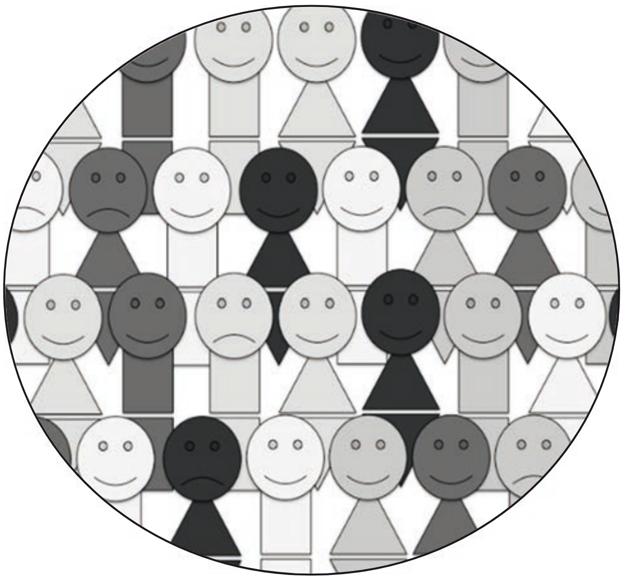
UNDERSTANDING SOCIAL CHANGE

A DECOMPOSITION APPROACH



PARFAIT M. ELOUNDOU-ENYEGUE
SARAH C. GIROUX
MICHEL TENIKUE

Preface

A story is often told of a drunk man who loses his key at night and only searches for it under a lamppost, because that is where the light is. Today, the social sciences may be in a similar predicament: Research on social change (a macro-phenomenon) often relies on (micro) survey data. Compared to macro-level regression, micro-level research is the 'methodological lamppost,' the better-lit area of quantitative social science, where analyses are deemed statistically most rigorous. Thus, to study how education affects national economies, a researcher might rely on estimates of how a worker's education level affects his or her income. In public health, to estimate how economic growth affects national obesity, our researcher might likewise extrapolate from the individual-level link between income and obesity. In demography, to capture the economic benefits of declining fertility, the researcher might explore how the economic fortunes of various family members depend upon the size of their family.

All these three examples show a discrepancy between the scale at which a problem occurs and the scale at which it is studied. Social change is a dynamic and macro-process, yet we often approach its study with micro cross-sectional data. This creates two fallacies. One, **ecological**, relates to unit of analysis, specifically, studying individual people rather than the entire society. The second, **historical**, fallacy amounts to "reading history sideways" i.e., reading cross-sectional data as indicating a historical trend. Most researchers acknowledge these fallacies: cross-sectional studies of interpersonal differences may reveal why some people have a higher propensity to experience a given condition (say, obesity) but they cannot explain why/how the social magnitude of this phenomenon changes over time. Nonetheless, many studies still directly infer macro-relationships from micro-results.

If an investigator wishing to avoid these biases turns to macro-level analysis, her peers may see this work as lacking statistical rigor or obscuring the differences between people in the same country (Rodrik 2005). The question, therefore, becomes, "How to avoid ecological and historical fallacies while also maintaining some statistical rigor?"

The challenge is to recognize within-country diversity but still show how the diverse behaviors of people in a country add up to produce a collective change. For this kind of analysis, decomposition is a useful tool. Even if it does not reveal ultimate causes, it points to main "sources" of change, i.e., the groups or proximate processes driving the social change.

Decomposition is widely applied across many disciplines, especially within the social sciences. However, its different variants remain insufficiently integrated and literature on the topic remains fragmented. Few textbooks offer a comprehensive introduction highlighting the method's wide range of possibilities. None to our knowledge clearly explains how to enrich elementary decompositions let alone combine them with other methods. We fill a gap with this monograph, which updates and augments an earlier version published in 2010.

Our work on this document was conducted under the auspices of the International Union for the Scientific Study of Population (IUSSP) with funding from the William and Flora Hewlett Foundation, and material support from Cornell University (USA), the Institute for Training and Demographic Research (IFoRD, Cameroon), and the Higher Institute of Population Sciences (ISSP, Burkina Faso). We received feedback from numerous colleagues, most prominently Françoise Vermeylen, Gervais Béninguissé, Jean-François Kobiané, Silvere Konan Yao, Crispin Mabika Mabika Gilles Gohi, Nina Trautmann Chaopricha and Latif Dramane as well as from dozens of participants in the various workshops conducted under this IUSSP project. We tested early versions of this

monograph with support from Serge Bahoken, Firmin Zinvi Charles Mouté, Justin Dansou, Sandrine Mbele, and Habibou Ouedraogo. Several researchers and students — Yédodé Ahokpossi, Jacques Emina, Edzengte Pascal, Kouam Nadège and Bela Foe Chrystelle. — were early volunteers to apply the method in their own research. Their experiences informed subsequent revisions, as did the insights of several non-scientists (Pie-Marie Belibi, Stephanie Hissoak, and Vincent Onguéné) whose comments helped make the document more accessible to a broader audience. In the end, however, the authors are solely responsible for the remaining shortcomings of this document.

Ithaca, August 2017

Sommaire

Chapter I
Introduction

Can scientific research be a useful guide for social policy? Whether reducing unemployment, expanding school enrollment, containing social inequality, or reducing mortality, the constant challenge facing those who plan and implement policies is to promote desirable social change. To meet this challenge, policymakers must understand the levers of social change — and, in this context; scientific research has a key role to play.

Social science has made great strides over the last half-century yet these strides were greatest in micro-level studies. Thanks to advances in computing and communication technology, researchers can now collect, share, and process statistical data on millions of households and individuals. Researchers can explore the detail and mix of factors shaping individual behaviors.

However, an analyst of societal change will not be satisfied with this micro-level detail alone. She may find the detail useful but still need to convert the micro-level information into valid inferences about social change. Over half a century ago, Robinson (1950) warned against ecological bias, noting that relationships observed macroscopically need not match those occurring at the individual level and vice versa.[1] Thornton (2001) likewise drew attention to an equally harmful bias, the tendency to "reading history sideways." In essence, it is wrong to use cross-country comparisons to draw conclusions about development trajectories. Researchers must thus be careful in navigating the space between micro-data and macro-issues, building on the detail and robustness of microlevel statistics as they aggregate them to inform macro-level questions. The decomposition methods presented here can facilitate this conversion.

---

[1] For example, richer *countries* may have higher rates of obesity but that does not imply that richer people creases with the level of wealth of a country does not necessarily imply that within countries, richer *people* have a greater propensity to become obese than poor people. Similarly, national rates of divorces increase with national crime rates, but that does not necessarily mean that people from divorced families are more prone to crime.

I.1 What is decomposition?

I.1.1 Definitions

To decompose something is to break it into its elementary components. In biology, decomposition is a process of organic material decay. In chemistry, it is the process of bursting a complex molecule into simpler molecules or atoms. In physics, it can describe the trajectory of a projectile, projecting its path into a three-dimensional space that includes vertical, horizontal, and transversal components. In the social sciences, decomposition can serve to estimate how several elementary processes (or groups) fuel an aggregate social change. For example, how do different regions of a country contribute to national wealth, or how do women from different education backgrounds contribute to change national fertility?

Several variants of decomposition methods exist in the literature but the fundamental idea is the same: how to partition a functional set into its elementary constituents. To appreciate the uniqueness of decomposition vis-à-vis other methods, we specify below some situations in which it applies.

I.1.2 Application areas

Within the social sciences, decomposition analysis is particularly useful for studying aggregate social change considered as any transformation — induced or spontaneous — in the structure, functioning or performance of a social community. The most relevant changes are quantifiable transformations resulting from an aggregation of individual behaviors.

Decomposition is thus not applicable in studies where individuals are the unit of analysis. Decomposition is also of little help when studying societal outcomes that do not result from the aggregation of individual behavior. These two exceptions aside, the method applies to a wide range of social processes in demography, economics, political science, and sociology (Kitagawa 1955, Oaxaca 1973; DasGupta 1993; Vaupel and Romo 2003). The only requirements are to have quantifiable social outcomes that reflect an aggregation of individual outcomes.

Quantification: The outcome under study should be quantifiable, i.e., captured as an absolute number, an average, a percentage, a ratio, or a measure of inequality. This excludes qualitative processes such as a country's political transition from autocratic to democratic rule, or its economic transition from subsistence to market orientation, or its demographic transition from extended to nuclear family systems. Even in these cases, dichotomies such as autocracy/democracy, subsistence/market, or extended/nuclear can be expanded into a continuum. For example, one can replace the subsistence/market dichotomy by a continuous outcome such as the percentage of people working for an employer other than family. With such reformulation, phenomena that may a first seem qualitative lend themselves to decomposition analysis.

Aggregation of individual behavior: Sociologists distinguish between social outcomes that an intrinsic feature of the whole society versus aggregate outcomes. The former have no correspondent at the individual level. An example might be a country's laws on abortion. The latter resulting from an aggregation of individual characteristics. Examples may include the fertility level of a country or its average consumption of tobacco. This tobacco consumption is an aggregated (not intrinsic) feature of the society because a country does not smoke; individual citizens do.

Gradual change: Decomposition methods are seldom applicable to rare or sudden, accidental processes (e.g., studying the number of casualties in an earthquake or a sudden outbreak of cholera). Rather, they best apply to studies of processes that change gradually over time.

I.1.3 Mode of explanation

Decomposition analysis is about the sources rather than the ultimate causes of change. It is better at determining the origin of a change rather than why the change occurred. It mechanically accounts for the sources of change, specifically, how different processes or groups contribute to generate a social change. While full causal analysis seeks to reveal the ultimate causes, decomposition merely reveals the proximate processes or groups from which the change occurred.

From what? (Proximate processes). In a basic demographic decomposition, a national outcome is cast as the weighted average of behaviors observed among different sub-populations in the country. The total change is set as coming from two proximate sources, namely (a) compositional changes such as changes in population composition like the relative weight of constitutive sub-populations and (b) behavioral changes such as changes in group behavior or outcomes.[2]

Decomposition can also apply to outcomes resulting from a chain of events where, say, process #1 leads to process #2, then process #3 … For instance, the amount of locally-produced food available in markets depends on a sequence of three processes: local food production, allocation of harvest between personal consumption and market, and food transportation into markets. Any change in amounts of food available in markets will reflect a mix of changes in these three elementary processes.

By whom? In addition to revealing the processes by which change occurs, researchers may seek to estimate the relative contribution of each group to the total change. How for instance how much do people from various regions, age categories or educational categories (for instance) contribute to the change. The decomposition and its identification of proximate processes and groups can pave the way for a deeper exploration of the ultimate causes of change.

I.2 Decomposition versus other methods

Imagine a researcher wishing to explain a recent decline in mortality in her country. In all likelihood, decomposition will not be the first or only tool considered. The researcher has multiple options including qualitative analysis, trend analysis, and regression.

She could use a qualitative approach based on key informants, focus group discussions, or archival data to shed light on historical events occurring during the decline. Did the country make new investments or implement new projects? Did it make recent scientific discoveries or put new drugs on the market? Was there an outbreak of unusual events, a rise of effective leaders, or an improvement in the water supply and medical services?

Our researcher might also decide to monitor recent trends in mortality, to pinpoint the exact moment when the decline began and to review other key events preceding this decline. She will ask about social processes that appear to co-vary with mortality. This approach can complement a qualitative reliance on key informants by systematically testing the hunches from the informants with trend data. However, the conclusions remain subjective if all she did was to eye-ball and visually compare various trends without attempting a formal statistical test.

---

[2] We will return to these two sources and can already offer one example here. Imagine a country today that experienced war in 1970. One can well-imagine that the views about war –specifically, whether the country ought to embark on another war today — would likely vary by generation. For instance, people born before the 1970s who have lived through a previous war may be less likely to support another war than the younger generations. Therefore, the results of a poll about a country’s readiness to go to war in 2000 and again in 2025 might change simply because the percentage of people born before 1970 would have dwindled as this group ages (a compositional effect). It could also be that the global political environment between 2000 and 2025 changes in ways that changes people’s attitudes, perhaps making them less supportive of war (a behavioral effect). Overall, the total change in national attitudes will be a combination of these two effects.

More formally, our researcher could use formal correlation/regression analysis. Unfortunately, her macro-correlations would obscure the detail of individual responses. On the other hand, regression using individual data is more detailed and rigorous but it does not address the right level of analysis for someone interested in social change.

Micro and macroscopic approaches thus complement each other: one being more rigorous and detailed, and the other more relevant for national policy design. This complementarity permits two possible kinds of integration (Figure 1). The first looks at how the 'macro' (society) affects the 'micro' (individual) and is known as multilevel regression (Luke 2004). The second conversely looks at how micro behaviors aggregate to shape macro-level outcomes and change therein; this is the essence of decomposition.

Table 1 summarizes the differences between the four approaches discussed above. In particular, they differ in their presumed drivers of change. The drivers, in demographic decomposition, are either compositional forces or behavioral factors. In some qualitative analyses, change can come from a single person or a key event in the study community.

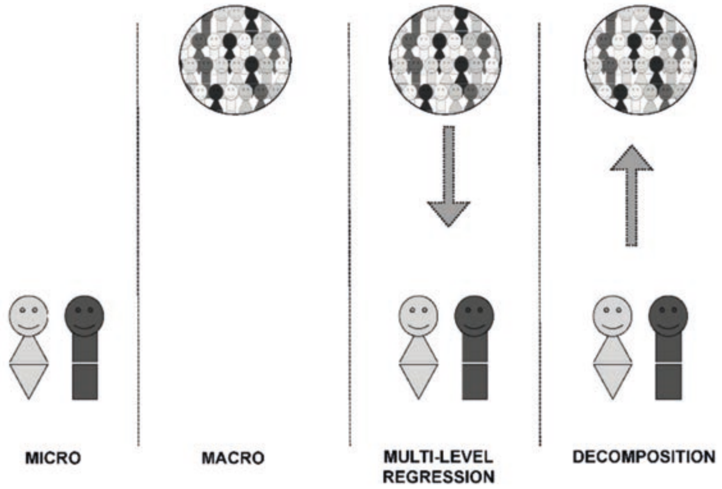
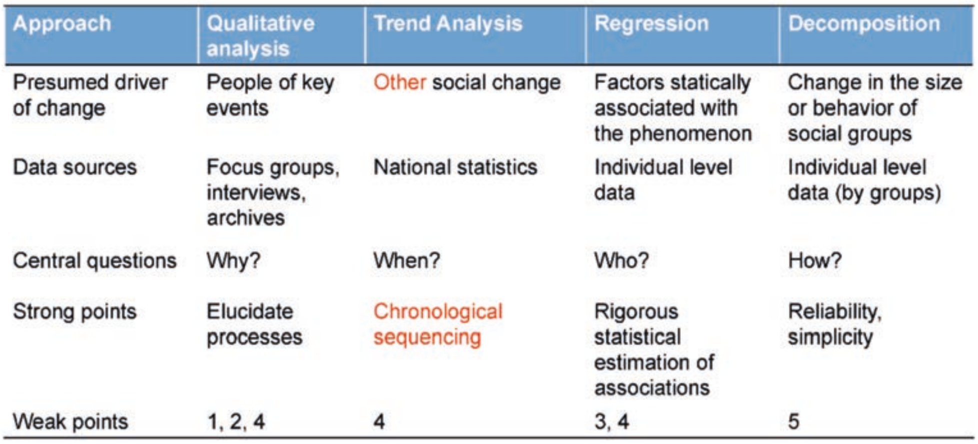


Figure 1. Four modalities of the integration of micro and macro analysis.



