On a Special Property of the Total Fertility Rate

Lu Lei *

21 April 2012

** The author is from China and was a former research fellow at the Institute of Population Research, Renmin University of China, Beijing, China. The author has a B.Sc. degree in computer science and applied mathematics, and an M.A. degree and a Ph.D. degree in demography.*

E-mail: lulei321@gmail.com

1. Introduction

In demographic analysis, the (period) total fertility rate (TFR) is one of the most widely used summary indicators for measuring the period fertility of a population. The total fertility rate is a period indicator, i.e. it is defined based on variables for the same period (e.g., a calendar year). However, when the demographic meaning of the total fertility rate is interpreted, demographers have to refer to a hypothetical (synthetic) birth cohort of women because of the nature of the definition of the total fertility rate. By definition, the total fertility rate is the sum of the age-specific fertility rates in a given year. The standard demographic interpretation of the total fertility rate is that it represents the average number of children that a hypothetical (synthetic) birth cohort of women would bear during their entire reproductive life span (normally between ages 15 and 50) if (i) all the women (of the birth cohort) survive through to the end of their reproductive life span, and (ii) they follow the age-specific fertility rates of the year in question.

So far, there has been a huge amount of demographic research on the total fertility rate. Now, researchers and demographers are very familiar with the advantages and disadvantages of the total fertility rate (e.g., Shryock, Siegel and Associates (1980); Ní Bhrolcháin (1992); Bogue, Arriaga, and Anderton (1993); Bongaarts and Feeney (1998); Ní Bhrolcháin (2007)). In demographic analysis, the total fertility rate is usually used for measuring and comparing (the level of) period fertility over time and/or across regions. The total fertility rate is also commonly used in population projections.

The present paper discusses a special property of the total fertility rate as a statistical indicator per se, i.e. the (quantitative) relationship between the total fertility rate of a total population and the total fertility rates of two sub-populations.

Suppose that we have a total population (denoted as P), which is divided into two sub-populations P_1 and P_2, and the two sub-populations satisfy the following conditions: $P_1 \cap P_2 = \varnothing$ (null) and $P_1 \cup P_2 = P$. Mathematically, it can be easily proved that the crude birth rate (CBR) of the total population for a given year is the weighted average of the crude birth rates of the two sub-populations for the same year, with the weights being the respective proportions of the two sub-populations in the total population. Now, the question is: Does the total fertility rate have a similar property?

Let TFR denote the total fertility rate of the total population (P), and TFR_1 and TFR_2 denote the total fertility rates of the two sub-populations (P_1 and P_2) respectively. In the present paper, we will look at the relationship between TFR and [TFR_1 and TFR_2].

Let $W(x)$ represent the number of women aged *x* at the midpoint of a given year *t* and $B(x)$ represent the number of live births delivered by women of age *x* in the same year, then the (period) age-specific fertility rate for age *x* of year *t* is defined as $f(x) = B(x)/W(x)$, $x = 15, 16, \ldots, 49$, and $\{f(x) \mid x = 15, 16, \ldots, 49\}$ is called the (period) age pattern of fertility. The corresponding total fertility rate is then defined as $TFR = \sum_{x=15}^{49} f(x)$. The total fertility rates of the two sub-populations are $TFR_1 = \sum_{x=15}^{49} f_1(x)$ and $TFR_2 = \sum_{x=15}^{49} f_2(x)$, where $f_1(x)$ and $f_2(x)$ are the (period) age-specific fertility rates of the two sub-populations of year *t*. Mathematically, the calculation of the total fertility rate is simple and straightforward.

2. A graphical analysis of the relationship between TFR and [TFR_1 and TFR_2]

Since the total fertility rate is the summation of the corresponding age-specific fertility rates, let's first look at the relationship between the age-specific fertility rates of the total population and the age-specific fertility rates of the two sub-populations. According to the definition of the age-specific fertility rate for age *x*, we can easily arrive at the following:

$$f(x) = u_1(x) \cdot f_1(x) + u_2(x) \cdot f_2(x), \tag{1}$$

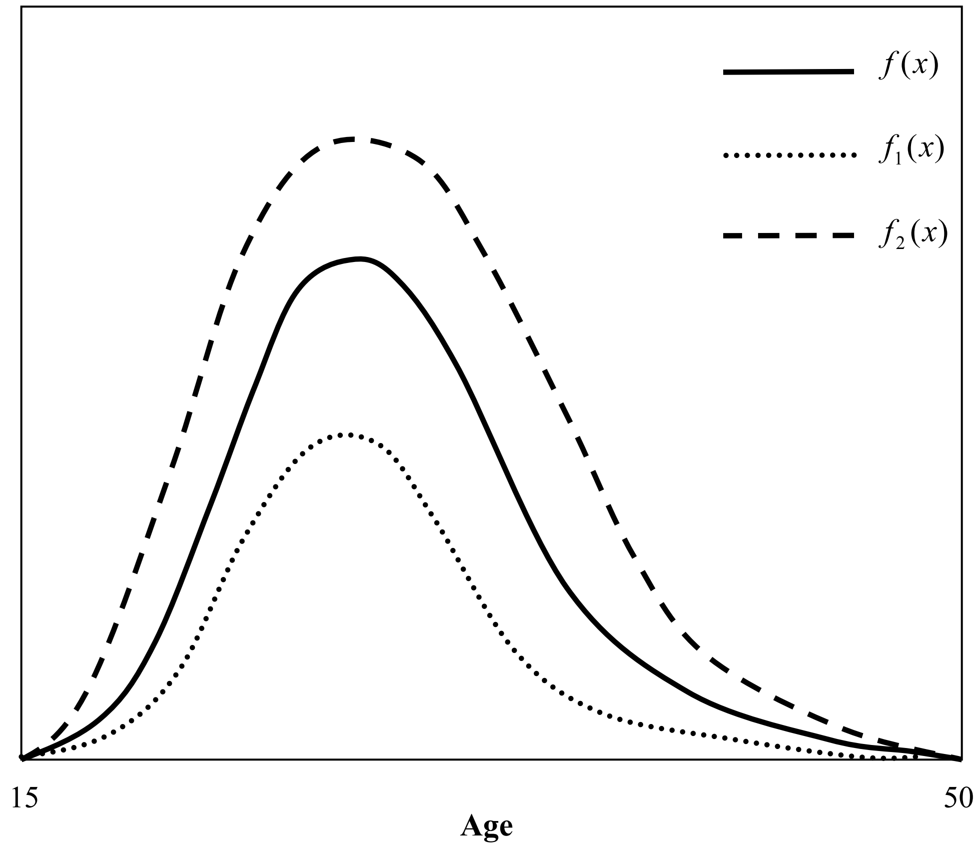
where $f(x)$ is the age-specific fertility rate of women aged *x* of the total population, $f_1(x)$ and $f_2(x)$ are the age-specific fertility rates of women aged *x* of the two sub-populations respectively, $u_1(x) = W_1(x)/W(x)$ and $u_2(x) = W_2(x)/W(x)$ are the proportions of women aged *x* of the two sub-populations in the women aged *x* of the total population. It is obvious that $u_1(x), u_2(x) > 0$ and $u_1(x) + u_2(x) = 1$, for all ages $x = 15, 16, \ldots, 49$. Therefore, equation (1) shows that for each age *x* ($x = 15, 16, \ldots, 49$), $f(x)$ is a weighted average of $f_1(x)$ and

