



FREDERICK S. PARDEE CENTER
FOR INTERNATIONAL FUTURES
EXPLORE UNDERSTAND SHAPE

WORKING PAPER 2014.03.05.b

IFs POPULATION MODEL DOCUMENTATION

Author: Barry B. Hughes
March 2014

Note: If cited or quoted, please indicate working paper status.



UNIVERSITY of
DENVER

JOSEF KORBEL SCHOOL OF INTERNATIONAL STUDIES

PARDEE.DU.EDU

IFs Population Model Documentation

Table of Contents

1.	Introduction.....	3
1.1	Overview.....	3
1.2	Dominant Relations	3
2.	Flow Charts.....	6
2.1	Demographic Overview	6
2.2	Fertility Detail.....	8
2.3	Mortality Detail.....	9
2.4	Nutrition.....	11
2.5	Migration.....	13
2.6	Urbanization.....	14
3.	Equations.....	15
3.1	Age Distribution.....	15
3.2	Fertility.....	16
3.3	Mortality: Life Expectancy and Infant Mortality.....	19
3.4	Mortality: The Legacy Formulation.....	20
3.5	Malnutrition: The Legacy Formulation	22
3.6	HIV/AIDs Mortality: The Legacy Formulation.....	23
3.7	Migration.....	23
3.8	Urbanization.....	24
3.9	Household Size	25
3.10	Demographic Indicators.....	26
3.11	Data Used.....	27
4.	References.....	29

1. Introduction

1.1 Overview

The population sub-model of IFs uses the cohort component analysis approach of many population models, including the studies done by the United Nations (United Nations, 1956 and 1977). The structure of the IFs population model drew initially on the World Integrated Model (WIM) or the second generation Mesarovic-Pestel Model (Hughes, 1980), but has changed much over time. In particular, José Solórzano and Randall Kuhn have made many contributions to its development.

The approach relies upon age, fertility, and mortality distributions for each country/region with 22 cohorts: one for infants, 20 of five-year size, and one for all individuals of age 100 or older. A major advantage of five-year cohorts is that data sources generally present demographic data in that form. Ideally, however, the cohort size should correspond to the model time step so as to avoid "numerical diffusion," the propagation of change from a five-year cohort to an adjoining cohort in a single year. To prevent such numerical diffusion, IFs actually runs an age distribution with 100 single-year cohorts and advances that over time, collapsing to 22 cohorts only for the calculations of births and deaths.

Because extensions of life expectancy are occurring steadily and there is at least the possibility of substantial breakthroughs, the IFs project has also created the option of extending the number of cohorts from 22 up to as many as 42 (allowing separate representation of age categories up to 200+). The capability is normally turned off, but this documentation will explain how to turn on extended aging.

1.2 Dominant Relations

The dominant population (POP) equation is a simple addition of births (BIRTHS) at the bottom of the cohort distribution, subtraction of deaths (DEATHS) from each population cohort, and advance of people to the next cohort over time.

The following key dynamics are directly linked to the Dominant Relations:

Births are primarily a function of the total fertility rate (TFR), which in the longer term responds especially to education level of the adult population. The model user has direct control over TFR with a multiplier (*tfrm*), but also much control for low fertility countries with a parameter specifying long-term stabilization level and lower boundary for fertility (*tfrmin*). There is also a secular trend reduction in fertility (controlled by *tfrr*).

Deaths are primarily a function of life expectancy (LIFEXP), itself computed within the IFs health model where, like fertility, it responds in the long run to adult education and also to GDP per capita and technology change. The model user has direct control over all deaths with a mortality multiplier (*mortm*) and over those specific to a cause of health with an alternative multiplier (*hlmortm*). There is also a secular trend reduction in mortality (controlled by *tmortr*).

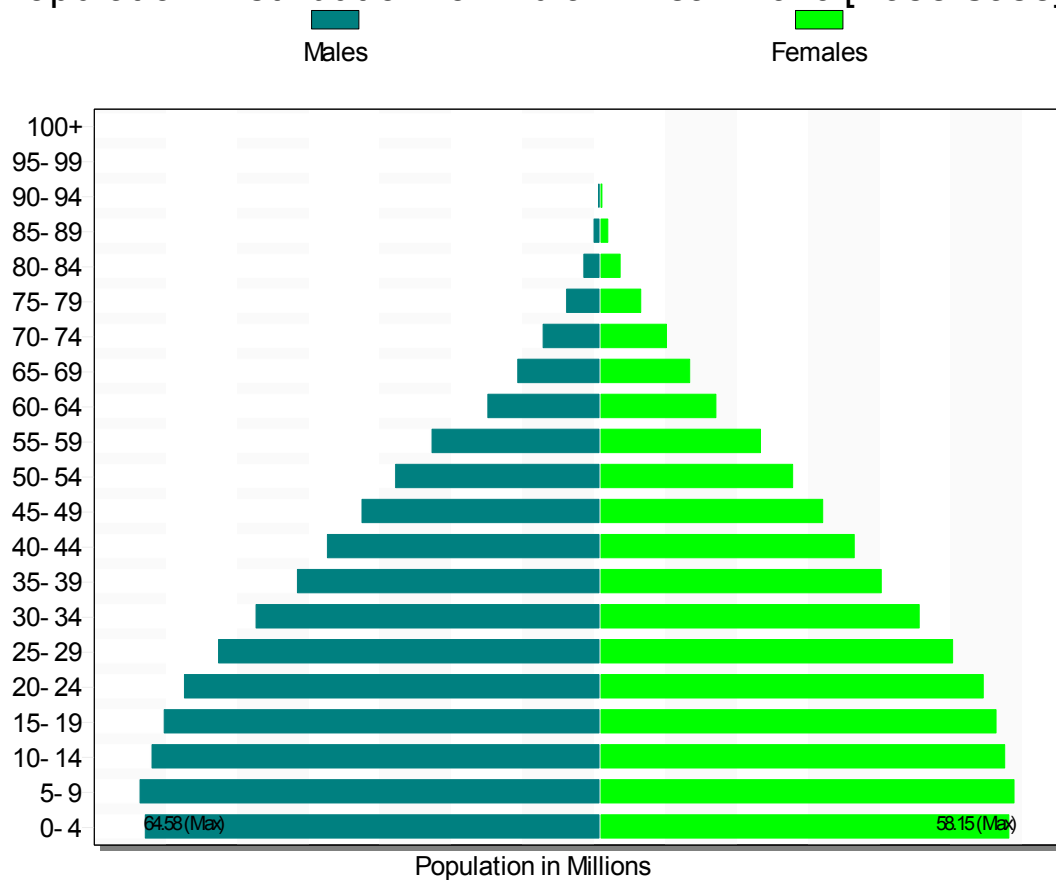
The larger demographic model in combination with the health model provides representation of and control over migration; the fertility impact of infant mortality and contraception use rates; and the mortality impact of many factors including undernutrition, smoking rates, and indoor air pollution from open burning of solid fuels.

