4

Applications of Decomposition Methods

4.1 Introduction

Many of the contributions which improved decomposition methods were developed through the study of particular areas of demography such as mortality, fertility and population growth. This chapter examines some of the studies that contributed to these decompositions. Several illustrative examples are presented here, which later are compared with direct vs. compositional decomposition presented in Part III.

As shown in the previous chapter, some authors analyze mortality by decomposing the difference between crude death rates into several components. Others instead look at the trends of mortality by examining the change in life expectancy and its decomposition over time or between populations or sexes. Among those that have contributed to these studies are Keyfitz (1985), Mitra (1978), a United Nations report (1982), Pollard (1982; 1988), Andreev (1982), Arriaga (1984), Pressat (1985), Vaupel (1986), Goldman and Lord (1986) and Valkovics (in Wunsch (2002)).

Demographic variables such as the total fertility rate, the crude birth rate, the mean parity and other measures of fertility have also been decomposed. The different components of these decompositions explain the change in fertility patterns. Among the authors that have studied the elements of the fertility change are Anderson (1975), Bongaarts (1978), Bongaarts and Potter (1983), Hobcraft and Little (1984), Cutright and Smith (1988), Nathanson and Kim (1989), Pullum et al. (1989), Zeng et al. (1991), Bongaarts (1993), Gertler and Molyneaux (1994), Smith et al. (1996) and Kohler and Ortega (2002).

Among the authors who have carried out research on the population growth rate and its components are Keyfitz (1985), Bennett and Horiuchi (1981), Preston and Coale (1982), Arthur and Vaupel (1984), Kim (1986), Horiuchi and Preston (1988), Kephart (1988), Caselli and Vallin (1990), Horiuchi (1995), Otani (1997) and Vaupel and Canudas Romo (2000).

Many applications of the decomposition methods have influenced the development of new methods of separating components. In the following sections we look at three areas of demography where these extensions of decomposition techniques have been developed.

4.2 Decomposition of Mortality Measures

Measures of the mortality level of a population such as the crude death rate, average age at death, or other age-aggregated indexes are affected by changes in the age structure of the population. In order to avoid confounding results due to the age structure when comparing over time, studies have analyzed mortality by applying decompositions of the kind shown in Chapter 3. These methods are generally characterized as the analysis of the difference of two age-aggregated indexes into several components.

Other attempts to describe the changes in mortality are, for example, found in the work of Keyfitz (1989). He applies a three-dimensional chart to account for past birth and cohort survivorship variation. Some studies have examined the changes in mortality based on parametric models. For example, Horiuchi and Wilmoth (1998) present a period and cohort decomposition of the gamma-Gompertz-Makeham models to demonstrate the force of mortality. Here the interest is focused on changes over time of demographic variables.

Other studies have analyzed the trends of mortality by examining the differences in life expectancy over time. For example, a United Nations report (1982) uses Kitagawa's approach to study the contribution by age group of the difference in life expectancy. The following section describes some of the studies of changes over time in life expectancy.

4.2.1 Decomposing Life Expectancy

This section starts with the definition of life expectancy and other lifetable functions. Then, four methods are shown together with their corresponding contributors and formulations. Next Arriaga's decomposition of the change in life expectancy and Pollard's cause-specific decomposition are applied.

Letting the radix of the lifetable be equal to one, $\ell(0,t) = 1$, life expectancy at birth at time t can be expressed as

$$e^o(0,t) = \int_0^\omega \ell(a,t)da, \tag{4.1}$$

where $\ell(a,t)$ is the lifetable probability at time t of surviving from birth to age a, and ω is the highest age attained. The lifetable probability of surviving from birth to age a is also a function of the sum of the force of mortality function between those ages

$$\ell(a,t) = e^{-\int_0^a \mu(x,t)dx}, \tag{4.2}$$

as before, $\mu(a,t)$ is the force of mortality, or hazard, at age a and time t. Because the radix of the lifetable is equal to one the integrated force of mortality, or cumulative hazard, $-\int_0^a \mu(x,t)dx$ acts as the growth rate of $\ell(0,t)$ from age 0 to age a.

Another term should be defined is $\rho(a,t)$, the rate of progress in reducing mortality rates, $\rho(a,t) = -\acute{\mu}(a,t)$, where the acute accent is the relative derivative as specified in Section 2.4.

The person-years lived between two ages x_a and x_{a+1} can be calculated as

$$L(x_a, x_{a+1}, t) = \int_{x_a}^{x_{a+1}} \ell(a,t)da, \tag{4.3}$$

and the person-years above a certain age a as

$$T(a,t) = \sum_{x_a=a}^{\omega} L(x_a, x_{a+1}, t). \tag{4.4}$$

The work of Keyfitz (1985) on the change over time in life expectancy corresponds to equal proportional changes in mortality at all ages. Keyfitz assumes constant improvements in mortality at all ages, $\rho(a,t) = \rho(t)$ for all a. The relative change in life expectancy is then

$$\acute{e}^o(0,t) = \rho(t)\mathcal{H}(t), \tag{4.5}$$

where $\mathcal{H}(t)$ denotes the entropy of the survival function

$$\mathcal{H}(t) = -\frac{\int_0^\omega \ell(a,t)\ln\left[\ell(a,t)\right]da}{\int_0^\omega \ell(a,t)da}. \tag{4.6}$$

The entropy function $\mathcal{H}(t)$ is a measure of the concavity of the survival curve. For example, if in a population deaths are concentrated in a narrow range of ages then $\mathcal{H}(t) = 0$. If age-specific death rates are constant at all ages then $\mathcal{H}(t) = 1$. Equation (4.5) shows how the logarithm of the survival curve $\ln\left[\ell(a,t)\right]$ contains information about the effects of changes in death rates on life expectancy.

Mitra (1978), Goldman and Lord (1986), Vaupel (1986) and Vaupel and Canudas Romo (2003) further developed this approach. A generalization of (4.5) by Vaupel and Canudas Romo (2003) is presented in Part III of this book.