Approach	Qualitative analysis	Trend Analysis	Regression	Decomposition
Presumed driver of change	People of key events	Other social change	Factors statically associated with the phenomenon	Change in the size or behavior of social groups
Data sources	Focus groups, interviews, archives	National statistics	Individual level data	Individual level data (by groups)
Central questions	Why?	When?	Who?	How?
Strong points	Elucidate processes	Chronological sequencing	Rigorous statistical estimation of associations	Reliability, simplicity
Weak points	1, 2, 4	4	3, 4	5

Weak points 1. Not statistically representative; 2. Statistical associations are not established; 3. Chronology not established; 4. Competing explanations are not eliminated; 5. Causal relationships are not clarified.

Table 1. Comparison of four explanation approaches to social change.

I.2.1 Advantages and limitations

Unlike causal analysis, decomposition does not establish causation. Returning to our earlier example about mortality, decomposition may reveal which sectors, groups, or causes of death accounted for a rise in overall mortality, but the results will not say why these declines occurred. This shortcoming is severe, because sound policies require a clear understanding of cause and effect. However, by revealing the proximate sources of change, decomposition analysis can usefully guide social intervention and targeting. Moreover, it has the advantage of being simple, flexible, easy to interpret, and compatible with other methods.

Simplicity. Decomposition methods are simple to apply and interpret. Their application requires no fancy statistical analysis, complex calculations, or advanced software. Much of the work is doable with a spreadsheet program such as Excel, and the input data are often readily available in reports and online tabulations.[3] When the input data is publicly available, the analyses are made easier and the process transparent because other researchers can easily check both the input data and the results.

Flexibility. The decomposition approach is quite flexible in its application. A reasonably creative analyst can move from basic forms to complex combinations, depending on her needs. The presentation and illustrations in this monograph seek to highlight this flexibility.

Easy interpretation. The interpretation of decomposition results is intuitive and easy. Compared to the results from regression analysis (ordinary least squares, logit, and odds ratios), the statistics generated by a decomposition analysis are readily expressed in plain language. The results simply indicate the percentage of social change coming from a given process or group.

Compatibility. The decomposition method is compatible with many other methods, including micro-regressions, multilevel analysis, simulations, and qualitative analysis. It can help aggregate the results from micro-regressions. It can combine fruitfully with multilevel analysis. It can serve as a prelude to a qualitative analysis.[4] For all these reasons, decomposition can apply eclectically to a wide range of fields and methodological traditions. Importantly, it does not replace or compete with other methods but, rather, it complements them to improve the quality of the findings.

Transparency. This is a major strength of decomposition. Contrasting with the opacity of most other methods, the average reader can easily check the internal consistency of results from decomposition analysis. In a multivariate regression, for example, the reader often has few opportunities to check the accuracy of the regression coefficients presented and s/he must trust the researcher.[5] In a decomposition analysis, by contrast, s/he can check the accuracy of final results by reviewing the initial information and following, within the table of findings, the subsequent data transformations and the plausibility of final results.

---

[3] See, for instance, the Demographic and Health Survey (www.statcompiler.com) or the World Bank (http://data.worldbank.org/data-catalog/world-development-indicators).

[4] For example, in our previous example about attitudes vis-a-vis war, the researcher's decomposition results may show that, contrary to expectations, much of the change in national attitudes vis-à-vis war reflects a true behavioral change rather than a compositional change. In other words, many people in the country actually changed their attitudes about war. Alternatively, she may find that most of the change came from people between the ages of 45 to 60. Armed with this information, she can now go ahead and conduct more investigations into why these people changed their minds during that period.

[5] The situation is improving, as researchers must increasingly make their data publicly available. However, the reader must have access to the computing program used by the researchers and be able to retrace all steps, from coding to the final modeling, which is rarely convenient.

I.2.2 Political relevance and applications

Decomposition methods can apply to many of the social changes underway across the globe. In particular, they can inform the study of several development goals pursued under the United Nations' Sustainable Development agenda (SDG), including poverty, health, inequality, gender, or basic education, for instance. Achieving these ambitious goals requires an efficient use of countries' scarce resources, which implies a clear understanding of the drivers of socioeconomic change. The rapid and uneven demographic changes occurring in many countries create fast-changing and diverse societies that can no longer be fully understood without careful attention to disaggregated evidence. Decomposition methods can help.

I.3 The main types of decomposition

Decomposition is not a single method but set of related methods. Its variants appear separately in different fields and they have not been sufficiently integrated. Below, we review some of these variants and their differences based on four criteria: (a) the type of independent variable, (b) the type dependent variable, (c) the link function, and (d) the complexity of the analysis.

Criterion 1: The independent variable type

One can distinguish demographic, regression, and temporal decompositions, depending on whether the independent variable is nominal, interval or ordinal.

- In a demographic decomposition, the independent variable is nominal, e.g., country region, age group, ethnicity, or marital status. For instance, a national outcome can be viewed as the population-weighted average of outcomes across all of the countries' regions. For example, the national support for a given cause or person (e.g., the president) will be a weighted average of support across all regions. This support will change if the relative size of the regions happens to change over time or if the views of people within any of the regions change.

  Note that demographic decomposition also applies to an ordinal variable (e.g., socioeconomic status) if it is treated nominally, i.e., with no explicit attention to the order of categories.
- In a temporal decomposition, the independent variable is ordinal, and the order between categories is explicitly considered. For instance, if age group is the independent variable, we keep in mind that people aged 15-19 are younger than people aged 20-29 years, who are themselves younger than people aged 30-39 years.
- In a regression decomposition, finally, the independent variable is quantitative. Examples include a person's years of education, the number of siblings, or income in dollars. The starting point in regression decomposition is to have an estimate of the statistical effect of an independent variable (e.g., years of education) on some outcome (e.g., health). Health outcomes then are expected to change either because the amount of education changes or because the average effect of education changes. The decomposition in this case seeks to determine how much of the total change in health reflects the change in the quantity of education versus the effectiveness of education.

Criterion 2: The dependent variable

For the simplest basic decomposition, the dependent variable is an average or a percentage. In more complex (derived) decomposition, the dependent variable may be a more complex measure such as inequality.

Criterion 3: The functional relationship

The main criterion here is the type of relationship linking the independent to the dependent variable. The type and complexity of this functional relationship vary (see Figure 2), and we can specify three types:

- A demographic relationship: Basically, the Y value of the entire country is a weighted (by demographic weight) average of prevailing values in the various subpopulations of the country (y_i);
- A statistical relationship, specifically, a regression relationship between Y and X; and
- A mathematical relationship: in this case, the dependent and independent variables are linked by a simple mathematical relationship (which involves a quotient, sum, product, or log). Unlike the statistical relationship, it is a true relationship that does not vary across countries or situations. The only change is in the values of these variables.

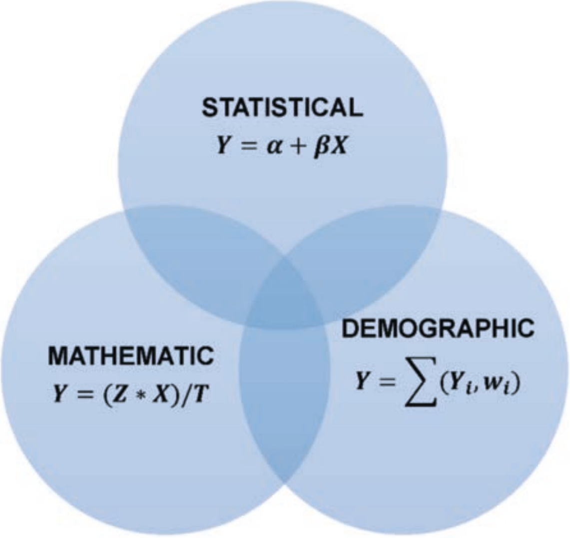


Figure 2. Basic types of decomposition.

Each of these basic functional relationships can become more complex. In regression analysis, for instance, one can move from a single independent variable and a linear specification to more elaborate forms that feature several independent variables, non-linear forms, or multi-level relations.

Criterion 4: The degree of complexity

This criterion distinguishes simple from advanced or mixed decompositions that combine two or more elementary forms. The basic forms shown in Figure 2 are thus the building blocks for generating complex forms. Therefore, as indicated in Figure 3, an elementary decomposition may yield a (derived) decomposition when its dependent variable or functional relationship becomes more complex. We obtain a mixed decomposition by combining two or three basic forms. This book's organization follows the typology shown in Figure 2. We begin with basic and derived forms and work our way up to mixed forms.[6]

---

[6] We should note here that a number of highly specialized decomposition formulas have been developed to study specific questions such as life expectancy (Vaupel, 2003), job discrimination (Oaxaca, 1973) and poverty or inequality (Shorrocks, 2013). The monograph does not show these detailed formulations but readers will easily see where these formulations fit into the taxonomy presented here.

I.4 Conventional notations

To facilitate the presentation, we use the following conventional notations for variables, indices, and historical change:

Variables. Following common practice in the social sciences, dependent and independent variables are indicated by the letters Y and X, respectively. In addition, we use capital letters when referring to national outcomes, and reserve lowercase letters when referring to individuals or sub-populations. Y thus designates a dependent, macro-level variable; y designates a dependent variable at the micro or meso-level; X designates an independent variable at the macro-level; and x designates an independent variable at the micro or meso-level. Therefore, for instance, if we study the effects of education on mortality, Y is the national mortality rate, while y is mortality rate in a subpopulation (e.g., for those between the ages of 15 to 19 years). In the study, X is the national education level, while x is the level of education in a subpopulation, e.g., for the 15-19 year olds.

Weighing. The letter w measuring the weights of different subgroups is mostly used in demographic decomposition or its close variants. The weights often reflect population size, i.e., the percentage of the national population in a given category.

Regression parameters. Regression decomposition includes conventional regression parameters such as:

- $\boldsymbol{\alpha}$, the intercept, which is the value of Y (or y) when X (or x) is 0;
- $\boldsymbol{\beta}$, the marginal increase of Y (or y) when X (or x) increases by one unit; the more complex regression analyses will integrate the case where β is a vector and take multiple values; and
- $\boldsymbol{e}$, the error term.

Indices. The presentation will also use the following indices:

- *j* indexes groups; e.g., y_j denotes the value of the dependent variable for the j group, while x_j indicate the value of the independent variable for the same group;
- *t* denotes time; e.g., Y_t indicate the value of the dependent variable for a given year t and for the entire population (e.g., the average mortality in Senegal as of 1990); and
- *a* indexes age; thus, Y_a is the value of the dependent variable for a given age group e.g., the specific fertility rate for the age group. Occasionally, we use the term + to show all ages above the referenced age; in fertility analysis Y_{39+} might be the average fertility for all ages above 39.

Historical change. Δ indicates the historical change. For example, ΔY represents the historical change in the outcome being studied, specifically, the difference in the Y values observed at two points in time (e.g., Y_{t1} –$Y_{t,}$ t_1>t $_{...}$).

Averages. The annotations will distinguish between two types of averages, whether cross-sectional or historical. A cross-sectional average is the average in the population at a specific time t. It is calculated over several groups at one point in time. Since the dependent variable in most basic decompositions is an average, these averages will simply be denoted Y (or y). A historical average is the average for the same group over two periods. We signal it by adding a bar over the corresponding letter. Thus, $\bar{Y}$ is an average between two periods for the value of the dependent variable at the national level. Likewise, $\bar{y}$ indicates the average value of two years of the dependent variable for one subpopulation.

$$\bar{Y} = (Y_{t1} + Y_t)/2$$

I.5 Structure of the monograph

Our presentation progresses from simple to complex decompositions. We began by defining and comparing decomposition with other methods. We then offered a typology of decomposition methods, depending on the type of data and the complexity of the analysis. Decompositions are labeled 'demographic,' 'temporal,' and 'regression' depending on whether the independent variable is nominal, ordinal, or quantitative. These elementary types are then modifiable to yield nested decompositions. Once again, elaborations may reflect greater complexity in the type of dependent variable (e.g., a measure of inequality instead of an average) or the link function between independent and dependent variables (e.g., a curvilinear or multivariate relationships instead of a simple linear relationship).

Finally, one can combine elementary types to produce advanced decompositions (Figure 3). We present a few examples of such advanced decompositions, but investigators can use their own imagination to generate additional forms.

To facilitate understanding, the document mixes verbal descriptions with mathematical formulas, illustrative examples, annotated charts, graphs, and figures. This pedagogical approach hopefully makes the material accessible to a wide pool of readers with different learning styles.

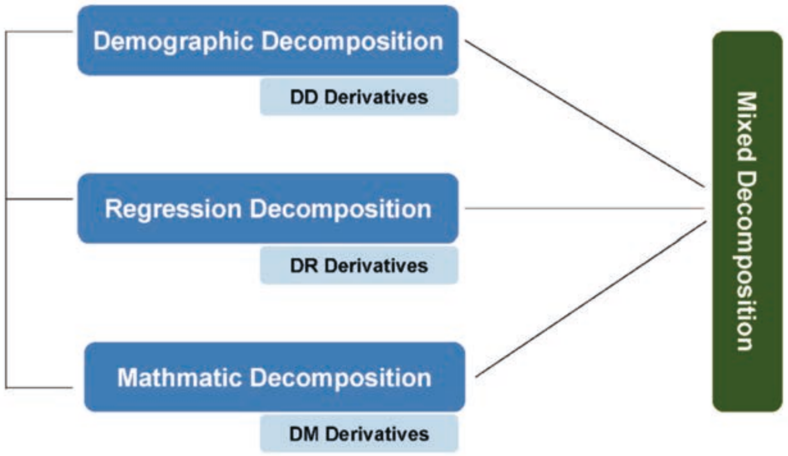
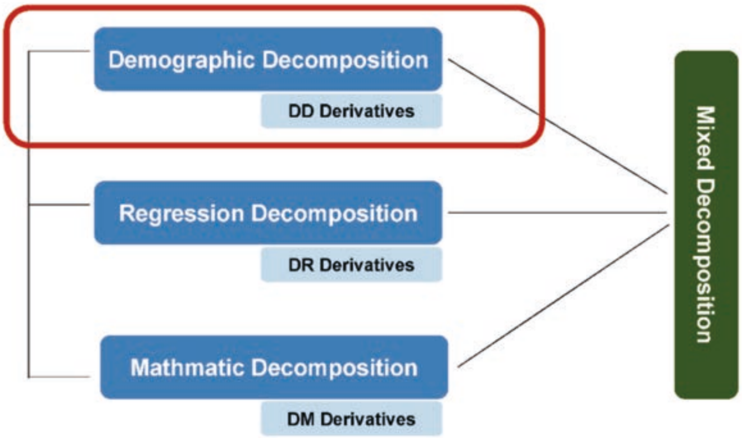


Figure 3. Decomposition types and combinations.