1.3 Structure and Agent System

System/Subsystem	Demographic
Organizing Structure	Cohort-component
Stocks	Population by age-sex
Flows	Birth, death, migration
Key Aggregate Relationships (illustrative, not comprehensive)	Life expectancy (from health model)
Key Agent-Class Behavioral Relationships (illustrative, not comprehensive)	Household fertility and migration

Humans as individuals within households interact in larger **demographic systems or structures**. The computer model should represent the behavior of such households, such as decisions to have children or to emigrate. And it should represent the larger demographic structures that incorporate the decisions of millions of such households. A typical approach to representing such demographic systems is through age-sex cohort distributions (see the figure below showing an example from the model). IFs also uses fertility and mortality distributions by age and sex and tracks migration across countries.

Population Distribution for India in Year 2010 [Base Case]



Demographers have widely accepted the representation of demographic systems and the development of demographic models with cohort-component structures. In fact, the United Nations, the U.S. Census Bureau, and the International Institute for Applied Systems Analysis (IIASA), pre-eminent demographic forecasting institutions, all use cohort-component modeling (O'Neill and Balk 2001).

2. Flow Charts

This section presents several block diagrams that are central to the population model: an overview, fertility, mortality, nutrition, migration, and urbanization.

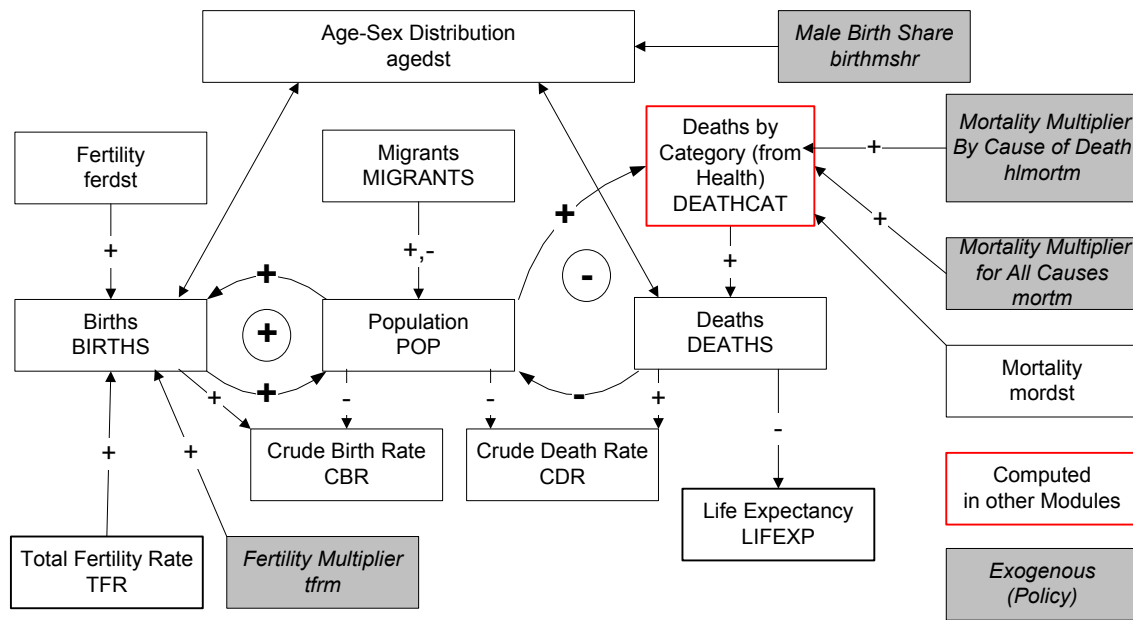
2.1 Demographic Overview

The demographic model represents the population of each geographic unit in terms of 22 cohorts (infants, five-year intervals up to age 99, and those aged 100 and older), separately for females and males. An age distribution records the population in each cohort and sex category. The sum across all cohorts in the age distribution and both sexes is the total population. A fertility distribution determines births, which are added to the bottom of the age distribution, while a mortality distribution determines deaths, which are subtracted from the appropriate cohort of the age distribution.

Those who might like to turn on the extension of age-cohort representation, to as many as 42, can do so by making changes in the IFsInit table of the IFsInit.mdb file. Specifically, the NCohorts field can be changed to as many cohorts as 42 and the NAgess field can be changed up to 200. Registering these changes requires a rebuild of the Base Case (see documentation of Extended Features).

The population model is central to many broader dynamics of IFs. Two key feedback loops drive its own dynamics. The first is a positive feedback loop around fertility, linking population and births (causing population to drive exponentially upward if nothing else changes), while the second is a negative loop around mortality, linking population and deaths (causing population to decline). This second loop actually runs through the health model of IFs where deaths are computed (switching the control parameter *hlmodelsw* from 1 to 0 would, however, cause the model to revert to an earlier formulation in which life expectancy was computed as function of GDP per capita and controlled the death rate and deaths; it would turn off the health model's impact). A Malthusian variation of the negative feedback loop involving deaths may be of interest to those who believe that food supplies do or will play an important role in population dynamics (as they clearly do in countries with very low nutritional levels) by raising mortality rates, especially of children. See the topic on nutrition. Whether population rises or falls depends on the relative strength of those two loops

The easiest and most often used scenario handles for the population model are a multiplier on the total fertility rate (the number of children borne by an average woman in a lifetime), namely *tfrm*, a multiplier on the total mortality rate, *mortm*, and a multiplier on mortality by cause, *hlmortm*.



A large number of indicators are calculated in IFs from the age distribution:

- the median age of the country's population (POPMEAGE)
- population aged 15 to 65 (POP15TO65)
- population above age 65 (POPGT65)
- population below age 15 (POPLE15)
- population pre working age (POPPREWORK), controlled by the parameter specifying the work starting age (*workageentry*)
- population post working age (POPRETIRED), controlled by the parameter specifying the retirement age (*workageretire*)
- population within the working years (POPWORKING)
- the potential support ratio, or the population from 15 to 64 over that above 65 (POTSUPRAT)
- an indicator of the youth bulge or the population from 15 to 29 as a portion of that 15 and above (YTHBULGE)

In addition there are a number of indicators calculated from the size of country populations:

- the growth rate of population (POPR)
- the world population (WPOP)
- growth in the world population (WOPR)

More description is available on the dynamics around fertility and mortality as well as several specialized topics on topics such as nutrition levels and migration.