$f_2(x)$. In other words, for each age x ($x = 15, 16, \ldots, 49$), $f(x)$ always falls between $f_1(x)$ and $f_2(x)$. Specially, if for an age x^*, there is $f_1(x^*) = f_2(x^*)$ (i.e. the two fertility curves $f_1(x)$ and $f_2(x)$ intersect at age x^*), then we have $f(x^*) = f_1(x^*) = f_2(x^*)$, i.e. curve $f(x)$ must pass through the point of intersection. Next, we discuss different situations based on the relative relations between the two fertility curves $f_1(x)$ and $f_2(x)$.

2.1 First type of relative relation between $f_1(x)$ and $f_2(x)$ (Figure 1)

In this situation, the age-specific fertility rate of all ages $(x = 15, 16, \ldots, 49)$ of the sub-population 1 is lower than the corresponding rate in sub-population 2, i.e. $f_1(x) < f_2(x)$, $x = 15, 16, \ldots, 49$. Therefore, we have $f_1(x) < f(x) < f_2(x)$, $x = 15, 16, \ldots, 49$.

Figure 1. First type of relative relation between $f_1(x)$ and $f_2(x)$



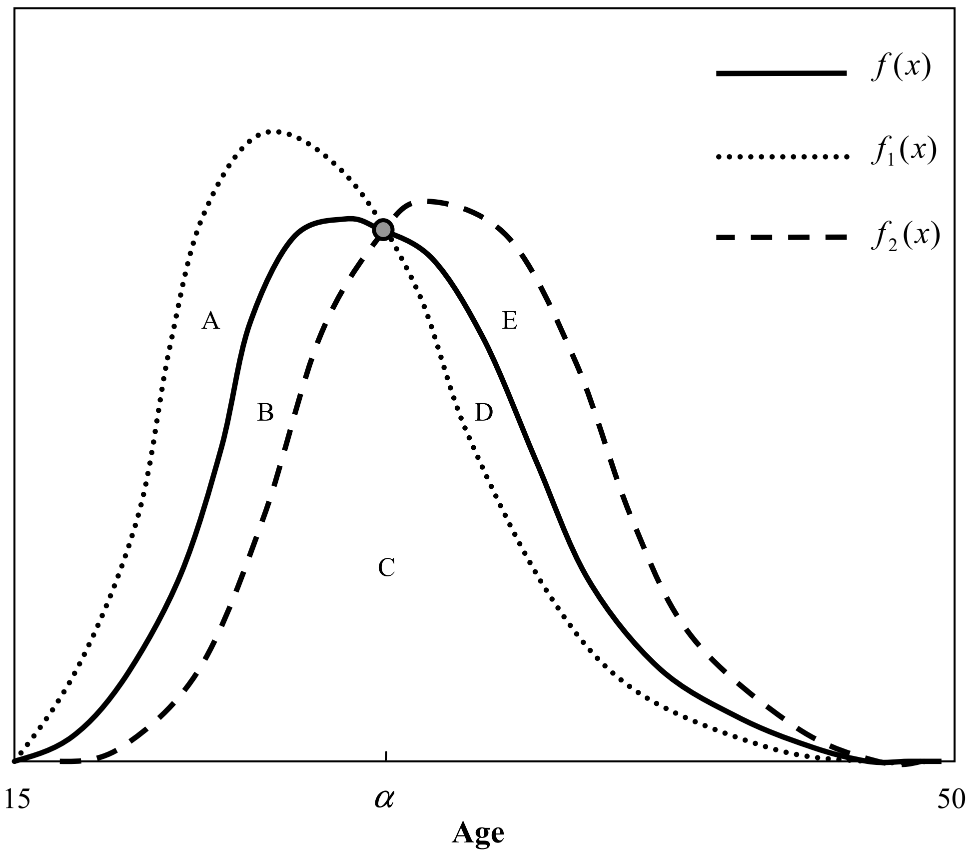
By taking summation with respect to age *x*, we obtain $\sum_{x=15}^{49} f_1(x) < \sum_{x=15}^{49} f(x) < \sum_{x=15}^{49} f_2(x)$, i.e. $TFR_1 < TFR < TFR_2$. It is obvious that, in this situation, the relationship $TFR_1 < TFR < TFR_2$

always holds regardless of the values of $u_1(x)$ or $u_2(x)$. However, it is not possible to know if TFR is a weighted average of [TFR_1 and TFR_2], as this will require knowledge of $u_1(x)$ or $u_2(x)$.

