Defining a sub-Saharan Fertility Pattern and a Standard for use with the Relational Gompertz Model

Thesis

Submitted in fulfilment of the requirements for the degree of Master of Philosophy (Demography) at the University of Cape Town

By

Reinier van Gijsen

Centre for Actuarial Research
Faculty of Commerce
University of Cape Town
Private Rag XS
Rondebosch
7701
South Africa

August 2009

PLAGIARISM DECLARATION

I know plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own. I have used the Harvard convention for citation and referencing. Each contribution to, and quotation in, this dissertation from the work(s) of other people has been attributed, and has been cited and referenced. This thesis is my own work. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

Signed by candidate

ABSTRACT

The relational Gompertz model is often used to obtain fertility estimates for sub-Saharan Africa populations. This indirect estimation technique is dependent on a fertility standard - the Booth standard. This standard was developed in 1979 using a selection of 33 Coale-Trussell schedules congruent with high fertility patterns. However, evidence from 61 Demographic and Health Surveys of sub-Saharan countries shows that fertility has decreased to levels that were considered medium fertility at the time the standard was developed. This raises concerns about the continued relevance of the (high fertility) Booth standard. In particular, the standard would appear to consistently underestimate fertility among African women aged 45-49. This understatement occurs irrespective of the level of total fertility and suggests that fertility may be generally higher at the older ages in Africa than can be constructed with the Coale-Trussell schedules. In addition, further investigation of the understatement highlighted that the patterns of fertility for 61 African DHS are broadly similar. This result has two important implications: First, a pattern of fertility that is distinctly African can be identified. Second, it suggests that an African standard be developed to utilize with the relational Gompertz model in the analysis of fertility data. To this end a number of alternatives are considered: 1) Two reworks of the Coale-Trussell model, 2) the Brass polynomial and 3) the Hadwiger function. Of these, the two restatements of the Coale-Trussell model are dismissed due to a continued misfit in the 45-49 age group. The appropriateness of the two remaining alternatives is assessed using least squares methodology and graphical graduation and both yield apparently reasonable results. However, the Brass polynomial has the advantage of being simpler than the Hadwiger function and allows the direct calculation of cumulative fertility rates. Furthermore, statistical tests show that the Brass polynomial is superior to the Hadwiger function, since the latter fails both the smoothness and goodness-of-fit graduation tests. As a result, the Brass polynomial is deemed to be the most suitable to model the African fertility pattern and is used to develop an African fertility standard.

TABLE OF CONTENTS

TABLE OF CONTENTS III
LIST OF TABLES V
LIST OF FIGURES VII
LIST OF APPENDIX TABLES AND FIGURES IX
ACKNOWLEDGEMENTS X
1. INTRODUCTION 1
2. FERTILITY MEASURES IN DEVELOPING COUNTRIES 3
2.1 DIRECT METHODS AND DATA QUALITY 3
2.2 INDIRECT TECHNIQUES 5
2.3 THE USE OF STANDARDS 6
2.4 THE COALE-TRUSSELL MODEL 7
2.5 RESTATEMENTS OF THE COALE-TRUSSELL MODEL 11
2.6 ALTERNATIVES TO THE COALE-TRUSSELL MODEL 13
3. THE BOOTH STANDARD IN THE RELATIONAL GOMPERTZ MODEL 16
3.1 THE BRASS RELATIONAL GOMPERTZ MODEL 16
3.2 ZABA'S RESTATEMENT OF THE RELATIONAL GOMPERTZ MODEL 20
3.3 DEVELOPMENT OF THE BOOTH STANDARD 25
3.4 APPARENT ANOMALIES 36
3.5 PROBLEMS OF MISFIT 43
3.6 THE AFRICAN PATTERN AND THE BOOTH STANDARD 47
3.7 DATA AND CRITERIA CONCERNS 52
4. ALTERNATIVES TO THE BOOTH STANDARD 55
4.1 AUTOMATING THE COALE-TRUSSELL MODEL 55
4.2 COMPARING THE COALE-TRUSSELL MODEL AND DHS POPULATIONS 56
4.3 COALE-TRUSSELL MODEL REWORKED 57
4.4 PROBLEMS WITH THE COALE-TRUSSELL MODEL 63
4.5 ALTERNATIVES TO THE COALE-TRUSSELL MODEL 68
4.6 COMPARING THE BRASS POLYNOMIAL AND HADWIGER FUNCTION 72
4.7 THE AFRICAN STANDARD AND THE RELATIONAL GOMPERTZ MODEL COEFFICIENTS 74
4.8 COMPARISON OF RELATIONAL GOMPERTZ RESULTS FROM THE BOOTH AND AFRICAN STANDARDS 76

5. CONCLUSION 79
5.1 METHODOLOGY 79
5.2 USES AND BENEFITS OF AFRICAN STANDARD 80
5.3 LIMITATIONS 81
5.4 LOOKING TO THE FUTURE 84
APPENDIX A 86
APPENDIX B 88
APPENDIX C 91
C.1 COALE-TRUSSELL MODEL 91
C.2 BRASS POLYNOMIAL 99
C.3 HADWIGER FUNCTION 103
APPENDIX D 108
APPENDIX E 112
APPENDIX F 115
F.1 GOODNESS-OF-FIT 115
F.2 THIRD DIFFERENCES 117
APPENDIX G 119
REFERENCES 122

LIST OF TABLES

Table 2.1: Age-Specific Legitimate Fertility Rates per 1000 10
Table 2.2: Comparison of Coale-Trussell *n(a)* with Calculated Average 10
Table 2.3: Standardized n(a) values for the Xie Independence and Coale-Trussell models 12
Table 2.4: Standardized v(a) values for the Coale-Trussell and Xie-Pimentel models 13
Table 3.1: Fertility cumulants - terms of Taylor expansion of øx"ß) (without shift) 22
Table 3.2: Fertility cumulants - terms of Taylor expansion of 1;"(3) (with ;4 year shift) 22
Table 3.3: Parameters and derived statistics of schedules identified by Booth 26
Table 3.4: Cumulative Fertility, F(x), by age group for the 33 schedules identified by Booth 28
Table 3.5: Y(x) values by 5 year age groups of the 33 schedules used by Booth 29
Table 3.6: AY(x) by 5 year age groups as calculated by Booth 30
Table 3.7: Identifying the schedules used in each tail 33
Table 3.8: Average AY(x) values by 5 year age groups 34
Table 3.9: Determination of Booth standard from the adjusted AY(x) values 36
Table 3.10: Comparison between the Booth Standard Rates and the Average Rates 37
Table 3.11: The 33 Coale-Trussell Schedules and the f(10-19) and f(35-49) values 38
Table 3.12: Comparative standards arising from alternative starting values 40
Table 3.13: Cumulative fertility arising from alternative starting values 40
Table 3.14: Average and Standard AY values and k1 implied by these values 42
Table 3.16: Ranking of ASFR for the 78 Sub-Saharan Africa DHS 47
Table 3.17: Booth Standard compared to African Pattern 50
Table 3.18: Evaluation of the Booth Criteria for inclusion in the dataset 52

Table 3.19: Comparison of the f(45-49) values 53
Table 4.1: Standardized n(a) values for the Xie Independence and Coale-Trussell models 57
Table 4.2: Xie single-year n(a) resulting from Beers 58
Table 4.3: Standardized v(a) values for the Coale-Trussell and Xie-Pimentel models 61
Table 4.4: Xie and Pimentel single-year v(a) resulting from Beers 61
Table 4.5: Model schedule using the Brass polynomial 69
Table 4.6: Model schedule using the Hadwiger function 71
Table 4.7: Comparison of Brass and Hadwiger standards with the African fertility pattern 73
Table 4.8: P(i), f(x, x+4) and F(x) for standard excluding the 10-14 age group 74
Table 4.9: Single-year ASFR for standard excluding the 10-14 age group 74
Table 4.10: P(i), f(x, x+4) and F(x) for standard including the 10-14 age group 75
Table 4.11: Single-year ASFR for standard including the 10-14 age group 75
Table 4.12: Standard relational Gompertz model parameters (10-14 age group excluded) 76
Table 4.13: Standard relational Gompertz model parameters (10-14 age group included) 76
Table 4.14: Standardised relational Gompertz model results for Kenya Census 1979 77
Table 4.15: Standardised relational Gompertz model results for Botswana Census 2001 78

LIST OF FIGURES

Figure 3.1: Divergence in the tails of transformed cumulative fertility, Y(x), from a straight line fitted to the middle ages 17

Figure 3.2: Transformed cumulative fertility against age and the stretched age scale 18

Figure 3.3: Effects of a change in a holding $p = 1$ 19

Figure 3.4: Effects of a change in p holding a constant at 0 20

Figure 3.5: Demonstration of the misfit in the tails of the Y(x) distribution 31

Figure 3.6: Comparative standard schedules arising from different starting values of Y(30) 41

Figure 3.7: Cumulative fertility, F(x), arising from different starting values of Y(30) 41

Figure 3.8: Standardised ASFR for the Booth standard, low and high fertility populations 44

Figure 3.9: Standardised ASFR of the Booth Standard compared to 61 African DHS 46

Figure 3.10: Standardised ASFR of the Booth standard compared to the African fertility pattern 51

Figure 3.11: Ratio of the 45-49 fertility of the sub-Saharan DHS to the Booth Standard 54

Figure 4.1: Comparison of average DHS fertility with two Coale-Trussell populations 57

Figure 4.2: Natural fertility schedule, n(a), as derived by Xie (1990) 59

Figure 4.3: Comparison of the African pattern with two schedules using changed n(a) 60

Figure 4.4: Schedule of fertility control, v(a), derived by Xie and Pimentel (1992) 62

Figure 4.5: Comparison of the African pattern with two schedules using changed v(a) 63

Figure 4.6: Proportion married by age, G(a), for schedules 15 and 28 identified by Booth 66
Figure 4.7: Proportion married by age, G(a), for different levels of k holding ao constant 67
Figure 4.8: Comparison of ASFR for the African pattern and Brass model schedule 70
Figure 4.9: ASFR of African pattern compared to Hadwiger model schedule 72
Figure 5.1: ASFR of the African standard and the HIV-positive population 82

LIST OF APPENDIX TABLES AND FIGURES

Table A 1: Single-year values of n(a) and v(a) 86
Table A 2: Standard e(x) and g(x) values (no shift) 86
Table A 3: Standard e(x) and g(x) values (1/2-year shift) 87
Table A 4: Standard e(i) and g(i) values 87
Table B 1: Comparison of the Booth and Criteria AY(x) 88
Table B 2: The Adjustment Factors used by Booth and based on the Criteria 88
Table B 3: Y(x) for Booth and Criteria 89
Table B 4: Adjusted Y(x) for Booth and Criteria 89
Table B 5: Derived cumulative fertility, F(x) 89
Table B 6: Derived standard age-group specific fertility, f(x) 90
Figure D. 1: A plot of G(a) and G(a) on the same x-axis 108
Figure D. 2: f(a) and G(a) plotted on the same x-axis 110
Table F. 1: Brass and Hadwiger differences used in the Signs and Grouping of Signs tests 116
Table F. 2: Brass and Hadwiger differences used in the Signs and Grouping of Signs tests 117
Table F. 3: First, second and third differences for Brass-based standard 118
Table F. 4: First, second and third differences for Hadwiger-based standard 118
Table G. 1: Original data from the 1979 Kenya Census 119
Table G. 2: Summary results using the Booth Standard 119
Table G. 3: Summary results using the African Standard 120
Table G. 4: Original data from the 2001 Botswana Census 120
Table G. 5: Summary results using the Booth Standard 121
Table G. 6: Summary results using the African Standard 121

ACKNOWLEDGEMENTS

The author would like to express his thanks in acknowledgement of the contributions made by several people to this thesis.

First, thanks must go to Associate Professor Tom Moultrie, for his invaluable contribution as supervisor. He was a sounding board for ideas and logic, as well as a source of knowledge. The hours of reasoning, explaining and checking this work are much appreciated. Second, my gratitude goes to the rest of the staff at the Centre for Actuarial Research (CARe) at the University of Cape Town for their assistance throughout this process.