Chapter II
Demographic decomposition



II.1 Basic demographic decomposition

II.1.1 Problem Type

This first type of decomposition applies to national outcomes (Y) that are an aggregated outcome from several subpopulations (y_i), each weighted by its relative size (w_j). Formally,

$$Y = f\ (y_j,\ w_j) \qquad [II.1]$$

For example, the mortality rate of a country is the weighted average of rates in different regions or socioeconomic groups. In this formulation, Y, the dependent variable, is quantifiable and X, the independent variable, is measured nominally. The variable X and its categories must meet at least four criteria:

Exhaustiveness: The full set of categories for the independent variable must be such that each member of the population belongs to one and only one category j. In other words, the set of categories must cover the entire population, with categories also being mutually exclusive.

Distribution: The number of categories for the classification variable should be neither too small (> 2) nor too large. With too few categories, the analyses are not detailed enough to be informative. Conversely, having too many categories spreads the data too thinly. Thus, variables such as sex (with too few categories) or age (too many categories if measured in singe years) are not ideal as classification variables.

Variability: The size of the individual categories must fluctuate over time. Otherwise, the compositional effect in the decomposition analysis will remain zero. For this reason, gender is, once again, a poor classification variable unless the researcher is dealing with a very dynamic population with rapidly-changing sex ratios.

Relevance: A good classification variable should be theoretically relevant or policy-relevant. Variables such as "region" for instance usually meet the criterion of policy-relevance, whereas education for instance is theoretically relevant if the outcome being studied is expected to depend on one's education level.

II.1.2 Visual representation

As stated in the introduction, decomposition methods seek to identify the sources of change, whether substantive (with processes driving the change) or sociological (with groups or people driving the change).

The figure below illustrates a basic decomposition with a study of gender parity in education. The left-hand side of the diagram shows five squares capturing the trend in parity for the country as a whole, with darker colors reflecting greater parity. The chart shows a steady progress at the national level from high educational inequality between boys and girls (the white square on the far left) to parity (the black square on the far right). Obviously, the country became more gender-equitable but for several reasons, a researcher may wish to understand how this evolution occurred, specifically, how the country's various social classes contributed to it.

On the right side of the diagram (Frame B) are two possible, opposite, scenarios of convergence. The first (B1) shows a horizontal convergence, with the educational gap closing at the same rate for all groups. Conversely, Frame B2 shows a vertical convergence, with change starting among the top income group before gradually trickling down. In year 2, the top income group had already achieved parity, while gender inequality remained prevalent in lower-income groups. If we ask about the groups that led the change, we get different answers from the two scenarios: in the first case, all groups evolved simultaneously, while in the second case the change came from above.

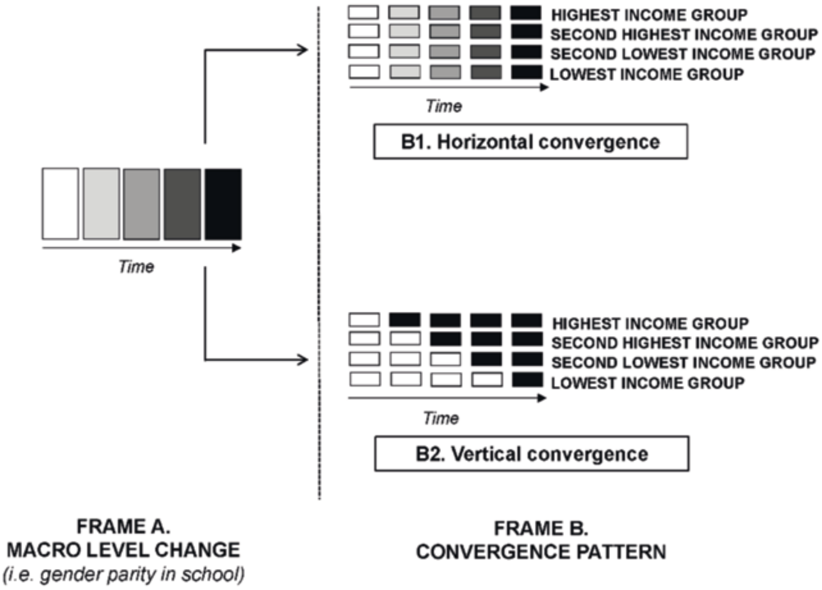


Figure 4. Vertical versus horizontal convergence in education.

II.1.3 Example

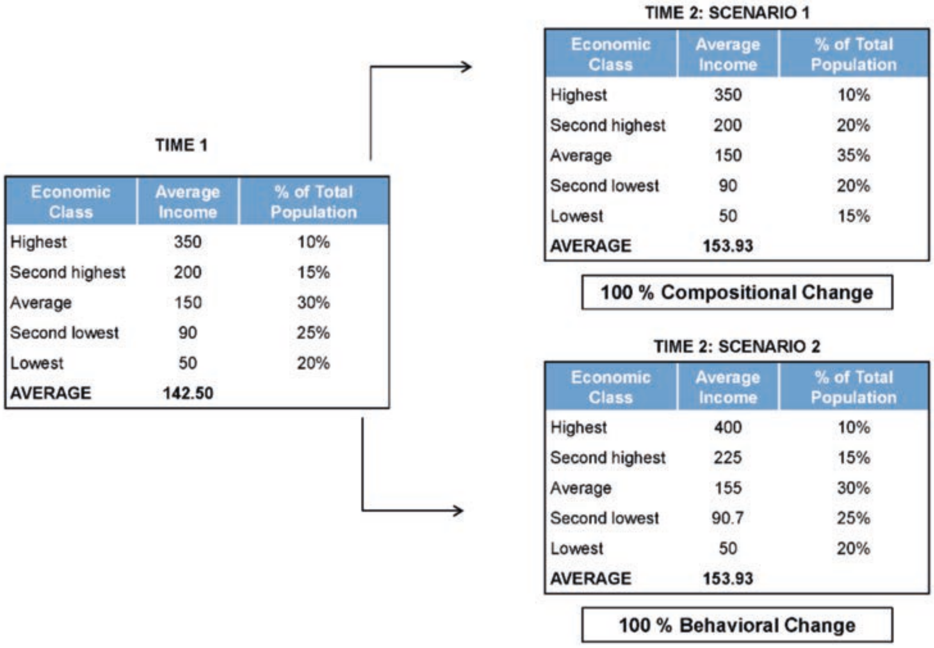
While the previous figure shows the groups leading the change, Table 2 below shows the leading processes. In particular, it shows the difference between compositional and behavioral effects. Imagine a country where the monthly income (in thousands of FCFA[7]) is initially 142.5. This income is the weighted average of incomes across all of the economic classes making up the national population, from the richest to the poorest. The table shows, for years 1 and 2, respectively, the average income and the size of each class. Thus, in year 1, the poorest category represented 20% of the population and earned an average of 50,000 FCFA.

---

[7] FCFA is the Cameroonian national currency, 1US$ ≈ 500 FCFA.

In year 2, the table shows two possible scenarios of change. Both scenarios lead to the same aggregate result, a growth in national income from 142.5 to 152.9. Yet they are qualitatively different. Under scenario 1, the average incomes of different classes did not change; what changed was the relative size of the different classes, including the poor's share of the national population, which fell from 20% to 15%. In contrast, scenario 2 does not involve a compositional change; what changed instead were the average incomes of some groups. The richest economic group (350 on average in year 1) became even richer (400), which fully explains the rise in the country's average income.

The two scenarios in Table 2 show extreme, textbook, illustrations (100% composition versus 100% behavior). In practice, compositional and behavioral change often occur simultaneously. One may thus end up with a situation where composition explains 30% of the change while behavioral change accounts for the rest. The value of a demographic decomposition is precisely to quantify these relative contributions.



TIME 1

Economic Class	Average Income	% of Total Population
Highest	350	10%
Second highest	200	15%
Average	150	30%
Second lowest	90	25%
Lowest	50	20%
AVERAGE	**142.50**	

TIME 2: SCENARIO 1

Economic Class	Average Income	% of Total Population
Highest	350	10%
Second highest	200	20%
Average	150	35%
Second lowest	90	20%
Lowest	50	15%
AVERAGE	**153.93**	

100 % Compositional Change

TIME 2: SCENARIO 2

Economic Class	Average Income	% of Total Population
Highest	400	10%
Second highest	225	15%
Average	155	30%
Second lowest	90.7	25%
Lowest	50	20%
AVERAGE	**153.93**	

100 % Behavioral Change

Table 2. Compositional vs behavioral change: An extreme case.

II.1.4 Mathematical formulation

In this first case, we focus on a national average; Y is expressed as a weighted average (by w_j) of the values of individual subpopulations (y_j).

$$\boldsymbol{Y_t = \sum w_{jt} * y_{jt}} \tag{II.2}$$

In this formula, a national change can be broken down into two components:

$$\boldsymbol{\Delta Y = \underbrace{\sum \bar{y}_j * \Delta w_j}_{\textit{Compositional effect}} + \underbrace{\sum \bar{w}_j * \Delta y_j}_{\textit{Behavioral effect}}} \tag{II.3}$$

This basic decomposition identifies two sources of change. The first, the compositional effect, measures the change in the relative representation of the various subpopulations. This change in composition affects the national average through a mechanical change in the weights and the importance of different subpopulations.

The second source of change, behavior, is less mechanical. It shows a real change of mortality within one or more groups. If the mortality of a group increases, other things being equal, the national mortality will increase. What changes here is not the relative size of groups, but rather the mortality levels within some or all subpopulations.

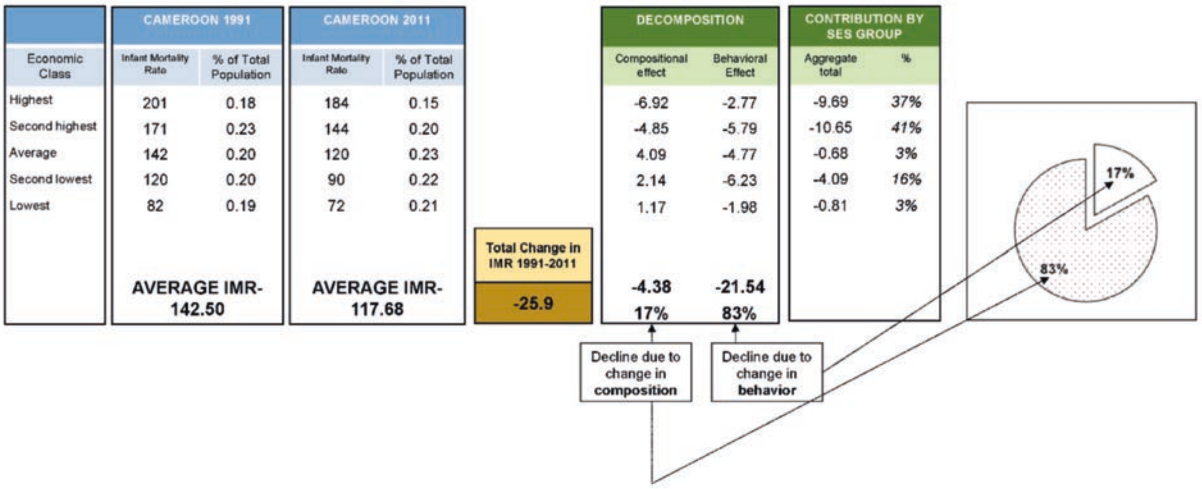
II.1.5 Application

In practice, one implements a decomposition analysis in four main steps: defining the problem; calculating national averages and change therein; decomposing the total change; and presenting/ discussing the findings.

Defining the problem: Here, the researcher specifies the nature of the substantive (dependent) variable, the classification (independent) variable, and the period. In our example below, the substantive variable is child mortality, the classification variable is socioeconomic status, and the study period is 1991-2001.

Calculating the national averages: The national averages are calculated using formula #1, the year-specific information about group size (w_j), and the value of the dependent variable (y_j). These data should be available for first and the last year of the period at least. The statistical procedures to generate these results are a simple frequency analysis (for w_j) and an equally simple comparison of means (for y_j). These basic data can come directly from published reports or automatic compilations available online (e.g., www.Statcompiler.com).

Doing the actual decomposition: Using the available annual information, one can simply apply the formulation in equation 3. Given the repetitive nature of the calculations, we recommend the use of spreadsheet software like Excel but other software or personal programs can be used. Table 3 summarizes the basic data for the calculations.



	CAMEROON 1991		CAMEROON 2011		DECOMPOSITION		CONTRIBUTION BY SES GROUP	
Economic Class	Infant Mortality Rate	% of Total Population	Infant Mortality Rate	% of Total Population	Compositional effect	Behavioral Effect	Aggregate total	%
Highest	201	0.18	184	0.15	-6.92	-2.77	-9.69	37%
Second highest	171	0.23	144	0.20	-4.85	-5.79	-10.65	41%
Average	142	0.20	120	0.23	4.09	-4.77	-0.68	3%
Second lowest	120	0.20	90	0.22	2.14	-6.23	-4.09	16%
Lowest	82	0.19	72	0.21	1.17	-1.98	-0.81	3%
	AVERAGE IMR- 142.50		AVERAGE IMR- 117.68		-4.38 17%	-21.54 83%		

Total Change in IMR 1991-2011

-25.9

Table 3. Decomposing a change in child mortality, Cameroon 1991-2011.

Again, the first four columns include the basic information needed for the analysis, including the child mortality rate for each of the social classes and the percentage of children in these classes for the years 1991 and 2011. Using these data, one can easily calculate the national average for each year and the change over the study period, from 143.6 in 1991 to 117.7 in 2011. The difference here is -25.9.