Pollard (1982) expands the exponential function in life expectancy through the powers of the force of mortality in equation (4.2): $\int_0^a \mu(x,t)dx$. Let the times be t and $t+h$. Pollard partitions the changes in the expectation of life as a result of mortality changes by age contributions as

$$\begin{aligned} e^o(0,t+h) - e^o(0,t) &= \int_0^\omega \left[\ell(a,t+h) - \ell(a,t)\right]da \\ &= \int_0^{x_1} \left[\mu(a,t) - \mu(a,t+h)\right]w(a)da + \\ &\quad ... + \int_{x_n}^{\omega} \left[\mu(a,t) - \mu(a,t+h)\right]w(a)da \end{aligned} \tag{4.7}$$

where 0, x_1,..., x_n, and ω are ages and the weights are defined as

$$w(a) = \frac{1}{2}\left[\ell\left(a,t\right)e^o\left(a,t+h\right) + \ell\left(a,t+h\right)e^o\left(a,t\right)\right]. \tag{4.8}$$

Pollard's equation (4.7) is useful to analyze the contribution of mortality improvements in the various age groups.

In the 1980s, Andreev (1982) (cited by Andreev et al. (2002)), Arriaga (1984) and Pressat (1985) independently developed a decomposition of the difference in life expectancies. We present them here as Arriaga's decomposition. The change in life expectancy is divided into direct, indirect and combined effects, by age categories as follows:

$$\begin{aligned} e^o(0,t+h) - e^o(0,t) &= [\Delta_D(0,x_1) + \Delta_I(0,x_1)] + ... \\ &\quad + [\Delta_D(x_n,\omega) + \Delta_I(x_n,\omega)], \end{aligned} \tag{4.9}$$

where the direct component from age x_a to x_{a+1} is defined as

$$\Delta_D(x_a,x_{a+1}) = \frac{\ell(x_a,t)}{\ell(0,t)} \left(\frac{L(x_a,x_{a+1},t+h)}{\ell(x_a,t+h)} - \frac{L(x_a,x_{a+1},t)}{\ell(x_a,t)} \right), \tag{4.10}$$

and the indirect, which also includes an interaction component, is

$$\Delta_I(x_a,x_{a+1}) = \frac{T(x_{a+1},t+h)}{\ell(0,t)} \left(\frac{\ell(x_a,t)}{\ell(x_a,t+h)} - \frac{\ell(x_{a+1},t)}{\ell(x_{a+1},t+h)} \right). \tag{4.11}$$

The direct effect (4.10) in the ages x_a to x_{a+1} is the change in life expectancy as a consequence of change in the number of years lived within that particular age group. The additional number of years is due to an alteration of mortality between ages x_a and x_{a+1}. Changes in mortality in each age group produce a different number of survivors at the end of the age interval. The indirect effect, here shown together with the interaction component, describes the changes in life expectancy due to the exposure of the additional survivors to new mortality conditions. Pollard (1988) shows that his formulation is analogous to Arriaga's.

Table 4.1 shows the application of equation (4.9) to the annual change in life expectancy at birth for the Swedish population from 1900 to 1905, 1950 to 1955 and 1995 to 2000. Three

Table 4.1: Life expectancy at birth, $e^o(0,t)$, and Arriaga's decomposition of the annual change over time from 1900 to 1905, from 1950 to 1955 and from 1995 to 2000, in Sweden.

t	1900	1950	1995
$e^o(0,t)$	52.239	71.130	78.784
$e^o(0,t+5)$	54.527	72.586	79.740
$\Delta e^o(0,t)$	0.458	0.291	0.191
Δ_D	0.008	0.010	0.007
Δ_I	0.450	0.282	0.184
$\Delta e^o(0,t) = \Delta_D + \Delta_I$	0.458	0.291	0.191

Source: Author's calculations based on equation (4.9). Lifetable data is derived from the Human Mortality Database (2002). Lifetable values from the years 1900 and 1905, 1950 and 1955, 1995 and 2000, were used to obtain the results.

six-year periods were chosen for the calculations, at the beginning of the twentieth century, 1900 - 1905, in the middle of the century, 1950 - 1955, and at the end of the century, 1995 - 2000. During all periods, gains in life expectancy (measured in years) were achieved. The greatest increases occurred in 1900-1905, followed by 1950-1955 while the most recent years registered the smallest increase. The main contributor to the changes is the component that combines the indirect and the interaction effects with minor participation of the direct effect.

Silber (1992) adapts the Gini concentration ratio to study age distribution of the causes of death. Three components are obtained: the inequality of age at death in each cause of death (the "within causes of death" element), the inequality of the average ages of the different causes of death (the "between causes of death" element), and lastly the overlapping between ages at death in the different causes of death. Similarly, Valkovics (in Wunsch (2002), pp. 79-94) suggests a decomposition of the male-female difference in life expectancies into the different causes of death categories. One last decomposition by cause of death is Pollard's proposed equation

$$e^o(0,t+h) - e^o(0,t) = \sum_{i=1}^{n} \int_0^{\omega} [\mu_i(a,t) - \mu_i(a,t+h)]\, w(a)da, \tag{4.12}$$

where $\mu_i(a,t)$ is the force of mortality at age a, time t and cause of death i, and $w(a)$ is as defined in equation (4.8).

Table 4.2 shows the decomposition obtained by combining the two formulas proposed by Pollard (4.7) and (4.12), by age and causes of death respectively. The decomposition is applied to Japan for the years 1980 and 1990. We divided the grouped data into five age groups

Table 4.2: Age and cause of death decomposition for the annual change over time in life expectancy, in percentages, for Japan for the period 1980-1990, following Pollard's decomposition methods.

Causes of death	$\Delta e^o(0,1980) = 2.880$					
$\mu_i(a,t)$	0 - 1	1 - 9	10 - 50	50 - 69	70 +	*All ages*
Malignant neoplasm	-0.001	0.014	0.077	0.053	0.005	0.148
Heart disease	0.002	0.000	0.033	0.119	0.270	0.424
Cerebrovascular disease	0.001	0.001	0.072	0.348	0.800	1.222
Infectious diseases	0.025	0.010	0.024	0.027	-0.036	0.050
Violent deaths	0.011	0.052	0.092	0.013	0.021	0.190
Stomach, liver and kidney disorders	-0.002	0.001	0.055	0.045	0.042	0.142
Senility without psychosis	0.000	0.000	0.000	0.003	0.260	0.263
Other causes	0.181	0.011	0.051	0.053	0.074	0.370
All causes of death	0.218	0.090	0.406	0.660	1.435	2.809

Source: Author's calculations based on equation (4.7) and (4.12), based on the Berkeley Mortality Database (2001).

and eight causes of death. The cause of death which contributes the most to the increase in life expectancy is cerebrovascular disease while infectious diseases are the most influential inhibitors in life expectancy in the oldest age group. The increase in life expectancy was mainly due to changes in the life expectancy of the population aged seventy and above whereas children between 1 and 9 experienced the smallest change.