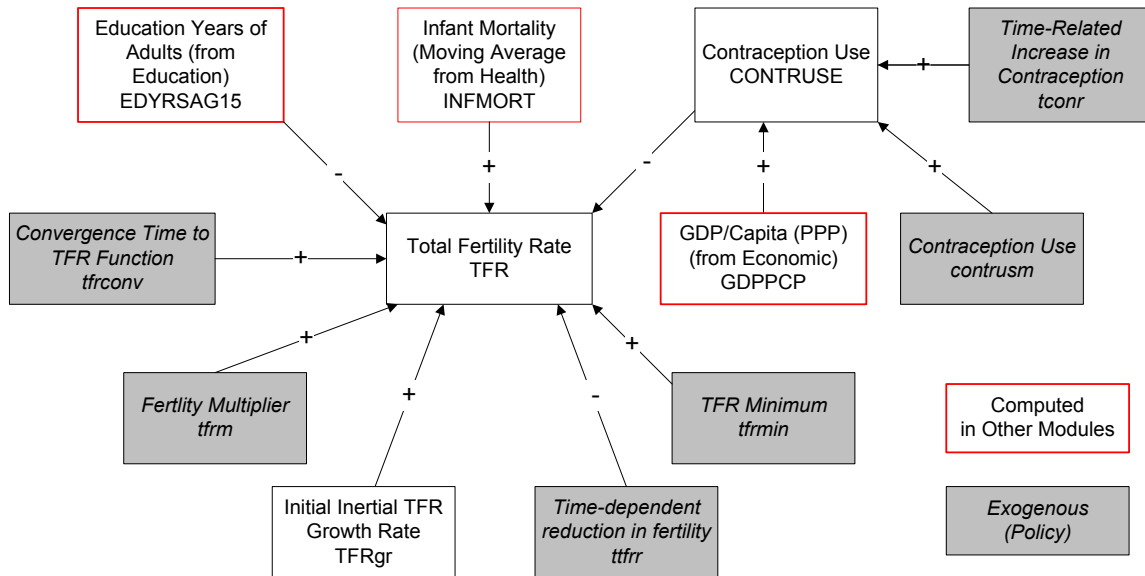
2.2 Fertility Detail

The central indicator of fertility is the total fertility rate (TFR), the number of children that the average woman will bear throughout her life. Fertility generally decreases in the long run as deep or distal drivers such as GDP per capita (from the economic model) or formal years of education of adults (EDYRSAG15 from the education model) increase; our own analysis suggested the use of education years, a result that Angeles (2010) reinforced.

In addition there are more proximate drivers, some of which can change more rapidly than GDP per capita or education levels and thereby affect fertility rate. We represent two, namely infant mortality rates (INFMOR) and contraception usage (CONTRUSE). The health model of IFs determines infant mortality rates. Sudden change in those do not, however, immediately affect fertility rates and we smooth changes in rates so as to introduce an approximately 10-year lag, consistent again with the findings of Angeles (2010). Based on cross-sectional analysis the population model of IFs forecasts the percentage of population using modern contraception as a function of GDP per capita at purchasing power parity (GDPPCP). We found, however, that there was additional growth over time and the parameter for time-related usage growth (*tconr*) controls that. The user can also change contraception use via an exogenous multiplier (*contrusm*).

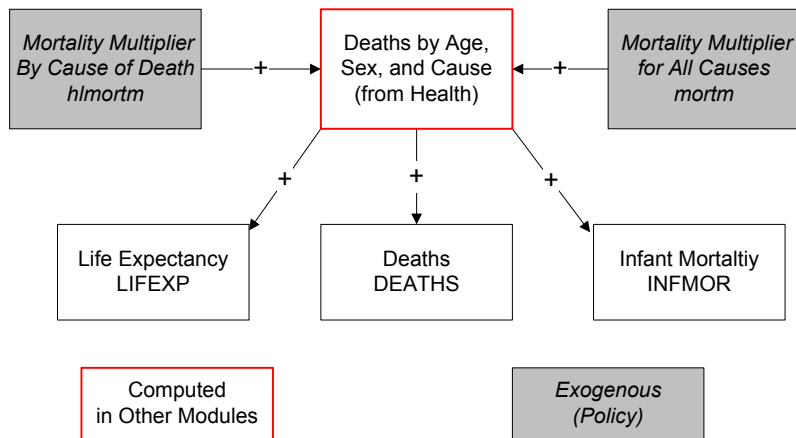
Although those three distal and proximate drivers substantially determine the forecasts of fertility rate, there are several additional elements that influence it. First, we calculate the historical growth rate of TFR (TFRgr) and use that internal variable in the first few years so as to maintain an inertial pattern of change in TFR consistent with history; we phase out that inertial element in favor of endogenously computed factors over a 10-year period. Second, we have used another time dependent parameter (*ttfrr*) to allow introduction of somewhat faster or slower growth rates in TFR. Mostly we have used that as a tuning parameter to adjust our long-term global population forecasts to be more consistent with those of others such as the UN Population Division or the International Institute of Applied Systems Analysis. Normally that parameter is very small or zero. Third, we provide the user with a direct multiplier on the fertility rate (*tfrm*). And finally, not knowing what the long-term minimum fertility rate might be in a world where for many countries rates have fallen very substantially below replacement rates, we provide such a minimum (*tfrmin*). Since some countries are below most expected minimums and therefore below common values of that parameter, we phase that minimum in over time with a convergence parameter (*tfrconv*), which serves double duty

by also marking the number of years of convergence of TFR itself to the values that the function with distal and proximate drivers produces.



2.3 Mortality Detail

The current default representation of mortality and life expectancy in IFs relies entirely on the health model so please see its documentation. That model computes deaths by age and sex and uses those to compute total deaths (DEATHS) as well deaths by category of cause (DEATHCAT). It also computes life expectancy (LIFEXP) and infant mortality (INFMOR), variables of importance to the population model. Two parameters in the health model allow multiplicative intervention with respect to total deaths (*mortm*) and those by cause (*hlmortm*).