The final step is to explain this decline in mortality. We do so by applying formula #3. The results in this case show that 17% of the decline stemmed from compositional change. Note, for example, how the percentage of children in poorer families dropped by 18 to 15%, while the percentage of children in the richest families rose from 19 to 21%. The remainder of the change (83%) reflects a behavioral effect. Notice in particular how mortality rates declined across all social classes, including the poorest.

Presenting the findings

For a scientific audience, the presentation can simply use a table like the one in Section II.2, with the discussion focusing on both the processes and groups driving the change.

Leading processes: The percentages at the bottom of the table indicate the contributions of the two competing processes (composition and behavior) to the total change in national mortality. In this case, these relative contributions are 17% for composition and 83% for behavior.

Leading groups. In addition to clarifying the contributions of various processes, a decomposition analysis reveals which groups drove the change. In our case, the contribution of the lower income group is -9.69, 37% of the total change. This total contribution reflected effects stemming from changes in group size (-6.92) and behavior (-2.77). The other income groups contributed 41%, 3%, 16%, and 3%, respectively.

In a decomposition analysis, the sum of these contributions is always 100%. However, individual contributions can be negative (less than 0%) or greater than 100%. A negative percentage indicates a contribution that goes in the opposite direction of the general change. For example, a group making a negative contribution to the national decline in mortality means that this group's effect worked to increase mortality, i.e., it worked against the prevailing trend. In contrast, a percentage greater than 100% indicates that group or process accounted for more than the total change observed at the national level. The national change would have been even greater if it had depended only on this group, and if it had not been offset by the opposite influences of other groups.

For a non-scientific audience, it might be useful to rely on graphs rather than a table. For instance, one might use pie charts or 100% (stacked area) histograms. Such diagrams offer nice summaries that clearly identify the dominant drivers of change.

Policy implications

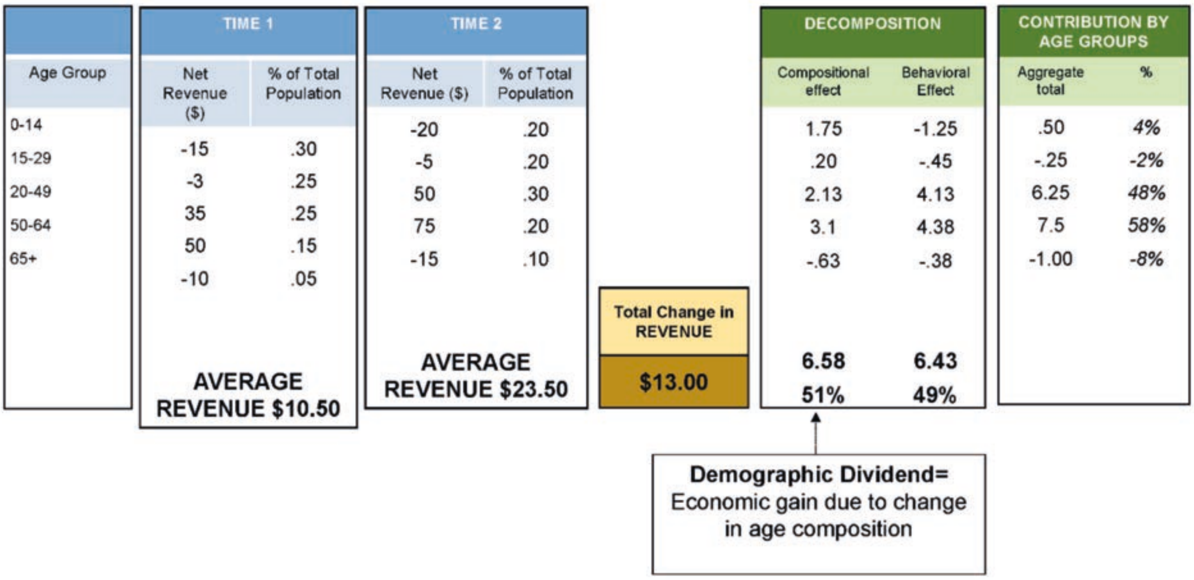
The policies to recommend will depend on whether a social change reflects a compositional or behavioral effect. If mortality is driven by a compositional effect, the appropriate response is to target the vulnerable segments of the population. If, on the other hand, the change reflects behavioral effects, targeting becomes less appropriate, and one would consider broader-based interventions to all families and their children. Later, we will see how to refine policy recommendations with more detailed decompositions.

II.1.6 Application to the demographic dividend

Basic demographic decomposition can apply to the study of demographic dividend. Indeed, it is the logic behind the National Transfer Accounts (NTA) methodology often used in this field (Mason and Lee 2005). At the heart of this method is the simple observation that income, consumption, and savings vary with age (Figure 5). In particular, the income/consumption balance tends to be negative among younger and older populations but positive among the middle-aged adult population. Intuitively, therefore, the greater the share of adults in the national population, the higher the national income.

The parallel between NTA and demographic decomposition is obvious. One only needs to note here that ‘age group’ is the classification variable and income-consumption is the substantive variable (see Table 4 below). Thus, the composition effect is the demographic dividend, or at least the mechanical component of demographic dividend. Indeed, the application of decomposition could enhance the standard NTA analysis by (a) considering cases where consumption/income profiles do change with age, rather than assuming that they remain fixed; (b) quantifying the dividend both in absolute terms and also in relative terms by comparing it to changes induced by historical evolution in income and consumption; and (c) exploring the contribution of different age groups to the dividend.

Figure 5. Income and consumption profiles by age. Hypothetical data.



	TIME 1		TIME 2			DECOMPOSITION		CONTRIBUTION BY AGE GROUPS	
Age Group	Net Revenue ($)	% of Total Population	Net Revenue ($)	% of Total Population		Compositional effect	Behavioral Effect	Aggregate total	%
0-14	-15	.30	-20	.20		1.75	-1.25	.50	4%
15-29	-3	.25	-5	.20		.20	-.45	-.25	-2%
20-49	35	.25	50	.30		2.13	4.13	6.25	48%
50-64	50	.15	75	.20		3.1	4.38	7.5	58%
65+	-10	.05	-15	.10		-.63	-.38	-1.00	-8%
	AVERAGE REVENUE $10.50		AVERAGE REVENUE $23.50		Total Change in REVENUE $13.00	6.58 51%	6.43 49%		

Demographic Dividend=
Economic gain due to change in age composition

Table 4. Estimating the demographic dividend with a demographic decomposition approach.

II.2 Derived demographic decompositions

II.2.1 Decomposing a difference

Let us return to our case study on mortality. Imagine a researcher who is not interested in historical change but, rather, in the mortality difference between two countries or provinces. Fortunately, the same approach works, as long as the research has data on the size and mortality rate of each social class within each of the two provinces. The calculations proceed in the same way as previously described.

II.2.2 Decomposing inequality

In our mortality case study, the dependent variable was an average—specifically, the average level of child mortality in the country. There may be cases where a researcher is interested not in the average but in the mortality inequality across regions, economic classes, education levels, or racial groups, for instance. At the global level, a researcher may seek to account for the historical change in global income inequality, and she can use basic demographic decomposition to do so. In Firebaugh and Goesling (2014), for instance, formulate economic inequality as a function of differences in national populations (p) and per capita income (i). Taking the mean log deviation (MLD) as a measure of inequality, they express the global level of inequality as follows:

$$\boldsymbol{I}_t = \sum w_{jt} * ln(^1/_{r_{jt}}) \qquad [\text{II.4}]$$

In this context, the historical change in global inequality is:

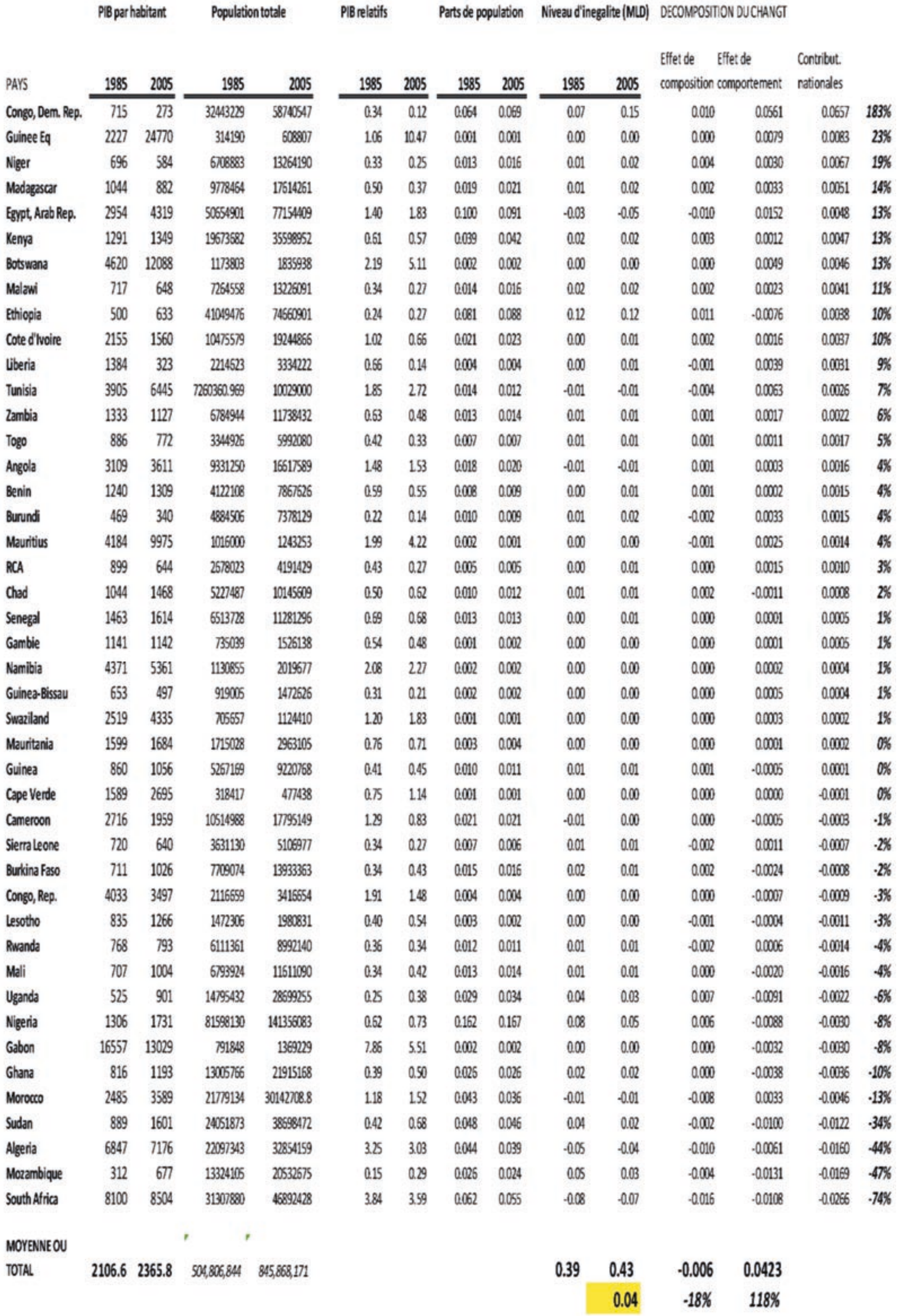
$$\Delta\boldsymbol{I} = \underbrace{\sum(\bar{r}_j - \overline{\ln r_j})\Delta w_j}_{\textit{Compositional effect}} + \underbrace{\sum(\overline{w_j r_j} - \overline{w_j})\Delta ln(r_j)}_{\textit{Behavioral effect}} \qquad [\text{II.5}]$$

We use this approach to analyze the historical change in income inequality between African countries from 1985 to 2005 (Table 5). Between these two years, economic inequality, as measured by the Mean Log Deviation in GDP per capita, increased from 0.391 to 0.425, or by about 9%. We apply a decomposition analysis to identify the sources of this economic divergence, looking at the respective contributions of compositional and behavioral forces. The analysis can also show individual country contributions, highlighting the countries that made the greatest contributions to increasing inequality and those that instead worked to reduce it. The results of this analysis are in Table 5.

The first four columns in the table show national data on GDP per capita and total population for 1985 and 2005, respectively. These nominal data are transformed into relative values: raw population numbers, for instance, are turned into the share of the total African population living in a country as obtained by dividing the population of each country's by the total population of Africa during the indexed of year. Similarly, GDP is converted into relative GDP by dividing the country's GDP by the weighted average of the African GDP during that year. Countries where the relative GDP value is greater than 1 are richer than average; they are poorer than average if that value is below 1. Using these relative values, one can calculate inequality, and the values obtained here are 0.39 and 0.43, i.e., a nominal increase of about 0.04. This nominal change is then decomposed into a compositional effect (-18%) and a behavioral effect (118%).

The compositional effect suggests that, during the investigated period, African populations grew at different rates, and this altered the relative sizes of individual countries. The pattern of these changes contributed to reduce inequality, suggesting that, perhaps, the countries at economic extremes (the very rich or the very poor) grew more slowly than others. At any rate, this differential population growth helped reduce economic inequality. However, this influence was overshadowed by the changes in economic circumstances, with countries' GDPs becoming even more unequal in 2005 than they were in 1985. At this stage, it is not possible to say why these differences in economic performance or population growth occurred. As we will see later, more detailed decompositions can shed light on these questions. For now, this first analysis is a good start.

It is also useful to examine the individual contributions of each country to the change in inequality as shown in the last column of Table 5.



