EXTENDED ABSTRACT

Death Distribution Methods for Estimating Adult Mortality: Sensitivity Analysis with Simulated Data Errors, Revisited

Rob Dorrington, Ian M Timaeus and Tom A Moultrie
Centre for Actuarial Research, University of Cape Town

Hill and Choi attempted to test how well the variable-r techniques for estimating adult mortality (Generalised Growth Balance and the Synthetic Extinct Generations methods) perform in the context where data deficiencies violate the assumptions underlying the models. From this work they concluded that each method had strengths and weaknesses and that superior estimates could be derived using a combination of the two methods. However, the version of the SEG method they tested did not account for differential coverage of the censuses. This paper tests the more general version of the model on the same set of scenarios and finds that the SEG method performs so well as to call into question the recommendation that one should use a combination of the two methods. In addition the paper explores the generalisation of these results to scenarios that represent mortality conditions in an African country with a generalised HIV/AIDS epidemic.

1. Introduction

The Generalised Growth Balance (GGB) method (Hill 1987) and the Synthetic Extinct Generations (SEG) method (Bennett and Horiuchi 1981, 1984) are two methods used to estimate completeness of reporting of deaths in developing countries where one can no longer assume that population has remained stable. However, as (Hill and Choi 2004)point out very little research seems to have been undertaken into how well these methods perform in the context of the sort of data deficiencies found in practice in developing countries.

(Hill and Choi 2004) attempted to do this measuring the extent of deviation of the estimate of ${}_{45}q_{15}$ produced by these methods under certain combinations of simulated data errors (see appendix). From this they concluded as follows.

> Reassuringly, both the GGB and the SEG approaches work very well when the errors for which they were developed are the only ones present in the data. GGB has some advantage here in that it allows for an additional systematic error, change in census coverage, but in this instance the SEG diagnostic plot shows very clearly that something has gone badly wrong – SEG diagnostic plots wear their hearts on their sleeves [more] than do those from GGB. Effects of age misreporting (as modelled in this exercise), whether in the populations or in the deaths, do not have huge adverse effects. A pronounced age pattern of population coverage (essentially high omission rates for young adults) has a large adverse effect on GGB results, but less so on SEG results. On the whole, SEG is somewhat less sensitive to age misreporting or differential coverage by age than GGB. Increasing or decreasing coverage of deaths by age affects GGB somewhat more than SEG, but both methods are generally an improvement over unadjusted

> data. Perhaps the nearest thing to a surprise is the different responses of the two methods to migration. Using emigration as the example, SEG is strongly affected, tending to underestimate death coverage and overestimate mortality. GGB, on the other hand, is little affected in terms of the final mortality estimate, because both the intercept (estimated census coverage) and the slope (death coverage) are distorted in compensating ways.

This led them and others (Hill, Choi and Timæus 2005) to conclude that a combined approach using the GGB method to estimated the extent of coverage of one census relative to the other, and the SEG method with censuses corrected for relative coverage (GGB+SEG), to estimate the completeness of the reporting of deaths was the best option.

Unfortunately the version of the SEG that Hill and Choi tested did not include an adaptation suggested by Bennett and Horiuchi in an end note to their 1981 paper, namely that, provided that the coverage of one census relative to the other was constant for all (relevant) ages, one could account for this by the simple addition of a constant (equivalent to $-\ln(c_2/c_1)/t$ where c_1 and c_2 represent the coverage of the first and second censuses respectively) to the age specific population growth rates.

In this paper we investigate whether one comes to the same conclusions applying this adaptation of the SEG method to the same data set[1]. In addition the paper investigates if the conclusions one arrives at hold when the data set is adapted to reflect the pattern and level of mortality likely to be experienced in African countries with a high level of HIV. Finally the paper considers the question of whether or not a combination of the two methods can be which might be preferable to either one of the methods and under what conditions this might be so.