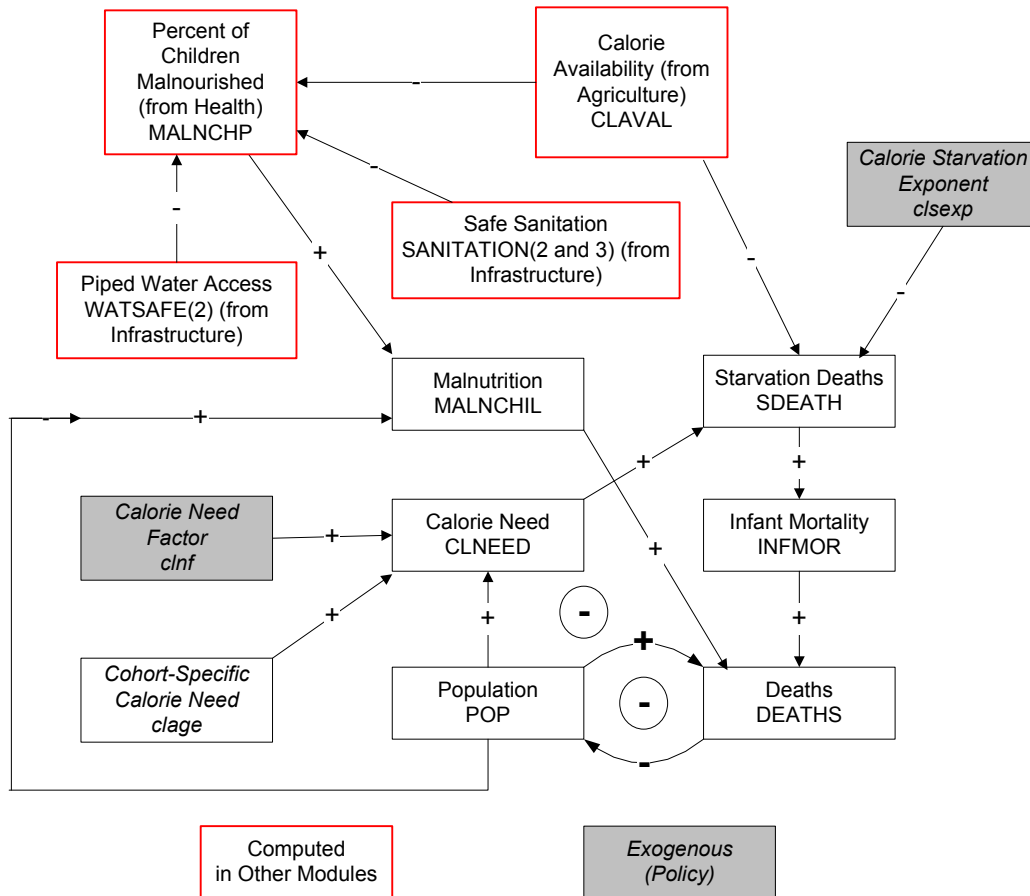
There is, however, a legacy representation of mortality in IFs (available if the health model is switched off with *hlmodels*=0) that reverses that logic and uses the model's calculation of an initial estimate of life expectancy to drive mortality by age and sex (not cause). Life expectancy normally increases and mortality normally decreases as GDP per capita rises (see the economic model) or as the income share of the poorest 20% of the population increases. In the legacy representation, the initial calculation of life expectancy is imposed on an initial mortality distribution that provides a country-specific age and sex profile of mortality.

A number of factors then further affect and alter the mortality distribution in the legacy mortality structure. These include deaths related to warfare (CIVDM), to AIDS, and possibly to starvation (via infant mortality, because it is primarily the very young who are at risk). In addition, the user of the model may introduce greater or lesser mortality via a mortality multiplier.

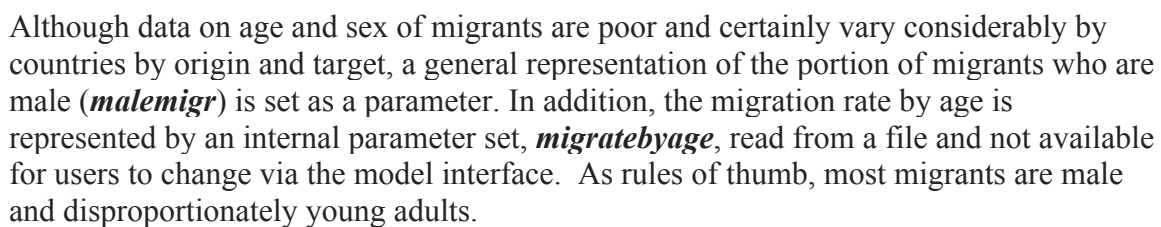
At the same time, however, increases in life expectancy shift the mortality distribution from its initial condition towards an ultimate life (survivor) table as life expectancy approaches that built into the ultimate life table (approximately 85 in the 1998 revision of the UN population tables).

Once the mortality distribution adjustments are made and deaths can be computed from it, the life expectancy is recomputed.

generate starvation deaths. Historical and contemporary data do not exist, however, to support calculation of starvation (the deaths of those who are severely malnourished are normally attributed to various diseases that prey on them, such as diarrhea among children). For this reason and because of the considerably greater sophistication of the health model we recommend leaving that health model engaged.



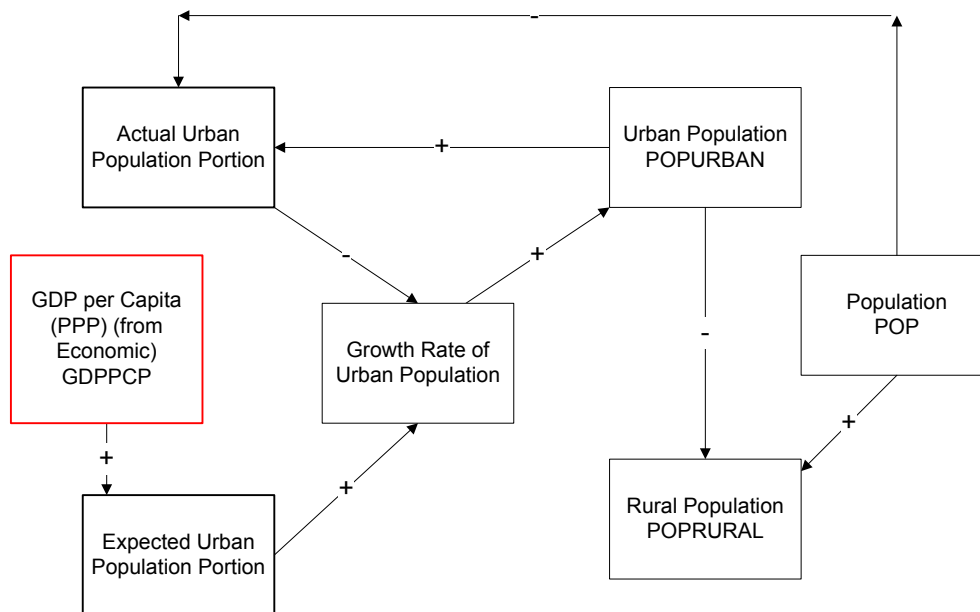
Although for most countries it is less important in determining population growth patterns than are fertility and mortality rate, migration is extremely important for some, such as those in the Arabian/Persian Gulf region. Because long-term migration is, however, very difficult to forecast, we rely on exogenous scenarios of migration rate (*migrater*) to drive forecasts in IFs. Users can also affect global migration patterns with a multiplier on global rates (*wmigrm*). On a global basis immigration and emigration are required to balance, and the number of migrants (MIGRANTS) in IFs is balanced, resulting in a computation also of an endogenous migration rate (MIGRATE).



