,5k+4) \\
&\quad = {}_{i}p_{b,b+4}(5k,5k+4) - {}_{i}p_{b,b+4}(5k-5,\ 5k-1)\cdot\left\{1-\frac{\sum_{i\in W}{}_{i}d_{b,b+4}(5k,5k+4)}{\sum_{i\in W}{}_{i}p_{b,b+4}(5k-5,5k-1)}\right\}
\end{aligned} \qquad (4)
$$

Here W is the set of all area numbers, in other words, means the whole area including all areas, e.g. the whole country if area i means Prefecture, Province, or State. If substituting survival ratios extracted from life table into the variable in parentheses in Equation (4), we can obtain the social increase.

Since the above assumption could induce an error, this method is less exact than the residual method. Nevertheless this method is much more practical than the above two methods because of requiring only life table data concerning the whole area W, and because of not requiring any vital statistics by area and by age class.

3-4. Intercensal Survival Ratio Method

The social increase estimated through this method is, strictly speaking, different in definition from that through the above three methods. The social increase through the above methods is based on all migration regarding area i, whereas that through this method is based only on the internal migration (i.e., the migration between area i and any other area inside the whole area W). In other words, the social increase through this method is not influenced by the external migration (i.e., the migration between area i and any other area outside the whole area W). For example, if the whole area is a country, the external migration means international migration[7]. It is both an advantage and a disadvantage that the social

[7] In some countries such as Japan, the social increase based on the international migration is extremely smaller than that based on the

increase is restricted to that based on the internal migration, whereas if we need to exclude the external migration from analysis, it is a large advantage.

In order to formulate the intercensal survival ratio method, we decompose the social increase into two parts as follows:

$$ {}_i s_{b,b+4}(5k,5k+4) = {}_i s^{\mathrm{INT}}{}_{b,b+4}(5k,5k+4) + {}_i s^{\mathrm{EXT}}{}_{b,b+4}(5k,5k+4) \qquad (5) $$

Here ${}_i s^{\mathrm{INT}}{}_{b,b+4}(5k,5k+4)$ and ${}_i s^{\mathrm{EXT}}{}_{b,b+4}(5k,5k+4)$ are components regarding the internal migration and the external migration, respectively.

By substituting equation (5) into equation (2) and modifying equation (2), we get,

$$ \begin{aligned} &{}_i s^{\mathrm{INT}}{}_{b,b+4}(5k,5k+4) \\ &\quad = {}_i p_{b,b+4}(5k,5k+4) - {}_i p_{b,b+4}(5k-5,5k-1) \cdot \left\{ 1 + \frac{{}_i s^{\mathrm{EXT}}{}_{b,b+4}(5k,5k+4) - {}_i d_{b,b+4}(5k,5k+4)}{{}_i p_{b,b+4}(5k-5,5k-1)} \right\} \end{aligned} \qquad (6) $$

Part of Equation (6), $\left\{1 - {}_i d_{b,b+4}(5k,5k+4) / {}_i p_{b,b+4}(5k-5,5k-1)\right\}$ corresponds to the survival ratio of area *i*. The method assumes that the ratio equals that of the whole area including area *i*. According to the assumption, equation (6) is transposed as follows:

$$ \begin{aligned} &{}_i s^{\mathrm{INT}}{}_{b,b+4}(5k,5k+4) \\ &\quad = {}_i p_{b,b+4}(5k,5k+4) - {}_i p_{b,b+4}(5k-5,5k-1) \cdot \left\{ 1 + \frac{\sum_{i \in W} {}_i s^{\mathrm{EXT}}{}_{b,b+4}(5k,5k+4) - \sum_{i \in W} {}_i d_{b,b+4}(5k,5k+4)}{\sum_{i \in W} {}_i p_{b,b+4}(5k-5,5k-1)} \right\} \end{aligned} \qquad (7) $$

---

internal migration, and if so, we can ignore the above difference in definition.

Meanwhile, if substituting equation (5) into equation (2) and summing up the both sides of equation (2) for all i $(i \in W)$, the following equation is derived because of $\sum_{i \in W} {}_i s^{\mathrm{INT}}{}_{b,b+4}(5k, 5k+4) = 0$:

$$\sum_{i \in W} {}_i s^{\mathrm{EXT}}{}_{b,b+4}(5k, 5k+4) = \sum_{i \in W} {}_i p_{b,b+4}(5k, 5k+4) - \sum_{i \in W} {}_i p_{b,b+4}(5k-5, 5k-1) + \sum_{i \in W} {}_i d_{b,b+4}(5k, 5k+4) \qquad (8)$$

Consequently, by substituting equation (8) into equation (7), the social increase estimated through the intercensal survival ratio method is written in equation (9).

$${}_i s^{\mathrm{INT}}{}_{b,b+4}(5k, 5k+4) = {}_i p_{b,b+4}(5k, 5k+4) - {}_i p_{b,b+4}(5k-5, 5k-1) \cdot \frac{\sum_{i \in W} {}_i p_{b,b+4}(5k, 5k+4)}{\sum_{i \in W} {}_i p_{b,b+4}(5k-5, 5k-1)} \qquad (9)$$

Equation (9) shows one of three kinds of the intercensal survival ratio method [8], i.e., the forward method such that all migration is assumed to occur immediately before the cohort reaches age 5*k* to 5*k*+4, in short, before the second census. The fraction in the right hand side means a cohort change ratio of the whole area.