2. Structure of the paper

Thus the paper comprises three analytical sections. The first section compares the results of the application of the three methods (GGB, SEG and the combination (GGB+SEG)) to the Hill and Choi data set and highlights the extent to which their conclusions are no longer valid. The second section repeats this exercise but on the data set with the true populations at the dates of the censuses and the true number of intecensal deaths changed to reflect and African population with a prevalence of HIV in adults aged 15-49 of around 17%, but with the same data distortions, relatively, as in the original data set. This exercise lets us examine the extent to which the conclusions from the first section can be generalised, in particular to African populations, where the need for such methods is probably most apparent.

[1] Choosing the constant which minimizes the mean of the absolute differences of the age-specific estimates of completeness from the median of these estimates.

The final section considers the question of how one might combine the two methods, which both make use of the same data, in order to produce results superior to those of one method on its own, and under what conditions this might work.

3. (Anticipated or known) Results

The figures below show the results of the above applications. Figure 1 gives the original results of Hill and Choi, Figure 2 the results of the SEG method are replaced by results using the method adapted to correct for differential coverage of the censuses (SEG+delta), and Figure 3 the application of these methods to the 'African (with AIDS)' data set.

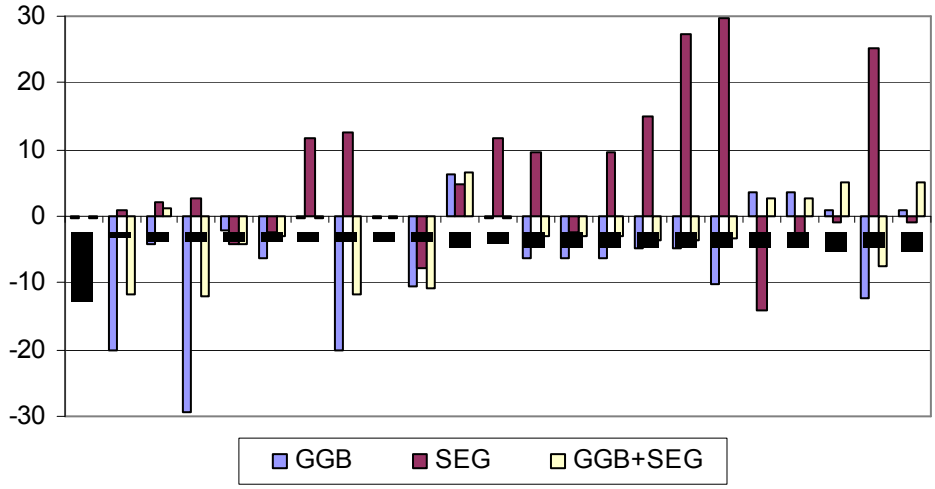

Figure 1 Percentage difference between the estimated and true ${}_{45}q_{15}$ by different methods for different error patterns: Hill and Choi

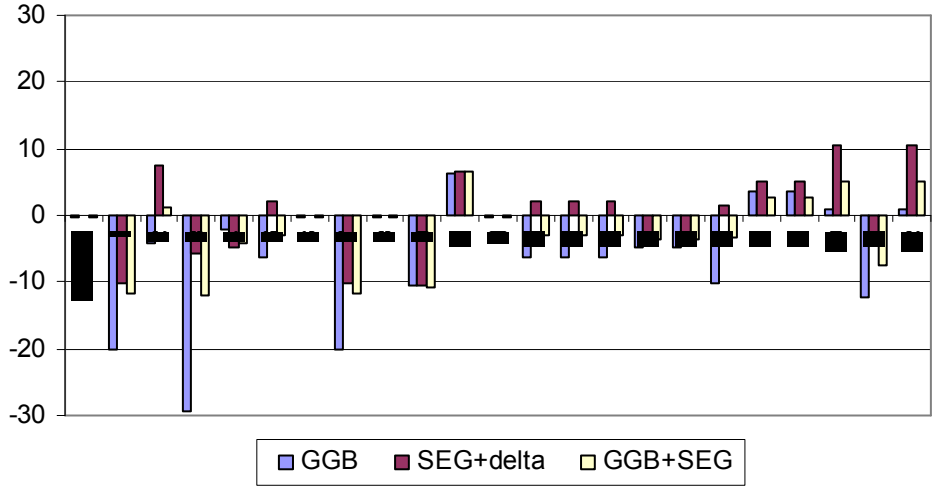

Figure 2 Percentage difference between the estimated and true ${}_{45}q_{15}$ by different methods for different error patterns: Adapted SEG method