13

2.6 Urbanization

IFs does not represent urban and rural populations by age and sex, but does forecast the division in total. The key driving variable in early model years, internal to the model, is the growth rate of the urban population. Because that variable is initialized with historical data, it initially introduces an inertial element into the forecast. But over time the rate of growth in urban population increases or decreases more and more in response to the gap between the actual portion of population that is urban and an expected urbanization rate based on a function driven by GDP per capita at purchasing power parity (GDPPCP). The growth rate also responds, however, to approaching very high levels of urban population (POPURBAN) as a percentage of the total by slowing down. Because the model dynamics are built around urban population, rural population (POPRURAL) is essentially a residual.



There is no parametric control over the growth of urban population in part because the model does not contain any significant forward linkages of urban population size or portion.

3. Equations

This section will present and discuss the equations that are central to the functioning of the population model.

3.1 Age Distribution

The basic structure of the population model is very simple, even if the implementation becomes more complex. The core of the model is fundamentally an accounting system around the age-sex distribution (AGEDST) with 5-year age categories—and an elaboration of that into single year categories (FAGDST)—in which people age over time, with births added into the bottom age category each year and deaths subtracted from the appropriate age and sex category. The key to long-term dynamics lies primarily within change in the fertility and mortality distributions, with migration playing a secondary role for most countries.

A 5-year cohort fertility distribution (FERDST) multiplies the age distribution (AGEDST) to produce births (BIRTHS). The total fertility rate (TFR), or total number of births expected to a woman during her lifetime, modifies the fertility distribution over time. The fertility distribution itself moves from the initial empirical, country-specific pattern to an ultimate fertility distribution (ULTIMATEFERTILITY) as GDP per capita (PPP) moves towards a specified level (currently \$45,000). The ultimate fertility distribution is exogenous to the model in a file and not available for the user to change via the model's interface. We will see the computation of TFR in our discussion of fertility.

$$BIRTHS_{r,p} = \sum_c AGEDST_{r,c,p} * FERDST_{r,c} * \frac{TFR_r}{TFR_{r,t=1}}$$

where

$$FERDST_{r,c} = F(FERDST_{r,c}, ULTIMATEFERTILITY_c, GDPPCP_{r,45})$$

In the above equation and other documentation of the population model:

r=region/country

c=age category

p=sex (because s is used elsewhere in the model for economic sector)

d=cause of death

Deaths (DEATHS) are computed in the health model of IFs, see the documentation of that model. They are the sum of age, sex, and cause-of-death specific mortality forecasts. Life expectancy (LIFEXP) is also computed in the health model.

There is, however, a legacy model of mortality and deaths that now is very rarely used but can be activated by changing the *hlmodelsw* parameter from 1 to 0. The legacy calculation of deaths parallels that of births in that it relies on a product of the age distribution with a mortality distribution (MORDST). As with fertility, the mortality distribution itself moves from the initial empirical, country-specific pattern to an ultimate mortality distribution as life expectancy moves towards a specified level (currently 85 years). In the legacy model, life expectancy is computed from the mortality distribution.

Most of the model uses the 5-year age categories of the age distribution (AGEDST). But 5-year categories can introduce a significant problem when we advance the model over time. Specifically, it can lead to diffusion of births or deaths too quickly up the distribution (for instance, if a surge of births entered the bottom 5-year category one year, 1/5 of those could potentially move up to the next category the following year already, because a model that only used 5-year categories would not recognize their recent arrival in the category).

Hence in the first year of the model we spread the 5-year categories that come to us from UN data into a 1-year or annual age distribution (FAGEDST) using a spline function and use that annual distribution for our accounting dynamics across time. One-fifth of deaths in each 5-year category reduce the appropriate annual age category and those in each age category advance to the next year (all surviving infants advance to age 1). We also add one-fifth of net migration by 5-year category into each underlying single-year category. Births enter the infant category of the age distribution.

Once the full age distribution has been advanced for the next year, it can also be collapsed back into the 5-year cohorts of the age distribution (AGEDST), which is used for the calculations of births and deaths in the next year and for display in the model. The population (POP) is a sum across this distribution.

3.2 Fertility

Change in fertility centers on the current value of the total fertility rate (TFR). IFs determines the TFR and then imposes that on the cohort-specific fertility distribution (FERDST) of the region/country.

IFs uses three key variables to drive TFR forecasts over time. One of those accounts for the change that typically accompanies long-term development and social evolution. The

two principal candidates to represent such change in the long term across all IFs models are GDP per capita at purchasing power parity (GDPPCP) and the years of formal education attained by adults (EDYRSAG15). Our own analysis and that by Angeles (2010) both suggest that the latter is the stronger predictor for TFR. Frequently in IFs we find (Hughes 2001) that the relationship between one of those two deep distal drivers and any specific element of social change is logarithmic (that is, social change happens especially rapidly at lower levels of income and education and then saturates) and this is the case in this instance also. You can see the approximate form of that relationship by examining a scattergram of TFR as a function of EDYRSAG15 in the initial model year, or you can look at the multivariate relationship that IFs actually uses (in Scenario Analysis/Change Selected Functions).

In addition to long-term development and the deep or distal variables associated with it, societies are subject to short-term factors, most of which are in turn influenced heavily over time by the distal variables. These more proximate variables do, however, exhibit patterns of change that are at least somewhat independent of the distal drivers and more dependent on societal choices and policies. In the case of fertility change, two such variables often identified to be important are the rate of mortality, often infant mortality in particular (INFMORT) and the rate of use of modern contraception (CONTRUSE).

$$TFR_r = TFR_{r,t-1} * F(\ln(EDYRSAG15_{r,p=total}), LagInfMor_r, CONTRUSE_r) * \mathbf{tfrm}_r * (1 + (t - 1) * \mathbf{tfrr})$$

In the equation above we have used a lagged form of infant mortality. The lag uses 10 percent of the new value for infant mortality and 90 percent of lagged (therefore actually moving average) value; the proportions are subject to change, but were chosen to capture roughly the 10-year lag to peak effect identified by Angeles (2010). When such a moving average is initiated with the value of the first year of the model run, rather than with a value computed over an historical period preceding that first year, it gives rise to a pattern of slow change in initial years (values of early years tend to be very close to those of the initial value) and then accelerating change over time up to the about the 10th year. We therefore phase in the effect of the moving average, also over 10 years.

The additional term involving the parameter $\mathbf{tfrr}$ is used to represent time change that is independent of the relationship estimated via cross-sectional analysis with recent data. There has been a global ideational change with respect to fertility that the term can represent; in addition, it can be a tuning parameter and normally the value is very low in IFs. Finally in the equation above, the user can adjust a multiplier parameter ($\mathbf{tfrm}$) from its default value of 1 so as to force higher or lower fertility.