On a more personal note, thanks to my parents - Nico and Yolande - and brother, Franscois, for the interest taken and support offered over the years. Also, my deepest love and appreciation goes to my wife, Suzette. She provided motivation and strength, listened while I sifted through ideas and patiently waited for the completion of this thesis.

Finally, praise and thanks to God for giving me the talents and opportunities to pursue my studies and blessing me with the support structure mentioned above.

1 INTRODUCTION

Demographic research focuses on the study of three variables - mortality, fertility and migration. As such, demography is essentially concerned with the effect of these variables on population dynamics and development.

For centuries the human fear of death focussed research on epidemiology and cause of death. This focus on death - added to a strong influence from the actuarial profession - centred demographic analysis primarily on modelling and understanding mortality until the 1950's. Such studies resulted in the production of model life tables such as the Coale-Demeny families (Coale and Demeny 1983).

However, during the 1960's and 1970's the high fertility rates and falling mortality rates resulted in high population growth rates (Bongaarts 1994; Demeny 2003). This lead to neo-Malthusian concerns of poverty resulting from over-population, further motivating research into fertility - the second significant demographic determinant (Livi-Bacci 1968; Merrick 2002). We now know that the population and poverty fears were never realised (Lam 2005), but many attempts were made to model fertility during this period.

Many of these models are based on standards such as the one developed by Booth (1979). In the period since the development of this standard, the levels and patterns of fertility have changed dramatically across the developing world. For example, in the late 1970s and early 1980s high fertility meant a total fertility rate (TFR) of 7 to 8 children per woman. However, by current standards a TFR of 5 to 6 constitutes high fertility which would have been a medium fertility population at the time Booth developed her standard.

Furthermore, other important factors have come to the fore: social and economic development in developing countries (particularly Africa) has gathered pace (Cohen 1993, 1998; Garenne and Joseph 2002) and HIV/AIDS has emerged as a global threat severely affecting both mortality and fertility (Lewis, Ronsman, Ezeh and Gregson 2004; Zaba and Gregson 1998).

These changes in the demographic landscape raise questions about the validity of the assumptions underlying the standard and the accuracy of methods based on the standard if these assumptions are violated. Specifically, investigation is required into the systematic underestimation of 45-49 fertility observed when using the Coale-Trussell schedules and the Booth standard to model African populations.

As a result, this dissertation focuses on maintaining and enhancing the relevance and utility of the relational Gompertz model. It re-evaluates the Booth standard and, in particular, considers the validity of the standard in different settings, the methodology used by Booth, the assumptions made and the relevance of the underlying data from the Coale-Trussell model.

To this end, a series of programs were developed allowing the automated production of fertility schedules using the Coale-Trussell fertility model. These programs are then evaluated to determine their usefulness in deriving new standards for low and medium fertility populations, particularly, in African societies.

Booth envisioned that each country or region may someday have an "internal" fertility standard or a standard appropriate to the area (Booth 1984, p. 496). This has not happened. This thesis renews research in this area as a step towards this vision. It specifically highlights the differences between African fertility and the current standard and emphasizes the need for a new standard to capture the shape, distribution and characteristics of fertility in sub-Saharan populations.

In order to achieve these aims the dissertation first reviews the available literature on data quality, fertility estimation techniques, the Coale-Trussell model and alternative fertility functions (Chapter 2). Chapter 3 discusses the use, limitations and restrictions of the Booth Standard within the framework of the relational Gompertz model and identifies an African fertility pattern. Attempts at addressing these problems and alternatives to the Booth standard are considered in Chapter 4 and an African fertility standard is developed. Chapter 5 draws conclusions from the preceding chapters and suggests possible areas of future research.

2 FERTILITY MEASURES IN DEVELOPING COUNTRIES

2.1 Direct Methods and Data Quality

Demographic work is largely focussed on the analysis, understanding and interpretation of quantitative data, although important qualitative information is collected. Since quantitative analysis lies at the core of demography, the first question that arises in all demographic research pertains to the quality of the data used (Brass 1996; United Nations 1983). The issue of data quality is so important that all substantial research work should assess and discuss the integrity of the data used in the analysis. According to Cleland (1996), the initial focus in developing countries was on improving data collection methods and not data analysis. The reason for this emphasis on data quality is not only to limit (and hopefully eliminate) errors and bias in the results, but good data also eases calculation of estimates.

In the case of perfect (or near-perfect) data the use of direct methods of analysis enable easy calculation and yield accurate results (Shryock and Siegel 1976). For example, directly calculated fertility rates at age a, are obtained by dividing the number of births to women aged a by the person-years lived by women age a in a given period.

Ideally data for measuring fertility would come from vital registration systems, but such registers are typically incomplete in developing countries and especially in sub-Saharan Africa. Consequently, data are usually obtained either from censuses or surveys (e.g. the Demographic and Health Surveys and the World Fertility Surveys) in the form of women's responses to questions on births in the last year and children ever born. Such data are also subject to many possible problems that may introduce bias into the fertility estimates. Generally, these errors will either affect lifetime fertility (parity) estimates or current fertility estimates.

2.1.1 Errors in lifetime fertility

Errors in lifetime fertility are characterised by increasing underreporting of parity as age increases. This is seen in the frequently severe underreporting - observed by many authors - for the 45-49 cohort and frequently, also, the 40-44 cohort (Brass 1996; Cohen 1993; Potter 1977; Sloggett 1994; Zaba 1981). Some authors argue that older women forget how many children they have had and that this causes the observed underreporting of parity (Potter 1977; Cohen 1993). Others, for example Zaba (1981) and Brass (1996), suggest the alternative - and possibly more plausible - explanation that the underreporting may result from age exaggeration among women in combination with other factors.

The omission of dead or absent children will reduce the reported lifetime fertility whereas the inclusion of adopted or step children and stillbirths will increase parity estimates (Brass 1996; Zaba 1981; Moultrie and Timæus 2002). In addition, the impact of other problems may be unknown save to say that there will be an effect. For example, the effect of the omission of the fertility of dead women from a survey will depend on whether the deceased had a higher or lower parity than the population surveyed. Similarly, the exclusion of emigrant women and the inclusion of immigrant women will alter parity, but the size and direction of the impact will depend on the relative fertility of migrant and non-migrant women. Excess mortality among women of higher parity will result in reported lifetime fertility decreasing with age. By contrast, excess mortality (and emigration) of lower parity women will increase reported lifetime fertility.

The last problem affecting parity is that census enumerators often mistakenly record parity zero as not stated, especially in the youngest age groups (El-Badry 1961; Sloggett 1994). There are primarily three methods of addressing this problem. The first is to exclude the not stated group from the analysis. However, if the group is large relative to the total number of women interviewed then relevant information may have been ignored and sample sizes could become an issue. Further, ignoring this group is equivalent to assuming that the women in the not stated category experience the same mean parity distribution as the remainder of the population. The second method of resolving the problem, proportional reallocation of the not stated group, has an essentially similar outcome. In most cases this assumption is probably not true and valuable information (of interest to demographers and policy-makers alike) may be hidden in the social forces guiding response.

El-Badry (1961) proposes a third method of resolving the problem which is useful when dealing with a large number of women with parity not stated. He argues that there is a linear relationship between the women who are true not stated cases and those with parity zero. Using this linear relationship the not stated group can be more suitably allocated and, in so doing, additional information is gained (El-Badry 1961).

The net-result of parity-affecting errors is that the shape of lifetime fertility is distorted since older women underreport lifetime fertility, whereas younger women are believed to accurately report parity.

2.1.2 Errors in current fertility

There are also numerous problems impacting the data on current fertility. First, younger children - those under 5 years old - are frequently and disproportionately omitted from censuses, surveys and vital registration (Cohen 1993; Heaton and Stanecki 2003; Potter 1977). The omission of these children will result in an underestimation of current fertility distorting both fertility estimates and trends. It also affects other demographic measures such as age structures, sex ratios and dependency ratios. Second, and similarly, the omission of twins and second births during the last year will understate the current fertility.

A third concern is the interpretation of the question on births during the last 12 months. This may be interpreted as births during "this year" which will result in underestimation since the year of census has not been completed yet. An alternative interpretation is births during the "last year". This will result in an underestimation if fertility is increasing and an overestimation if fertility is decreasing.

The errors altering current fertility are believed to result in net-underreporting that is constant across all ages, affecting only the level of fertility estimates but not the shape.

2.2 Indirect Techniques

The problems discussed above illustrate that, in almost every instance, the data are not perfect and the direct analysis techniques usually cannot be correctly applied (Brass 1996; Cleland 1996; Potter 1977). Fortunately, many indirect techniques of estimation have been developed to analyse imperfect data and, although these methods do not explicitly address the data problems, they provide methods of estimation using the available information (Brass 1996). However, it must be stressed that the existence of better analytical techniques applied to poor data can never replace good quality data. The problems in the data remain irrespective of the method used and these techniques merely attempt to utilise the available information to obtain fertility estimates in the absence of better quality data.

These indirect techniques of fertility analysis have a long history. Brass developed the P/F ratio method to analyse fertility using data on children ever born and births in the last year (Brass 1968). The method compares lifetime fertility with information on current fertility in order to evaluate the internal consistency of the data based on the assumption that fertility is not changing (Brass 1968; United Nations 1983).

As discussed, the errors affecting parity alter the shape of the parity distribution, but the level is assumed to be correctly reported for younger women. Given the assumption of constant fertility, the level implied by the parity of younger women can be used to ascertain an indication of true lifetime fertility in the population. By contrast, the errors impacting current fertility change the level, but not the shape of the fertility distribution. Hence, by using the level implied by the parity data and the pattern from the current fertility data a more accurate estimate is derived than by using only a single data set.

Following on the development of the Brass P/F ratio Hobcraft, Goldman and Chidambaram (1982) applied the method to birth histories demonstrating its versatility. In order to maintain the relevance of the technique, in times of falling fertility, it was restated by Feeney to relax the underlying assumption of constant fertility required in the original formulation (Feeney 1998). However, Moultrie and Dorrington (2008) show that the Feeney restatement does not work well when fertility is falling. Instead, they suggest using the adjustment recommended in Manual X (United Nations 1983) whereby reported fertility is scaled using the P/F ratio at 20-24.

A more recent successor to the P/F method is the relational Gompertz model proposed by Brass and further developed by Booth (1979) and Zaba (1981). As one of the most powerful techniques available to estimate fertility from limited or defective data, the relational Gompertz model continues the long history of analysis techniques based upon the ratio of lifetime fertility (or parity), P, to cumulative fertility, F, and benefits from the use of a schedule capturing a standard fertility pattern (Booth 1984; Brass and Airey 1988; Zaba 1981). It is this standard that is the principle focus of this dissertation.

2.3 The Use of Standards

The use of a standard population is a common theme in demographic research since it is often necessary to compare the experiences of different populations or population groups. Two methods of using a standard stand out.

The first, is to use a standard to eliminate the effects of confounding factors - through standardisation - as explained by Shryock and Siegel (1976). According to Shryock and Siegel, a suitable standard must reflect a large, clearly defined population. This population can be real or theoretical provided it is constant and known (Shryock and Siegel 1976). An example of a physical standard population is the fertility experience of women in the British National Healthcare System while that of a hypothetical (constructed) standard is the Coale-Demeny family of schedules (often used in mortality analysis).

The second, an alternative to standardisation (and the one of greater importance in this research), is the idea of comparing a population with a standard. Using this relational logic, one can express the observed population in terms of its correspondence to (or deviation from) the standard. This idea, of relating populations to a standard, lies at the core of many indirect methods of estimation. For example, the Coale-McNeil nuptiality function (Coale and McNeil 1972) is widely used to determine first marriage frequencies and the cumulative proportion married by age.

The Coale-Trussell model is an example of both a procedure to develop a standard as well as a method using a standard. The Coale-Trussell model uses the Coale-McNeil nuptiality function to derive a series of standard marital fertility distributions (Coale and Trussell 1974). The Coale-Trussell schedules were subsequently used to derive a standard for use with the relational Gompertz model (Booth 1979; Brass and Airey 1988).