According to equation (9), it is obvious that all required for calculation is census data, and that the social increase is restricted to that based on the internal migration. These facts show that the CoCSIR

---

[8] There are the backward and average methods besides the forward method.

defined through the intercensal survival ratio method is the most versatile, and especially in the case of focusing the internal migration, is reasonably available.

3-5. Practical Mathematical Formulas

Among the above four methods, only the direct calculation naturally does not need a practical mathematical formula. As regards the other three methods, if replacing ${}_i s_{b,b+4}(5k,\ 5k+4)$ in equation (1) by equations (2), (4), (9), the practical mathematical formulas are given by equations (10), (11), (12), respectively.

$$
{}_iCoCSIR_{b,b+4}(a,a+4) = \frac{\sum_{k=3}^{a/5}\left\{{}_i p_{b,b+4}(5k,5k+4) - {}_i p_{b,b+4}(5k-5,5k-1) + {}_i d_{b,b+4}(5k,5k+4)\right\}}{{}_i p_{b,b+4}(10,14)} \quad (10)
$$

for $a = 15, 20, 25, \cdots$

$$
{}_iCoCSIR_{b,b+4}(a,a+4) = \frac{\sum_{k=3}^{a/5}\left[{}_i p_{b,b+4}(5k,5k+4) - {}_i p_{b,b+4}(5k-5,5k-1)\cdot\left\{1-\frac{\sum_{i\in W} {}_i d_{b,b+4}(5k,5k+4)}{\sum_{i\in W} {}_i p_{b,b+4}(5k-5,5k-1)}\right\}\right]}{{}_i p_{b,b+4}(10,14)} \quad (11)
$$

for $a = 15, 20, 25, \cdots$

$$
{}_iCoCSIR_{b,b+4}(a,a+4) = \frac{\sum_{k=3}^{a/5}\left\{{}_i p_{b,b+4}(5k,5k+4) - {}_i p_{b,b+4}(5k-5,5k-1)\cdot\frac{\sum_{i\in W} {}_i p_{b,b+4}(5k,5k+4)}{\sum_{i\in W} {}_i p_{b,b+4}(5k-5,5k-1)}\right\}}{{}_i p_{b,b+4}(10,14)} \quad (12)
$$

for $a = 15, 20, 25, \cdots$

Equation (10), (11), (12) show the practical mathematical formulas of the CoCSIR in the case of using the residual method, the life table survival ratio method and the intercensal survival ratio method, respectively.

4. Conclusion

This paper derived the basic mathematical formula of the Cohort Cumulative Social Increase Ratio (CoCSIR) from its definition (equation (1)), considered its mathematical properties, and showed its practical mathematical formulas (equations (10), (11), (12)). The CoCSIR is a theoretical measure to estimate the long-term influence of the social increase on an area, and shows what proportion of the native place population of the area conforms to the social increase accumulated for the long term. The discussion in this paper suggests the versatility and availability of the CoCSIR as mentioned below.

1) The CoCSIR can be calculated for any size of area, and if using the practical formula written in equation (12), requires only census data by age class for calculation. Therefore the measure has an advantage especially for analyzing the population of developing countries that are not providing small area statistics and/or exact vital statistics.
2) The CoCSIR has a potential to detect the influence of social phenomena such as baby and economic booms on the internal migration.
3) The practical formula written in equation (12) is reasonably available especially in the case of focusing the internal migration, because of restricting the social increase to that of the internal migration.

Acknowledgements

This paper is a result of the study which received a financial support of the Japan Society for the Promotion of Science by grant-in-aid (grant number: 25370919). The outline of this paper was presented in IGU (International Geographical Union) 2013 Kyoto Regional Conference by the author. The author deeply appreciates Professors Shinichi Takahashi, Kiyosi Hirosima and Takashi Abe as well as his important colleague, Mr. Masato Shimizu for their valuable comments.

References

Inoue, T., 2002, "The Migration Turnarounds in Japan from the Demographic Viewpoint," Arai, Y., T. Kawaguchi and T. Inoue eds., *Migration in Japan: Life Course and Regionality*, Tokyo, Kokon-Shoin, pp.53-70.

Ishikawa, Y. ed., 2001, *Studies in the Migration Turnarounds*, Kyoto, Kyoto University Press.

Kawabe, H., 1985, "Some Characteristics of Internal Migration Observed from the Cohort-by-Cohort Analysis," *Journal of Population Problems* (*Jinko Mondai Kenkyu*), No. 175, pp.1-15.

Shimizu, M., 2006, "On the Quantum and Tempo of Cumulative Net Migration," *Journal of Population Problems* (*Jinko Mondai Kenkyu*), Vol. 62, No. 4, pp.41-60.

Shimizu, M., 2009, "The Cohort Cumulative Social Increase Ratio by Municipality: The Case of Nagano Prefecture," *Journal of Population Studies* (*Jinkogaku Kenkyu*), No. 44, pp.33-42.