A comparison of the Figures 1 and 2 demonstrates clearly the improvement in estimates which result from the adaptation of the SEG method. The mean and median of the estimates of this method applied in all 22 scenarios is closer to the true value with a much smaller root mean square error than either of the other two methods. It outperforms or was joint best in 15 out of the 23 scenarios and was second best in a further scenario. In only three of the cases where the SEG+delta was not the best was the difference between it and the best greater than 5% of the mortality rate – all involved age misreporting in the censuses, and the two extreme cases, where GGB was by far the best, included 2% decline in census coverage and immigration, all not accounted for. The GGB+SEG combination was only best in five scenarios, when an overestimate of mortality by the SEG method was more or less cancelled by an underestimate by the GGB method, typically where there was age misreporting in the census or where there was immigration or emigration, situations where the SEG performed worst. All methods are vulnerable to violation of the assumptions that census coverage and the completeness of death registration are constant for all (relevant) ages.

In order to apply the SEG method one needs (as described by Bennett and Horiuchi) an estimate of the life expectancy at the age of the open interval (in the above data set this is age 85). Hill and Choi derived this from the ratio of ${}_{30}d_{10}$ to ${}_{20}d_{40}$ and a look-up table based on regressions for the West Princeton life table ((Coale, Demeny and Vaughan 1983) produced by (Bennett and Horiuchi 1984). Unfortunately this relationship and any others based on deaths in the age range in which AIDS deaths mostly occur, are not applicable to a life table which reflects the ravages of HIV/AIDS and thus an alternative, of approximating the life expectancy from the Gompertz curve fitted to the mortality rates over age 60 after correcting for under-reporting, was used.

Application of this alternative method produced very similar results on the Hill and Choi data set, identifying SEG+delta as best in the same 15 scenarios and producing a mean and median of all scenarios only 1 per mille lower than was previously the case, with exactly the same variation. Thus it was decided that this method of estimating the life expectancy at the age of the open interval could be applied to the second (African) data set.

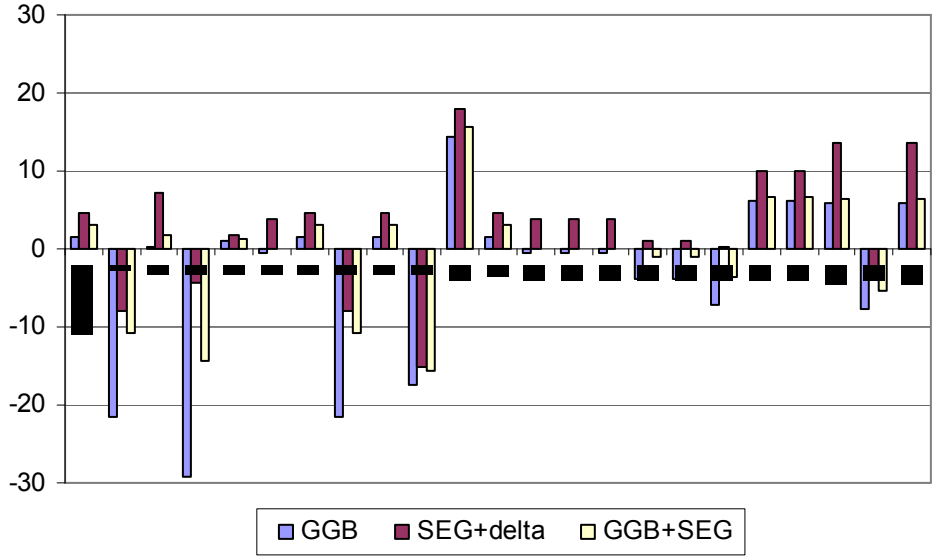

Figure 3 Percentage difference between the estimated and true 45q15 by different methods for different error patterns: African data set