In the case of the Coale-Trussell model the model relates age-specific *(marital)* fertility, f(a), to the cumulative proportion ever married, *G(a),* derived from the standard Coale-McNeil function using a multiplicative equation. By contrast, in other settings the idea of this relational construct is to express the population rates as a linear transform of the standard rates of the form

$$F(x) = \alpha + \beta * F_S(x)$$ **Equation 2.1**

where F,(x) and F(x) are the standard and (modelled) population rates respectively, a is the intercept parameter and p the slope parameter. The principle behind this formulation is that - if such a transform exists - the selection of different intercept and slope parameters can express the entire range of possible populations.

2.4 The Coale-Trussell Model

The Coale-Trussell schedules are frequently used as standards in fertility analysis, for example to parameterise coefficients in the P/F method. They were developed - using the model proposed by Coale and Trussell - to capture human fertility patterns (Coale and Trussell 1974). The schedules describe various fertility patterns but do not attempt to represent fertility levels, since these can be found by multiplying the standard rates at each age by the actual population TFR. Coale and Trussell reason that age-specific marital fertility, f(a), can be modelled as the product of the cumulative proportion ever married, *G(a),* and the probability of married women having a live birth, *r(a),* both at age *a* (Equation 2.2).

$$f(a) = G(a) * r(a)$$ **Equation 2.2**

The relationship in Equation 2.2 assumes homogeneity in marital fertility rates and the proportion ever married at age *a.* It is also implicitly assumed that marriage rates are high and out-of-marriage (illegitimate) fertility is trivial relative to total fertility.

According to Coale and Trussell, the formulation in Equation 2.2 allows the model to be parameterised by three variables. Two parameters define the model schedule of marriage, *G(a),* and the third parameter specifies the model marital fertility schedule, *r(a).* They deem it reasonable to define *G(a)* by only two parameters on the basis that first-marriage frequencies tend to have the same shape, but vary in respect of the age at which first marriage begins, a_0, and the pace at which first marriage occurs as indicated by the scale factor, ***k***. Although no explicit expression of *G(a)* exists, Coale and Trussell state that it can be found by integrating McNeil's equation, ,g(a), which is reproduced in Equation 2.3 and parameterised by *a,* and ***k*** (Coale and Trussell 1974; Coale and McNeil 1972)1.

$$g(a) = \frac{0.19465}{k} e^{[-0.174W - e^{(-0.2881W)}]}$$ **Equation 2.3**

where

$$W = \frac{a - a_0 - 6.06k}{k}$$ **Equation 2.4**

[1] By using the standard McNeil nuptiality equation, g(a), to calculate G(a) transforms equation 2 into the multiplicative relation mentioned in section 2.3.

Coale and Trussell (1974) explain that the scaling factor, k, models the deviation in the pace of first marriage from the pace of the 19 th century Swedish population used as basis. The model is constructed so that k = 1 signifies the same pace of marital progression - in the modelled population - as the Swedish base-population. Similarly, k = 0.5 indicates twice the pace observed in the standard population. Furthermore, Coale and Trussell note that the pace of marriage in the Swedish population determines that 50 per cent of the population ever-married will do so within 10 years of the starting age, a_o. As a result, they assert that k = 0.5 implies that "one-half the cohort has experienced marriage five years after a," (Coale and Trussell 1974, p. 187).

The second function, $r(a)$, is the schedule of marital fertility. Coale and Trussell argue that marital fertility can be parameterised by one parameter, m, where in is the extent of departure from natural fertility and is treated as constant for all ages in a modelled population. In the absence of voluntary birth control $r(a)$ follows natural *(marital)* fertility. However, in populations practising deliberate birth limitation fertility deviates from natural fertility according to a typical pattern as age increases (Coale and Trussell 1974). This pattern of deviation is captured by the vector, $v(a)$, in Equation 2.5. Coale and Trussell give the ratio at each age of marital fertility, $r(a)$, to natural fertility, $n(a)$, as:

$$\frac{r(a)}{n(a)} = M * e^{(m * v(a))}$$ **Equation 2.5**

The scale factor, M, is constant at a level such that *1),(4* = 0.0 for age 20-24 for each schedule i. This means that M equals the ratio $r(20\text{-}24)/n(20\text{-}24)$. Since m is constant for all ages, the selection of M at age a, equates $v(a)$ - the propensity to reduce fertility at age a - to zero for that age.

Given that M is a constant scale factor applicable at all the relevant ages it is not important when determining the fertility schedules (Coale and Trussell 1974). The scale variable can be removed from each age - as a common factor - resulting in a pattern of fertility independent of level.

The two constant vectors, $n(a)$ and $v(a)$, are derived by Coale and Trussell from empirical data [2]. The vector $n(a)$ is obtained - for the age groups 20-24 to 45-49 - from ten of the schedules identified as natural by Henry (1961). Table 2.1 reproduces the data from Henry (1961 p. 148).

[2] The values of $n(a)$ and *U(C1)* are presented in Table A 1 in Appendix A.

Table 2.1: Age-Specific Legitimate Fertility Rates per 1000

Population	20-24	25-29	30-34	35-39	40-44	45-49	Included in *n(a)*
Hutterites marriages 1921-1930	550	502	447	406	222	61	
Canada	509	495	484	410	231	30	
Hutterites marriages before 1921	475	451	425	374	205	29	
Bourgeoisie of Geneva wives of men born 1600-1649	525	485	429	287	141	16	
Europeans of Tunis	468	430	402	324	190	13	
Sotteville-les-Rouen	480	450	410	315	125	10	
Crulai	440	420	375	280	140	10	
Norway	396	380	341	289	180	41	
Iran	395	370	325	255	130	20	
Bourgeoisie of Geneva wives of men born before 1600	389	362	327	275	123	19	
Taiwan	365	334	306	263	114	8	
India	323	288	282	212	100	33	
Guinea	357	320	273	183	74	32	

Source: Henry (1961), 1).148

Coale and Trussell (1974) do not state which 10 schedules were used to determine ***n(a).*** However, investigation of these schedules indicates that averaging the rows marked by "t" very closely approximates the results by Coale and Trussell. Table 2.2 below compares the rates calculated from the schedules identified in Table 2.1 with the n(a)-values reported by Coale and Trussell (1974, p. 188). The differences are trivial and no other combination of 10 schedules yields more accurate results.

Table 2.2: Comparison of Coale-Trussell *n(a)* with Calculated Average

Age Group	20-24	25-29	30-34	35-39	40-44	45-49
Average of rows marked	0.460	0.431	0.395	0.322	0.167	0.024
n(a) from Coale-Trussell (1974)	0.460	0.431	0.396	0.321	0.167	0.024

To determine *v(a)* Coale and Trussell (1974) state that the marital fertility schedules from the 1965 United Nations Demographic Yearbook (United Nations 1966) were assessed to determine if they were subject to errors such as age misreporting. The vector v(a) - also for the age groups 20-24 to 45-49 - was then calculated as an average of the 43 schedules believed to be free of such errors. For both vectors the values derived were then extended to the two lower age groups - 10-14 and 15-19. According to Coale and Trussell (1974, p. 190) the downward extension of ***n(a)*** is based on biomedical information. They contend that the dominant role of *G(a)* means that the specific *n(a)*-values selected for ages 20 years is of little significance. Also, they state that non-zero values of *v(a)* start at age 20 in order to avoid sudden changes around age 25 (Coale and Trussell 1974, p. 190). That is, the propensity to limit or control fertility effectively begins at age 20.

Finally, the relational formulation (discussed in section 2.3) produces a family of standard schedules by inserting different values of the three parameters into the FORTRAN program provided by Coale and Trussell (1974). The benefit of this family of schedules is that it models a wide range of fertility experience using demographic reasoning and features. By modelling such a wide range of experience the schedules can be used, as Coale and Trussell suggested, to approximate single-year fertility rates when only five-year rates are available. However, the Coale-Trussell schedules can also be put to other uses such as basis for comparison or as a standard to be used with indirect techniques like the relational Gompertz model. This is discussed in greater detail in the next chapter.

2.5 Restatements of the Coale-Trussell Model

The model developed by Coale and Trussell (1974) depends on the accuracy of the two vectors *n(a)* and *v(a)*. However, Henry (1961) and Coale and Trussell (1974) are both subject to mounting criticism. Blake (1985) and Wilson, Oeppen and Pardoe (1988) question the original formulation of natural fertility and the manner in which the Coale-Trussell model is used. Faced by these criticisms Xie (1990) and Xie and Pimentel (1992) attempt to maintain the relevance of the Coale-Trussell model by reformulating the method.

Both investigations focus on the manipulation of the r(a)-ffinction to bring about changes to the model. More specifically, since Al and m are assumed constant in the Coale-Trussell model (for a particular schedule), a change in *r(a)* can only be affected by adjusting the vectors *n(a)* and *v(a)*.

2.5.1 Coale-Trussell using Xie n(a)

Xie (1990) addresses the concept of natural marital fertility, *n(a),* introduced by Henry (1961). Henry argued that although the levels can differ the age pattern of natural fertility should be fairly constant. However, Xie notes that this hypothesis, although reasonable, had not been tested before. Coale and Trussell (1974) used the Henry data to determine the standard natural marital fertility pattern, *n(a),* as discussed in section 2.4. Their development of a natural fertility standard exacerbated the problem of the unverified theory since, as Xie states, *n(a)* has subsequently "been treated as exactly known" (Xie 1990, p. 656).

To test the data identified by Henry, Xie uses log-linear methods and maximum-likelihood estimation. Three models were developed: 1) A Homogeneity model, 2) an Independence model and 3) a Fertility Control model (Xie 1990).

The homogeneity model argues that the levels and patterns are the same for all the populations under investigation and that differences can be ascribed to random error. However, Xie echoes Henry in rejecting this hypothesis (Xie 1990).

The second model assumes that a population-specific (level) factor exists that is independent of age. This independence model is contrasted with the fertility control model which postulates that the natural fertility populations from Henry (1961) are ordinary populations. The comparison of these models by Xie shows that the independence version provides a good fit to the data and Xie notes that the additional benefit gained by the fertility control model is negligible. As such, the independence version is accepted as the final form of the log-linear model (Xie 1990).

Xie, like Henry, concludes from the independence model that a common age pattern exists between the identified natural fertility populations but that these populations have different levels (Xie 1990, p. 662). He further argues that the maximum likelihood method of parameter estimation gives weighting to larger samples making estimates derived superior to those derived by the simple average used by Coale and Trussell (1974). The new *n(a)* values derived for the age groups 20-24, 25-29, ..., 40-44 and 45-49 are given in Table 2.3.

Table 2.3: Standardized n(a) values for the Xie Independence and Coale-Trussell models

Age Group	20-24	25-29	30-34	35-39	40-44	45-49
Xie n(a)	0.460	0.436	0.392	0.333	0.199	0.043
Coale-Trussell n(a)	0.460	0.431	0.395	0.322	0.167	0.024

Obtained from Table 2 by Xie (1990, p. 660)

It is clear from Table 2.3 that the differences between the Coale-Trussell n(a)-levels and the value derived by Xie (1990) are only large at the oldest ages. The effect of this increase in n(45-49) is expected to result in an increase in f(45-49).

2.5.2 Coale-Trussell with alternative v(a)

Xie and Pimentel (1992) reformulate the Coale-Trussell model as a statistical method and deriving a new series of v(a)-values (Xie and Pimentel 1992, p. 982). They use data from the World Fertility Surveys (WFS) to obtain the revised *v(a)*. This has two benefits: a) the WFS includes more recent data - 1974 to 1982 - than used by Coale and Trussell and b) are of higher quality than preceding surveys (Xie and Pimentel 1992)

To establish a new v(a)-series, Xie and Pimentel compare the results from a natural fertility schedule, a log-linear and a log-multiplicative model. They applied these models to the natural fertility standard developed by Coale and Trussell (1974) and the *n(a)* determined by Xie (1990). The results were then compared using the Bayesian criterion, Log-likelihood statistic and the Pearson Chi-square statistic. Xie and Pimentel (1992) establish that the log-multiplicative model using the Xie n(a)-values yields better estimates of *1)(a)* (given in Table 2.4) than were originally derived by Coale and Trussell.