The contributions to the increase in the Japanese life expectancy by the different causes of death can be divided into four types. First we have death decreases concentrated mainly in the very young ages in the category of "other causes of death"; then are the improvements located in the age group 1 to 50 years, as seen in violent deaths; another category is concentrated in the age group 10 years and above, as seen in malignant neoplasm, and stomach, liver and kidney disorders. Another category consists of the causes of death which show great improvements in old age groups, among which are heart disease, cerebrovascular disease, and senility without psychosis.

4.2.2 Conclusion

Life expectancy is among the most frequently used demographic measures. Still there is a ongoing debate regarding misleading indications of life expectancy (Bongaarts and Feeney (2002) and Vaupel (2002)). In the absence of a better measure of the current conditions of mortality, life expectancy will continue to play a central role in demography.

The above section has covered some of the studies which have contributed to the decomposition in life expectancy. All these methodologies share the common interest of separating the changes in life expectancy into components of mortality variation, age at death distribution and cause of death distribution.

Criteria about the choice of methodology for our analysis and the priority of a particular decomposition method over the others are provided in Part IV. We also compare the methods of this section with direct vs. compositional decomposition.

4.3 Decomposition of Fertility Measures

This section includes decompositions associated with fertility measures. We start with a discussion on the crude birth rate and how to decompose this measure's change over time. Then we look at changes over time in the total fertility rate.

Crude rates are influenced by the age structure of a population, due to the pervasive association between age and demographic rates.

In Chapter 3, on *Standardization and Decomposition Techniques,* the crude death rate was decomposed in order to separate the confounding effects of the age structure. The decompositions of mortality measures shown in Chapter 3 have also been applied to the study of components of fertility change. Nathanson and Kim (1989), for example, used the decomposition proposed by Kim and Strobino (1984) to study adolescent fertility. Smith et al. (1996) used Das Gupta's (1978) separation techniques to study the fertility of unmarried women.

The crude birth rate (CBR), as expressed in equation (2.8), has the same problem as the crude death rate of confounding effects due to the age structure. To take account of this the CBR has also been decomposed. Zeng et al. (1991) explained the increase in the Chinese CBR between 1984 and 1987 by applying a decomposition. The CBR, as defined in equation (2.8), was modified by Zeng and colleagues to include only the fertility of married women (because in China most births occur in marriage). Let the proportion of married women at age a be denoted as $\pi_{ma}(t)$, and the marital fertility rate as $b_{ma}(t)$, the CBR of married women $CBR_m(t)$ is

$$CBR_m(t) = \frac{\sum_{a=0}^{\omega} b_{ma}(t)\pi_{ma}(t)N_{f,a}(t)}{\sum_{a=0}^{\omega} N_a(t)} = \sum_{a=0}^{\omega} b_{ma}(t)\pi_{ma}(t)\pi_{fa}(t), \tag{4.13}$$

where $\pi_{fa}(t)$ is the proportion of women at age a in the total population, $\pi_{fa}(t) = \frac{N_{fa}(t)}{\sum_{a=0}^{\omega} N_a(t)}$. By using a variant of *polar* decomposition shown in Section 3.4.4, Zeng and colleagues decompose the change in $CBR_m(t)$ into three main effects. The change in $CBR_m(t)$ is due to changes in marital fertility rates Δb_{ma}, changes in the proportion of married women $\Delta\pi_{ma}$, and changes in the proportion of women in the total population $\Delta\pi_{fa}$,

$$CBR_m(t+h) - CBR_m(t) \approx \sum_{a=0}^{\omega} \Delta\left[b_{ma}\right]\pi_{ma}(t+h)\pi_{fa}(t+h)$$

$$+\sum_{a=0}^{\omega} b_{ma}(t+h)\Delta\left[\pi_{ma}\right]\pi_{fa}(t+h) + \sum_{a=0}^{\omega} b_{ma}(t+h)\pi_{ma}(t+h)\Delta\left[\pi_{fa}\right]. \tag{4.14}$$

equation (4.14) is an approximation because the terms of interactions between the differences Δb_{ma}, $\Delta\pi_{ma}$ and $\Delta\pi_{fa}$ are ignored. The conclusion of the research by Zeng et al. (1991) is that the age structure, $\Delta\pi_{fa}$, and the declining age at marriage, seen in $\Delta\pi_{ma}$, are the two main contributors to the increase in the crude birth rate in China. Table 4.3 presents the CBR of married women and the decomposition of the annual change over time, as suggested by Zeng and colleagues, for Denmark, the Netherlands and Sweden from 1992 to 1997.

Table 4.3: Crude birth rate of married women and Zeng's decomposition of the annual change over time for Denmark, the Netherlands and Sweden from 1992 to 1997.

	Denmark	Netherlands	Sweden
$CBR_m(1992)$	0.702	1.134	0.716
$CBR_m(1997)$	0.702	0.997	0.470
ΔCBR_m	0.000	-0.027	-0.049
Δb_{ma}	0.013	0.015	-0.020
$\Delta\pi_{ma}$	-0.013	-0.037	-0.024
$\Delta\pi_{fa}$	0.000	-0.005	0.000
$\Delta CBR_m = \Delta b_{ma} + \Delta\pi_{ma} + \Delta\pi_{fa}$	0.000	-0.027	-0.044

Source: Author's calculations from formula (4.14), based on Eurostat (2000).

The first three rows are for the observed CBR of married women in the selected countries, and the observed change in CBR during this period. The lower part of the table is for the estimated components of the decomposition. Table 4.4 below contains similar calculations but these are applied to the CBR of unmarried women.