	PIB par habitant		Population totale		PIB relatifs		Parts de population		Niveau d'inegalite (MLD)		DECOMPOSITION DU CHANGT			
PAYS	1985	2005	1985	2005	1985	2005	1985	2005	1985	2005	Effet de composition	Effet de comportement	Contribut. nationales	
Congo, Dem. Rep.	715	273	32443229	58740547	0.34	0.12	0.064	0.069	0.07	0.15	0.010	0.0561	0.0657	*183%*
Guinee Eq	2227	24770	314190	608807	1.06	10.47	0.001	0.001	0.00	0.00	0.000	0.0079	0.0083	*23%*
Niger	696	584	6708883	13264190	0.33	0.25	0.013	0.016	0.01	0.02	0.004	0.0030	0.0067	*19%*
Madagascar	1044	882	9778464	17614261	0.50	0.37	0.019	0.021	0.01	0.02	0.002	0.0033	0.0051	*14%*
Egypt, Arab Rep.	2954	4319	50654901	77154409	1.40	1.83	0.100	0.091	-0.03	-0.05	-0.010	0.0152	0.0048	*13%*
Kenya	1291	1349	19673682	35598952	0.61	0.57	0.039	0.042	0.02	0.02	0.003	0.0012	0.0047	*13%*
Botswana	4620	12088	1173803	1835938	2.19	5.11	0.002	0.002	0.00	0.00	0.000	0.0049	0.0046	*13%*
Malawi	717	648	7264558	13226091	0.34	0.27	0.014	0.016	0.02	0.02	0.002	0.0023	0.0041	*11%*
Ethiopia	500	633	41049476	74660901	0.24	0.27	0.081	0.088	0.12	0.12	0.011	-0.0076	0.0038	*10%*
Cote d'Ivoire	2155	1560	10475579	19244866	1.02	0.66	0.021	0.023	0.00	0.01	0.002	0.0016	0.0037	*10%*
Liberia	1384	323	2214623	3334222	0.66	0.14	0.004	0.004	0.00	0.01	-0.001	0.0039	0.0031	*9%*
Tunisia	3905	6445	7260360.969	10029000	1.85	2.72	0.014	0.012	-0.01	-0.01	-0.004	0.0063	0.0026	*7%*
Zambia	1333	1127	6784944	11738432	0.63	0.48	0.013	0.014	0.01	0.01	0.001	0.0017	0.0022	*6%*
Togo	886	772	3344926	5992080	0.42	0.33	0.007	0.007	0.01	0.01	0.001	0.0011	0.0017	*5%*
Angola	3109	3611	9331250	16617589	1.48	1.53	0.018	0.020	-0.01	-0.01	0.001	0.0003	0.0016	*4%*
Benin	1240	1309	4122108	7867626	0.59	0.55	0.008	0.009	0.00	0.01	0.001	0.0002	0.0015	*4%*
Burundi	469	340	4884506	7378129	0.22	0.14	0.010	0.009	0.01	0.02	-0.002	0.0033	0.0015	*4%*
Mauritius	4184	9975	1016000	1243253	1.99	4.22	0.002	0.001	0.00	0.00	-0.001	0.0025	0.0014	*4%*
RCA	899	644	2678023	4191429	0.43	0.27	0.005	0.005	0.00	0.01	0.000	0.0015	0.0010	*3%*
Chad	1044	1468	5227487	10145609	0.50	0.62	0.010	0.012	0.01	0.01	0.002	-0.0011	0.0008	*2%*
Senegal	1463	1614	6513728	11281296	0.69	0.68	0.013	0.013	0.00	0.01	0.000	0.0001	0.0005	*1%*
Gambie	1141	1142	735039	1526138	0.54	0.48	0.001	0.002	0.00	0.00	0.000	0.0001	0.0005	*1%*
Namibia	4371	5361	1130855	2019677	2.08	2.27	0.002	0.002	0.00	0.00	0.000	0.0002	0.0004	*1%*
Guinea-Bissau	653	497	919005	1472626	0.31	0.21	0.002	0.002	0.00	0.00	0.000	0.0005	0.0004	*1%*
Swaziland	2519	4335	705657	1124410	1.20	1.83	0.001	0.001	0.00	0.00	0.000	0.0003	0.0002	*1%*
Mauritania	1599	1684	1715028	2963105	0.76	0.71	0.003	0.004	0.00	0.00	0.000	0.0001	0.0002	*0%*
Guinea	860	1056	5267169	9220768	0.41	0.45	0.010	0.011	0.01	0.01	0.001	-0.0005	0.0001	*0%*
Cape Verde	1589	2695	318417	477438	0.75	1.14	0.001	0.001	0.00	0.00	0.000	0.0000	-0.0001	*0%*
Cameroon	2716	1959	10514988	17795149	1.29	0.83	0.021	0.021	-0.01	0.00	0.000	-0.0005	-0.0003	*-1%*
Sierra Leone	720	640	3631130	5106977	0.34	0.27	0.007	0.006	0.01	0.01	-0.002	0.0011	-0.0007	*-2%*
Burkina Faso	711	1026	7709074	13933363	0.34	0.43	0.015	0.016	0.02	0.01	0.002	-0.0024	-0.0008	*-2%*
Congo, Rep.	4033	3497	2116659	3416654	1.91	1.48	0.004	0.004	0.00	0.00	0.000	-0.0007	-0.0009	*-3%*
Lesotho	835	1266	1472306	1980831	0.40	0.54	0.003	0.002	0.00	0.00	-0.001	-0.0004	-0.0011	*-3%*
Rwanda	768	793	6111361	8992140	0.36	0.34	0.012	0.011	0.01	0.01	-0.002	0.0006	-0.0014	*-4%*
Mali	707	1004	6793924	11611090	0.34	0.42	0.013	0.014	0.01	0.01	0.000	-0.0020	-0.0016	*-4%*
Uganda	525	901	14795432	28699255	0.25	0.38	0.029	0.034	0.04	0.03	0.007	-0.0091	-0.0022	*-6%*
Nigeria	1306	1731	81598130	141356083	0.62	0.73	0.162	0.167	0.08	0.05	0.006	-0.0088	-0.0030	*-8%*
Gabon	16557	13029	791848	1369229	7.86	5.51	0.002	0.002	0.00	0.00	0.000	-0.0032	-0.0030	*-8%*
Ghana	816	1193	13005766	21915168	0.39	0.50	0.025	0.026	0.02	0.02	0.000	-0.0038	-0.0036	*-10%*
Morocco	2485	3589	21779134	30142708.8	1.18	1.52	0.043	0.036	-0.01	-0.01	-0.008	0.0033	-0.0046	*-13%*
Sudan	889	1601	24051873	38698472	0.42	0.68	0.048	0.046	0.04	0.02	-0.002	-0.0100	-0.0122	*-34%*
Algeria	6847	7176	22097343	32854159	3.25	3.03	0.044	0.039	-0.05	-0.04	-0.010	-0.0061	-0.0160	*-44%*
Mozambique	312	677	13324105	20532675	0.15	0.29	0.026	0.024	0.05	0.03	-0.004	-0.0131	-0.0169	*-47%*
South Africa	8100	8504	31307880	46892428	3.84	3.59	0.062	0.055	-0.08	-0.07	-0.016	-0.0108	-0.0266	*-74%*
MOYENNE OU TOTAL	**2106.6**	**2365.8**	*504,806,844*	*845,868,171*					**0.39**	**0.43**	**-0.006**	**0.0423**		
										0.04	*-18%*	*118%*		

Table 5. Decomposition of change in economic inequality between African countries (1985-2005).

II.2.3 Ordinal decomposition

The ordinal decomposition is very similar to the demographic decomposition, except that the classification variable is ordinal. We illustrate this variant with a study of total fertility rate (TFR). TFR values, roughly the average number of births per woman in a country, are the sum of all of the fertility rates found in all of the relevant age groups—in this case, the ages between 15 and 49 during which women are assumed to be of reproductive age:

$$Y_t = \Sigma\, y_{at} \tag{7}$$

Suppose a country's TFR were to fall over time. A researcher may wish to understand why the decline occurred, specifically whether the decline occurred among all groups (a quantum effect) or whether women mostly altered the ages at which they bear children, perhaps postponing childbearing (a tempo effect).

Using temporal decomposition, the researcher can express age-specific fertility rates relative to the fertility observed in the oldest age group . For example, the fertility among the 15-19 year olds will be expressed in terms of the average fertility for all women aged 20 to 49 ($y_{15\text{-}19} = r_{15\text{-}19} * y_{20\text{-}49}$);

$$y_a / y_{a+} = r_a \rightarrow y_a = r_a * y_{a+}$$

The TFR can be re-expressed as follows:

$$Y_t = \Sigma\, r_{at} * y_{a+t} \tag{II.6}$$

This new formulation helps decompose the change in the TFR into two terms that reflect the quantum and tempo, respectively.

$$\Delta \boldsymbol{Y} = \sum \overline{\boldsymbol{y}}_{\boldsymbol{a}+} * \Delta \boldsymbol{r}_{\boldsymbol{a}} + \sum \overline{\boldsymbol{r}}_{\boldsymbol{a}} * \Delta \boldsymbol{y}_{\boldsymbol{a}+} \tag{II.7}$$

Illustration

A paper by Ouedraogo (2012) (data not shown here) looked at patterns of change in age-specific fertility curves in Cameroon from 1991, 1998, and 2004. An examination of these curves shows some parallelism. If the change occurs at the same rate across all age groups, the fertility curves should be perfectly parallel. Otherwise, it becomes difficult to succinctly describe the differences between the various curves. For a researcher wishing to assess how much the decline in births occurs disproportionately among younger age groups, the temporal decomposition is a good tool.

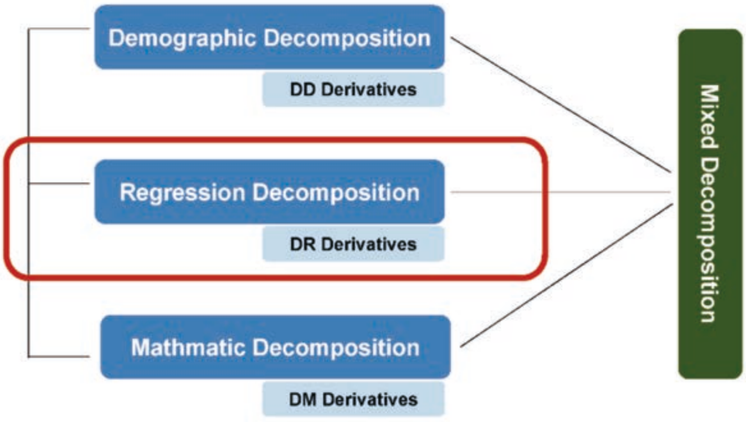
Calculations here show a decline driven by a quantum effect (90%): the decline affected all age groups, but, as the chart shows, it was larger among younger women, suggesting some postponement of births (10%). Note that this analysis could be repeated using SES, for example, as a classification variable to study the dispersion of reproduction across social classes.

II.2.4 Nested decomposition

This extension offers more detail by expressing demographic change (the composition component) in terms of its constitutive elementary processes such as fertility, mortality, and migration. To apply this decomposition, one obviously needs much more data than for a simple decomposition.

In addition, one can also envision an extension based on nesting categories of two classification variables. For instance, the researcher might be interested in the intersection of age and SES and would group people in terms of these two variables. Thus, instead of distinguishing only between the poor, middle class, and rich, one would further distinguish between the poor who are young, middle-aged, or old. Thus, if each of the two classification variables had three categories, the researcher would end up with a categorization of her population including six categories.

Chapter III
Regression decomposition



III.1 Simple regression decomposition

III.1.1 Problem type

The main difference between demographic decomposition and simple regression decomposition is in the kind of independent variable involved. It is nominal in a demographic decomposition but quantitative here. Moreover, the independent variable is linked to the dependent variable through a statistical relationship (derived from regression analysis) rather than an arithmetic one. The equation has a regression coefficient (the effect of the independent variable on Y) and an intercept. The error term is omitted here for now.

III.1.2 Formulation

The generic form is Y = f (α, β, X) but a more explicit formulation is as follows:

$$y_t = \alpha_t + \beta_t X_t \qquad [III.1]$$

In this case, the decomposition seeks to explain the change in the dependent variable based on the change in the various parameters of the regression equation. This change is expressed as follows:

$$\Delta Y = \underset{\uparrow \atop \boxed{\text{Change in baseline}}}{\Delta\alpha} + \underset{\uparrow \atop \boxed{\text{Change in effect magnitude of } X \text{ level}}}{\bar{\beta}\Delta X} + \underset{\uparrow \atop \boxed{\text{Change in effect of } X}}{\bar{X}\Delta\beta} \qquad [III.2]$$

Once again, the same procedure can apply to both cross-sectional analysis (the difference between two groups in a given year) and longitudinal analysis (the change experienced by one group between years). The approach is the same; only the interpretations differ.

Illustration

A classic realm of application for regression decomposition is the study of income/wage differentials (say, between men and women) and the extent to which they reflect discrimination in the labor market. Differences could stem from discrimination (different returns to education for men and women) but it is also possible that men and women enter the workforce with different education levels. One must therefore estimate how the wage differentials reflect discrimination rather than differences in human capital. These two possibilities are explored below. The formal analysis (III.3) consists of writing the earning equations for males (h) and females (f) and then taking the difference between these two equations.

$$Y_h = \alpha_h + \beta_h * X_h$$

$$Y_f = \alpha_f + \beta_f * X_f$$

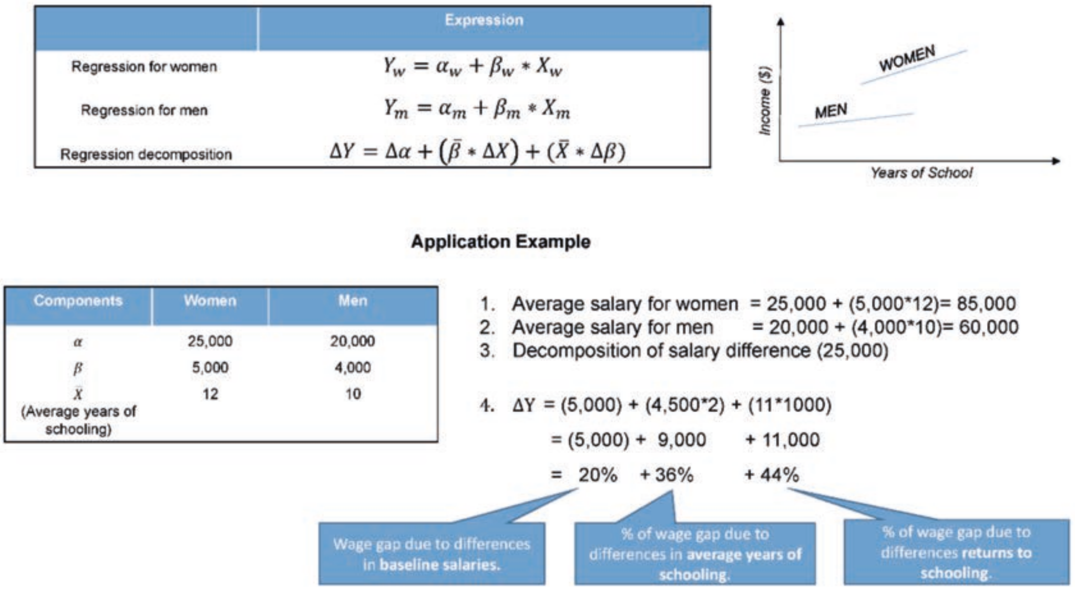
$$\Delta Y = \Delta\alpha + \bar{\beta}\Delta X + \bar{X}\Delta\beta \qquad [III.3]$$

III.1.3 Application

Table 6 illustrates with a numerical example showing earnings by level of education. In this case, the average return on a year of education is 4,000 FCFA for men and 5,000 FCFA for women. The base salaries are 20,000 and 25,000 for men and women, respectively. With the average education levels being 10 and 12 years for men and women, respectively, the average salary of a man is 60,000, while the average woman would earn 85,000 FCFA, yielding an average salary difference of 25,000 between women and men. A decomposition analysis will show that this pay gap reflects a complex mix of forces:

- 20% of the gap comes from the gender difference in base salaries (Δα),
- 36% of the gap comes from the gender difference in levels of education, and
- 44% of the gap comes from the gender difference in returns to education.