There are also, however, three important algorithmic elements that wrap this equation in more extensive model code. First, we compute in the model preprocessor the historical growth rate of TFR (TFRgr) and use that to help drive year-to-year change in TFR. In fact, in the first year the change in TFR is fully driven by that internal variable, but attention to it is phased out over 10 years. Second, we have captured in the first year of the model forecast the difference between TFR from the function and TFR from the data. This difference or shift could be viewed as a country-specific fixed effect dependent on variables such as historical paths and cultural factors. We choose, however, to phase it out over a fairly long period of time specified by the parameter *tfrconv*. Often in IFs the reduction of such shift factors is done over a half century or more, and, at the time of this writing, the parameter's value was 100. Third, total fertility rate is unlikely to shift indefinitely toward zero. In fact, it requires a value of about 2.1 simply to maintain a steady population (unless life expectancies are growing). TFR is therefore bound by a minimum that responds to a global parameter (*tfrmin*). The equation below represents that long-term bound which is again phased in over a very long period of time and algorithmically raises the fertility of countries below the minimum.

$$TFR_r = AMAX(TFR_r, tfrmin)$$

The use of modern contraceptives (CONTRUSE) is itself a function of a key distal driver, in this case GDP per capita at purchasing power parity (GDPPCP). The reader may wish to use the model to look also at a scattergram of CONTRUSE against GDPPCP in the initial year. The "actual" level of contraception use depends not only on GDPPCP, but on an exogenous multiplier (*contrusm*), and on a temporal (t) upward drift in contraception use related to ideational change again, as well as related technological innovation and diffusion (controlled by *tconr*).

$$CONTRUSE_r = F(GDPPCP_r) * contrusm_r + t * tconr$$

Once we have computed the total fertility rate (TFR), the number of births in a given year is a simple function of the fertility distribution and the TFR.

If advances in health very substantially affect life expectancy, they may also affect fertility patterns. Parameters in IFs allow control of the onset age of fertility (*hltfrageinit*), the peak age of it (*hltfragepeak*), the age of menopause (*hltfragestop*), and the rate of decline from peak to menopause (*hltfragehalflife*). If child-bearing age were greatly extended, it would necessarily lead at some point to a change not only in the peak age of child-bearing, but the rate of child-bearing at that age (*hltfrpeaklevel*), changed from current patterns at a rate controlled in the model by a final fertility parameter (*hltfrconv*).

3.3 Mortality: Life Expectancy and Infant Mortality¹

The health model calculates the mortality distribution by country/region, age category, sex, and cause of death (modmordstdet). This distribution allows the specification of key variables in the population model, including life expectancy (LIFEXP) and infant mortality (INFMOR). Life Expectancy is computed as a mean average number of years of life given the survival rates in each age group. First we find total mortality by country/region (r), age (c) and gender (p) by adding all 15 types of mortality (d) using modmordstdet(c,a,g,t). (Note with respect to model code: we actually combine the gender and mortality type subscript into one, with the odd type values representing males and the even type values for females).

Second we find the average years lived (nax), within the age group, by those who die (per Coale and Demeny 1983, using parameters that came from the arithmetic mean of the separate male and female parameters shown in Preston, Heuveline, and Guillot 2001):

Infants with mortality $\geq 0.107 = 0.34$ years

Infants with mortality $< 0.107 = 0.049 + 2.742 * (\text{mortality})$

Children 1-4 where infant mortality $\geq 0.107 = 1.356$ years

Children 1-4 where infant mortality $< 0.107 = 1.587 - 2.167 * (\text{infant mortality})$

Everybody else lives 2.5 years (out of 5 possible years).

Third we compute the probability of death nqx for each country c, group age a, and gender g (this is the probability of dying between ages x and x + N, which is period a):

$$nqx_{r,c,p} = \frac{N_c * TotalMortality_{r,c,p}}{1 + (N_c - nax_{r,c,p}) * TotalMortality_{r,c,p}}$$

where N is the number of years in the age category (1 for infants, 4 for children 1-4, and 5 for everybody else), and nax is the number of years lived by those who died, described in the previous step. We're assuming $nqx = 1$ when we reach our maximum age category (100+ in general).

Fourth we start adding years for each age category a in the following way:

$$LifeEx_{r,p} = LifeEx_{r,p} + (lx_{r,c,p} * (1 - nqx_{r,c,p}) * N_c) + (lx_{r,c,p} * nqx_{r,c,p} * nax_{r,c,p})$$

Where the first term added to life expectancy is the total number of years (N) lived by those who survive this age category (1 - nqx) given they have survived all previous ages (lx). The second term is the number of years (nax) lived by those who die in this age category (nqx) given they have survived all previous ages (lx).

¹ Thanks to José Solórzano and Randall Kuhn for help on this section and broader help on the population model.

The probability of surviving until age a is computed as:

$$lx_{r,c,p} = lx_{r,c-1,p} * (1 - nqx_{r,c-1,p})$$

where lx at birth is 1.

Infant Mortality is simply calculated as the sum of all our 15 mortality types from internal variable `modmordstdet` but only for age 0 (infants).

3.4 Mortality: The Legacy Formulation

In the legacy population model (should the use of the health model ever be turned off) an initial value of life expectancy (LIFEXP) is computed first and used to determine the mortality distribution (`mordst`, dimensioned by region, age cohort, and sex).

Adjustments are made to the mortality distribution by a number of factors and then life expectancy is recomputed.

The initial calculation of life expectancy is based on long-term development, namely GDP per capita at purchasing power parity (GDPPCP). The logarithmic function is modified by an additive term related to the extent of government spending on health (GDS), although that term is very minor in the calculation.

$$LIFEXP_{r,p} = F(\ln(GDPPCP_r)) + F(GDS_{r,g=health})$$

The calculation of life expectancy is wrapped in a substantial algorithmic structure. For instance, should the formulation suggest decrease in life expectancy over time, the decrease is smoothed via use of a moving average. The impact of government spending is also limited algorithmically.

We impose this initial calculation of life expectancy on the mortality distribution (with the movement towards an ultimate life table that is discussed in connection with the overall logic of the age distribution), by calculating a mortality factor (MFACTOR) that, when applied to all cohorts of the mortality distribution would generate the calculated life expectancy. The multiplier is computed so that the cumulative mortality to the age of life expectancy will ultimately be 0.5.

$$MFACTOR_r = .5 / \sum_{c=1}^{LIFEXP} MORDST_{r,c,p}$$

We then further modify the mortality distribution and therefore the life expectancy (which we will need to recompute below) by the specification of several additional mortality factors. These include three of the four horsemen of the apocalypse, which tend

to have a more immediate, shorter-term impact: starvation deaths, plague or in this case AIDS deaths (AIDSDTHS), and war deaths (using the civilian damage variable, CIVDM, calculated in the social-political module). We build starvation deaths in a recalculated infant mortality (INFMOR), because the youngest are most vulnerable to calorie shortages.

The additional mortality factors also include a parameter that reflects a time-related shift in mortality from medical advance (**tmortr**); long-term development (as reflected by GDP per capita) does not capture this additional influence on mortality. Finally, it includes a multiplier on mortality (**mortm**) that the user can set as desired to introduce further factors into a scenario.