2.2 Second type of relative relation between $f_1(x)$ and $f_2(x)$ (Figure 2)

In this situation, the three fertility curves $f(x)$, $f_1(x)$ and $f_2(x)$ intersect at age α and form five regions, i.e. *A*, *B*, *C*, *D* and *E*. In age interval $(15, \alpha)$, we have $f_2(x) < f(x) < f_1(x)$, and in age interval $(\alpha, 50)$, we have $f_1(x) < f(x) < f_2(x)$. Let *A*, *B*, *C*, *D* and *E* represent the areas of the corresponding regions. When the two curves $f_1(x)$ and $f_2(x)$ are fixed, areas *A*, *B*, *D* and *E* change with $u_1(x)$ and $u_2(x)$, while area *C* remains constant. It is obvious that, no matter what values $u_1(x)$ and $u_2(x)$ take, $A+B$ and $D+E$ are constant. Let $X = A+B$ and $Y = D+E$, then *X* and *Y* are not affected by $u_1(x)$ and $u_2(x)$.

Figure 2. Second type of relative relation between $f_1(x)$ and $f_2(x)$



The following table provides a summary of the theoretical upper and lower limits of areas *B* and *D*, depending on $u_1(x)$ and $u_2(x)$.

Table 1. Theoretical upper and lower limits of areas *B* and *D*

Area		Age Interval $(15, \alpha)$	Age Interval $(\alpha, 50)$
B	Upper limit	***X*** (when $u_1(x) \to 1$ or $u_2(x) \to 0$)	
	Lower limit	**0** (when $u_1(x) \to 0$ or $u_2(x) \to 1$)	
D	Upper limit		***Y*** (when $u_1(x) \to 0$ or $u_2(x) \to 1$)
	Lower limit		**0** (when $u_1(x) \to 1$ or $u_2(x) \to 0$)

Since

$$TFR = B + C + D \tag{2}$$

$$TFR_1 = A + B + C = X + C \tag{3}$$

$$TFR_2 = C + D + E = Y + C \tag{4}$$

we have

$$TFR = TFR_1 + TFR_2 - (A + C + E) \tag{5}$$

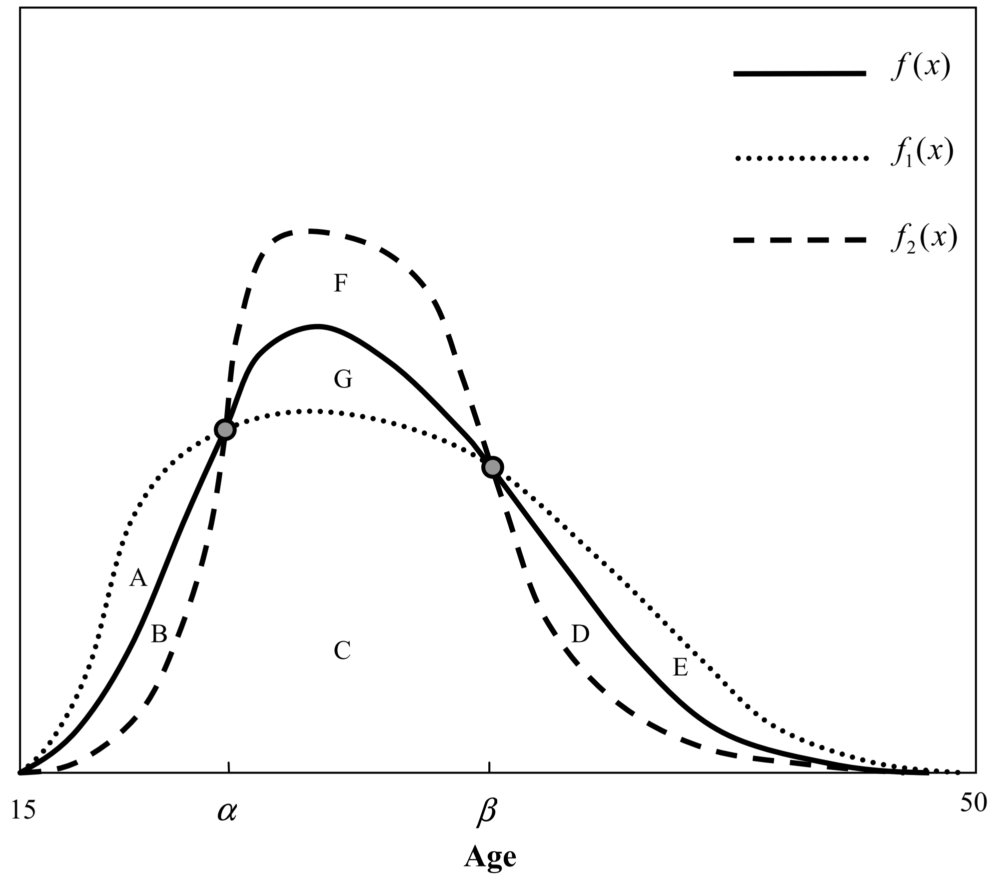
It is obvious that area *A* is affected by $u_1(x)$ and $u_2(x)$ $(15 < x < \alpha)$ and area *E* is affected by $u_1(x)$ and $u_2(x)$ $(\alpha < x < 50)$, while area *C* is not affected by $u_1(x)$ and $u_2(x)$ $(15 < x < 50)$. From equation (2) and Table 1, we know that the TFR has a theoretical upper limit of $X + Y + C$ and a theoretical lower limit of *C*. From equations (3) and (4), we can obtain $X + Y + C = TFR_1 + TFR_2 - C$. Thus, the theoretical upper limit for the TFR is $TFR_1 + TFR_2 - C$.