A comparison of Figure 3 with Figure 2 shows that while the relative pattern of errors holds, the magnitudes are surprisingly somewhat different. Generally the effect is to increase the GGB estimates closer to the correct value and to increase the errors in the estimates for the SEG+delta and GGB+SEG methods, and to increase the variation in all cases. The increase in error in the latter two methods is particularly noticeable where completeness of vital registration is assumed to increase or decrease with age (scenarios 9 and 10).

As a result of this the number of situations where the GGB and GGB+SEG outperform increases to eight while the number reduces for the SEG+delta method from 15 to seven. None of the methods perform as well in the 'no error' scenario with the error for the SEG+delta being nearly 5%. Interestingly the error is exactly the same for scenarios 6 and 8, i.e. where there is uniform differential coverage of the censuses or uniform omission of deaths by age – which are the conditions the methods are designed to allow for, but separately and not together. When these errors are combined with one another or others errors (e.g. scenarios 11-14) then the SEG+delta method does not perform very well.

The GGB method does not perform well in scenarios 1, 3 and 7 (i.e. scenarios where census coverage varies with age) or scenarios 15-17 (i.e. scenarios with unaccounted for emigration). The SEG+delta method performs better, particularly for scenario 3 (which includes age misreporting in the census) and scenarios 15-17.

All methods perform poorly, with the SEG+delta method performing worst, in scenarios 18-20 and 22 (i.e. those involving immigration), and scenario 21 (i.e. emigration in combination with other factors), with SEG+delta method performing best.

These results are despite the use of an objective method of determining life expectancy at the age of the open interval that in all scenarios produces life expectancies at age 85 higher than that for the level 25 West model life table. Application using the life expectancy derived from the mortality rates used to create the scenarios lead to an increase in the estimated mortality rates resulting in them being closer to the correct rates on average for the GGB+SEG method and nearly as close in the SEG+delta method, with the variance little changed. In this case the GGB method outperformed in nine cases, the SEG+delta in eight and the GGB+SEG in only six, although the pattern of distortion was consistent with that shown in Figure 3. Interestingly, the performance in the 'no error' scenario was somewhat worse with the error ranging from 7 (GGB) to 11 (SEG+delta) per mille.

4. Relevance

This research is relevant for several reasons. Estimating adult mortality has gained in importance in the era of HIV/AIDS as a means of measuring and tracking the extent of the epidemic in addition to estimating HIV prevalence from antenatal survey results. However, methods that rely on the use of model life tables are by and large no longer applicable since traditional model life tables do not include HIV/AIDS and even model life tables such as those published by INDEPTH, are strictly only applicable to situations having experienced similar levels of prevalence in the past as the average of the sites used in the sample used to produce the model life table. Thus increasing attention needs to be focussed (in situations where there is limited or no VR data, as is the case in many, if not most, countries with high HIV prevalence) on deaths reported by households in censuses as a source of mortality data. In order to convert these data into reliable estimates of mortality one needs to be able to adjust them for errors in completeness within the context of the data errors one might encounter in such countries. This research helps inform that process.

In addition, three other benefits are realised. First, this research corrects the conclusions drawn by others about the (in)applicability of the SEG method arising from their testing a less general form of the SEG model. Second, related to this, the research indicates that the applicability GGB+SEG combination is more restricted than has been suggested and that – in several instances – the correct interpretation of SEG should prevail (the full paper will seek to identify the specific conditions under which the corrected version of SEG is both better and

worse than the alternative suggestion). Third, this research suggests how one might produce the best estimates of adult mortality from limited data of the form typically collected in censuses in developing countries.