In other words, even if one achieved parity in education, only 36% of the current wage inequality between men and women would disappear. The rest of the inequality that comes from discriminatory inequalities in basic salary and returns to schooling would remain.



	Expression
Regression for women	$Y_w = \alpha_w + \beta_w * X_w$
Regression for men	$Y_m = \alpha_m + \beta_m * X_m$
Regression decomposition	$\Delta Y = \Delta\alpha + (\bar{\beta} * \Delta X) + (\bar{X} * \Delta\beta)$

Application Example

Components	Women	Men
α	25,000	20,000
β	5,000	4,000
$\bar{X}$ (Average years of schooling)	12	10

1. Average salary for women = 25,000 + (5,000*12)= 85,000
2. Average salary for men = 20,000 + (4,000*10)= 60,000
3. Decomposition of salary difference (25,000)

4. ΔY = (5,000) + (4,500*2) + (11*1000)
= (5,000) + 9,000 + 11,000
= 20% + 36% + 44%

Wage gap due to differences in **baseline salaries.**

% of wage gap due to differences in **average years of schooling.**

% of wage gap due to differences **returns to schooling.**

Table 6. Regression decomposition regression for analyzing wage differentials between men and women.

III.2 Other regression decompositions

III.2.1 Curvilinear regression

Most processes of interest to social scientists do not fit the linear bivariate pattern posited in the previous section. A curvilinear, rather than a linear, specification is often more realistic. Fortunately, regression decomposition analysis easily extends to these situations. For instance, earnings can be written as a quadratic (rather than linear) function of education:

$$Y = \alpha + \beta_1 X + \beta_2 X^2 \qquad \text{(III.4)}$$

The decomposition analysis is as follows:

$$\Delta Y = \Delta\alpha + (\bar{X} * \Delta\beta_1) + (\overline{\beta_1} * \Delta X) + (\overline{X^2} * \Delta\beta_2) + (\overline{\beta_2} * 2\Delta X)$$

$$\Delta Y = \Delta\alpha + (\bar{X} * \Delta\beta_1) + (\overline{X^2} * \Delta\beta_2) + ((\overline{\beta_1} + 2\overline{\beta_2}) * \Delta X) \qquad \text{(III.5)}$$

To give an example, one can return to our study of wage differences between men and women, only this time, we set the effects of education on income to be curvilinear.

III.2.2 Multivariate regression

In other cases, the researcher wishes to estimate the contributions of multiple independent variables. This is feasible. The decomposition will simply include more than a single factor (and the additional factors will be other independent variable rather than the square term of the main independent variable). In the simplest of these situations, the researcher has two variables, rather than a single one. For instance, she might wish to examine how income is affected by both the level of education and the number of years of experience:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 \qquad \text{(III.6)}$$

The decomposition then writes as follows:

$$\Delta Y = \Delta\alpha + (\overline{X_1} * \Delta\beta_1) + (\overline{\beta_1} * \Delta X_1) + (\overline{X_2} * \Delta\beta_2) + (\overline{\beta_2} * \Delta X_2) \qquad \text{(III.7)}$$

III.2.3 Multilevel regression

This type of regression involves factors at two or more levels, e.g., individual and community levels. Research on schooling might for instance examine how the academic performance of students depends on their individual characteristics (level 1) but also the characteristics of the schools (level 2). The equations to estimate this model are as follows:

Level 1: $Y_{jk} = \beta_{0k} + \beta_{1k} * x_{jk} + r_{jk}$

Level 2: $\beta_{0k} = \gamma_{00} + \gamma_{01} Z_k + \mu_{0k}$

$B_{1k} = \gamma_{10} + \gamma_{11} Z_k + \mu_{1k}$

An integration of the values from level 2 into the level 1 equation yields a mixed equation that expresses individual performance based on the individual characteristics and those of the group at a particular time and their interactions:

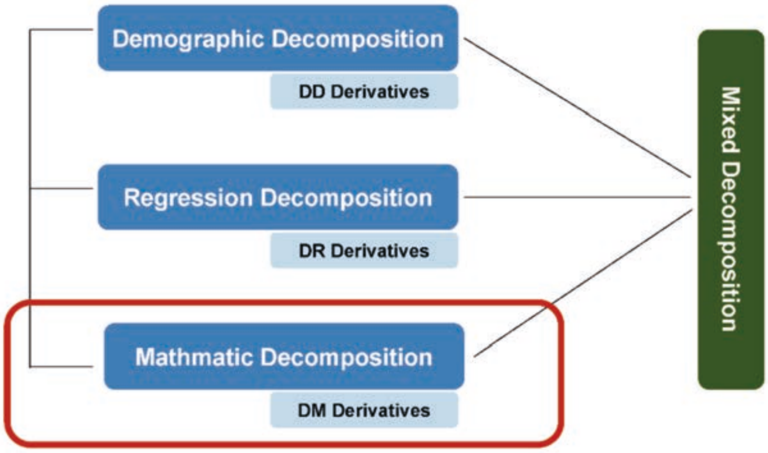
$$Y_{jk} = \gamma_{00} + \gamma_{10} x_{jk} + \gamma_{01} Z_k + \gamma_{11} Z_k x_{jk} + \mu_{0k} + \mu_{1k} x_{jk} + r_{jk} \qquad \text{(III.8)}$$

To investigate the change of Y_{jk} over time, one simply needs to differentiate the above formula and incorporate it into the second term of the equation [1].

III.3 Application to the demographic dividend

The study of demographic dividends is about how changes in national birth rates affects national levels of schooling. In essence, the researcher must investigate a relationship between two macro-outcomes. For instance, imagine a country where average family size declines from 6 to 4 children between 1990 and 2010 at the same time as the percentage of 6-10 year olds enrolled in school rises from 60 to 75%. A question one might ask is whether or how much these improvements in education came from the observed reduction in average family size. It is possible that the smaller family sizes played a role, but other factors could have also contributed. A macro-level correlation cannot apply because we are dealing with a single country with only two points in time. Furthermore, even if a researcher could implement, the macro-regression is open to criticism for lacking in detail and rigor. Instead, one can use decomposition to look at how the percentages of children living with different sibsizes has changed but also how the effects of sibsize itself has changed. Through decomposition analysis, one can estimate the relative contributions of several factors, but the researcher will need reliable information on how the statistical relationship between education and family size changed over time. In a recent study (Eloundou-Enyegue and Giroux 2012), we show how to combine the micro relation between education and fertility with information about fertility change to estimate the implications of demographic change on national school enrollment.

Chapter IV
Mathematical decomposition



IV.1 Simple mathematical decomposition

We have so far covered demographic and regression decompositions. In demographic decomposition, the dependent outcome is an average, and the function linking it to the independent variable is simply a weighing function. In regression decomposition, the dependent variable is a relationship linked to the independent variables through a statistical relationship, estimated via regression analysis. Next, we want to explore yet another case, where there is an exact mathematical relationship between the independent variables and the outcome. For example, the average per-capita income in a country equals the total income divided by the total population. A change in average income can only result from a change in total income (numerator) or in the size of the population (denominator).

IV.1.1 Problem type

This type of analysis applies to processes that involve a mathematical relationship between two or more social variables. GDP per capita is one example. It combines an economic component (GDP) and a demographic component (number of inhabitants), and it can be broken down into simple terms that show how the GDP per capita changes as either one of these two components varies. Other similar variables measure individual welfare by relating available resources to the population served.

IV.1.2 Mathematical formulation

For this example, we will use a slightly more complex example than GDP per capita. Consider the public education spending per child (r). This expenditure is positively related to the total level of national resources (g) and the percentage of resources allocated to education (k). On the other hand, it is inversely proportional to the number of school-aged children in the country (p).

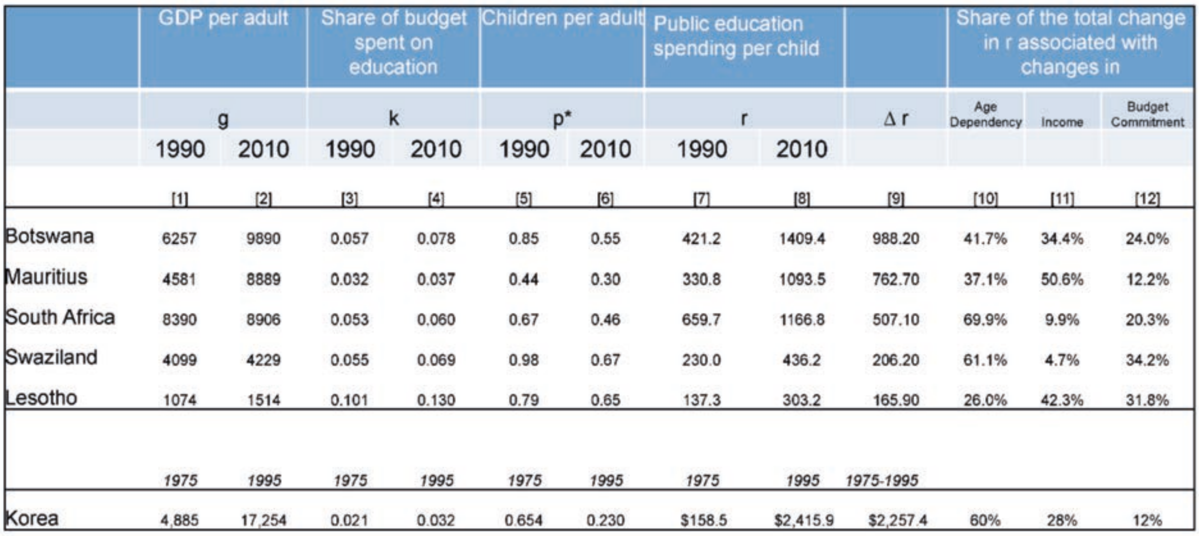
$$r = {}^{gk}/_{p} \tag{IV.1}$$

In this case, the historical change in this expense can be broken down as follows:

$$\Delta r \cong -[\underbrace{(\overline{kp/p^2})}_{\boxed{\text{Effect of change in population}}} * \Delta p] + [\underbrace{\overline{(k/p)}}_{\boxed{\text{Effect of change in national income}}} * \Delta g] + [\underbrace{\overline{(g/p)}}_{\boxed{\text{Effect of change in share of budget to education}}} * \Delta k] \tag{IV.2}$$

IV.1.3 Application

Using World Bank statistics (2014), Eloundou, Tenikue, and Ryu (2014) decompose the changes in public expenditure per child among Southern Africa countries between 1990 and 2010 and compare the results to the results of South Korea between 1975 and 1995. The results show that, over their respective study periods, South Korea's economy cumulatively grew by 250%, much faster than the growth observed in Southern African nations (3-94%). Its ratio of youth to adults also fell faster, from 0.65 to 0.23, compared to South Africa's decline from 0.67 to 0.46. Despite such demographic and economic differences, the relative contributions of demographic change to the gains in r values were similar (60% for Korea versus 70 % and 61% in South Africa and Swaziland, respectively). The study also noted an often-overlooked fact: many African countries allocate a larger portion of their national budget to education (e.g., 10% in Lesotho and almost 6% in Botswana versus 2% in Korea). It is therefore unsurprising that budget decisions made larger contributions to improving the r values in Lesotho (32%) and Botswana (24 %) than in South Korea (12%).



	GDP per adult		Share of budget spent on education		Children per adult		Public education spending per child			Share of the total change in r associated with changes in		
	g		k		p*		r		Δ r	Age Dependency	Income	Budget Commitment
	1990	2010	1990	2010	1990	2010	1990	2010				
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
Botswana	6257	9890	0.057	0.078	0.85	0.55	421.2	1409.4	988.20	41.7%	34.4%	24.0%
Mauritius	4581	8889	0.032	0.037	0.44	0.30	330.8	1093.5	762.70	37.1%	50.6%	12.2%
South Africa	8390	8906	0.053	0.060	0.67	0.46	659.7	1166.8	507.10	69.9%	9.9%	20.3%
Swaziland	4099	4229	0.055	0.069	0.98	0.67	230.0	436.2	206.20	61.1%	4.7%	34.2%
Lesotho	1074	1514	0.101	0.130	0.79	0.65	137.3	303.2	165.90	26.0%	42.3%	31.8%
	1975	*1995*	*1975*	*1995*	*1975*	*1995*	*1975*	*1995*	*1975-1995*			
Korea	4,885	17,254	0.021	0.032	0.654	0.230	$158.5	$2,415.9	$2,257.4	60%	28%	12%

Table 7. Mathematical decomposition of trends in public spending per child (South Korea versus the 'vanguard' countries of Africa).

IV.2 Derived mathematical decomposition

IV.2.1 Extended mathematical chain

Let us return to the mathematical decomposition of GDP per capita as introduced in section IV.1.1. This first decomposition usefully describes the contributions two components (population and GDP). However, is not very informative because its two components do not refer to key decision variables but also because these variables themselves need fuller exploration. With only a light transformation of these initial mathematical expressions, we can get a formula that is slightly longer but conceptually richer. Specifically, the initial formula of per capita GDP

$Y = {}^{G}/_{P}$ was differentiated as

$$\Delta Y = \left(\bar{G} * \Delta(1/P)\right) + ((\overline{1/P}) * \Delta G) \tag{IV.3}$$

which can be rewritten more interestingly as

$$Y = \frac{G}{P} = \frac{G}{A} * \frac{A}{P} = \pi * \alpha \tag{IV.4}$$

where G and P represent mean the national income and the total national population, respectively. The new term introduced is A, the active (adult) population in the country. In this new formula, G/A (or) refers to adult productivity, and it is a conceptually interesting variable. It is frequently cited in economic growth theories and may be shaped through specific policies to raise productivity through education, research, and technological development. Also, the new term A/P (or α) refers to the population's age structure—specifically, the ratio of adults to the total population, a core variable in demographic dividend theory. Thus, our analyst now has two very interesting variables (and α), and she can break down the changes in per capita GDP in terms of these two variables.

$$\Delta Y = (\bar{\pi} * \Delta\alpha) + (\bar{\alpha} * \Delta\pi) \tag{IV.5}$$

Obviously, this new expression can itself expand further to yield a more detailed decomposition, with better potential to inform policy decision-making. For example, adult productivity can be split into two components: the adult unemployment rate and the productivity of adult workers:

$$Y = \frac{G}{P} = \frac{G}{E} * \frac{E}{A} * \frac{A}{P} = \rho * \varepsilon * \alpha \tag{IV.6}$$

And its historical change can be represented as follows:

$$\Delta Y = (\bar{\varepsilon}\bar{\rho} * \Delta\alpha) + (\bar{\rho}\,\overline{\alpha} * \Delta\epsilon) + (\bar{\varepsilon}\,\overline{\alpha} * \Delta\rho) \tag{IV.7}$$

Similarly, the adult ratio of the total population can be expressed as a function of the youth and the elderly populations, respectively. How far the researcher extends this development depends on how much detail she needs, and on data availability, the policy relevance and the tractability of the added terms.