In a country like China where most births occur in marriage, equation (4.14) provides good estimates. For European countries it is more appropriate to add a CBR for married women and another for unmarried women in an equation similar to (4.13). This is possible because the total number of births is equal to the births of married and unmarried women. As such, the crude birth rate is equal to the CBR of married women plus the CBR of unmarried women as follows:

$$\begin{aligned} CBR(t) &= \frac{B(t)}{N(t)} = \frac{B_m(t)}{N(t)} + \frac{B_u(t)}{N(t)} \\ &= CBR_m(t) + CBR_u(t), \end{aligned} \tag{4.15}$$

where $B(t)$, as before, denotes the births, and $B_m(t)$ and $B_u(t)$ are births of married and unmarried women, respectively. The expression for the CBR of unmarried women, $CBR_u(t)$, is

$$CBR_u(t) = \frac{\sum_{a=0}^{\omega} b_{ua}(t)\pi_{ua}(t)N_{fa}(t)}{\sum_{a=0}^{\omega} N_a(t)} = \sum_{a=0}^{\omega} b_{ua}(t)\pi_{ua}(t)\pi_{fa}(t), \tag{4.16}$$

where $\pi_{ua}(t)$ corresponds to the proportion of unmarried women at age a.

Table 4.4 presents the CBR for unmarried women and the decomposition of the annual change over time, as suggested by Zeng and colleagues, which is also applied to data from Denmark, the Netherlands and Sweden from 1992 to 1997.

Table 4.4: Crude birth rate of unmarried women and Zeng's decomposition of the annual change over time for Denmark, the Netherlands and Sweden from 1992 to 1997.

	Denmark	Netherlands	Sweden
$CBR_u(1992)$	0.608	0.161	0.701
$CBR_u(1997)$	0.578	0.236	0.553
ΔCBR_u	-0.006	0.015	-0.030
Δb_{ua}	-0.008	0.012	-0.041
$\Delta \pi_{ua}$	0.006	0.007	0.012
$\Delta \pi_{fa}$	-0.003	-0.003	-0.002
$\Delta CBR_u = \Delta b_{ua} + \Delta \pi_{ua} + \Delta \pi_{fa}$	-0.005	0.016	-0.031

Source: Author's calculations from formula (4.14), based on Eurostat (2000).

The decomposition by Zeng et al. (1991) is an approximation. Adding the results of Tables 4.3 and 4.4 gives us the observed components (first three rows of the tables) and the estimated components of the decomposition in the CBR for both married and unmarried women. For example, the total crude birth rate arising from the addition of the results in Tables 4.3 and 4.4 is compiled for the Netherlands in Table 10.4. The Netherlands is the only country which experiences an increase in the CBR for unmarried women. Sweden had the largest decrease for both groups, and the Netherlands had an important decline in marital CBR. Some general dynamics in the fertility change of these European countries should be noted. The proportion of married women decreases while the proportion of unmarried women increases. The Netherlands experienced an increase in both marital and non-marital age-specific fertility rates. In Sweden, both groups, married and unmarried, experienced a decline in fertility. For the three countries and both populations of women, the change in the age structure, $\pi_{fa}(t)$, is responsible for only a minor contribution to the total change, in contrast to Zeng's results (1991), for China, where it was the most important factor.

Anderson (1975) and Anderson et al. (1977) also present a variant of *polar* decomposition for age-specific birth rates. The age-specific fertility rates are separated into three components for each education class in terms of fertility of married women $b_{ema}(t)$, the proportion of married women $\pi_{ema}(t)$, and the proportion of women of level of education e, $\pi_{ea}(t)$,

$$b_{ma}(t) = \sum_{e \in E} b_{ema}(t)\, \pi_{ema}(t)\, \pi_{ea}(t), \tag{4.17}$$

where $e \in E$ indicates that all the education classes in E are considered. The change over time is then calculated for one of the components while the other terms remain constant as in time t,

$$b_{ma}(t+h) - b_{ma}(t) = \sum_{e \in E} \Delta\left[b_{ema}\right] \pi_{ema}(t)\, \pi_{ea}(t)$$

$$+ \sum_{e \in E} b_{ema}(t)\, \Delta\left[\pi_{ema}\right] \pi_{ea}(t) + \sum_{e \in E} b_{ema}(t)\, \pi_{ema}(t)\, \Delta\left[\pi_{ea}\right] + \epsilon_{ma}, \tag{4.18}$$

where the term ϵ_{ma} is simply the residual between observed and estimated change in $b_{ma}(t)$. The result of all the differences in age-specific fertility rates is summarized in the total fertility rate (TFR).

Analogous to the study of change over time in life expectancy, most of the research has been concentrated on the change in the total fertility rate. Similar to life expectancy, the TFR is an indicator of a hypothetical cohort, which does not depend on the structure of the population. The TFR is the average number of children per woman if the age-specific birth rates remain constant over a long period,

$$TFR = \int_{\alpha}^{\beta} b(a,t)da, \tag{4.19}$$

where α and β are the lower and upper limits of childbearing. In the following section we look at the changes over time of the TFR.

4.3.1 Decomposing the Total Fertility Rate

Bongaarts (1978) presented a decomposition called "the proximate determinants of fertility" of the total fertility rate. The TFR at time t is expressed as an identity involving indexes

$$TFR(t) = C_m(t)C_c(t)C_a(t)C_i(t)TF(t), \tag{4.20}$$

where C_m, C_c, C_a, and C_i represent indexes of the proportion of married women, contraception use, induced abortion and postpartum infecundability. The last component is the total fertility TF, which comprises natural fecundability, spontaneous intrauterine mortality and permanent sterility. Hobcraft and Little (1984) tested Bongaarts' decomposition (4.20) using an individual-level analysis, and arrived at an additive decomposition.

Any change in a population's level of fertility is necessarily caused by a change in one or more of the proximate determinants. Bongaarts and Potter (1983) present a decomposition of the change over time of Bongaarts' proposal (4.20). The trend in the TFR is expressed as

$$\frac{TFR(t+h)}{TFR(t)} = \frac{C_m(t+h)C_c(t+h)C_a(t+h)C_i(t+h)TF(t+h)}{C_m(t)C_c(t)C_a(t)C_i(t)TF(t)}, \tag{4.21}$$

which is then observed as a relative difference expressed in additive terms as

$$T\grave{F}R = \grave{C}_m + \grave{C}_c + \grave{C}_a + \grave{C}_i + T\grave{F} + \epsilon, \tag{4.22}$$

where the grave accent denotes the relative difference as defined in Section 2.4. For example, the proportional change of the TFR between year $t+h$ and year t with respect to year t is $T\grave{F}R(t) = \frac{TFR(t+h)-TFR(t)}{TFR(t)}$. On the right hand side of the equation the other components are similarly defined.