It is obvious that area *C* is completely determined by the relative relations of the two fertility curves $f_1(x)$ and $f_2(x)$. Specially, when $C = 0$ (i.e. the two fertility curves $f_1(x)$ and $f_2(x)$ do not have an overlap), the TFR has a theoretical upper limit of $TFR_1 + TFR_2$ and a theoretical lower limit of zero.

2.3 Third type of relative relation between $f_1(x)$ and $f_2(x)$ (Figure 3)

In this situation, the three fertility curves $f(x)$, $f_1(x)$ and $f_2(x)$ intersect at ages α and β and form seven regions, i.e. *A*, *B*, *C*, *D*, *E*, *F* and *G*. In age interval $(15, \alpha)$, we have $f_2(x) < f(x) < f_1(x)$; in age interval (α, β), we have $f_1(x) < f(x) < f_2(x)$; and in age interval $(\beta, 50)$, we have $f_2(x) < f(x) < f_1(x)$. Let *A*, *B*, *C*, *D*, *E*, *F* and *G* represent the areas of the corresponding regions. When the two curves f_x^1 and f_x^2 are fixed, areas *A*, *B*, *D*, *E*, *F* and *G* change with $u_1(x)$ and $u_2(x)$, while area *C* remains constant. It is obvious that, no matter what values $u_1(x)$ and $u_2(x)$ take, $A+B$, $D+E$ and $F+G$ are constant. Let $X = A+B$, $Y = D+E$ and $Z = F+G$, then *X*, *Y* and *Z* are not affected by $u_1(x)$ and $u_2(x)$.

Figure 3. Third type of relative relation between $f_1(x)$ and $f_2(x)$



The following table provides a summary of the theoretical upper and lower limits of areas *B*, *D* and *G*, depending on $u_1(x)$ and $u_2(x)$.

Table 2. Theoretical upper and lower limits of areas *B*, *D* and *G*

Area		Age Interval $(15, \alpha)$	Age Interval (α, β)	Age Interval $(\beta, 50)$
B	Upper limit	**X** (when $u_1(x) \to 1$ or $u_2(x) \to 0$)		
	Lower limit	**0** (when $u_1(x) \to 0$ or $u_2(x) \to 1$)		
D	Upper limit			**Y** (when $u_1(x) \to 1$ or $u_2(x) \to 0$)
	Lower limit			**0** (when $u_1(x) \to 0$ or $u_2(x) \to 1$)
G	Upper limit		**Z** (when $u_1(x) \to 0$ or $u_2(x) \to 1$)	
	Lower limit		**0** (when $u_1(x) \to 1$ or $u_2(x) \to 0$)	

Since

$$TFR = B + C + D + G \tag{6}$$

$$TFR_1 = A + B + C + D + E = X + Y + C \tag{7}$$

$$TFR_2 = C + F + G = Z + C \tag{8}$$

we have

$$TFR = TFR_1 + TFR_2 - (A + C + E + F) \tag{9}$$

It is obvious that area *A* is affected by $u_1(x)$ and $u_2(x)$ $(15 < x < \alpha)$, area *F* is affected by $u_1(x)$ and $u_2(x)$ $(\alpha < x < \beta)$, and area *E* is affected by $u_1(x)$ and $u_2(x)$ $(\beta < x < 50)$, while area *C* is not affected by $u_1(x)$ and $u_2(x)$ $(15 < x < 50)$. From equation (6) and Table 2, we know that the TFR has a theoretical upper limit of $X + Y + Z + C$ and a theoretical lower limit of *C*. From equations (7) and (8), we can obtain $X + Y + Z + C = TFR_1 + TFR_2 - C$. Thus, the theoretical upper limit for the TFR is $TFR_1 + TFR_2 - C$.

It is obvious that area *C* is completely determined by the relative relations of the two fertility curves $f_1(x)$ and $f_2(x)$. Specially, when *C* tends to 0, the TFR approaches the theoretical upper limit of $TFR_1 + TFR_2$.

From the above graphical analysis, we can conclude that the TFR has a theoretical upper limit of $TFR_1 + TFR_2$ and a theoretical lower limit of zero. Here, we have an interesting observation. Although the TFR is standardized for the age-sex structure of the total population, it is affected by the relative age distributions of the women of reproductive ages of the two sub-populations. In a special situation where $TFR_1 = TFR_2$ (i.e. the two sub-populations have the same total fertility rate), we cannot guarantee that $TFR = TFR_1 = TFR_2$ because the TFR is affected by $u_1(x)$ and $u_2(x)$, while TFR_1 and TFR_2 are not.

3. A mathematical analysis of the relationship between TFR and [TFR_1 and TFR_2]

From the above graphical analysis, we have noticed that the relationship between the TFR and the [TFR_1 and TFR_2] is complex. For given TFR_1 and TFR_2, the TFR may range between zero and $TFR_1 + TFR_2$, depending on $u_1(x)$ and $u_2(x)$. Now, let's look at the relationship between the TFR and the [TFR_1 and TFR_2] from a mathematical perspective.