In the second stage of mortality calculation we compute deaths by cohort.

$$\begin{aligned}
 DEATHS_{r,c=1,p} &= \sum^G AGEDST_{r,c=1,p} \\
 &\quad * (MORDST_{r,c>1,p} + CIVDM_r + CLSF_r + AIDSDTHSCOH_{r,c,p}) \\
 &\quad * \mathbf{mortm}_r * (1 + (t - 1) * \mathbf{tmortr})
 \end{aligned}$$

$$\begin{aligned}
 DEATHS_{r,c>1,p} &= \sum^G AGEDST_{r,c>1,p} * (MORDST_{r,c>1,p} + CIVDM_r + AIDSDTHSCOH_r) \\
 &\quad * MFACTOR_r * \mathbf{mortm}_r * (1 + (t - 1) * \mathbf{tmortr})
 \end{aligned}$$

where

$$AIDSDTHSCOH_{r,c,p} = AIDSDEATHS_r * \mathbf{maleageportion} * \mathbf{aidsdeathsbyage}_{c,p}$$

The computation of civilian war damage/deaths (CIVDM) is shown in the international political module.

One of the factors above that affects infant deaths is a calorie starvation factor (CLSF). It depends on the ratio of calories available (CLAVAL) from the agricultural model to the calories needed (CLNEED). Details are available with the discussion of the legacy approach to nutrition/malnutrition.

We can now recompute the actual infant mortality, based on the actual infant deaths:

$$INFMOR_r = DEATHS_{r,c=1,p}$$

Finally, we recompute life expectancy based on the entire patterns of deaths across age categories.

$$LIFEXP_r = F(DEATHS_{r,c,p})$$

3.5 Malnutrition: The Legacy Formulation

The health model has replaced the legacy formulation for child malnutrition rate or percent of population (MALNCHP) with a representation tied not just to calorie availability but also to access to safe water and sanitation. See documentation on that relationship. This section documents the earlier and simpler formulation tied only to calories per capita.

In the legacy model IFs has estimated a relationship between calorie availability per capita (CLPC) and the percentage of children (MALNCHP) between the ages of 0-5 who are malnourished. In some countries, notably India, Bangladesh, and Nepal, initial values for this percentage are far from the value predicted by the analytical function representing this relationship. IFs assumes that outliers will converge towards the table function relationship over time (as controlled by the conversion parameter, *polconv*).

IFs uses that relationship to update the percentage malnourished over time and to compute the actual number of malnourished children (MALNCHIL) in population cohorts 1 (infants) and 2 (0-4 years of age).

$$MALNCH_r = F(CLPC_r, \text{polconv})$$

$$MALNCHIL_r = (AGEDST_{r,c=1,p} + AGEDST_{r,c=2,p}) * MALNCH_r / 100$$

Relatively few models attempt to close the loop between food availability and mortality. (See, for example, Meadows, et. al., 1974 and Mesarovic and Pestel, 1974). IFs does so, while recognizing that little is actually known about the linkage. IFs treats calories as the basis for severe malnutrition- or starvation-related deaths. Regional calorie need (CLNEED) is computed by a sum across the age distribution (AGEDST), considering age specific calorie requirements (CLAGE) and an exogenous factor *clnf* with which the user can introduce regional variation in needs (or assumptions of regional differences in ability to respond to calorie shortages).

$$CLNEED_r = \sum_c^C \sum_g^G AGEDST_{r,c,p} * clage_c * clnf_r$$

$$CLSF_r = (1 - \frac{CLAVAl_r}{CLNEED_r})^{clexp}$$

Once the calorie-based starvation factor (CLSF) is computed, with admitted arbitrariness in specification, it is possible to compute actual starvation death levels (SDEATH) in the youngest two cohorts,

$$SDEATH_r = \sum_{c=1,2}^G \sum_{p} agedst_{r,c,p} * CLSF_r$$

3.6 HIV/AIDs Mortality: The Legacy Formulation

In the legacy version of the population model, mortality from HIV/AIDs was treated separately from other mortality, which was related largely to income growth and increasing life expectancy. HIV/AIDS was seen to be a special plague-like disease with a likely rise and fall in coming years that should be represented additionally to other mortality. The HIV/AIDS formulation is still in the legacy code and would be activated in the health model switch (*hlmodels*) were turned off. But normally HIV/AIDS is represented (with fundamentally the same logic) in the health model and those with interest should look at that documentation.

3.7 Migration

Migration is treated with a pooled approach, which means that the model does not determine the flows between any two countries, but rather the net inward migration (MIGRANTS) to each country, making sure that new inflows and outflows balance globally. It is driven by an exogenous parameter (*migrater*), which we derive from the migration forecasts of other organizations such as the UN Population Division or the International Institute of Applied Systems Analysis, specifying the net percentage of the population migrating each year (negative values indicate immigration and positive values indicate emigration). The user can increase or decrease global migration as a whole with a world migration multiplier, *wmigrm*. The first step is to swap the parameter values into an internal model calculation of the migration rate (MIGRATE).

$$MIGRATE_r = migrater_r * wmigrm$$

The full global set of migration rates is unlikely, however, to provide a balanced global total of immigrants and emigrants. The next step is thus to calculate those totals, even though they are likely to be unequal.

$$\text{if } MIGRATE_r > 0 \text{ then SUMIM} = \sum^R MIGRATE_r * POP_r$$

$$\text{if } MIGRATE_r \leq 0 \text{ then } SUMEM = \sum^R MIGRATE_r * POP_r$$

After calculation of the world sums of immigrants and emigrants, the total world migration is assumed to be the average of the two. Then that total world migration is imposed on net immigrant and net emigrant regions through normalization.

$$WORLDIMEM = \frac{SUMIM + SUMEX}{2}$$

$$\text{if } MIGRATE_r > 0 \text{ then } MIGRANTS_r = \frac{MIGRATE_r * POP_r * WORLDIMEM}{SUMIM}$$

$$\text{if } MIGRATE_r \leq 0 \text{ then } MIGRANTS_r = \frac{MIGRATE_r * POP_r * WORLDIMEM}{SUMEM}$$

Although the above equation assures that the global sum of migrants will be zero (immigration equals emigration), it is important to recompute the actual migration rate, so that it represents the true inflow or outflow of migrants after that balancing. Note that the computed migration rates (MIGRATE) will almost certainly be a bit different from the input parameter (migrater).

$$MIGRATE_r = \frac{MIGRANTS_r}{POP_r}$$

The migration specification in IFs is, as indicated above, basically exogenous. Different series can be pulled from IFsHistSeries.mdb to drive it. The active series is determined by specification within IFsInit.mdb, Table IFsInit, variables MigrantsTbl and MigrationRateTbl. For instance, those two variables have values of SeriesForecastNetMigrationUNPD and SeriesForecastNetMigrationRateUNPD to pull in the migration data from the UN Population Division

3.8 Urbanization

The size of urban population (POPURBAN) in the very near future is probably best forecast by using a growth rate (POPURBGR) computed initially from historic data, but gradually coming to represent the dynamic growth rate of urbanization calculated by the model. The growth rate applied to past urban population provides an initial estimate of urban population each year PopUrbanGro).