Cutright and Smith (1988) also present a decomposition of ratios for the comparison of different subpopulations. Their suggestion for a decomposition is to look at the logarithm of the ratios in equation (4.21).

Table 4.5 presents the results of applying Bongaarts and Potter's equation (4.22) in the analysis of the change over the period 1975 to 1993/94 of the proximate determinants of the TFR, for Bangladesh. Table 4.5 differs from the rest of the tables in the book in terms of

Table 4.5: Total fertility rate as a product of five factors, $TFR(t)$, and Bongaarts and Potter's decomposition of the annual relative change over the period 1975-1993/94 for Bangladesh.

t	1975	1993/94	*Relative change*
TFR	7.360	4.503	-0.388
C_m	0.850	0.761	-0.105
C_c	0.937	0.610	-0.349
C_a	0.604	0.653	0.081
C_i	1.000	0.971	-0.029
TF	15.300	15.300	0.000
ϵ			0.013
$T\grave{F}R = \grave{C}_m + \grave{C}_c + \grave{C}_a + \grave{C}_i + T\grave{F} + \epsilon$			-0.388

Source: Author's calculations from formula (4.22), based on the data from Islam et al. (1998).

information about the proximate determinants of fertility in 1975 and 1993/94 in the first two columns of values. In the third column are the contributions of each proximate to the total change over the period, $\grave{C}_m$, $\grave{C}_c$, $\grave{C}_a$, $\grave{C}_i$, $T\grave{F}$ and ϵ. The change in use of contraception, $\grave{C}_c$, is the major component of the change in Bangladesh's TFR. A second component which constitutes an important contribution is the change in the proportion of married women. In contrast to the decline in TFR is the increase in the proportion of induced abortions.

Based on the proximate determinants in equation (4.20) Gertler and Molyneaux (1994) present a simplified decomposition, which they applied to study the reduction in Indonesian fertility. The TFR is decomposed into three main components: indexes of marriage and contraception, and a residual for the remaining factors, $\epsilon(t)$,

$$TFR(t) = C_m(t)C_c(t)\epsilon(t). \tag{4.23}$$

To study the change over time Gertler and Molyneaux suggest calculating the derivatives of the logarithms of equation (4.23) with respect to time,

$$\frac{\partial \ln(TFR)}{\partial t} = \frac{\partial \ln(C_m)}{\partial t} + \frac{\partial \ln(C_c)}{\partial t} + \frac{\partial \ln(\epsilon)}{\partial t}. \tag{4.24}$$

Equation (4.24) can be further developed to obtain a general formula. When studying change over time of multiplicative models, similar to expressions (4.20) and (4.23), it is generally advisable to look at the relative change. The change is then decomposed into the contribution of components that explain the total change. In general terms, let a variable be the product of components $v = v_1v_2...v_n$. The intensity of this variable is

$$\acute{v} = \acute{v}_1 + \acute{v}_2 + ... + \acute{v}_n. \tag{4.25}$$

As this decomposition is a particular case of the proposed direct vs. compositional decomposition presented in Part III, more details are provided there.

Applying equation (4.25) to (4.20) results in the decomposition of the change over time of the TFR as the addition of relative changes of the indexes,

$$T\acute{F}R = \acute{C}_m + \acute{C}_c + \acute{C}_a + \acute{C}_i + T\acute{F}. \tag{4.26}$$

Table 4.6 presents the results of equation (4.26) by analyzing the change over time of the proximate determinants of the TFR, as given in equation (4.20), for Bangladesh in the period 1975-1993/94. Table 4.6 has the same structure as Table 4.5. Table 4.6 also leads to the conclusion that the major component of change for the TFR in Bangladesh is the contraception index. It is important to distinguish the differences between Tables 4.5 and 4.6, since there is no residual term in the last table. A residual term is the impossibility of allocating change to the components of the decomposition. Equation (4.26) contains an inherent advantage over (4.22). The relative change is further studied in Part III as an element of direct vs. compositional decomposition.

Another study which included a decomposition of fertility measures is the study by Pullum et al. (1989). These authors looked at the cohort TFR^c, or mean parity of a cohort that has completed fertility, defined as

$$TFR^c = \sum_i i\pi_i^c, \tag{4.27}$$

Table 4.6: Total fertility rate as a product of five factors, $TFR(t)$, and Gertler and Molyneaux's decomposition of the relative change over time in the period 1975-1993/94 for Bangladesh.

t	1975	1993/94	*Relative change*
TFR	7.360	4.503	-2.729
C_m	0.850	0.761	-0.614
C_c	0.937	0.610	-2.385
C_a	0.604	0.653	0.433
C_i	1.000	0.971	-0.163
TF	15.300	15.300	0.000
$\acute{TFR} = \acute{C}_m + \acute{C}_c + \acute{C}_a + \acute{C}_i + \acute{TF}$			-2.729

Source: Author's calculations using formula (4.26) described in Chapter 9, based on the data from Islam et al. (1998).

where π_i^c is the proportion of women at parity i. The mean parity was decomposed by Pullum and colleagues by using the Delta method which is examined in Chapter 5 entitled *Alternative Decomposition Methods*.

Kohler and Ortega (2002) looked at the age-specific and parity-specific fertility rates and proposed a decomposition of these measures into two components. The first component is a level term for each parity during the period, which increases or decreases childbearing at all ages. The second component is a schedule of fertility that determines the age pattern of each parity.

4.3.2 Conclusion

Analogous to the study of mortality in which life expectancy is frequently used, the total fertility rate is the fertility measure that is used more often. The question of how to properly adjust the TFR has sparked recent debate. Among researchers involved in the development of the TFR are Bongaarts and Feeney (1998), Kim and Schoen (2000), Van Imhoff and Keilman (2000) and Kohler and Ortega (2002). Future decompositions of the TFR are expected to account for each of the components of this adjusted TFR.