Let $g_1(x) = f_1(x)/TFR_1$ and $g_2(x) = f_2(x)/TFR_2$, $x = 15, 16, \ldots, 49$, then it is obvious that $g_1(x), g_2(x) \geq 0$, $\sum_{x=15}^{49} g_1(x) = 1$ and $\sum_{x=15}^{49} g_2(x) = 1$. Sequences $\{g_1(x) \mid x = 15, 16, \ldots, 49\}$ and $\{g_2(x) \mid x = 15, 16, \ldots, 49\}$ are called the standardized (period) age patterns (schedules) of fertility of the two sub-populations. From the above definitions, we have $f_1(x) = TFR_1 \cdot g_1(x)$ and $f_2(x) = TFR_2 \cdot g_2(x)$. By taking summation on the two sides of equation (1) with respect to *x*, we obtain

$$TFR = TFR_1 \cdot \sum_{x=15}^{49} [u_1(x) \cdot g_1(x)] + TFR_2 \cdot \sum_{x=15}^{49} [u_2(x) \cdot g_2(x)] \qquad (10)$$

Equation (10) gives the general mathematical relationship between TFR and [TFR_1 and TFR_2], which provides the theoretical basis for the discussions below. Let $k_1 = \sum_{x=15}^{49}[u_1(x) \cdot g_1(x)]$ and $k_2 = \sum_{x=15}^{49}[u_2(x) \cdot g_2(x)]$. Since $0 < u_1(x) < 1$, and $0 < u_2(x) < 1$, it follows that $0 < k_1 < 1$ and $0 < k_2 < 1$. Equation (10) can be written as

$$TFR = k_1 \cdot TFR_1 + k_2 \cdot TFR_2 \tag{11}$$

Equation (11) shows that the relationship between TFR and [TFR_1 and TFR_2] is completely determined by the two coefficients k_1 and k_2. Suppose that $TFR_1 \le TFR_2$, then from equation (11), we have $(k_1 + k_2) \cdot TFR_1 \le TFR \le (k_1 + k_2) \cdot TFR_2$.

Let

$$u_1^{\max} = \max\{u_1(x) \mid x = 15,\ 16,\ ...,\ 49\} \tag{12}$$

$$u_1^{\min} = \min\{u_1(x) \mid x = 15,\ 16,\ ...,\ 49\} \tag{13}$$

and

$$u_2^{\max} = \max\{u_2(x) \mid x = 15, 16, ..., 49\} \tag{14}$$

$$u_2^{\min} = \min\{u_2(x) \mid x = 15,\ 16,\ ...,\ 49\} \tag{15}$$

Since $u_1(x) + u_2(x) = 1$, $x = 15, 16, \ldots, 49$, it can be proved that $u_1^{\max} + u_2^{\min} = 1$ and $u_1^{\min} + u_2^{\max} = 1$, and further $u_1^{\max} - u_1^{\min} = u_2^{\max} - u_2^{\min}$. In addition, it can also be proved that $u_1^{\min} \le k_1 \le u_1^{\max}$ and $u_2^{\min} \le k_2 \le u_2^{\max}$. Therefore, we have $u_1^{\min} + u_2^{\min} \le k_1 + k_2 \le u_1^{\max} + u_2^{\max}$.

According to the definition of the two coefficients k_1 and k_2, we have

$$k_1 + k_2 = 1 + \sum_{x=15}^{49}\{u_1(x) \cdot [g_1(x) - g_2(x)]\} = 1 + \sum_{x=15}^{49}\{u_2(x) \cdot [g_2(x) - g_1(x)]\} \tag{16}$$

Equation (16) shows that the two coefficients k_1 and k_2 do not necessarily constitute a pair of weights. Whether the two coefficients k_1 and k_2 form a pair of weights depends on the second term on the right-hand side of equation (16). Therefore, the TFR may not be a weighted average of TFR_1 and TFR_2 with respect to k_1 and k_2.

Let $k = \sum_{x=15}^{49}\{u_1(x)\cdot[g_1(x)-g_2(x)]\}$, then we have $k_1 + k_2 = 1 + k$. It is obvious that whether the two coefficients k_1 and k_2 constitute a pair of weights depends on the value of *k*. From the discussion above, we have $-1 < k < 1$. Specially, when $k = 0$, then the TFR is a weighted average of TFR_1 and TFR_2, with k_1 and k_2 being the two respective weights.

Next, we look at the relationship between TFR and [TFR_1 and TFR_2] under three special situations.

3.1 First situation

If the two sub-populations have the same standardized age pattern (schedule) of fertility (denoted as $\{g^*(x) \mid x = 15, 16, \ldots, 49\}$), i.e. $g_1(x) = g_2(x) = g^*(x)$, $x = 15, 16, \ldots, 49$, then from the definition of *k*, we have $k = 0$. In this situation, the TFR is a weighted average of TFR_1 and TFR_2, with $k_1 = \sum_{x=15}^{49}[u_1(x)\cdot g^*(x)]$ and $k_2 = \sum_{x=15}^{49}[u_2(x)\cdot g^*(x)]$ being the two respective weights. Obviously, in this situation, $u_1(x)$ and $u_2(x)$ do not affect the relationship, but affect the two weights.

3.2 Second situation

If the two sub-populations have the same total fertility rate (denoted as TFR^*), i.e. $TFR_1 = TFR_2 = TFR^*$, then from equation (11), we have $TFR = (k_1 + k_2)\cdot TFR^*$ $= (1+k)\cdot TFR^*$. Therefore, the relationship between the TFR and the TFR^* is dependent on *k*. Specifically, when $k > 0$, we have $TFR > TFR^*$; when $k = 0$, we have $TFR = TFR^*$; and when $k < 0$, we have $TFR < TFR^*$. Here, we see that even if the two sub-populations have the same total fertility rate, the TFR (of the total population) may not necessarily be equal to the total fertility rate of the sub-populations.