Students of economic growth can apply the same logic in decomposing popular formulations of economic performance such as the Cobb-Douglas function, which expresses growth based on physical capital (K), human capital (h), employment (L), and total factor productivity (A):

$$Y = AK^{\alpha}(hL)^{\beta} \tag{IV.8}$$

From this formula, one can decompose economic growth as

$$\Delta y_t \cong \Delta A_t + \alpha\Delta k_t + (1-\alpha)\Delta h_t + \Delta l_t + \Delta w_t \tag{IV.9}$$

where y_t is the GDP per capita (Y/P)

k is the stock of physical capital per employee (K/L)

h is human capital

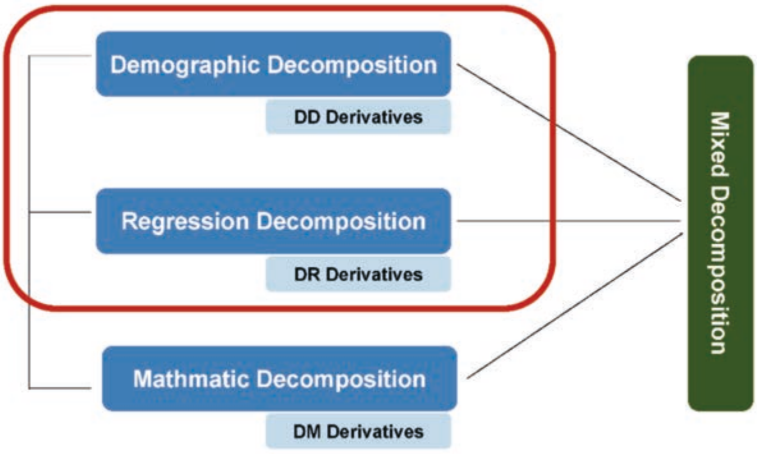
l is the employment rate (L/W)

w is the population's age structure (W/P)

A is the total factor productivity (TFP)

Clearly, our analyst has several options to decompose economic growth and education. As long as these decompositions include the effects of population age structure, then they help assess the magnitude of the demographic dividend.

Chapter V
Combination of demographic and regression decompositions



Elementary decompositions have the merit of being simple, but they lack in detail. This problem is solved with mixed decompositions combining two or more elementary forms of decomposition. One can combine demographic and regression decompositions to get a more detailed form as described in this chapter. Starting from a simple demographic decomposition, we will see how to expand each of the main components, starting with the behavioral effect (Δy_j) and then the composition effect (Δw_j).

V.1 Extension of behavior effect

V.1.1 General presentation

Let us start with the basic formula (3), which expresses a change in any national outcome in terms of the composition and behavior of various groups:

$$\Delta \boldsymbol{Y} = \underbrace{\left[\sum \bar{y}_j * \Delta w_j\right]}_{\textit{Compositional effect}} + \underbrace{\left[\sum \bar{\boldsymbol{w}}_{\boldsymbol{j}} * \Delta \boldsymbol{y}_{\boldsymbol{j}}\right]}_{\textit{Behavioral effect}}$$

This formula's extension can express the behavior of any given group (y_j) as a function of one or more other variables. In a simple regression analysis,

$$y_j = \alpha + \beta x_j + \mu^j \quad \text{where,}$$

- α (intercept) is the baseline performance, when x = 0,
- β is the increased mortality associated with a unit increase in X, and
- μ_j is the error term (the over-performance or under-performance of the group or the residual effect of factors other than X, that are not considered in the analysis).

The change in the value of y_j between two periods is thus:

$$\Delta y_j = \Delta\alpha + \bar{\beta}\Delta x_j + \bar{x}_j\Delta\beta + \Delta\mu_j \qquad [V.1]$$

By inserting [V.1] in the basic equation (3), we get

$$\Delta \boldsymbol{Y} = \underbrace{\sum \bar{\boldsymbol{y}}_j \Delta \boldsymbol{w}_j}_{\substack{A \\ \uparrow \\ \boxed{\textit{Compositional Effect}}}} + \underbrace{\underbrace{\sum \bar{\boldsymbol{w}}_j \Delta\boldsymbol{\alpha}}_{B1} + \underbrace{\sum \bar{\boldsymbol{w}}_j \bar{\boldsymbol{\beta}}\Delta \boldsymbol{x}_j}_{B2} + \underbrace{\sum \bar{\boldsymbol{w}}_j \bar{\boldsymbol{x}}_j \Delta\boldsymbol{\beta}}_{B3} + \underbrace{\sum \bar{\boldsymbol{w}}_j \Delta\boldsymbol{\mu}_j}_{B4}}_{\substack{\uparrow \\ \boxed{\textit{Behavioral effect}}}} \qquad [V.2]$$

- A (the composition effect) remains unchanged from the previous situation.
- B (the behavioral effect) is now divided into four sub-components that respectively reflect changes in
  - baseline performance (B1),
  - the level of the independent variable (B2),
  - the effect of the independent variable (B3) , and
  - the residual effect (B4).

V.1.2 Illustration

To illustrate this mixed decomposition, we can return to our first example in which we used simple demographic decomposition to study changes in infant mortality between 1990 and 2011 (Table 2). To recall, this simple decomposition showed a nearly 26-point decline in infant mortality, which reflected a mix of compositional change (17%) and behavioral change (83%). We can refine this analysis by considering that mortality varies with one's education level. We can thus express the mortality within each group as a function of the group's average education level. If one can estimate this relationship, and get a reliable estimate for the values of α (baseline mortality), β (the effect of education on mortality), and μ (the residual term), it becomes possible to refine our initial decomposition and get more detail about the forces driving change at the national level. This decomposition has the potential to reveal, in greater detail, the policy areas that were most influential in driving the change. Thus, in equation V.2,

- B1 (the baseline performance) reflects the improvement in the public sanitation and public health conditions that raise the minimum health standard of the population, regardless of education level;
- B2 measures the health improvements associated with the gains in the national level of education, assuming that the payoffs to education remain the same;
- B3 measures the improvement in the educational effects on health; and
- B4 measures the residual effect of other variables not considered.

Although this first example limits itself to a linear bivariate regression model (a single independent variable modeled linearly), one can easily imagine how this analysis can extend to cases where multiple independent variables or curvilinear influences are considered. The only difference is that the corresponding equations become longer and longer!

V.1.3 Comparison with national transfer accounts

The method of National Transfer Accounts is a core method for studying demographic dividends (Mason and Lee 2005). It builds on the simple idea that economic behavior (income, savings, transfers) varies systematically with age, even if the exact age profiles of economic behavior vary across countries (Figure 5). One can use a country's economic profile to estimate the balance between income and consumption at each age and combine the data for all age groups to calculate a cumulative balance for the entire country. Assuming a constant age pattern of economic behavior, any change in a population's age structure automatically changes the national balance between income and consumption.

However, it is not realistic to assume a constant age profile of economic behavior, especially during a demographic transition, given the large changes during a life course (e.g., age at marriage, duration of schooling, and the time spent bearing children versus participating in the labor force). It is therefore useful to consider situations where a country experiences changes in both the age structure and the consumption profile. Such situations are easily managed in a decomposition framework, and a full NTA approach can be seen as an example of mixed decomposition (Eloundou, Tenikue, Giroux 2014). Specifically, the national surplus in a given year is the weighted average of the specific surpluses for each age group.

$$\boldsymbol{S_t} = \sum \boldsymbol{w_{jt} s_{jt}} \qquad [V.3]$$

Change is thus expressed as

$$\Delta \boldsymbol{S} = \sum \bar{\boldsymbol{s}}_{\boldsymbol{j}} \Delta \boldsymbol{w_j} + \sum \bar{\boldsymbol{w}}_{\boldsymbol{j}} \Delta \boldsymbol{s_j} \qquad [V.4]$$

Now, we can additionally consider, as indicated above, that consumption/income profiles can change over time, perhaps even in response to a change in the age structure. It is therefore useful to explore the additional possibility that a change in the age structure can also have an effect on behavior. If the analyst can obtain, through regression analysis, a reliable estimate of the effect of the age structure of (J) on the economic behavior of each age group [Equation V.5 below], then the change in the national surplus can be estimated through a mixed decomposition.

$$S_{JT} = a_{jt} + b_{jt}J_t + e_{jt} \qquad [V.5]$$

$$\Delta \boldsymbol{s_j} = \Delta \boldsymbol{a_j} + \bar{\boldsymbol{J}} \Delta \boldsymbol{b_j} + \bar{\boldsymbol{b}} \Delta \boldsymbol{J} + \Delta \boldsymbol{e_j} \qquad [V.6]$$

$$\Delta \boldsymbol{S} = \sum \underset{\substack{\uparrow \\ \underbrace{\boxed{\textit{m-dividend}}}}}{\bar{\boldsymbol{s}}_{\boldsymbol{j}} \Delta \boldsymbol{w_j}} + \sum \bar{\boldsymbol{w}}_{\boldsymbol{j}} \Delta \boldsymbol{a_j} + \sum \bar{\boldsymbol{w}}_{\boldsymbol{j}} \bar{\boldsymbol{J}} \Delta \boldsymbol{b_j} + \sum \underset{\substack{\uparrow \\ \boxed{\textit{s-dividend}}}}{\bar{\boldsymbol{w}}_{\boldsymbol{j}} \bar{\boldsymbol{b}} \Delta \boldsymbol{J}} + \sum \bar{\boldsymbol{w}}_{\boldsymbol{j}} \Delta \boldsymbol{e_j} \qquad [V.7]$$

In this last equation, the change in the age structure has both a mechanical effect (the first term) and a substantive effect on behavior (the penultimate term). The question, of course, is whether one can get a reliable estimate for the coefficient b—the effect of age structure on economic behavior.

V.2 Extension of composition effect

Just as with the behavior effect, the composition effect can also be disaggregated. The disaggregation can focus on primary demographic groups, on the age structure of the sub-populations, or on the demographic processes shaping the size of the groups (fertility, mortality and migration).

(A) Extension according to primary groups

In this case, the size of the study group is expressed as a function of the size of a primary group that generates the members of the group being studied. As one example, the number of children in poor families (w_j) might be expressed as a function of the number of poor families (n_j) and the relative fertility of poor families (f_j), i.e., the fertility of the poor compared to the national average.

$$w_j = n_j * f_j \tag{V.8}$$

The change in the percentage of children from each class will depend on the change in (1) the proportion of families that belong to that social class and (2) the relative fertility of those families.

$$\Delta w_j = \bar{f}\Delta n_j + \bar{n}\Delta f_j \tag{V.9}$$

One can thus insert [V.9] in [3], yielding

$$\boldsymbol{\Delta Y} = \underbrace{\underbrace{\left[\sum \bar{\boldsymbol{y}}_j * \bar{\boldsymbol{f}}\Delta \boldsymbol{n}_j\right]}_{\text{A1}} + \underbrace{\left[\sum \bar{\boldsymbol{y}}_j * \bar{\boldsymbol{n}}\,\Delta \boldsymbol{f}_j\right]}_{\text{A2}}}_{\boxed{\textit{Compositional effect}}} + \underbrace{\left[\sum \bar{\boldsymbol{w}}_j * \Delta \boldsymbol{y}_j\right]}_{\overset{\text{B}}{\boxed{\textit{Behavioral effect}}}} \tag{V.10}$$

In this equation, A1 represents the change in the percentage of poor families in a society, and A2 represents the change in the relative fertility of the various social classes. These two variables are conceptually richer, more informative, and more policy-relevant. The first (A1) might guide policies to stimulate growth and reduce poverty. The second might be related to the ability of family planning programs to reduce fertility inequality, especially if that inequality stems from differential access to family planning.

IV.3 Double extension

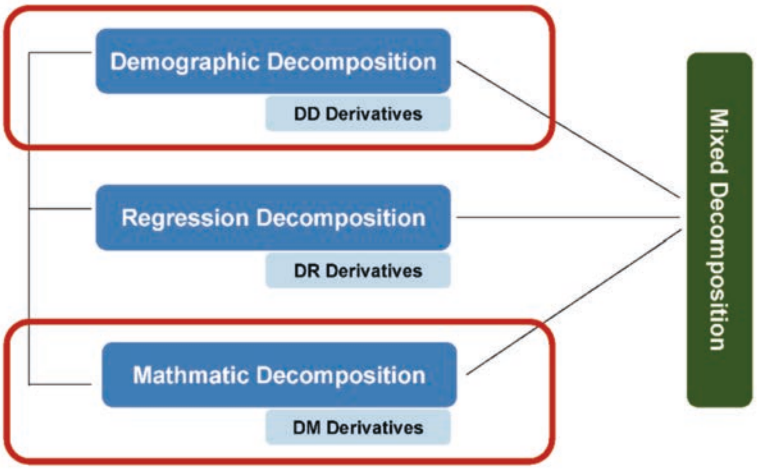
Finally, one can merge equations [V10] and [3] and get a detailed system that refines the analysis of both the compositional and behavioral sides. The result of this combination (below) yields an even more detailed expression where

- A1 = change in the distribution of mothers by social class
- A2 = change in the relative fertility of mothers
- B1 = change in baseline health
- B2 = change in income levels
- B3 = change in the health benefits of income
- B4 = residual effect of the factors omitted from the regression equation.

$$\Delta Y = \underbrace{\underset{\boxed{\text{B1}}}{\left[\sum \bar{w}_j\,\Delta\alpha\right]} + \underset{\boxed{\text{B2}}}{\left[\sum \bar{w}_j\,\bar{\beta}_j\Delta x_j\right]} + \underset{\boxed{\text{B3}}}{\left[\sum \bar{w}_j\,\bar{x}_j\Delta\beta\right]} + \underset{\boxed{\text{B4}}}{\left[\sum \bar{w}_j\,\Delta\mu_j\right]}} + \underbrace{\underset{\boxed{\text{A1}}}{\left[\sum \bar{y}_j\bar{f}_j\,\Delta n_j\right]} + \underset{\boxed{\text{A2}}}{\left[\sum \bar{y}_j\bar{n}_j\Delta f_j\right]}}$$

Again, both of the primary components of a basic demographic decomposition (compositional or behavioral) can be expanded for greater detail. With this more detailed accounting, a planner can get more nuanced insights into policy priorities.