Table 2.4: Standardized v(a) values for the Coale-Trussell and Xie-Pimentel models

Age Group	20-24	25-29	30-34	35-39	40-44	45-49
Xie-Pimentel v(a)	0	0.329	0.713	1.194	1.671	1.082
Coale-Trussell v(a)	0	0.279	0.677	1.042	1.414	1.671

Notably, the v(a) are not monotonic over the entire age range as originally found by Coale and Trussell (1974). Instead the *v(a)* increases over the age range 25-44, but decreases again for 45-49. Xie and Pimentel attribute this to the fact that more recent data are used to determine the new *v(a)*-estimates (Xie and Pimentel 1992, p.982). Like the *n(a)* derived by Xie (1990), the lower fertility control at the oldest ages is expected to increase 45-49 fertility.

2.6 Alternatives to the Coale-Trussell Model

When developing or testing a standard it is important to also consider other options. For example, Hoem et al. (1981) analyse Danish fertility data using a number of different fertility distributions. They use the Coale-Trussell model, cubic splines, the Gamma density function, the Hadwiger function, and the Brass polynomial and conclude that the cubic spline provides the best fit. The Hadwiger function, Gamma density and Coale-Trussell model are determined joint second best and the Brass polynomial is deemed less accurate (Hoem, et al. 1981).

Gage (2001) also considers cubic splines, the Gamma density, the Brass polynomial and the Hadwiger function for the analysis of seven mammalian populations. However, he dismisses cubic splines for mammalian population since it "requires good underlying empirical data" (Gage 2001, p.490). By the same argument cubic splines are dismissed for the sub-Saharan populations due to the enduring problems around data quality.

Among the seven mammalian populations investigated by Gage are two human populations - Sweden 1967 and Costa Rica 1966. Using the remaining three methods Gage concludes that all three yield good fits, but that the Gamma function appears to fit best based on absolute mean square error (Gage 2001, p. 492). However, Gage proceeds to conduct other tests of fit since mean square error does not allow for the relative differences in complexity arising from the different parameterisations of the methods. From the additional statistical tests Gage concludes that, for six of the seven populations, the Brass polynomial cannot be rejected (at a 5 per cent level) in favour of the more complex Gamma or Hadwiger functions (Gage 2001). He further states that if the Bonferroni adjustment is applied then the simpler Brass polynomial cannot be rejected even for the seventh population. However, Hoem et al. (1981) successfully used the Hadwiger function to model early Danish fertility and Gage (2001) found a good fit for the 1967 Swedish data.

Another alternative is one of the two mathematical formulae modelling age-specific fertility rates that Pevistera and Kostaki (2007) proposed. They find that the exponential formulae proposed perform better than the Hadwiger function, Gamma density model, Beta function and cubic splines on data from the United States and European countries like the United Kingdom, Denmark, Spain, Italy, Norway and Sweden. However, these countries all have fertility rates far below those currently observed in sub-Saharan Africa and different patterns of fertility to those under consideration throughout this thesis.

As a result, only the Brass polynomial and Hadwiger functions are investigated in this research for its usefulness in measuring and capturing African fertility patterns.

2.6.1 Brass polynomial

Brass (1975) suggests using the polynomial given in Equation 2.6 to model fertility.

$$F(z) = c \;_{s} (\mathbf{x - s})\, \boldsymbol{s} + 33 - x)^{2}\, dx \qquad \textbf{Equation 2.6}$$

Here c is a constant relating to the total fertility rate, s is the starting age and is the age to which fertility is cumulated. Brass explains that the constant 33 arises from the length of the fertility period which tends to range between 30 and 36 years with an average of about 33 years (Brass 1975).

Gage (2001) generalises this polynomial by replacing the constant 33 with a parameter, w, representing the length of the fertility period for the population or sub-population under observation (Equation 2.7).

$$F(z) = c\int_{s}^{z} (x - s)(s + w - x)^2 \, dx$$ **Equation 2.7**

However, integrals are cumbersome to work with and Appendix E uses the properties of integrals and fertility patterns to establish a convenient expression to calculate F(z) directly (Equation 2.8).

$$1 - F(z) = 4p^3 - 3p^4$$ **Equation 2.8**

where p is the portion of the fertility period remaining at age z as given by $p = \frac{t}{w} = \frac{s+w-z}{w}$.

2.6.2 Hadwiger function

Hoem et al. (1981) and Gage (2001) indicate that the second reasonable alternative to the Coale-Trussell model is the Hadwiger function (Equation 2.9).

$$m_x = \left(\frac{ab}{c}\right) * \left(\frac{c}{x}\right)^{\frac{3}{2}} * e^{\left[-b^2 * \left(\frac{c}{x} + \frac{x}{c} - 2\right)\right]}$$ **Equation 2.9**

The Hadwiger function has four parameters - *a, b, c* and x. The parameter x is age and allowed to vary between starting age *s* and maximum age *u*. The parameter, *a,* is a measure of total fertility. According to Hoem et al. (1981) and Gage (2001) the other two parameters, *b* and *c,* have no clear demographic interpretation. In addition, unlike the Brass polynomial discussed in section 2.6.1, the more complex Hadwiger function does not have a convenient simplification to ease calculation.

However, despite these drawbacks the Hadwiger function is investigated since both Hoem et al. and Gage have successfully used the Hadwiger function to model human populations.

3 THE BOOTH STANDARD IN THE RELATIONAL GOMPERTZ MODEL

As discussed in Chapter 2, the indirect techniques of fertility analysis are of great importance since it gives access to important information even when faced with limited, incomplete or poor quality data.

3.1 The Brass Relational Gompertz Model

While working on indirect techniques of fertility analysis, Brass proposed the use of a relational model to represent fertility (Brass 1974). This challenge was subsequently taken up by Booth (1979) - a PhD supervised by Brass and further elaborated by Brass and Airey (1988). Brass hoped that the method could be applied to poor quality data to yield reasonable fertility estimates. He found that the ratio of cumulative fertility at age x to the total fertility rate can be quite well represented by the Gompertz distribution as in Equations 3.1.

$$\frac{F(x)}{T} = A^{B^{x - x_0}}$$ **Equation 3.1**

F(x) is the cumulative fertility up to age *T* is the total fertility rate, x,, is the origin of the age range and two constants, *A* and *B,* describe the pattern of fertility. A fertility function, Y(x), is generated by performing a Gompertz transformation on Equation 3.1 and identifying the resulting equation as a straight line (Equations 3.2, 3.3 and 3.4).

$$Y(x) = -\ln\left(-\ln\left(\frac{F(x)}{T}\right)\right)$$ **Equation 3.2**

$$= -\ln[-\ln(A)] + [-\ln(B)][x - x_0]$$ **Equation 3.3**

$$= a + b(x - x_0)$$ **Equation 3.4**

where a = - In (-1n (A)) and b = -In B and Y(x) is a linear function of age only.

However, Brass further refines the method. He notes that Y(x) fits observed fertility well over the middle ages (25-39) but not for the tails of the distribution. For the younger age groups Y(x) is too high and for the oldest ages Y(x) is too low (Figure 3.1).

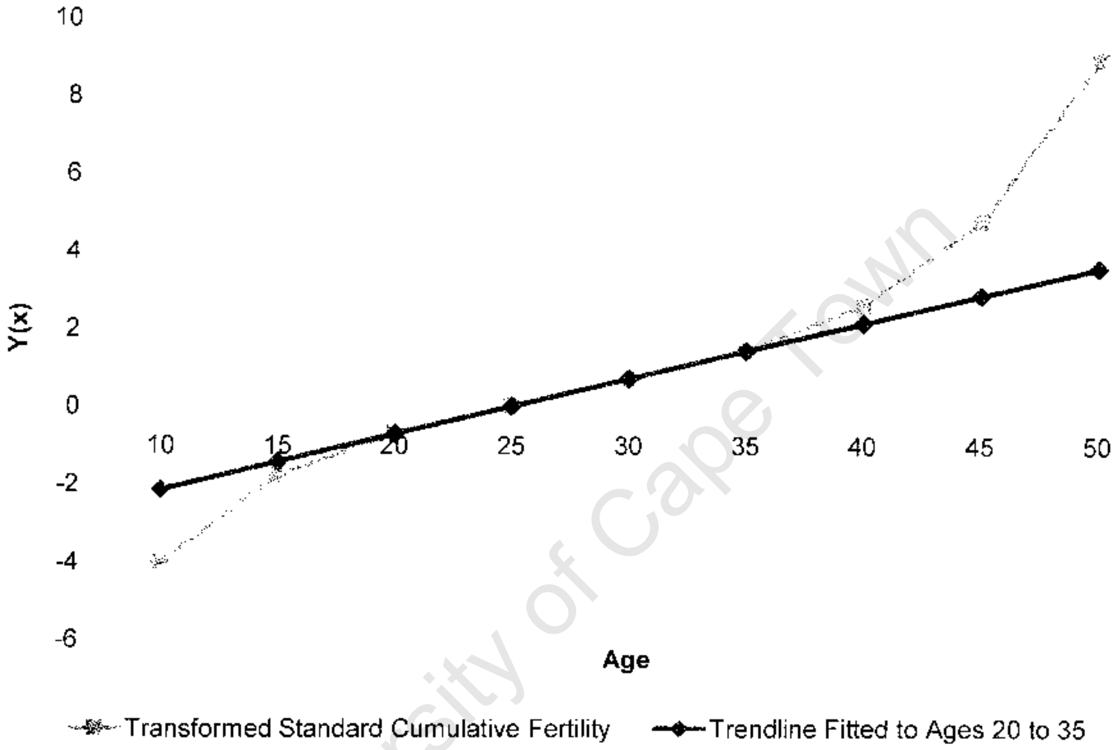

Figure 3.1: Divergence in the tails of transformed cumulative fertility, Y(x), from a straight line fitted to the middle ages

Brass argues that a better fit is possible by transforming the age scale (Brass 1974, Brass and Airey 1988). That is, by replacing the natural age scale with a stretched age scale, Z(x), will result in a better tit, since Y(x) will lie on a straight line as shown by Equation 3.5 (Booth 1979).

$$Y\ x = a + bZ\ (x)$$ **Equation 3.5**

Figure 3.2 illustrates this graphically. It shows the transformed cumulative fertility, Y(x), plotted against age with the stretched age scale, Z(x), on the right hand y-axis.

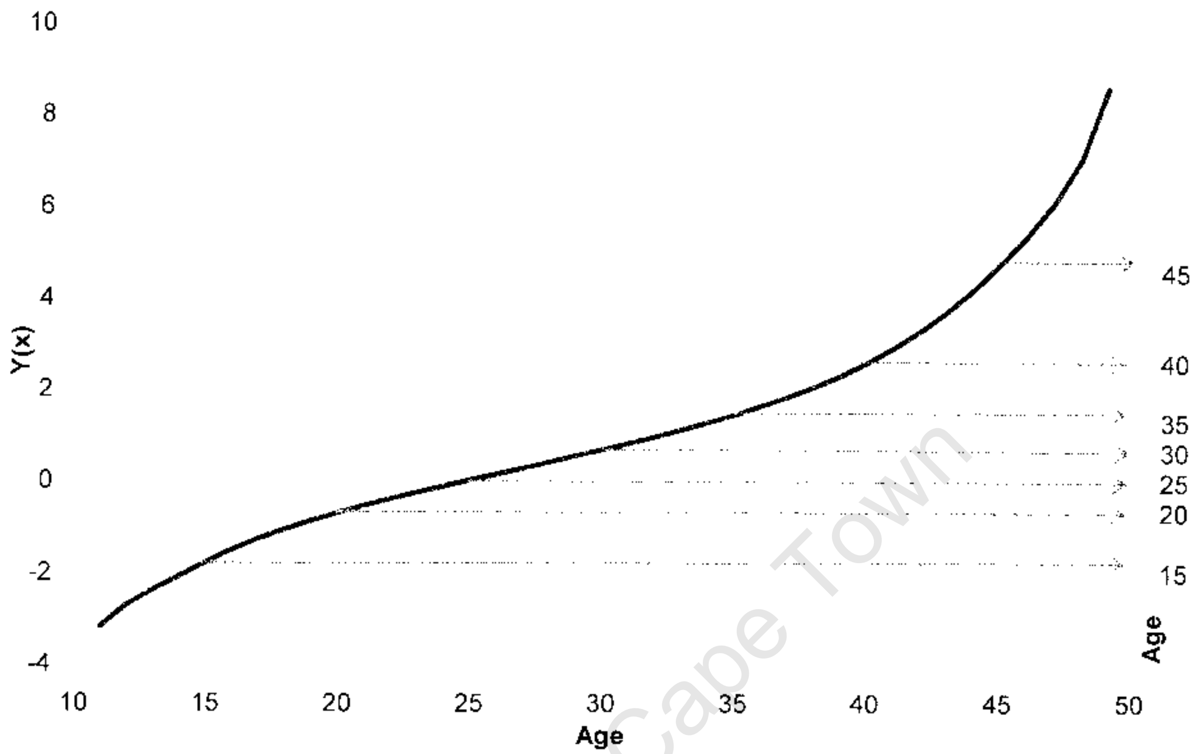

Figure 3.2: Transformed cumulative fertility against age and the stretched age scale

According to Booth, the quantity by which to expand the age scale is equivalent to the distances between the ages on the right hand axis (Booth 1979, p. 30). However, each of the ages of the expanded age scale, Z(x), intersects the original age scale at the point Y(x). That is, Z(x) = Y(x) at the point of intersection for each age, Therefore, since Y(x) is the transformed fertility pattern, the expanded age scale is itself also equal to the transformed fertility pattern.