3.3 Third situation

Suppose that $u_1(x)$ can be represented by the following n^{th}-degree polynomial of *x*:

$$u_1(x) = \lambda_0 + \lambda_1 \cdot x + \lambda_2 \cdot x^2 + \cdots + \lambda_n \cdot x^n = \sum_{i=0}^{n} (\lambda_i \cdot x^i) \tag{17}$$

where *n* is a non-negative integer. Then from the definition of *k*, we have

$$k = \sum_{x=15}^{49} \left\{ [g_1(x) - g_2(x)] \cdot \sum_{i=0}^{n} (\lambda_i \cdot x^i) \right\} = \sum_{i=0}^{n} \left\{ \lambda_i \cdot \sum_{x=15}^{49} \left[x^i \cdot [g_1(x) - g_2(x)] \right] \right\} \tag{18}$$

Define the r^{th} absolute moment (about zero or origin) of $g_1(x)$ and $g_2(x)$ as follows:

$$\hat{M}_r\langle g_1 \rangle = \sum_{x=15}^{49} [x^r \cdot g_1(x)] \text{ and } \hat{M}_r\langle g_2 \rangle = \sum_{x=15}^{49} [x^r \cdot g_2(x)] \tag{19}$$

Then equation (18) can be rewritten as $k = \sum_{i=0}^{n} [\lambda_i \cdot (\hat{M}_i\langle g_1 \rangle - \hat{M}_i\langle g_2 \rangle)]$. It is obvious that $\hat{M}_0\langle g_1 \rangle = 1$ and $\hat{M}_0\langle g_2 \rangle = 1$. Define the mean age of $g_1(x)$ and $g_2(x)$ as follows:

$$\mu\langle g_1 \rangle = \sum_{x=15}^{49} [x \cdot g_1(x)] \text{ and } \mu\langle g_2 \rangle = \sum_{x=15}^{49} [x \cdot g_2(x)] \tag{20}$$

then we have $\mu\langle g_1 \rangle = \hat{M}_1\langle g_1 \rangle$ and $\mu\langle g_2 \rangle = \hat{M}_1\langle g_2 \rangle$. Define the variance of $g_1(x)$ and $g_2(x)$ as follows:

$$v\langle g_1 \rangle = \sum_{x=15}^{49} [(x - \mu\langle g_1 \rangle)^2 \cdot g_1(x)] \text{ and } v\langle g_2 \rangle = \sum_{x=15}^{49} [(x - \mu\langle g_2 \rangle)^2 \cdot g_2(x)] \tag{21}$$

then we have $v\langle g_1 \rangle = \hat{M}_2\langle g_1 \rangle - (\mu\langle g_1 \rangle)^2$ and $v\langle g_2 \rangle = \hat{M}_2\langle g_2 \rangle - (\mu\langle g_2 \rangle)^2$.

Next, we look at three special cases.

Case (1): $u_1(x)$ is a constant with respect to age *x*. This is equivalent to taking $n = 0$ in equation (18), i.e. $u_1(x) = \lambda_0$, $x = 15, 16, \ldots, 49$. In this case we have

$$k = \lambda_0 \cdot \sum_{x=15}^{49} [g_1(x) - g_2(x)] = \lambda_0 \cdot \left[\sum_{x=15}^{49} g_1(x) - \sum_{x=15}^{49} g_2(x) \right] = 0 \qquad (22)$$

Therefore, in this case, the TFR is a weighted average of TFR_1 and TFR_2, with $k_1 = \lambda_0$ and $k_2 = 1 - \lambda_0$ being the two respective weights.

Case (2): $u_1(x)$ is a linear function of age *x*. This is equivalent to taking $n = 1$ in equation (18), i.e. $u_1(x) = \lambda_0 + \lambda_1 \cdot x$, where $\lambda_1 \neq 0$. In this case we have

$$k = \lambda_1 \cdot (\mu\langle g_1 \rangle - \mu\langle g_2 \rangle) \qquad (23)$$

If $\mu\langle g_1 \rangle = \mu\langle g_2 \rangle$ (denoted as μ), then we have $k = 0$. In this case, the TFR is a weighted average of TFR_1 and TFR_2, with $k_1 = \lambda_0 + \lambda_1 \cdot \mu$ and $k_2 = 1 - (\lambda_0 + \lambda_1 \cdot \mu)$ being the two weights respectively. It is interesting to note that in this situation, the value of *k* is affected by the relative positions (as measured by the mean age of fertility) of the two standardized fertility curves $g_1(x)$ and $g_2(x)$, but not affected by the shapes of the two curves.

Case (3): $u_1(x)$ is a quadratic function of age *x*. This is equivalent to taking $n = 2$ in equation (18), i.e. $u_1(x) = \lambda_0 + \lambda_1 \cdot x + \lambda_2 \cdot x^2$, where $\lambda_2 \neq 0$. In this case we have

$$k = \lambda_1 \cdot (\mu\langle g_1 \rangle - \mu\langle g_2 \rangle) + \lambda_2 \cdot \left\{ [(\mu\langle g_1 \rangle)^2 + v\langle g_1 \rangle] - [(\mu\langle g_2 \rangle)^2 + v\langle g_2 \rangle] \right\} \qquad (24)$$

Equation (24) shows that, in this situation, the value of *k* is not only affected by the relative positions (as measured by the mean age of fertility) of the two standardized fertility curves $g_1(x)$ and $g_2(x)$, but also affected by the shapes of the two curves (as measured by the variance). It is obvious that if the two standardized fertility curves $g_1(x)$ and $g_2(x)$ have the same mean (denoted as μ) and the same variance (denoted as *v*), then $k = 0$. The TFR is therefore a weighted average of TFR_1 and TFR_2, with $k_1 = \lambda_0 + \lambda_1 \cdot \mu + \lambda_2 \cdot (\mu^2 + v)$ and

$k_2 = 1-[\lambda_0 + \lambda_1 \cdot \mu + \lambda_2 \cdot (\mu^2 + v)]$ being the two weights respectively.