This section contains some of the works that have contributed to the decomposition methods of the TFR and also of the crude birth rate. The decomposition methods of the CBR have a common interest in separating the changes into components of fertility and the composition of the population. For the TFR we have mainly concentrated on those studies that cover the decomposition of relative changes in the proximate determinants of fertility.

In Part IV these methods are compared with direct vs. compositional decomposition.

4.4 Decomposition of Growth Measures

One of the basic aspects in demography is the balancing equation of population change. There are only four possible ways of entering or leaving a population. The population at time t, $N(t)$, is the result of the population at the beginning, time 0, $N(0)$, plus the births occurring between these years, $B(t)$, minus the deaths, $D(t)$, and plus the net migration (immigration minus emigration), $I(t)$. The changes in the size of the population must be attributable to the magnitude of these flows, expressed as follows:

$$N(t) = N(0) + B(t) - D(t) + I(t). \tag{4.28}$$

Expressing equation (4.28) in term of rates yields

$$r(t) = CBR(t) - CDR(t) + i(t). \tag{4.29}$$

The crude growth rate, $r(t) = \acute{N}(t) = \frac{dN(t)/dt}{N(t)}$, is equal to the crude birth rate, $CBR(t)$, minus the crude death rate, $CDR(t)$, plus the crude net migration, $i(t)$. In other words, the change in population size is decomposed into change due to the flows of births, deaths and net migration.

Table 4.7 presents the figures obtained by applying (4.29) in the period 1985-1995 for France, Japan and the USA. Because the change occurs continuously from the initial point of 1985 to the final moment of 1995, we choose to study the change at the mid-point or mid-year, which is 1990. In all the tables of this section the values of the averages at all these years are displayed together with the observed change at the mid-year. Japan has the lowest relative

Table 4.7: The population growth rate, $r(t)$, rates in percentages, and the balancing equation in 1985-1995 for France, Japan and the USA.

	France	*Japan*	*USA*
$N(1990)$	56,570,545	121,907,530	248,940,972
$N(1985)$	55,157,223	119,740,316	236,872,781
$N(1995)$	58,020,081	124,113,968	261,624,012
$r(1990)$	0.506	0.359	0.994
$CBR(1990)$	1.333	1.050	1.589
$CDR(1990)$	0.938	0.664	0.871
$i(1990)$	0.111	-0.027	0.275
$r(1990) = CBR - CDR + i$	0.506	0.359	0.994

Source: Author's calculations from formula (4.29) described in Chapter 9. Data from Human Mortality Database (2002).

increase in population, $r(1990)$, while the highest increase is seen in the USA. Among the components of the balancing equation, the CBR accounts for the greatest rate in all countries. Again the country with the lowest level, of CBR is Japan and USA has the highest CBR level. The second most important component is the crude death rate which tempers the increase in the growth rate, due to the negative sign in (4.29). Here again, we find that Japan is the country with the lowest level of mortality while France has the highest. The growth rate of the Japanese population is also reduced due to emigration. Both the French and the American populations are increasing due to a positive net migration rate.

New research has been developed in the area of decomposition of the growth rate, the start of which can be attributed to Bennett and Horiuchi (1981). Instead of looking at the population growth rate, $r(t)$, they propose looking at the age-specific growth rates, $r(a,t) = \frac{\partial N(a,t)}{N(a,t)\partial t}$. The age-specific growth rate at age a and time t, $r(a,t)$, is the relative change over time in the population size. Another way of calculating this relative change over time is to measure the relative change over age and the relative change over cohort together. The age-specific growth rate can then be expressed as

$$r(a,t) = \nu(a,t) - \mu(a,t), \tag{4.30}$$

where $\nu(a,t)$ is the relative change in the rate as age increases $\nu(a,t) = -\frac{\partial N(a,t)}{N(a,t)\partial a}$, and $\mu(a,t)$ is, as above, the force of mortality or change over cohort in the population size, $\mu(a,t) = -\frac{\partial N(a+c,t+c)}{N(a,t)\partial c}$. In this equation and others of this section, net migration is omitted. It could easily be included as a way of exiting (or entering) the population at any age, that is, treated in the same way as the force of mortality.

Two articles look at this new idea. Preston and Coale (1982) generalized Lotka's (1939) fundamental equations for any population by using equation (4.30). Their base identity is the relation between the populations at age 0 and a, both at time t, so that

$$\begin{aligned} N(a,t) &= N(0,t)e^{-\int_0^a r(x,t)dx}e^{-\int_0^a \mu(x,t)dx} \\ &= B(t)R(a,t)s(a,t), \end{aligned} \tag{4.31}$$

where $R(a,t) = e^{-\int_0^a r(x,t)dx}$ is the accumulated growth rate from age 0 to a, and $s(a,t) = e^{-\int_0^a \mu(x,t)dx}$ is the period survival function from 0 to a.

Arthur and Vaupel (1984) developed a second system of formulas also relating populations at age 0 and a by

$$N(a,t) = N(0,t)e^{-\int_0^a r(0,t-x)dx}e^{-\int_0^a \mu(x,t-a+x)dx}, \tag{4.32}$$

where $\mu(x, t-a+x)$ is the cohort force of mortality, and $r(0, t-x)$ is the growth rate of births at time $t-x$. For calculations of these systems Kim (1986) presented the formulas in discrete time and ages which are reviewed in Chapter 9.

The systems of Preston and Coale, and Arthur and Vaupel are decompositions of the population size. While equation (4.31) uses a period decomposition, (4.32) uses a cohort decomposition. An application of their methods was performed by Otani (1997) who substituted (4.32) and a variant of (4.31) in the difference of age distribution and in the average age of the Japanese population.

Arthur and Vaupel also presented a decomposition of the age-specific growth rates by combining both systems as follows,

$$\begin{aligned} r(a,t) &= r(0,t-a) - \int_0^a \frac{\partial \mu(x,u)}{\partial u}|_{u=t-a+x}dx \\ &= r(0,t-a) - \varphi(a,t). \end{aligned} \tag{4.33}$$

Horiuchi and Preston (1988) interpreted this formula as the legacy of past population dynamics. The age-specific growth rates are decomposed into two terms: the growth rate of the number of births $t-a$ years earlier and the accumulation of changes in the cohort age-specific mortality rates up to age a at time t, denoted by $\varphi(a, t)$.