Booth (1979) states that if F(x) is the standard pattern of cumulative fertility then this transformed age scale can be represented by the Gompertz transform of cumulative fertility, Y,(x). Then, in any population, the transformed observed fertility rates, Y(x), are linearly related to the transformed standard rates, Y,(N), through Equations 3.6 and 3.7.

$$Y_s(x) = -\ln\{-\ln[F_s(x)]\}$$ **Equation 3.6**

$$Y(x) = \alpha + \beta Y_s(x)$$ **Equation 3.7**

Combining Equations 3.2 and 3.7, simplifying and reorganising to make F(x) the subject, results in Equation 3.8 below.

$$F(x) = Te^{\{-e^{[-(\alpha + \beta Y_s(x))]}\}}$$ **Equation 3.8**

In these equations, a and p are the location and spread parameters, respectively. Changes in these two parameters reflect the timing and distribution of fertility for the population. A change in a varies the timing of childbirth in a population (Figure 3.3).

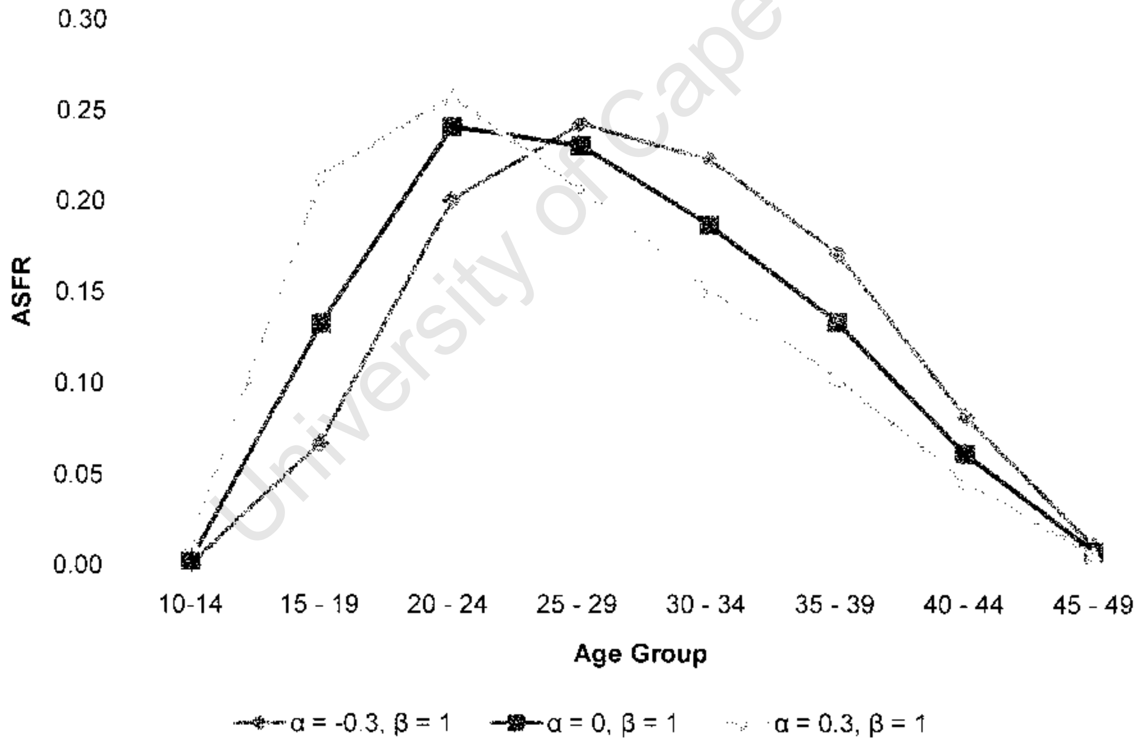

Figure 3.3: Effects of a change in α holding β = 1

For example, a negative a is equivalent to delaying births and a positive a brings childbearing forward. However, the shape of the fertility distribution is also modified, since an early start to fertility is associated with a fast rise in ASFR at younger ages and delayed fertility steepens old-age rates.

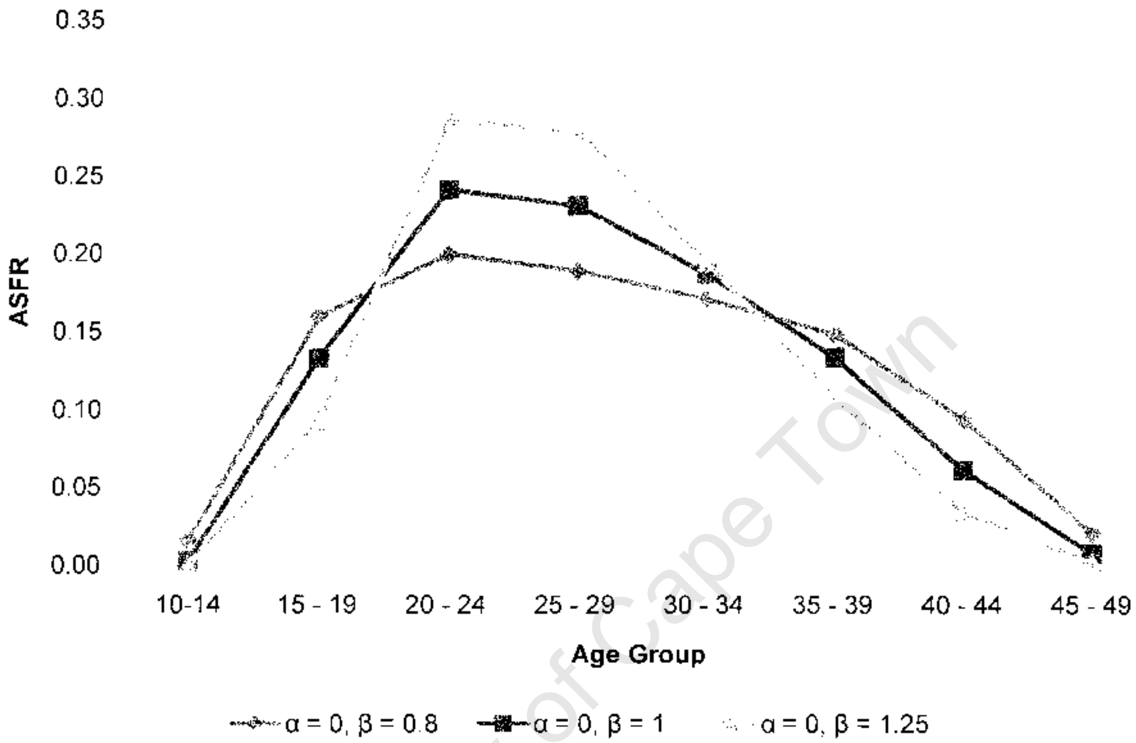

Figure 3.4: Effects of a change in β holding α constant at 0

By contrast, Figure 3.4 shows that a rise in β narrows the spread of the fertility distribution and a decrease flattens it. Clearly, this also has the effect of changing the shape of the fertility distribution.

3.2 Zaba's Restatement of the Relational Gompertz Model

A drawback of this formulation is that the original model requires knowledge of the total fertility rate which will, in most applications of the method, not be known. This drawback resulted in Zaba (1981) formulating the relational Gompertz model such that the TFR need not be known. Zaba suggests two modifications to the relational Gompertz model. The first restated model is expressed in terms of the ratio of cumulative fertility at ages and $1v+5$ (Equation 3.9 below).

$$Y(x) = -\ln\left[-\ln\left(\frac{F(x)}{F(x+5)}\right)\right]$$ **Equation 3.9**

Equation 3.9 does not depend on a known TFR. Brass and Airey (1988) note that the Gompertz transformation relating F(x)/F(x+5) to the standard, Y(x), is not exactly linear, but generally offers a good approximation. Substituting Equation 3.8 into 3.9 and rearranging the terms results in Equation 3.10 below.

$$Y(x) = \alpha - \ln\left\{e^{[-\beta Y_s(x)]} - e^{[-\beta Y_s(x+5)]}\right\}$$ **Equation 3.10**

Equation 3.11 gives $\Phi_x(\beta)$ which is defined by Zaba as the second term of Equation 3.10. A Taylor expansion of $\Phi_x(\beta)$ around $\beta = 1$ yields the approximation in Equation 3.12.

$$\Phi_x(\beta) = -\ln\left\{e^{[-\beta Y_s(x)]} - e^{[-\beta Y_s(x+5)]}\right\}$$ **Equation 3.11**

$$\Phi_x(\beta) = \Phi_x(1) + (\beta - 1)\Phi_x'(1) + \frac{(\beta-1)^2}{2}\Phi_x''(1) + \ldots$$ **Equation 3.12**

The primes, in the second and subsequent terms of the right hand side of the equation, signify differentiation with respect to β. Zaba (1981) evaluates the expressions for $\Phi_x(1)$, $\Phi_x'(1)$ and $\Phi_x''(1)$ and obtains the results tabulated in Tables 3.1 and 3.2.

It can be seen that $\Phi_x''(1)$ is fairly constant for the groups (20, 25); (25, 30) and (30, 35). Similarly, for (19½, 24½); (24½, 29½) and (29½, 34½) in Table 3.2, the $\Phi_x''(1)$ values are relatively consistent.

Table 3.1: Fertility cumulants - terms of Taylor expansion of $\Phi_x''(\beta)$ (without shift)

(x, x+5)	$\Phi_x(1)$	$\Phi_x'(1)$	$\Phi_x''(1)$
(15, 20)	-1.32718	-2.31376	0.91127
(20, 25)	-0.02139	-1.37530	0.95825
(25, 30)	0.73793	-0.67479	0.96295
(30, 35)	1.31432	0.03933	0.95096
(35, 40)	1.86070	0.94501	0.89717
(40, 45)	2.74551	2.34887	0.68203
(45, 50)	4.80970	4.80970	0.00006

Average of $\Phi_x''(1)$ 0.9574

Source: Zaba 1981

Table 3.2: Fertility cumulants - terms of Taylor expansion of $\Phi_x''(\beta)$ (with ½ year shift)

(x, x+5)	$\Phi_x(1)$	$\Phi_x'(1)$	$\Phi_x''(1)$
(14½, 19½)	-1.42600	-2.40198	0.90930
(19½, 24½)	-0.11373	-1.45013	0.95673
(24½, 29½)	0.67545	-0.74298	0.96338
(29½, 34½)	1.25957	-0.03818	0.95319
(34½, 39½)	1.80260	0.83562	0.90759
(39½, 44½)	2.61577	2.16491	0.71962
(44½, 49½)	4.50266	4.45641	0.18742

Average of $\Phi_x''(1)$ 0.9578

Source: Zaba 1981

Due to its evenness $\Phi_x''(1)$ is then replaced by constant, c. The values for the three age groups in each table are then averaged and a value of about 0.96 is calculated for $\Phi_x''(1)$. Equation 3.13 is identified through the substitution of terms and replacing the constant c gives rise to Equation 3.14 (Zaba 1981).

$$Y(x) - \Phi_x(1) + \Phi'_x(1) = \alpha + (\beta + 1)^2 \frac{c}{2} + \beta\Phi'_x(1)$$ **Equation 3.13**

Alternatively,

$$Y(x) - e(x) = \alpha + 0.48(\beta - 1)^2 + \beta g(x)$$ **Equation 3.14**

The expression in Equation 3.13 is a linear relationship between the left hand side and $\Phi_x'(1)$ with slope β and intercept $\alpha+(\beta-1)^2\frac{c}{2}$. Equation 3.14 is a further simplification using standard tabulated values of *e(x)* and *g(x)* (Tables A 2 and A 3 in Appendix A). However, Zaba warns that the parameter values must be within the range $-0.3 < \alpha < 0.3$ and $0.8 < \beta < 1.25$. If β lies outside these bounds then the third and subsequent differences in the Taylor expansion assumption used in Equations 3.12 to 3.14 become significant.