Bennett, N. G. and Horiuchi, S. 1981. "Estimating the completeness of death registration in a closed population", *Population Index* **47**(2):207-221.

Bennett, N. G. and Horiuchi, S. 1984. "Mortality estimation from registered deaths in less developed countries", *Demography* **21**(2):217-233.

Coale, A. J., Demeny, P. and Vaughan, B. 1983. *Regional Model Life Tables and Stable Populations.* New York: Academic Press.

Hill, K. and Choi, Y. 2004. "Death Distribution Methods for Estimating Adult Mortality: Sensitivity Analysis with Simulated Data Errors," Paper presented at Adult Mortality in Developing Countries Workshop. The Marconi Center, Marin County, California, 8 - 11 July, 2004.

Hill, K., Choi, Y. and Timæus, I. M. 2005. "Unconventional approaches to mortality estimation ", *Demographic Research* **S4**

Hill, K. H. 1987. "Estimating census and death registration completeness", *Asian and Pacific Population Forum* **1**(3):8-13.

Appendix 1: List of data error scenarios

Description	Scenario number
No error	No error
Age varying census coverage (based on net-undercount of male black population from the 1980 United States Census)	1
Age misreporting in censuses (derived from a matrix of transfers between 5-year age groups estimated for India)	2
Age misreporting in censuses + Age varying census coverage	3
Age misreporting in VR (derived from a matrix of transfers between 5-year age groups estimated for India)	4
Age misreporting in censuses + Age misreporting in VR	5
Census coverage decline of 2%	6
Census coverage 2% decline + Age varying census coverage	7
VR 20% omission	8
VR 20% omission + increasing completeness with age (linearly to 100% by age 85+)	9
VR 20% omission + decreasing completeness with age (linearly to 70% at age 85+)	10
Census coverage 2% decline + VR 20% omission	11
Age misreporting in censuses + Age misreporting in VR + Census coverage 2% decline	12
Age misreporting in censuses + Age misreporting in VR + VR 20% omission	13
Age misreporting in censuses + Age misreporting in VR + Census coverage 2% decline + VR 20% omission	14
Emigration (based on a pattern of age-specific in- migration to the U.S. of Mexican males 1980-1990)	15
Emigration + Census coverage 2% decline	16
Emigration + Census coverage 2% decline + Age misreporting in censuses	17
Immigration (based on a pattern of age-specific in- migration to the U.S. of Mexican males 1980-1990)	18
Immigration + Census coverage 2% decline	19
Immigration + Census coverage 2% decline + Age misreporting in censuses	20
Emigration + Age misreporting in censuses + Age misreporting in VR + Census coverage 2% decline + VR 20% omission	21
Immigration + Age misreporting in censuses + Age misreporting in VR + Census coverage 2% decline + VR 20% omission	22

Appendix 2: Creating of African AIDS scenarios

1. Age-varying census coverage

Assumed the same pattern of differential coverage as used in original scenario.

2. Age-misreporting in the censuses

Applied the same ratio of erroneous classifications at a particular age to the numbers in the other ages as found in the original scenario, rebalanced to sum to the original total.

3. Age-misreporting in the vital registration

Similar to age-misreporting in the censuses

4. Census coverage decline by 2% and 20% omission in vital registration

Applied change to the complete data

5. Increasing completeness with age (to 100% by age 85+, and to 70% by age 85+)

Applied the same adjustment process as used in the original data set (this is designed to ensure linearly changing completeness while maintaining the same overall level of completeness).

6. Emigration

True population after emigration = true population without + migration rate per true population from the original data set times the African true population. True period deaths allowing for emigrants = true deaths without allowing for emigrants – death rate from the original time the number of emigrants in each age.

7. Immigration

Same as 6, but opposite

8. Combinations

As might be expected when combining the above.