Chapter VI
Combination of demographic and mathematical decompositions



This variant also begins with a simple demographic decomposition, but the expansion of the behavioral component uses an exact mathematical (rather than statistical) expression. To illustrate, let us consider a study of GDP change in Africa. The region's GDP per capita is a weighted average across all African countries, and the decomposition of its change can be expressed as usual.

$$\boldsymbol{\Delta Y} = \underbrace{\left[\sum \bar{\boldsymbol{y}}_j * \Delta \boldsymbol{w}_j\right]}_{\textit{Compositional effect}} + \underbrace{\left[\sum \bar{\boldsymbol{w}}_j * \Delta \boldsymbol{y}_j\right]}_{\textit{Behavioral effect}}$$

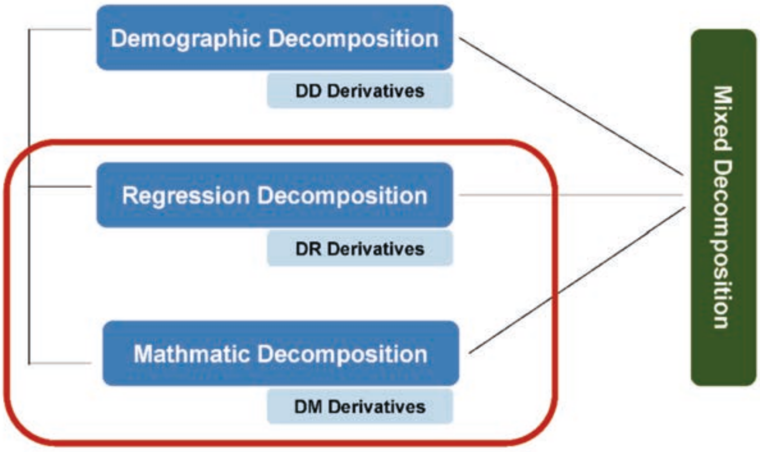
Next, we now express the countries' GDPs as a function of adult productivity (π) and age structure (α). The change of individual country GDP is as follows:

$$\Delta y_j = \left(\bar{\pi}_j * \Delta\alpha_j\right) + \left(\bar{\alpha}_j * \Delta\pi_j\right) \qquad \text{[VI.1]}$$

By inserting VI.1 into the basic equation of the demographic decomposition (II.3), the change in the average GDP of Africa becomes

$$\boldsymbol{\Delta Y} = \underbrace{\left[\sum \bar{\boldsymbol{y}}_j \Delta \boldsymbol{w}_j\right]}_{\textit{Effect of population size}} + \underbrace{\left[\sum \bar{\boldsymbol{w}}_j \bar{\boldsymbol{\pi}}_j \Delta \boldsymbol{\alpha}_j\right]]}_{\textit{Effect of age structure}} + \underbrace{\left[\sum \bar{\boldsymbol{w}}_j \bar{\boldsymbol{\alpha}}_j \Delta \boldsymbol{\pi}_j\right]}_{\textit{Effect of adult productivity}} \qquad \text{[VI.2]}$$

Chapter VII
Combination of regression and mathematical decompositions



The combination covered here is less common. Consider the link between money and happiness. Assume, for the sake of simplicity, that this relationship is linear.

$$Y = \alpha + \beta\$ + e \qquad [VII.1]$$

Next, consider that people have multiple income sources, perhaps wages and transfers, and that they derive different levels of satisfaction from each. If that is true, then the formula VII.1. is simplistic and should ideally incorporate this differential satisfaction.

Alternatively, one can begin with a mathematical relationship and incorporate a regression relationship. Consider, for instance, the factors shaping education spending per child. Suppose, as is likely, that the share of a family budget allocated to its children education depends on parental income , so the researcher can refine the analysis by integrating details from the regression study.

The possibilities depend only on the researcher's own imagination. Still, even where further expansion is possible, the researcher must compromise between detail and parsimony. The combination opportunities suggested are not ready-made recipes to apply mechanically. Rather, they are tools to be used selectively by researchers in their quest to understand social change.

Chapter VIII
Summary and conclusions

This monograph offers an introduction to decomposition, a cluster of methods that can enrich the methodological toolbox of social scientists. The adoption of these methods advances the study of social change beyond the current limits of micro and macro-regressions. While decomposition does not address causation, it can reliably locate the social or geographic sources of social change. In doing so, decomposition reduces the margin of error in understanding societal change and in designing related policies. Within these limitations, decomposition is a reliable and transparent tool that, alone or with other methods, can usefully guide the allocation of policy resources. Policymakers faced with several options can use a decomposition approach to narrow down the list of most promising options.

The flexibility of the decomposition method makes it possible to imagine creative extensions. The few options presented here are not an exhaustive list, and researchers are encouraged to imagine other possibilities. In addition, many of the examples used here draw from the fields of population or economics, but the methods apply to a wide range of social phenomena.

One strength of the decomposition approach is its compatibility with other methods. When carefully combined with other methods, it can enrich understanding. It is therefore not a substitute but a complement. It complements other methods by identifying key processes and groups. At the same time, it leaves room for other methods (say causal analysis or informant interviews) to elucidate questions about causation, processes, key events, and key actors. Such complementarities facilitate a more complete understanding of social change. In a sense, and going back to our initial story of the proverbial drunk looking exclusively under the lamp-post, the contribution of decomposition is twofold: it widens the search area by considering all possible sources of change; and it can help begin the search closer to where the key—rather than the light—is. Having identified the groups of processes that lead the change, researchers can launch additional investigations into the reasons why these groups did change. Ultimately, those who seek to change the world can do so more effectively with a better understanding of the drivers of social change.

References

Adams Jr, R. H. (1994). Non-farm income and inequality in rural Pakistan: A decomposition analysis. *The Journal of Development Studies, 31*(1), 110–133.

Barsky, R., Bound, J., Charles, K. K., & Lupton, J. P. (2002). Accounting for the Black–White Wealth Gap. *Journal of the American Statistical Association, 97*(459), 663–673. https://doi.org/10.1198/016214502388618401

Bauer, T., Göhlmann, S., & Sinning, M. (2007). Gender differences in smoking behavior. *Health Economics, 16*(9), 895–909.

Bechtel, W., & Richardson, R. C. (2010). *Discovering complexity: Decomposition and localization as strategies in scientific research.* MIT Press.

Bertrand, M., & Hallock, K. F. (2001). The gender gap in top corporate jobs. *ILR Review, 55*(1), 3–21.

Blau, F. D., & Kahn, L. M. (1992). The Gender Earnings Gap: Learning from International Comparisons. *The American Economic Review, 82*(2), 533–538.

Blau, F. D., & Kahn, L. M. (2006). The US gender pay gap in the 1990s: Slowing convergence. *ILR Review, 60*(1), 45–66.

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human Resources,* 436–455.

Bourguignon, F. (1979). Decomposable income inequality measures. *Econometrica: Journal of the Econometric Society,* 901–920.

Bourguignon, F., & Ferreira, F. (2005). Decomposing changes in the distribution of household incomes: methodological aspects. *The Microeconomics of Income Distribution Dynamics in East Asia and Latin America, Ed. by F. Bourguignon, FH Ferreira, and N. Lustig,* 17–46.

Bourguignon, F., Ferreira, F. H., & Leite, P. G. (2008). Beyond Oaxaca–Blinder: Accounting for differences in household income distributions. *The Journal of Economic Inequality, 6*(2), 117–148.

Cho, L.-J., & Retherford, R. D. (1973). Comparative analysis of recent fertility trends in East Asia.

Eloundou-Enyegue, P. M., & Giroux, S. C. (2012). Fertility transitions and schooling: From micro-to macro-level associations. *Demography, 49*(4), 1407–1432.

Fairlie, R. W. (2005). An extension of the Blinder-Oaxaca decomposition technique to logit and probit models. *Journal of Economic and Social Measurement, 30*(4), 305–316.

Firebaugh, G., & Goesling, B. (2004). Accounting for the recent decline in global income inequality. *American Journal of Sociology, 110*(2), 283–312.

Fortin, N., Lemieux, T., & Firpo, S. (2011). Chapter 1 - Decomposition Methods in Economics. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 4, pp. 1–102). Elsevier. https://doi.org/10.1016/S0169-7218(11)00407-2

Gage, T. B. (1993). The decline of mortality in England and Wales 1861 to 1964: Decomposition by cause of death and component of mortality. *Population Studies, 47*(1), 47–66.

Gómez-Redondo, R., & Boe, C. (2005). Decomposition analysis of Spanish life expectancy at birth: evolution and changes in the components by sex and age. *Demographic Research, 13*, 521–546.

Gupta, P. D. (1978). A general method of decomposing a difference between two rates into several components. *Demography, 15*(1), 99–112.

Gupta, P. D. (1991). Decomposition of the difference between two rates and its consistency when more than two populations are involved. *Mathematical Population Studies, 3*(2), 105–125.

Gupta, P. D. (1993). *Standardization and decomposition of rates: A user's manual.* US Department of Commerce, Economics and Statistics Administration, Bureau of the Census.

Horiuchi, S., Wilmoth, J. R., & Pletcher, S. D. (2008). A decomposition method based on a model of continuous change. *Demography, 45*(4), 785–801.

Kaufman, R. L. (1983). A structural decomposition of black-white earnings differentials. *American Journal of Sociology, 89*(3), 585–611.

Kim, C. (2010). Decomposing the change in the wage gap between white and black men over time, 1980-2005: An extension of the Blinder-Oaxaca decomposition method. *Sociological Methods & Research, 38*(4), 619–651.

Kim, Y. J., & Strobino, D. M. (1984). Decomposition of the difference between two rates with hierarchical factors. *Demography, 21*(3), 361–372.

Kim, Y.-M., & Shirahase, S. (2014). Understanding intra-regional variation in gender inequality in East Asia: Decomposition of cross-national differences in the gender earnings gap. *International Sociology, 29*(3), 229–248.

Kitagawa, E. M. (1955). Components of a difference between two rates. *Journal of the American Statistical Association, 50*(272), 1168–1194.

Krieg, J. M., & Storer, P. (2006). How much do students matter? Applying the Oaxaca decomposition to explain determinants of adequate yearly progress. *Contemporary Economic Policy, 24*(4), 563–581.

Lemieux, T. (2006). Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill? *American Economic Review, 96*(3), 461–498.

Liao, T. F. (1989). A flexible approach for the decomposition of rate differences. *Demography, 26*(4), 717–726.

Luke, D. A. (2004). Multilevel Modeling, Quantitative Applications in the Social Sciences. *London: A Sage University Paper Series.*

Martikainen, P., Valkonen, T., & Martelin, T. (2001). Change in male and female life expectancy by social class: decomposition by age and cause of death in Finland 1971–95. *Journal of Epidemiology & Community Health, 55*(7), 494–499.

McEwan, P. J., & Marshall, J. H. (2004). Why does academic achievement vary across countries? Evidence from Cuba and Mexico. *Education Economics, 12*(3), 205–217.

Mookherjee, D., & Shorrocks, A. (1982). A decomposition analysis of the trend in UK income inequality. *The Economic Journal, 92*(368), 886–902.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review,* 693–709.

Oaxaca, R. L., & Ransom, M. R. (1994). On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, *61*(1), 5–21.

Oaxaca, R. L., & Ransom, M. R. (1999). Identification in detailed wage decompositions. *Review of Economics and Statistics*, *81*(1), 154–157.

Page, H. J. (1977). Patterns underlying fertility schedules: A decomposition by both age and marriage duration. *Population Studies*, *31*(1), 85–106.

Pollard, J. H. (1988). On the decomposition of changes in expectation of life and differentials in life expectancy. *Demography*, *25*(2), 265–276.

Ponnapalli, K. M. (2005). A comparison of different methods for decomposition of changes in expectation of life at birth and differentials in life expectancy at birth. *Demographic Research*, *12*, 141–172.

Powers, D. A., & Yun, M.-S. (2009). Multivariate decomposition for hazard rate models. *Sociological Methodology*, *39*(1), 233–263.

Pradhan, J., & Arokiasamy, P. (2010). Socio-economic inequalities in child survival in India: a decomposition analysis. *Health Policy*, *98*(2), 114–120.

Retherford, R. D., & Ogawa, N. (1978). Decomposition of the Change in the Total Fertility Rate in the Republic of Korea, 1966–1970. *Social Biology*, *25*(2), 115–127.

Retherford, R. D., & Rele, J. R. (1989). A decomposition of recent fertility changes in South Asia. *Population and Development Review*, 739–747.

Robinson, W. S. (1961). Ecological correlations and the behavior of individuals. *Studies in Human Ecology (Row Peterson, Evanston, Ill.)*, 115–120.

Rodrik, D. (2012). Why we learn nothing from regressing economic growth on policies.

Sandberg, J. F., & Hofferth, S. L. (2001). Changes in children's time with parents: United States, 1981–1997. *Demography*, *38*(3), 423–436.

Shorrocks, A. F. (1980). The class of additively decomposable inequality measures. *Econometrica: Journal of the Econometric Society*, 613–625.

Shorrocks, A. F. (1984). Inequality decomposition by population subgroups. *Econometrica: Journal of the Econometric Society*, 1369–1385.

Shorrocks, A. F. (2013). Decomposition procedures for distributional analysis: a unified framework based on the Shapley value. *Journal of Economic Inequality* 11, 99-126

Shorrocks, A., & Wan, G. (2005). Spatial decomposition of inequality. *Journal of Economic Geography*, *5*(1), 59–81.

Smith, H. L., Morgan, S. P., & Koropeckyj-Cox, T. (1996). A decomposition of trends in the nonmarital fertility ratios of blacks and whites in the United States, 1960–1992. *Demography*, *33*(2), 141–151.

Solow, R. M. (1957). Technical change and the aggregate production function. *The Review of Economics and Statistics*, 312–320.

Thornton, A. (2001). The developmental paradigm, reading history sideways, and family change. *Demography*, *38*(4), 449–465.

UNICEF. (2008). *The State of Africa's Children, 2008: Child Survival*. UNICEF.

Vaupel, J. W., & Romo, V. C. (2003). Decomposing change in life expectancy: A bouquet of formulas in honor of Nathan Keyfitz's 90th birthday. *Demography, 40*(2), 201–216.

Wan, G., & Zhou, Z. (2005). Income Inequality in Rural China: Regression-based Decomposition Using Household Data. *Review of Development Economics, 9*(1), 107–120.

Yamaguchi, K. (2017). Decomposition Analysis of Segregation. *Sociological Methodology, 47*(1), 246–273.