The second restated model that Zaba uses is concerned with correcting errors in mean parities. She defines *P(i)* as the parity in the i'th group where $i = 0, 1, 2, \ldots, 7$ and refers to the ages 10-14, 15-19, …, 45-49. In particular, errors reducing parity for women over 40 - for example omission of children, excess mortality among older women with higher parity and others are to be corrected. First, note that *P(i)* can be defined in a similar fashion as F(x) (Equation 3.15 below) and *T*, α and β as before.

$$P(i) = Te^{\left(-e^{[-(\alpha+\beta Y_s(i))]}\right)}$$ **Equation 3.15**

Second, define the equation for $Y_s(i)$.

$$Y_s(i) = -\ln\left[-\ln\left(P_s(i)\right)\right]$$ **Equation 3.16**

$P_s(i)$ is the standard mean parity for group *i*. Using the ratio of consecutive parities in a way similar to the cumulative fertilities above gives Equation 3.17 below.

$$Y(i) = -\ln\left[-\ln\left(\frac{P(i)}{P(i+1)}\right)\right]$$ **Equation 3.17**

By substituting Equation 3.15 into Equation 3.17 and simplifying, *Y(i)* can be expressed in the same fashion as Y(x) in Equation 3.12.

$$Y(i) = \alpha - \ln\left\{e^{[-\beta Y_i(i)]} - e^{[-\beta Y_i(i+1)]}\right\}$$ **Equation 3.18**

As before, $\Phi_i(\beta)$ follows as the substitute for the second term on the right-hand side of Equation 3.18.

$$\Phi_i(\beta) = -\ln\left\{e^{[-\beta Y_i(i)]} - e^{[-\beta Y_i(i+1)]}\right\}$$ **Equation 3.19**

Zaba uses the same logic as for cumulative fertility and applies a Taylor expansion around β=1 to $\Phi_i(\beta)$. Again, using substitution, reshuffling the terms and identifying the constant $c = \Phi_i''(\beta)$ establishes Equation 3.20.

$$Y(i) - \Phi_i(1) + \Phi_i'(1) = \alpha + (\beta - 1)^2 \frac{c}{2} + \beta\Phi_i'(1)$$ **Equation 3.20**

Or,

$$Y(i) - e(i) = \alpha + 0.48(\beta - 1)^2 + \beta g(i)$$ **Equation 3.21**

As for cumulative fertility above, Equation 3.21 follows directly for mean parities and the functions *e(i* and *g(i)* are tabulated in Table A 4 in Appendix A. Again, the caveat holds that the parameters a and p must fall within the specified ranges, -0.3 < a < 0.3 and 0.8 < p < 1.25, to prevent the third and subsequent differences becoming significant.

Applying this method and using the parity ratios that lie on a straight line, we can find an estimate of the unknown TFR. The strength of this method lies in the fact that it uses the best properties of both the fertility and parity data, thereby allowing the strengths of one data set to correct errors in the other. Further, this method provides more flexibility since the analyst can decide which data set to use thereby allowing the investigator to give preference to the data source believed to be superior in quality. This results in better estimates and additional flexibility.

3.3 Development of the Booth Standard

Both variants of the relational Gompertz model still require a standard fertility pattern when deriving estimates. Brass envisioned using the method in high fertility populations where data quality is poor (Brass 1974). This application of the relational Gompertz model to high fertility populations is later reaffirmed by Booth (1979; 1984) and Zaba (1981). In fact, Booth states that the method will be employed "to detect and correct the kinds of errors that are found in data from such *(high fertility)* populations" (Booth 1984, p. 496). Given this emphasis - on high fertility populations - Booth (1979) develops a standard pattern for use with the Brass variant of the relational Gompertz model and this standard is also the foundation for the work by Zaba (1981). This section describes in detail the methodology used to develop this standard - as well as the results obtained - so that it may be analysed, modified and improved on in Sections 3.4 to 3.7 and Chapter 4.

In order to develop the standard, Booth uses the Coale-Trussell model since it provides a method of easily obtaining a wide range of fertility patterns (Booth 1979; 1984). She sets out a number of criteria to distinguish between the high and the low fertility patterns. First, she allows the three parameter inputs required by the Coale-Trussell model to vary within specified ranges. The age at first marriage, a_o, is started at 10 years and increased in steps of 0.5 years to a maximum of 15 years. The parameter ranges for k and m - which determine the shape of the fertility distribution - were set to $0.1 < k < 1.3$ and $0 < P_i < 1$. However, Booth notes that m can be restricted further to m 0.6, since, according to Booth, the other combinations of m and k are unlikely (Booth 1979, p. 49).

She also restricts the singulate mean age at marriage, M, to 21 years and uses the simplification, $M = a_o + 11.37k$, given by Coale and McNeil (1972). This restriction further constrains k for any selected value of a_o and, in particular, k must be less than 1 since a_o has a starting value of 10 years. The resulting schedules produced have mean, p, and standard deviation, a, such that 27c p 29 and $6.75 < o < 8$.

These criteria result in a selection of 33 Coale-Trussell schedules. Table 3.3 lists the parameter combinations and derived statistics for each of these schedules (Booth 1979, p. 55).

Table 3.3: Parameters and derived statistics of schedules identified by Booth

No.	Generating Parameters a0	k	m	Derived Statistics p	a	M
1	10.0	0.7	0.2	28.65	7.54	17.96
2	10.0	0.7	0.6	27.41	7.21	17.96
3	10.0	0.9	0.4	28.79	7.25	20.23
4	10.5	0.6	0.2	28.45	7.56	17.32
5	10.5	0.6	0.6	27.20	7.22	17.32
6	10.5	0.8	0.4	28.63	7.24	19.60
7	11.0	0.5	0.2	28.22	7.59	16.68
8	11.0	0.6	0.6	27.42	7.14	17.82
9	11.0	0.8	0.4	28.85	7.16	20.10
10	11.5	0.4	0.4	27.34	7.46	16.05
11	11.5	0.5	0.6	27.19	7.16	17.18
12	11.5	0.6	0.2	28.87	7.39	18.32
13	12.0	0.4	0.2	28.18	7.56	16.55
14	12.0	0.5	0.6	27.41	7.08	17.68
15	12.0	0.7	0.4	28.91	7.07	19.96
16	12.5	0.3	0.2	27.93	7.62	15.91
17	12.5	0.4	0.6	27.17	7.10	17.05
18	12.5	0.5	0.2	28.86	7.33	18.18
19	13.0	0.2	0.4	27.03	7.51	15.27
20	13.0	0.4	0.2	28.61	7.37	17.55
21	13.0	0.5	0.6	27.90	6.91	18.69
22	13.5	0.2	0.2	27.87	7.61	15.77
23	13.5	0.3	0.6	27.13	7.06	16.91
24	13.5	0.4	0.2	27.84	7.27	18.05
25	14.0	0.2	0.2	28.07	7.51	16.27
26	14.0	0.3	0.6	27.37	6.97	17.41
27	14.0	0.5	0.4	28.99	6.89	19.69
28	14.5	0.1	0.2	27.80	7.62	15.64
29	14.5	0.2	0.6	27.08	7.03	16.77
30	14.5	0.3	0.2	28.81	7.23	17.91
31	15.0	0.1	0.4	27.37	7.33	16.14
32	15.0	0.2	0.2	28.52	7.31	17.27
33	15.0	0.3	0.6	27.88	6.78	18.41

Source: Booth (1979, p 55); reproduced by the author

The cumulative fertilities, F(x), of these schedules are transformed to Y(x) - using the Brass variant of the relational Gompertz model (Equation 3.22) - for each of the five-year age bands.

$$Y(x) = -\ln(-\ln\left(\frac{F(x)}{T}\right))$$ **Equation 3.22**

Where $F(x)$ is cumulative fertility and T is the total fertility rate as before. However, noting that T = 1 in the Coale-Trussell model reduces Equation 3.22 to Equation 3.23:

$$Y(x) = -\ln(-\ln(F(x)))$$ **Equation 3.23**

Booth (1979) then uses the calculated Y(x) values to obtain the first differences, ΔY(x), through Equation 3.24.

$$\Delta Y(x, x+4) = Y(x+5) - Y(x)$$ **Equation 3.24**

The values of F(x) for the 33 schedules identified by Booth are listed in Table 3.4. This process also produces 33 series of Y(x) and ΔY(x) for age groups 15-19, 20-24, 25-29, 30-34, 35-39 and 40-44 shown in Tables 3.5 and 3.6.

Table 3.4: Cumulative Fertility, F(x), by age group for the 33 schedules identified by Booth

	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
1	0.01150	0.13846	0.35821	0.58006	0.77254	0.91884	0.98991	1
2	0.01371	0.16504	0.41974	0.65328	0.82882	0.94442	0.99373	1
3	0.00792	0.11728	0.34164	0.57956	0.78016	0.92475	0.99106	1
4	0.01147	0.14707	0.37133	0.59086	0.77892	0.92119	0.99020	1
5	0.01362	0.17454	0.43340	0.66354	0.83433	0.94627	0.99394	1
6	0.00728	0.12159	0.35202	0.58951	0.78638	0.92707	0.99135	1
7	0.01165	0.15778	0.38555	0.60176	0.78513	0.92345	0.99049	1
8	0.00976	0.16106	0.42114	0.65576	0.83043	0.94500	0.99380	1
9	0.00513	0.10976	0.33826	0.57950	0.78090	0.92516	0.99112	1
10	0.01327	0.18599	0.43248	0.64933	0.81937	0.93860	0.99273	1
11	0.00941	0.17235	0.43642	0.66664	0.83610	0.94687	0.99401	1
12	0.00549	0.12295	0.34739	0.57413	0.76970	0.91789	0.98979	1
13	0.00785	0.15816	0.38991	0.60548	0.78727	0.92422	0.99058	1
14	0.00591	0.15750	0.42355	0.65869	0.83215	0.94559	0.99386	1
15	0.00275	0.10181	0.33539	0.58001	0.78192	0.92565	0.99119	1
16	0.00812	0.17368	0.40551	0.61607	0.79302	0.92628	0.99084	1
17	0.00533	0.17121	0.44050	0.67006	0.83790	0.94747	0.99408	1
18	0.00279	0.11887	0.34926	0.57713	0.77171	0.91866	0.98989	1
19	0.01003	0.20747	0.45255	0.66220	0.82603	0.94087	0.99300	1
20	0.00221	0.13034	0.36591	0.58961	0.77868	0.92116	0.99020	1
21	0.00163	0.12594	0.39443	0.64049	0.82310	0.94265	0.99353	1
22	0.00398	0.17820	0.41058	0.61944	0.79484	0.92692	0.99092	1
23	0.00192	0.17182	0.44553	0.67359	0.83969	0.94805	0.99414	1
24	0.00086	0.11542	0.35233	0.58057	0.77379	0.91942	0.98999	1
25	0.00106	0.16392	0.40010	0.61267	0.79119	0.92562	0.99076	1
26	0.00050	0.15388	0.43195	0.66554	0.83573	0.94677	0.99400	1
27	0.00017	0.08499	0.33242	0.58317	0.78478	0.92680	0.99133	1
28	0.00053	0.18374	0.41498	0.62228	0.79637	0.92747	0.99099	1
29	0.00013	0.17506	0.45103	0.67700	0.84136	0.94859	0.99420	1
30	0.00005	0.11319	0.35662	0.58425	0.77586	0.92016	0.99008	1
31	0.00000	0.18434	0.43686	0.65252	0.82105	0.93917	0.99280	1
32	0.00000	0.13139	0.37577	0.59695	0.78272	0.92260	0.99038	1
33	0.00000	0.11498	0.40070	0.64693	0.82658	0.94380	0.99366	1