If $\mu\langle g_1 \rangle = \mu\langle g_2 \rangle$, then from equation (24), we have $k = \lambda_2 \cdot (v\langle g_1 \rangle - v\langle g_2 \rangle)$. In this case, $k = 0$ is equivalent to $v\langle g_1 \rangle = v\langle g_2 \rangle$. If $\mu\langle g_1 \rangle \neq \mu\langle g_2 \rangle$, then $k = 0$ is equivalent to $\frac{[(\mu\langle g_1 \rangle)^2 + v\langle g_1 \rangle] - [(\mu\langle g_2 \rangle)^2 + v\langle g_2 \rangle]}{\mu\langle g_1 \rangle - \mu\langle g_2 \rangle} = -\frac{\lambda_1}{\lambda_2}$.

Next, we establish the general criteria regarding the relationship between TFR and [TFR_1 and TFR_2]. From equation (11), we can obtain the following:

(a) $TFR_1 < TFR < TFR_2$ is equivalent to

$$\frac{TFR_1}{TFR_2} < \min\left(\frac{k_2}{1-k_1}, \frac{1-k_2}{k_1}\right) \text{ or } \frac{TFR_2}{TFR_1} > \max\left(\frac{k_1}{1-k_2}, \frac{1-k_1}{k_2}\right) \tag{25}$$

(b) $TFR = TFR_1 = TFR_2$ is equivalent to

$$TFR_1 = TFR_2 \text{ and } k_1 + k_2 = 1 \tag{26}$$

(c) $TFR > TFR_1$ and $TFR > TFR_2$ is equivalent to

$$\frac{1-k_2}{k_1} < \frac{TFR_1}{TFR_2} < \frac{k_2}{1-k_1} \tag{27}$$

(d) $TFR < TFR_1$ and $TFR < TFR_2$ is equivalent to

$$\frac{k_2}{1-k_1} < \frac{TFR_1}{TFR_2} < \frac{1-k_2}{k_1} \tag{28}$$

From (c) and (d) above, we notice that, theoretically speaking, it is possible that the TFR is larger or smaller than both TFR_1 and TFR_2. This seems to be a "paradox". Regarding Simpson's paradox in demography, Cohen (1986) has made a comprehensive analysis of the commonly used crude rates in demography, e.g. the crude death rate (CDR).

Here, we need to add one point regarding the relationship between TFR and [TFR_1 and

TFR_2]. For convenience of discussion, we suppose that $TFR_1 < TFR_2$.

When $k_1 + k_2 > 1$, from equation (11), we have $TFR > (k_1 + k_2) \cdot TFR_1 > TFR_1$. But it cannot be guaranteed that there is also $TFR > TFR_2$. For example, if $TFR_1 = 2$, $TFR_2 = 5$, $k_1 = 0.8$ and $k_2 = 0.4$, then we have $k_1 + k_2 = 1.2 > 1$ and $TFR = 3.6$. In this case, the TFR falls between TFR_1 and TFR_2.

When $k_1 + k_2 < 1$, from equation (11), we have $TFR < (k_1 + k_2) \cdot TFR_2 < TFR_2$. But it cannot be guaranteed that there is also $TFR < TFR_1$. For example, if $TFR_1 = 2$, $TFR_2 = 5$, $k_1 = 0.3$ and $k_2 = 0.5$, then we have $k_1 + k_2 = 0.8 < 1$ and $TFR = 3.1$. In this case, the TFR falls between TFR_1 and TFR_2.

From the discussions above, we know that even if coefficients k_1 and k_2 do not constitute a pair of weights (i.e. $k_1 + k_2 \neq 1$), it is still possible that TFR falls between TFR_1 and TFR_2. It is obvious that when $k_1 + k_2 = 1$, TFR always falls between TFR_1 and TFR_2.

4. A real case with regard to the relationship between TFR and [TFR_1 and TFR_2]

Now, let's look at a real case vis-à-vis the "paradox". The data used are from the 1% Population Sampling Survey of China 1987, which was conducted by the National Bureau of Statistics of China. The total population here is the population of Shanghai Municipality and the two sub-populations are urban Shanghai and rural Shanghai. Table 3 shows the numerical values of the total fertility rates of Shanghai in 1986.

Table 3. The total fertility rate of Shanghai, China, 1986

Shanghai (TFR)	Urban Shanghai (TFR_1)	Rural Shanghai (TFR_2)
1.371	1.255	1.356

It is obvious from Table 3 that the 1986 total fertility rate of Shanghai (as a whole) was larger than the total fertility rates of both urban and rural shanghai. Table 4 shows the corresponding numerical values of the related factors.

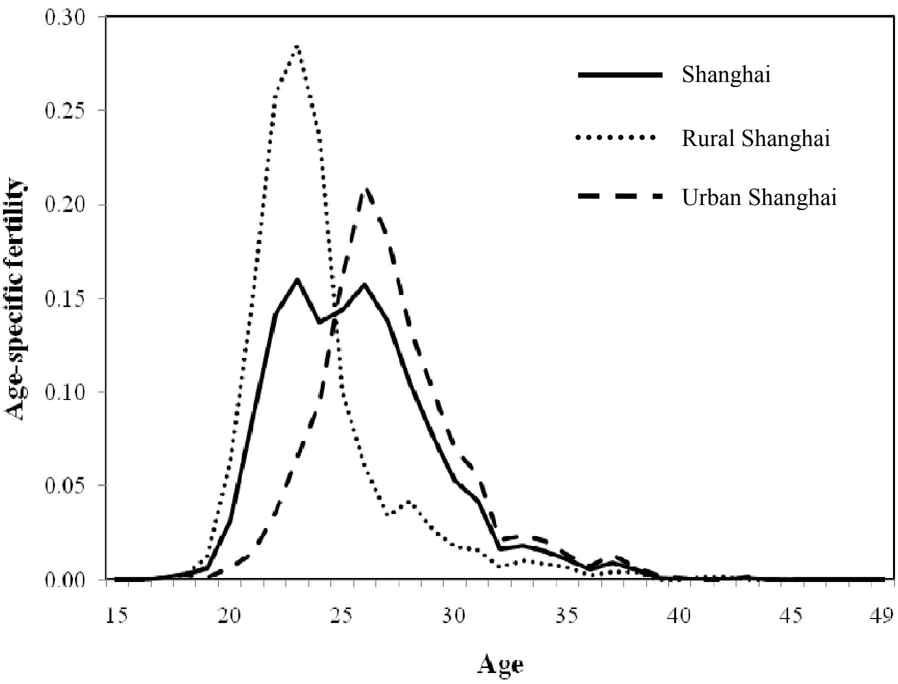
Table 4. Numerical values of related factors, Shanghai, China, 1986