Table 4.8 presents the age-specific growth rates for the ages 20, 50 and 80 years for France and the decompositions shown in equations (4.30) and (4.33). From Table 4.7 we learned

Table 4.8: Age-specific growth rates, $r(a, t)$, rates in percentages, and decompositions as suggested in the systems of Preston and Coale, and Arthur and Vaupel for France at ages 20, 50 and 80 in 1990.

a	20	50	80
$N(a, 1990)$	8,555,162	5,920,990	1,820,266
$N(a, 1985)$	8,531,224	6,202,743	1,598,866
$N(a, 1995)$	8,579,167	5,652,035	2,072,324
$r(a, 1990)$	0.056	-0.930	2.594
$\nu(a, 1990)$	0.061	-0.474	8.600
$\mu(a, 1990)$	0.005	0.456	6.006
$r(a, 1990) = \nu - \mu$	0.056	-0.930	2.594
$r(0, 1990 - a)$	-1.114	3.349	-2.707
$\varphi(a, 1990)$	-1.170	4.278	-5.301
$r(a, 1990) = r - \varphi$	0.056	-0.930	2.594

Source: Author's calculations of equations (4.30) and (4.33) as described in Chapter 9. Age groups are of ten years starting from the age indicated. Data based on the Human Mortality Database (2002).

that in 1990 France had a low population growth rate. In Table 4.8 one can see the different levels of age-specific growth rates for the French population. France has an elderly population. We see this in the minor relative increase of the population size of persons aged 20 while the population size of those aged 80 exhibits a very high growth rate. However the use of the relative change in population size as age increases, $\nu(a, 1990)$ is still not completely established in demography. We could say that is a measure of the difference in cohort sizes and past mortality. The $\mu(a, 1990)$ accounts for attrition due to mortality and migration. The age group 80-90 years shows a very high mortality rate compared to the other age groups. From the second decomposition we see that those aged 20 and 80 display decreasing birth growth rates. In other words, they come from cohorts of babies $B(t-a)$ that were smaller than their neighboring cohort $B(t-a+1)$. The change in cohort mortality is seen in the results of $\varphi(a, 1990)$. The age group between 50 and 60 years is from a cohort that experienced cohort survival situations that were worse than successive cohorts. As a result of this the $\varphi(50, 1990)$ implies a decline in the age-specific growth rate $r(50, 1990)$.

Decompositions (4.30) and (4.33) can be substituted whenever the age-specific growth rates are present. For example, Caselli and Vallin (1990) substitute equation (4.33) in the derivative over time of the age distribution of the population. In Part III it is shown that direct vs. compositional decomposition involves age-specific growth rates, which allows substitutions of equations (4.30) and (4.33) at any time.

4.4.1 Decomposing the Crude Growth Rate

The relation between the crude growth rate and age-specific growth rates is an average of the age-specific rates weighted by the population size $N(a,t)$,

$$\bar{r}(t) = \frac{\int_0^\omega r(a,t)N(a,t)da}{\int_0^\omega N(a,t)da} \tag{4.34}$$

Keyfitz (1985) assumes that the age-specific growth rates are fixed over time, and by examining the derivative over time of equation (4.34) he obtains

$$\dot{\bar{r}}(t) = \frac{\int_0^\omega r^2(a)N(a,t)da}{\int_0^\omega N(a,t)da} - \left(\frac{\int_0^\omega r(a)N(a,t)da}{\int_0^\omega N(a,t)da}\right)^2 = \sigma^2(r). \tag{4.35}$$

Equation (4.35) informs us that the change in the mean rate of increase at time t is equal to the variance among the age group rates of increase. Kephart (1988) uses equation (4.35) to study the heterogeneous character of aggregate rates of spatial units of analysis.

Table 4.9: Population growth rate of the world, $r(t)$, and Keyfitz's estimation of the annual change, around January 1, 1979 and around January 1, 1982.

t	1979	1982
$\bar{r}(t-1.5)$	1.717 %	1.697 %
$\bar{r}(t+1.5)$	1.741 %	1.722 %
$\dot{\bar{r}}(t)$	0.798 *	0.832 *
$\dot{\bar{r}} = \sigma^2(r)$	0.798 *	0.832 *

Source: Author's calculations of formula (4.35) as described in Chapter 9. Data is based on the U.S. Census Bureau (2001). Note: * denotes per 10,000. Growth rates for every country were assumed fixed over intervals of 8 years (1975-1983 and 1978-1986) for the columns of 1979 and 1982, respectively. The growth rates of the world for 1977.5, 1980.5 and 1983.5 were calculated as the averages in formula (4.34). Growth rates were estimated based on data for all the countries of the world for which data were available.

As shown in Table 4.9, it is possible to obtain a good approximation of the change in the growth rate by using Keyfitz's formulation. Nevertheless, the assumption of fixed growth rates for the examined countries is somewhat problematic. For example, the value in the row of $\bar{r}(t+1.5)$ for the column 1979 and in the row $\bar{r}(t-1.5)$ for the column 1982 should have the same estimates of the year 1980.5, but instead are different. In Part III we show a generalization of the formula proposed by Keyfitz for countries' growth rates that change over time.

The total population size comes from the addition of the population sizes at all ages, $N(t) = \int_0^\omega N(a,t)da$. By substituting the population size at age a for the Preston and Coale system and recalling the definition of the population growth rate, $r(t) = \acute{N}(t)$, we obtain a decomposition. By applying the general equation (4.25) Vaupel and Canudas Romo (2000) show that the current population growth rate can be decomposed into three components,

$$r(t) = \acute{B}(t) + \acute{e}_o(t) + R^*(t), \tag{4.36}$$

where $\acute{B}(t)$ is the intensity of births, $\acute{e}_o(t)$ is the intensity of life expectancy and $R^*(t)$ is a residual equal to

$$\begin{aligned} R^*(t) &= \frac{\int_0^\omega \left[\acute{R}(a,t) + \acute{s}(a,t)\right] N(a,t)da}{\int_0^\omega N(a,t)da} - \frac{\int_0^\omega \dot{s}(a,t)da}{\int_0^\omega s(a,t)da} \\ &= \bar{\acute{R}}(t) + \bar{\acute{s}}(t) - \acute{e}_o(t). \end{aligned} \tag{4.37}$$

Equation (4.36) permits a decomposition of the current population growth rate into three components. The first two are the current intensity of change in births and the current intensity of change in period life expectancy (which captures the impact of current mortality change). The third is a residual term which reflects the influence of historical fluctuations. These fluctuations result in a population size and structure that is different from the stationary population size and structure implied by current mortality and birth rates.