Source: Booth (1979, p. 58); reproduced by the author

Table 3.5: Y(x) values by 5 year age groups of the 33 schedules used by Booth

	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
1	-1.49636	-0.68167	-0.02629	0.60766	1.35452	2.46931	4.59114	
2	-1.45620	-0.58866	0.14143	0.85390	1.67263	2.86148	5.06884	
3	-1.57658	-0.76230	-0.07139	0.60608	1.39329	2.54808	4.71273	
4	-1.49695	-0.65068	0.00938	0.64212	1.38691	2.49995	4.62045	
5	-1.45773	-0.55710	0.17901	0.89119	1.70856	2.89630	5.10301	
6	-1.59384	-0.74531	-0.04312	0.63778	1.42580	2.58063	4.74586	
7	-1.49345	-0.61332	0.04805	0.67748	1.41921	2.53026	4.65064	
8	-1.53244	-0.60212	0.14527	0.86284	1.68302	2.87227	5.08010	
9	-1.66253	-0.79275	-0.08060	0.60589	1.39712	2.55376	4.71950	
10	-1.46378	-0.52002	0.17648	0.83976	1.61335	2.75883	4.92035	
11	-1.54030	-0.56431	0.18735	0.90262	1.72033	2.90784	5.11466	
12	-1.64959	-0.74002	-0.05573	0.58897	1.34035	2.45716	4.57926	
13	-1.57841	-0.61202	0.05992	0.68969	1.43052	2.54078	4.66019	
14	-1.63532	-0.61428	0.15189	0.87347	1.69422	2.88336	5.08985	
15	-1.77430	-0.82621	-0.08843	0.60750	1.40241	2.56059	4.72745	
16	-1.57141	-0.55992	0.10246	0.72486	1.46142	2.56944	4.68831	
17	-1.65525	-0.56807	0.19864	0.91532	1.73242	2.91951	5.12645	
18	-1.77185	-0.75599	-0.05063	0.59840	1.35036	2.46700	4.58915	
19	-1.52653	-0.45284	0.23211	0.88628	1.65483	2.79770	4.95833	
20	-1.81071	-0.71178	-0.00535	0.63810	1.38567	2.49956	4.62045	
21	-1.85929	-0.72849	0.07223	0.80851	1.63641	2.82920	5.03734	
22	-1.70955	-0.54514	0.11633	0.73618	1.47135	2.57850	4.69712	
23	-1.83345	-0.56606	0.21259	0.92853	1.74456	2.93092	5.13667	
24	-1.95424	-0.76973	-0.04228	0.60928	1.36080	2.47679	4.59914	
25	-1.92417	-0.59243	0.08769	0.71350	1.45151	2.56017	4.67958	
26	-2.02827	-0.62678	0.17501	0.89856	1.71786	2.90591	5.11299	
27	-2.16099	-0.90228	-0.09654	0.61753	1.41736	2.57679	4.74354	
28	-2.02057	-0.52723	0.12837	0.74578	1.47976	2.58634	4.70490	
29	-2.19143	-0.55539	0.22788	0.94139	1.75599	2.94165	5.14699	
30	-2.29289	-0.77872	-0.03061	0.62096	1.37127	2.48642	4.60822	
31	undefined	-0.52530	0.18857	0.85117	1.62368	2.76846	4.93006	
32	undefined	-0.70783	0.02145	0.66180	1.40658	2.51876	4.63908	
33	undefined	-0.77149	0.08933	0.83122	1.65832	2.85006	5.05770	

Source: Booth (1979, p. 59); reproduced by the author

Table 3.6: AY(x) by 5 year age groups as calculated by Booth

	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
1	undefined	0.81469	0.65538	0.63395	0.74686	1.11479	2.12183	
2	undefined	0.86754	0.73008	0.71248	0.81873	1.18884	2.20736	
3	undefined	0.81428	0.69091	0.67747	0.78722	1.15478	2.16465	
4	undefined	0.84626	0.66006	0.63274	0.74479	1.11304	2.12050	
5	undefined	0.90064	0.73611	0.71218	0.81737	1.18774	2.20671	
6	undefined	0.84853	0.70219	0.68091	0.78802	1.15483	2.16522	
7	undefined	0.88013	0.66137	0.62943	0.74173	1.11105	2.12038	
8	undefined	0.93032	0.74738	0.71758	0.82018	1.18925	2.20783	
9	undefined	0.86979	0.71214	0.68649	0.79123	1.15664	2.16574	
10	undefined	0.94376	0.69650	0.66328	0.77359	1.14548	2.16152	
11	undefined	0.97599	0.75166	0.71527	0.81771	1.18751	2.20682	
12	undefined	0.90957	0.68429	0.64469	0.75138	1.11681	2.12210	
13	undefined	0.96639	0.67194	0.62977	0.74084	1.11025	2.11941	
14	undefined	1.02104	0.76617	0.72158	0.82075	1.18915	2.20649	
15	undefined	0.94809	0.73778	0.69594	0.79491	1.15818	2.16686	
16	undefined	1.01148	0.66239	0.62239	0.73656	1.10802	2.11888	
17	undefined	1.08718	0.76671	0.71668	0.81710	1.18710	2.20694	
18	undefined	1.01586	0.70536	0.64904	0.75196	1.11664	2.12216	
19	undefined	1.07369	0.68495	0.65416	0.76856	1.14286	2.16064	
20	undefined	1.09893	0.70642	0.64346	0.74757	1.11388	2.12090	
21	undefined	1.13080	0.80072	0.73628	0.82790	1.19279	2.20814	
22	undefined	1.16441	0.66147	0.61985	0.73517	1.10714	2.11863	
23	undefined	1.26739	0.77864	0.71595	0.81603	1.18636	2.20575	
24	undefined	1.18452	0.72745	0.65156	0.75153	1.11599	2.12235	
25	undefined	1.33174	0.68012	0.62580	0.73801	1.10866	2.11940	
26	undefined	1.40148	0.80180	0.72354	0.81930	1.18805	2.20708	
27	undefined	1.25871	0.80574	0.71407	0.79984	1.15943	2.16674	
28	undefined	1.49334	0.65560	0.61740	0.73399	1.10658	2.11855	
29	undefined	1.63603	0.78327	0.71352	0.81460	1.18566	2.20534	
30	undefined	1.51416	0.74811	0.65157	0.75031	1.11514	2.12181	
31	undefined	undefined	0.71387	0.66260	0.77251	1.14477	2.16161	
32	undefined	undefined	0.72928	0.64035	0.74478	1.11218	2.12032	
33	undefined	undefined	0.86083	0.74189	0.82710	1.19174	2.20764	

Source: Booth (1979, p. 60); reproduced by the author

However, Booth like Brass (1974) - discussed in section 3.1 - finds that Y(x) for the middle age groups (20-24, 25-29, 30-34 and 35-39) lies on a fairly straight line, but that there is a divergence in the tails of the distribution (Figure 3.5).

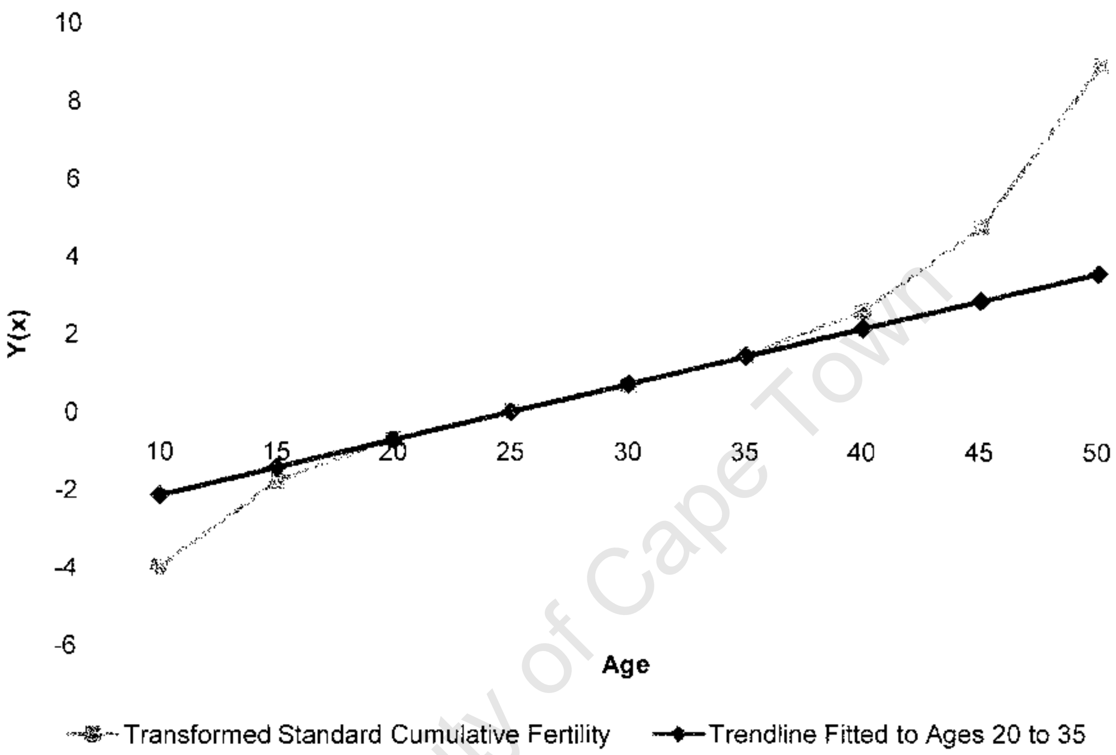

Figure 3.5: Demonstration of the misfit in the tails of the Y(x) distribution

As a result of this misfit in the tails, Booth recognizes the need to weight the tails of the fertility distributions to ensure that the resulting standard will more adequately reflect patterns observed for high fertility populations. She notes that, given the close fit over the central ages, any transformation should leave the values in this range relatively unchanged while still altering the tails (Booth 1979).

To do so, Booth follows a three step procedure. First, she determines the correct level for the central ages. She calculates the average AY(x) of all 33 schedules identified for the age groups 25-29, 30-34 and 35-39. These average values are then used as the standard for each age group - AY,(25-29), AY,(30-34) and AY,(35-39).

Second, Booth determines the schedules to be used for the upper and lower tails and calculate the AY(x) of these schedules. She sets out two criteria to determine which of the schedules to use for calculating the lower and upper tails. Booth reasons that, since f(10-14) and f(45-49) are small, the age groups must be grouped to ensure that at least ten percent of fertility will occur in each categorisation (Booth 1979, p. 61). As a result, for the first criterion Booth combines the 10-14 and 15-19 age groups and determines that a schedule must have f(10-19) > 0.15 to qualify for the lower tail.

Similarly, Booth groups the fertility for the 40-44 and 45-49 age group to obtain f(40-49). However, f(40-49) is still less than 10 per cent of total fertility and, consequently, the 35-39 fertility must also be included in the grouping for the upper tail. This allows Booth to set the second condition - f(35-49) > 0.21 (Booth 1979, p. 56).

Based on the criteria, 17 schedules are chosen for use in the lower tail and 16 schedules are used for the upper tail. Table 3.7 below shows the values of f(10-19) and f(35-49) for the 33 Coale-Trussell schedules as well as the tail that the schedule was used for as indicated in column 7.