In the long-term future, urbanization must saturate as the portion of the population urbanized approaches 100%. Moreover there is a relationship between income levels of countries and urbanization level that should affect the growth of urban population. Thus a function estimated cross-sectionally against GDP/capita at PPP was used to provide a target (PopUrbanTar) for urbanization that could gradually replace the value of growing urban population (PopUrbanGro) calculated by use of the growth rate –countries with very high levels of GDP/capita have already begun to approach saturation; algorithmic modifications help assure that the target is reasonable and also that it approaches saturation smoothly.

$$POPURBAN_r = ConvergeOverTime(PopUrbanGro, PopUrbanTar)$$

where

$$PopUrbanGro = POPURBAN_{r,t-1} * (1 + POPURBGR_{r,t-1})$$

and

$$PopUrbanPor = AnalFunc(GDPPCP_r)$$

$PopUrbanTar = POP_r * PopUrbanPor$ with algorithmic modifications for smooth behavior over time and as saturation is approached.

Once the urban population has been updated in each time cycle, it is possible to compute the actual growth rate (POPURBGR), which will then be the starting point for growing urban population (PopUrbanGro) in the next time cycle.

$$POPURBGR_r = \left(\frac{POPURBAN_r}{POPURBAN_{r,t-1}} - 1 \right) * 100$$

3.9 Household Size

Household size (HHSIZE) is a function of the portion of the population that is of pre-work-force-entry age (POPWORK); the bigger that population that has not begun to work is as portion of the population, the larger is household size.

$$HHSIZE_r = \frac{1}{F\left(\frac{POPWORK_r}{POP_r}\right)} + HHSIZEShift_r$$

Internal to the model the denominator of this equation is referred to as the household intensity, which falls as the pre-work age term rises. Thus the household size rises with the pre-work age term.

There is an additive shift factor calculated in the first year of the model run to assure a match of the calculated and empirical values; that shift factor decays to zero over 100 years.

3.10 Demographic Indicators

Among the indicators computed in the population submodel of IFs are the crude birth rate (CBR) and crude death rate (CDR).

$$CBR_r = \frac{BIRTHS_r}{POP_r} * 1000$$

$$CDR_r = \frac{DEATHS_r}{POP_r} * 1000$$

Population growth rate (POPR) follows easily from crude death and birth rates.

$$POPR_r = \frac{CBR_r - CDR_r}{1000}$$

Regional population (POP) is simply a sum across age cohorts.

$$POP_r = \sum^c AGEDST_{r,c}$$

For information and use elsewhere in the model, three computations of sub-portions of the population by age are made (POPLE15, POP15TO65, and POPGT65). [Note: each one of these is slightly misnamed.]

$$POPLE15_r = \sum_0^{14} AGEDST_{r,c}$$

$$POP15to65_r = \sum_{15}^{64} AGEDST_{r,c}$$

$$POPGT65_r = \sum_{65}^{oldest} AGEDST_{r,c}$$

More recently the IFs model has recognized that the working life span is not uniformly from 15 to 65 across countries or time and has designated country/region specific parameters for age of work entry (*workageentry*) and retirement (*workageretire*).

These are used to compute POPPREWORK, POPWORKING, and POPRETIRED. They also allow the computation of a potential support ratio for the retired population (POTSUPRAT), which is the ratio of those of working population to those of retirement age.

$$POTSUPRAT_r = \frac{POPWORKING_r}{POPRETIRED_r}$$

Another useful indicator is the youth bulge (YTHBULGE), defined as the ratio of the population between ages 15-29 to that aged 15 and above. In general, a ratio of more than 0.4 and especially 0.5 suggests a particularly youthful society and may indicate potential for social instability.

$$YTHBULGE_r = \frac{\sum_{15}^{29} AGEDST_{r,c}}{\sum_{15}^{Oldest} AGEDST_{r,c}}$$

Median age (POPMEDAGE) is another useful indicator and the age distribution (fagedst) can be used to determine that age at which there are equal numbers of people older and younger.

World population (WPOP) and world population growth rate (WPOPR) are simple functions across countries/regions.

$$WPOP = \sum^R POP_r$$

$$WPOPR = \frac{\sum^R POP_r * POPR_r}{WPOP}$$

3.11 Data Used

Our data for the population model come from the United Nations Population Division revisions of data and forecasts, released every second year. We take population by age and sex from that source, as well as historical series for life expectancy, total fertility rate, and infant mortality. We also pull in their migration data.

Often they present their data values in 5 year categories (1950-1955, . . . , 2095-2099). To obtain values at specific 5-year intervals, such as 1960 or 2010, we average the values for the two 5-year categories that bracket that year. To estimate annual values for all years from such categories, for instance for migration numbers, we used a Sprague

algorithm to spread the 5 year data (1950-1955, . . . , 2095-2099). With respect to migration, to obtain net migration rates we divided their annualized numbers by annual population data.

4. References

- Angeles, Luis. 2010. "Demographic Transitions: Analyzing the Effects of Mortality on Fertility", *Journal of Population Economics* 23: 99-120. DOI 10.1007/s00148-009-0255-6.
- Coale, Ansley and Paul Demeny with Barbara Vaughan. 1983. *Regional Model Life Tables and Stable Populations*. New York: Academic Press.
- Hughes, Barry B. 1980. *World Modeling*. Lexington, Mass: Lexington Books.
- Hughes, Barry B. 2001. "Global Social Transformation: The Sweet Spot, the Steady Slog, and the Systemic Shift," *Economic Development and Cultural Change* 49, No. 2 (January 2001): 423-458.
- Meadows, Dennis L. et al. 1974. *Dynamics of Growth in a Finite World*. Cambridge, Mass: Wright-Allen Press.
- Meadows, Donella H., Dennis L. Meadows, Jørgen Randers, and William W. Behrens III. 1972. *The Limits to Growth*. New York: Universe Books.
- Mesarovic, Mihajlo D. and Eduard Pestel. 1974. *Mankind at the Turning Point*. New York: E.P. Dutton & Co.
- Preston, Samuel H., Patrick Heuveline, and Michel Guillot. 2001. *Demography: Measuring and Modeling Population Processes*. Oxford: Blackwell Publishing.
- United Nations, Department of Economic and Social Affairs. 1956. *Methods of Population Projections by Sex and Age*. New York: United Nations, ST/SOA Series A.