k_1	k_2	$k_1 + k_2$	k	$\frac{TFR_1}{TFR_2}$	$\frac{1-k_2}{k_1}$	$\frac{k_2}{1-k_1}$
0.657	0.403	1.060	0.060	0.926	0.908	1.176

This is a case that meets criterion (c) of section 3, i.e. $\frac{1-k_2}{k_1} < \frac{TFR_1}{TFR_2} < \frac{k_2}{1-k_1}$.

Now, let's take a look at the age patterns of fertility of Shanghai in 1986 (Figure 4). The two (period) fertility curves of urban Shanghai and rural Shanghai intersect at around age 25. An interesting phenomenon from Figure 4 is that the age patterns of fertility of urban and rural Shanghai are both single-peaked curves, while the age pattern of fertility of Shanghai (as a whole) has two main peaks, which correspond to the rural peak (at age 23) and the urban peak (at age 26) respectively. The fertility curve of Shanghai (as a whole) has a local valley at ages 24 and 25.

Figure 4. The age patterns of fertility of Shanghai, China, 1986



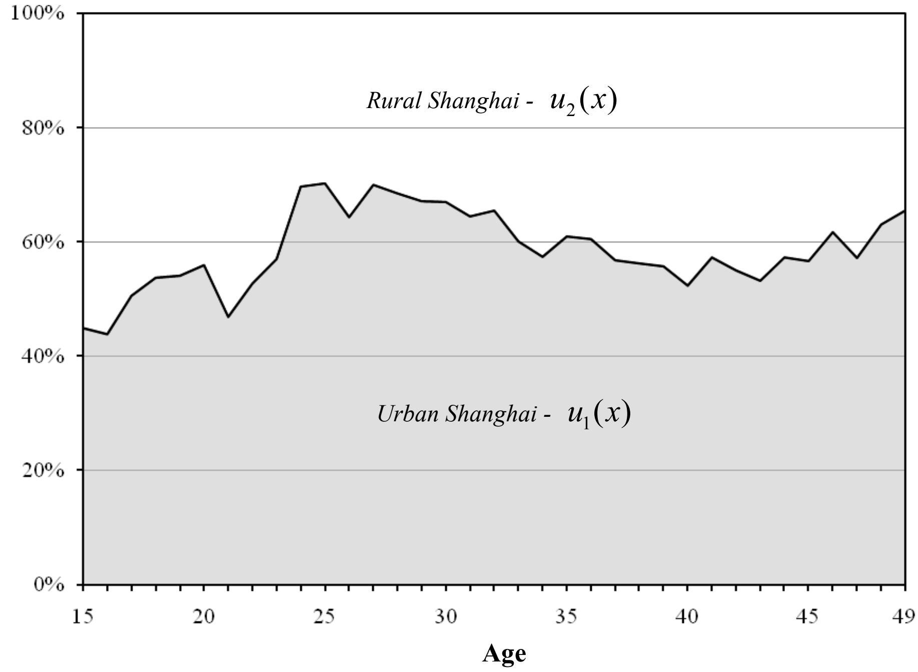
The means and the standard deviations of the age at childbearing are as follows:

Table 5. Mean age at childbearing and standard deviation, 1986

	Shanghai	Urban Shanghai	Rural Shanghai
Mean age of childbearing	26.20	27.65	24.35
Standard deviation	3.66	3.37	3.19

Figure 5 shows the relative age structures of urban and rural Shanghai in 1986.

Figure 5. Relative age structures of urban and rural Shanghai, China, 1986



5. Summary

In demography, the total fertility rate (TFR) is a very important measure of period fertility. In this paper, we analyzed the relationship between the total fertility rate of a total population and the total fertility rates of two sub-populations. The analysis shows that the relationship is complex and a “paradox” may occur in the relationship. Although we have only discussed a scenario of dividing a total population into two sub-populations, the results of this paper should apply to situations with multiple sub-populations.

References

Bogue, Donald J.; Arriaga, Eduardo E.; and Anderton, Douglas L. (1993). Reading in Population Research methodology, Volume 3, Fertility Research. Chicago: United Nations Population Fund and Social Development Center.

Bongaarts, John; and Feeney, Griffith. (1998). "On the Quantum and Tempo of Fertility". *Population and Development Review*, Vol. 24, No. 2.

Cohen, Joel E. (1986). "An Uncertainty Principle in Demography and the Unisex Issue." *The American Statistician*, Vol. 40, No. 1.

Ní Bhrolcháin, Máire. (1992). "Period Paramount? A Critique of the Cohort Approach to Fertility". *Population and Development Review*, Vol. 18, No. 4.

Ní Bhrolcháin, Máire. (2007). "Why we measure period fertility". *Applications & Policy Working Paper A07/04*, Southampton Statistical Sciences Research Institute, University of Southampton.

Shryock, Henry S; Jacob S. Siegel and Associates. (1980). *The Methods and Materials of Demography*. US Bureau of the Census.