Table 4.10 presents the application of equation (4.36) to France, Japan and the USA. These data were also used in Table 4.7. In the period under study, France and Japan both experienced

Table 4.10: The population growth rate, $r(t)$, rates in percentages, and Vaupel and Canudas Romo's decomposition of the growth rates in the period 1985-1995 for France, Japan and the USA.

	France	*Japan*	*USA*
$N(1990)$	56,570,545	121,907,530	248,940,972
$N(1985)$	55,157,223	119,740,316	236,872,781
$N(1995)$	58,020,081	124,113,968	261,624,012
$r(1990)$	0.506	0.359	0.994
$\acute{B}(1990)$	-0.777	-1.450	0.498
$\acute{e}_o(1990)$	0.325	0.263	0.146
$R^*(1990)$	0.958	1.546	0.349
$r(1990) = \acute{B} + \acute{e}_o + R^*$	0.506	0.359	0.994

Source: Author's calculations of formula (4.36) as described in Chapter 9. Data from the Human Mortality Database (2002).

a low fertility rate but the growth rate remained positive. In contrast to the negative intensity of change in birth is the positive change in life expectancy. By applying the Vaupel-Canudas decomposition one learns that new births are being substituted by longer life. For the USA the same conclusion cannot be reached since all three components contribute to an increase in the population size.

By using a variant of equation (4.32), Horiuchi (1995) shows a similar application of the general equation (4.25). Horiuchi also obtains a decomposition of the population growth rate.

The population size at age a and time t is now the product of the total population at time $t-a$, the crude birth rate at time $t-a$ and the cohort survival (mortality and net migration)

$$\begin{aligned} N(a,t) &= N(t-a)\left[\int_{\alpha}^{\beta} \pi_f(x,t-a)b(x,t-a)dx\right] e^{-\int_0^a \mu(x,t-a+x)dx} \\ &= N(t-a)CBR(t-a)s_c(a,t), \end{aligned} \tag{4.38}$$

where $s_c(a,t) = e^{-\int_0^a \mu(x,t-a+x)dx}$ is the cohort survival or probability that a person born in $t-a$ will live to attain age a. As before, $\pi_f(x,t-a)$ is the proportion of women at age x in the total population at time $t-a$. From this equation the age-specific growth rates are calculated as $r(a,t) = \acute{N}(a,t)$. By using the relative change of a product (4.25), a decomposition similar to (4.33) is obtained,

$$r(a,t) = \acute{N}(t-a) + C\acute{B}R(t-a) + \acute{s}_c(a,t). \tag{4.39}$$

As a last step the age-specific growth rates are substituted in the crude growth rate or average growth rate, as seen in equation (4.34). The result is a crude growth rate at time t expressed as the effects of past changes in the population size, fertility, and mortality (migration is also included here),

$$r(t) = \overline{\acute{N}} + \overline{C\acute{B}R} + \overline{\acute{s}_c}, \tag{4.40}$$

where the elements on the right hand side are the average of the relative change of population size, fertility and survivorship. The fertility component is separated into two other components: the female age-distribution and age-specific fertility rates.

In comparison to the decompositions obtained in (4.29) and (4.36) which only required data for the period of interest, equation (4.40) is more demanding. The optimal situation for using equation (4.40) is when countries have enough historical data, not only on mortality but also on fertility, age distribution and migration, such as the Scandinavian countries. Horiuchi notes that his proposal requires data with long time series, but it is also possible to adjust the method to countries with less data available. By using his adjustment and equation (4.40) the population growth rate for France, Japan and the USA, was calculated and compiled in Table 4.11. As previously mentioned, this decomposition is the result of the averages of past changes. Two of these components contribute to the increase in growth rates, while the past change in births neutralizes this increase. The average past change in population size is the major component accounting for the change in France and USA. In Japan, on the other hand, the average change in birth rates is the major component. The average change in cohort survivorship helped to sustain the Japanese growth rate at positive levels.

4.4.2 Conclusion

Several decompositions of the growth rate of the population have been presented in this section. The crude growth rate and the age-specific growth rates are measures of change of the population size. Therefore, it is possible to decompose the change in the growth rates or simply to decompose it into components of the past and present changes in fertility, mortality and migration.

Table 4.11: The population growth rate, $r(t)$, rates in percentages, and Horiuchi's decomposition of the growth rates in the period 1985-1990 for France, Japan and the USA.

	France	*Japan*	*USA*
$N(1990)$	56,570,545	121,907,530	248,940,972
$N(1985)$	55,157,223	119,740,316	236,872,781
$N(1995)$	58,020,081	124,113,968	261,624,012
$r(1990)$	0.506	0.359	0.994
$r(1990)$ *	0.504	0.357	0.994
$\overline{\acute{N}}$	0.547	1.129	1.260
$\overline{C\acute{B}R}$	-0.412	-1.246	-0.668
$\overline{\acute{s}_c}$	0.370	0.474	0.402
$r(1990) = \overline{\acute{N}} + \overline{C\acute{B}R} + \overline{\acute{s}_c}$	0.504	0.357	0.994

Source: Author's calculations of formula (4.40) as described in Chapter 9. Note: * means that the growth rates were calculated as an average of growth rates which is explained in Chapter 9. Data is derived from the Human Mortality Database (2002). Historical data for Japan is based on Japan Statistical Association (2002), and the USA data is from the U.S. Census Bureau (2001).

Formulations by Preston and Coale (1982) and Arthur and Vaupel (1984) have helped us to understand the relationships among the demographic variables. These systems are generalizations of population models that hold for any population, which look at changes over time, age and cohort in a continuous way. Therefore, they are related to the formulations of direct vs. compositional decomposition, in Part III, that have the purpose of accounting for changes over time in a continuous way. Furthermore, the population growth rates are measures of the changes in the age structure, and therefore they are included in one of the components of direct vs. compositional decomposition.

Chapter 9, on estimation procedures, includes a section on population growth rates. Explanations of the estimations of the tables of the present chapter are found in that chapter.