

The Demographer's Toolbox

Population Projection as a Consistency Check

Griffith Feeney

July 2003

The Problem

It is the mid-1970s and the detailed results of the 1972 census of Pakistan have just been released. The press was rife with speculation about over statement of the population in different places. Everyone knows the reasons why this would have happened, if it did happen.

Following the breakdown of the Federation of Pakistan, the political principle that divided government benefits equally between the former East and West Pakistan has been applied to provinces in West Pakistan. It is rumored that the government of Sindh Province is developing a plan to systematically overstate it's population, and that other provinces are developing similar plans in "self defense", so to speak.

Most people don't know how much truth, if any, there is to these rumors, and those that do know are not talking, but now the 1972 census results are out, and it is possible to calculate the intercensal growth rate between 1961 and 1972. The two censuses enumerated 42,880 and 65,309 thousand persons, respectively, which gives a growth rate (taking account of the census dates, which were February 1, 1961 and September 16, 1972) of 3.6 percent.

If there was over enumeration in 1972, and none, or less, in 1961, this growth rate should be too high. It is indeed very high, but what is "too high"? One can't make a persuasive argument on the basis of the growth rate alone, but other evidence may be brought into play.

If we knew the birth rate and the death rate (the so-called "crude" rates), for example, and if we knew that international migration was negligible as a source of inter-censal population change, we could calculate the growth rate as the birth rate minus the death rate and see whether this is lower than the inter-censal growth rate.

The projection calculations described below are an extension of this simple idea. They show beyond reasonable doubt that the 1972 census count is too high relative to the 1961 census count.

Projection is as Easy as 1-2-3

Population projection originated as a technique for forecasting future population. It turns out to be very important to take account of the population *age distribution* when doing this. It is not so easy to understand the reason for this, but it has to do with something called *population momentum*, the tendency of population growth rates to lag behind changes in childbearing behavior and longevity. For more on this subject see my entry on in the recently published *Encyclopedia of Population* (edited by Paul Demeny and

Geoffrey McNicoll, Macmillan, 2003). A final prepublication version is available to persons attending this presentation.

We are going to use population projection not to forecast future population, but to forecast population in the more recent past (1972) from population in the more remote past (1961). Our aim is but to test the consistency of (1) the two census counts and (2) estimated intercensal levels of fertility and mortality.

To introduce the idea of population projection (by the so-called “component method”, if you have heard this jargon), consider how we might forecast the population 5 years from now given the current age distribution of the population.

The population five years into the future consists of two kinds of persons, those who are alive now (and survived) and those who are born during the next five years (and survived). To calculate the population five years in the future therefore involves the following three steps.

?? Calculate the number of survivors 5 years in the future of persons living now. We do this by breaking down the population by age and sex and applying “survivorship ratios” to the number of persons in each age and sex group. To get the survivorship ratios we need a life table. This gives us the number of persons age 5 and over 5 years into the future, distributed by 5 year age group and sex.

?? Calculate the births over the next five years by applying age-specific birth rates to the “person years lived” by women in the age groups 15-19, 20-24, ..., 45-49. This gives us total births. Since we are distributing the population by sex, we need to apply the sex ratio at birth to get the numbers of male and female births.

?? Calculate the number of female and male births during the period who survive to the end of the period, at which time they are necessarily in the 0-4 age group.

There is a very handy type of diagram for visualizing these various sets of persons and events. If you are interested, see my entry on “Lexis Diagrams” in the *Encyclopedia of Population* (as above, final pre-publication version is available to persons attending this presentation).

Spreadsheet Implementation

Implementation is a bit harder than 1-2-3. The accompanying series of spreadsheets, numbered 1 through 7, starts for the very beginning and adds progressively more detail. I’ll work through them in the presentation, but not give a blow-by-blow here.

For those of you who are technically inclined, the most complicated things here are the VBA functions *GetLxFemale* and *GetLxMale* used to calculate the model life table ${}_5L_x$

values needed for survivorship. These are in the spreadsheet *un-general-llx.xls*, which must be open for the other spreadsheets to work.

If you copy these files to your computer, you will also have to fiddle a bit with the paths in the links to get them to work properly. See your IT people in case of problems, or email me at feeney@gfeeney.com.

The 1961-72 Inter-censal Growth Rate is Too High

With these tools in hand we're ready to address the question about the 1961-72 inter-censal growth rate, namely, is it too high? We have very strong evidence from fertility surveys that the level of fertility during the inter-censal period was constant at 6.8 children per woman. We combine this with the age pattern of fertility rates given by Blacker, which is as follows.

Age Group	Rate
15-19	0.1352
20-24	0.3203
25-29	0.3406
30-34	0.2882
35-39	0.2081
40-44	0.0943
45-49	0.0118

The evidence on mortality level is less satisfactory, for we must extrapolate levels during the 1961-72 intercensal period back from estimates for the 1972-81 intercensal period (John Blacker) and from the Pakistan Demographic Survey from 1984 forward. Since the trend is pronounced, however, the result may be satisfactory. We get the following estimates of life expectancy at birth.

Period	Female	Male
1961-65	43.3	44.4
1966-70	47.1	47.7
1971-75	50.8	51.1

With these statistics in hand we project the 1961 age-sex distribution forward 15 years and interpolate between the projections for 10 and 15 years to get the projected age-sex distribution for the time of the 1972 census.

The projection number for females is 26,782 thousand, versus the census number of 30,476. The projection is low by 13.8 percent. If our fertility and mortality levels are correct, this shows that the 1972 count is too high relative to the 1961 count. Either the 1972 count is high in absolute terms, or the 1961 count is low in absolute terms, or some

combination of the two. In any case, the accuracy of enumeration in the two censuses is not the same, and the calculated inter-censal growth rate is too high.

But what if our levels of fertility and mortality are not correct? This is a possibility we must consider. We should ask, for example, what level of fertility would be required to lift the 1972 projection up to the level of the 1972 census count. Using “Tools > Goal Seek” in Excel we see that the answer is 9.4 children per woman (don’t anybody tell Paul Erlich).

Similarly, we must ask what level of life expectancy at birth would be required to bring the 1972 projection value up to the level of the 1972 census count. In this case there is a sharp increase in values over the three projection periods, so we ask how much all three values would have to be raised to accomplish the result. The answer is 22.7 years, which gives life expectancies at birth for the three periods of 66.0, 69.8 and 73.5 years.

As these levels of fertility and mortality are patently absurd for the 1961-72 inter-censal period, we may safely dismiss them and conclude that natural increase cannot account for the population growth rate observed between the 1961 and 1972 censuses. The levels of fertility and mortality would have given a growth rate of only 2.6 percent, a full percentage point below the observed value.

Similar observations apply to males. We may go through them in the presentation, but will not here.

The only remaining explanation, aside from differential accuracy of census enumeration, is migration, but it would have taken a net international in-migration of 3.6 million females between the 1961 and 1972 censuses to generate the observed growth for females. Another 5 million male in-migrants would be required. Even absent data on migration, it is clear that these are most improbable numbers.

Analyzing Age Distribution Errors

Projection is a powerful tool for analyzing errors in age distributions. We will discuss this during the presentation if time permits. There is a glaring anomaly in the 1961 census age-sex distribution. We can test whether it is real or an artifact of data errors by comparing the projected age distribution for 1972 to the census age distribution.