Table 3.7: Identifying the schedules used in each tail

Schedule	ao	k	m	f(10-19)	f(35-49)	Used For
1	10	0.7	0.2	0.1385	0.2275	Upper Tail
2	10	0.7	0.6	0.1650	0.1712	Lower Tail
3	10	0.9	0.4	0.1173	0.2198	Upper Tail
4	10.5	0.6	0.2	0.1471	0.2211	Upper Tail
5	10.5	0.6	0.6	0.1745	0.1657	Lower Tail
6	10.5	0.8	0.4	0.1216	0.2136	Upper Tail
7	11	0.5	0.2	0.1578	0.2149	Both Tails
8	11	0.6	0.6	0.1611	0.1696	Lower Tail
9	11	0.8	0.4	0.1098	0.2191	Upper Tail
10	11.5	0.4	0.4	0.1860	0.1806	Lower Tail
11	11.5	0.5	0.6	0.1724	0.1639	Lower Tail
12	11.5	0.6	0.2	0.1230	0.2303	Upper Tail
13	12	0.4	0.2	0.1582	0.2127	Both Tails
14	12	0.5	0.6	0.1575	0.1679	Lower Tail
15	12	0.7	0.4	0.1018	0.2181	Upper Tail
16	12.5	0.3	0.2	0.1737	0.2070	Lower Tail
17	12.5	0.4	0.6	0.1712	0.1621	Lower Tail
18	12.5	0.5	0.2	0.1189	0.2283	Upper Tail
19	13	0.2	0.4	0.2075	0.1740	Lower Tail
20	13	0.4	0.2	0.1304	0.2213	Upper Tail
21	13	0.5	0.6	0.1259	0.1769	Neither Tail
22	13.5	0.2	0.2	0.1782	0.2052	Lower Tail
23	13.5	0.3	0.6	0.1718	0.1603	Lower Tail
24	13.5	0.4	0.2	0.1154	0.2262	Upper Tail
25	14	0.2	0.2	0.1639	0.2088	Lower Tail
26	14	0.3	0.6	0.1539	0.1643	Lower Tail
27	14	0.5	0.4	0.0850	0.2152	Upper Tail
28	14.5	0.1	0.2	0.1838	0.2036	Upper Tail
29	14.5	0.2	0.6	0.1751	0.1586	Lower Tail
30	14.5	0.3	0.2	0.1132	0.2241	Upper Tail
31	15	0.1	0.4	0.1844	0.1789	Lower Tail
32	15	0.2	0.2	0.1314	0.2173	Upper Tail
33	15	0.3	0.6	0.1150	0.1734	Neither Tail

Source: Booth (1979, p. 58); reproduced by the author

Once the schedules were selected the average AY(x) were calculated. Table 3.8 gives the average AY(x) values for all 33 schedules - used for the middle section of the standard - in column 2. The 17 lower tail schedules are listed in column 3 and the 16 upper tail schedules are shown in column 4.

Table 3.8: Average AY(x) values by 5 year age groups

	All Schedules	Lower Tail	Upper Tail
AY (10-14)			
AY (15-19)	1.07356	1.09120	1.03088
AY (20-24)	0.72354	0.72320	0.70338
\Y (25-29)	0.67436	0.67976	0.65492
\Y (30-34)	0.77872	0.78638	0.76042
\Y (35-39)	1.14730	1.15692	1.12689
AY (40-44)	2.15989	2.17299	2.13486
AY (45-49)	∞	∞	∞

It is clear from the table that the AY(x) values are different for the three columns. Of particular concern are the differences in the middle age groups where fertility should be relatively constant and this indicates disparity in the fertility levels and distribution.

Since the intention was to alter the tails while leaving the middle age groups largely unchanged, Booth uses adjustment factors to bring the middle age groups of all three columns (2, 3 and 4) to the same level (Booth 1979, p. 61). The two factors k_1 and k_2 are calculated using Equations 3.25 and 3.26:

$$k_1 = \frac{average \quad \Delta Y(25-39) \quad for \quad all \quad schedules}{average \quad \Delta Y(25-39) \quad for \quad lower \quad tail}$$ **Equation 3.25**

$$= \frac{0.67436 + 0.77872 + 1.14730}{0.67976 + 0.78638 + 1.15692}$$

$$= 0.99135$$

Similarly,

$$k_2 = \frac{average \quad \Delta Y(25-39) \quad for \quad all \quad schedules}{average \quad \Delta Y(25-39) \quad for \quad upper \quad tail}$$ **Equation 3.26**

$$= \frac{0.67436 + 0.77872 + 1.14730}{0.65492 + 0.76042 + 1.12689}$$

$$= 1.02287$$

The adjustment factor *k,* is used to adjust the average AY(15-19) and AY(20-24) values, whilst *k,* is used to adjust AY(40-44) to give the standard values AY,(15-19), AY,(20-24) and AY,(40-44) in Table 3.9 (Booth 1979, p.61).

In order to convert the standard AY,(x) values into a useful age-specific fertility schedule a fixed starting point is required and Booth selects Y,(30) = 0.7. She states that the selection of Y,(30) = 0.7 is "arbitrary" (Booth 1979, p. 63) and that other selections will result in other age-specific schedules being obtained (Booth 1979; 1984). The standard Y,(x) is derived from AY,(x) using the fixed starting point - $Y_s(30) = 0.7$ - and recursive Equations 3.27(for ages under 30) and 3.28 (for ages over 30).

$$Y_s(x) = Y_s(x+5) - \Delta Y_s(x \quad to \quad x+4)$$ **Equation 3.27**

$$Y_s(x+5) = Y_s(x) + \Delta Y_s(x \quad to \quad x+4)$$ **Equation 3.28**

For example:

$$Y_s(25) = Y_s(30) - \Delta Y_s(25 - 29)$$
$$= 0.7 - 0.67436$$
$$= 0.02564$$

Similarly,

$$Y_s(35) = Y_s(30) + \Delta Y_s(30 - 34)$$
$$= 0.7 + 0.77872$$
$$= 1.47872$$

The Y,(x) values obtained are then converted back to a standard cumulative fertility schedule, F,(x), by means of an anti-gompit (columns 6 and 7 in Table 3.9). These are then differenced to obtain the standard fertility schedule, $f_s(x)$, by five year age group (column 8). The standard obtained in this fashion applies to exact ages 15, 20, ..., 45 and 50.

Table 3.9: Determination of Booth standard from the adjusted $\Delta Y(x)$ values

Age group (1)	$\Delta Y(x)$ (2)	Weight (3)	Adjusted $\Delta Y(x)$ (4) = (2)*(3)	Exact age (5)	$Y_s(x)$ (6)	$F_s(x)$ (7)	$f_s(x)$ (8)
10-14	∞			15	-1.77306	0.00277	0.00277
15-19	1.09120	0.99135	1.08176	20	-0.69130	0.13584	0.13307
20-24	0.72320	0.99135	0.71694	25	0.02564	0.37731	0.24147
25-29	0.67436	1	0.67436	30	0.70000	0.60861	0.23124
30-34	0.77872	1	0.77872	35	1.47872	0.79618	0.18757
35-39	1.14730	1	1.14730	40	2.62602	0.93019	0.13401
40-44	2.13486	1.02287	2.18368	45	4.80970	0.99188	0.06169
45-49	∞			50		1	0.00812

Source: Booth (1979, p. 62); reproduced by the author

Booth (1979), however, generalizes the five-year standard by producing a single-year standard in the same fashion. Finally, after calculating both the 5-year and single-year standards, Booth provides a function[3] (given in Equation 3.29) to obtain average parity values, F_s (x, x + 4) . This average parity function yields values that, according to Booth, "refer to the ages at which average and actual parities are equal" and not to the exact midpoint of the usual age intervals (Booth 1979, p.80).

$$\bar{F}_S(x, x+4) = F_S(x) + \frac{1}{5}[4.5 f_S(x) + 3.5 f_S(x+1) + 2.5 f_S(x+2) + 1.5 f_S(x+3) + 0.5 f_S(x+4)]$$

Equation 3.29

3.4 Apparent Anomalies

As discussed in section 3.3, Booth (1979) identifies 33 Coale-Trussell schedules with high fertility patterns and recognises that the tails need to be weighted to more accurately reflect the pattern of fertility observed for high fertility populations.

She splits the 33 schedules into two groups using the predefined criteria f(10-19) > 0.15 for the lower tail and f(35-49) > 0.21 for the upper tail (Booth 1979, p. 56). That is, Booth uses schedules with the highest early fertility to model the lower tail and the highest late fertility for the upper tail. The first impression is that this methodology (of using the schedules with the highest early fertility to model the fertility for the 10-14, 15-19 and 20-24 age groups) will inflate early fertility rates. Further, seeing as fertility rates are cumulative, an early inflation of fertility must be

[3] This function can be derived from the trapezium rule approximation to an integral (Zaba 1981).

offset by later compression. That is, the fertility rates for the older age groups - in particular the 45-49 ASFR - must be suppressed below what would otherwise have been the case.

If this is the case then the impact of Booth's procedure defeats her reasons for using the high fertility schedules to calculate the upper tail. However, contrary to first impressions, Table 3.10 shows that the method results in lower ASFR for the age groups below 25.

Table 3.10: Comparison between the Booth Standard Rates and the Average Rates

Age Group	Arithmetic Average of 33 Schedules	Standard as derived by Booth
10 – 14	0.00536	0.00277
15 – 19	0.14286	0.13307
20 – 24	0.24586	0.24147
25 – 29	0.22711	0.23130
30 – 34	0.18181	0.18757
35 – 39	0.12936	0.13401
40 – 44	0.05955	0.06169
45 – 49	0.00811	0.00812

By contrast, there is an increase in ASFR for the age groups above 25 and negligible effect in the 45-49 age group. That is, the method shifts fertility slightly from the earlier ages to the later age groups. This shift results from the selection of Y(30) = 0.7 which is lower than the average value observed for the 33 schedules used by Booth - Y(30) = 0.74208. The lower Y(30) translates into a smaller portion of fertility completed by age 30. Put differently, there is a larger portion of fertility remaining after age 30 for lower values of Y(30) which means higher levels of fertility in the older age groups.[4]

Despite the seemingly reasonable nature of the results in Table 3.10, some apparent irregularities came to light while replicating Booth's results and these are recorded here. Table 3.11 shows the values of f(10-19) and f(35-49) for the 33 Coale-Trussell schedules used by Booth.

[4] The effect of alternative selections of Y(30) is discussed more fully later.

Table 3.11: The 33 Coale-Trussell Schedules and the f(10-19) and f(35-49) values

Schedule (1)	ao (2)	K (3)	M (4)	f(10-19) (5)	f(35-49) (6)	Based on Criteria (7)	Used by Booth (8)	(9)
1	10	0.7	0.2	0.1385	0.2275	Upper Tail	Upper Tail	
2	10	0.7	0.6	0.1650	0.1712	Lower Tail	Lower Tail	
3	10	0.9	0.4	0.1173	0.2198	Upper Tail	Upper Tail	
4	10.5	0.6	0.2	0.1471	0.2211	Upper Tail	Upper Tail	
5	10.5	0.6	0.6	0.1745	0.1657	Lower Tail	Lower Tail	
6	10.5	0.8	0.4	0.1216	0.2136	Upper Tail	Upper Tail	
7	11	0.5	0.2	0.1578	0.2149	Both Tails	Both Tails	
8	11	0.6	0.6	0.1611	0.1696	Lower Tail	Lower Tail	
9	11	0.8	0.4	0.1098	0.2191	Upper Tail	Upper Tail	
10	11.5	0.4	0.4	0.1860	0.1806	Lower Tail	Lower Tail	
11	11.5	0.5	0.6	0.1724	0.1639	Lower Tail	Lower Tail	
12	11.5	0.6	0.2	0.1230	0.2303	Upper Tail	Upper Tail	
13	12	0.4	0.2	0.1582	0.2127	Both Tails	Both Tails	
14	12	0.5	0.6	0.1575	0.1679	Lower Tail	Lower Tail	
15	12	0.7	0.4	0.1018	0.2181	Upper Tail	Upper Tail	
16	12.5	0.3	0.2	0.1737	0.2070	Lower Tail	Lower Tail	
17	12.5	0.4	0.6	0.1712	0.1621	Lower Tail	Lower Tail	
18	12.5	0.5	0.2	0.1189	0.2283	Upper Tail	Upper Tail	
19	13	0.2	0.4	0.2075	0.1740	Lower Tail	Lower Tail	
20	13	0.4	0.2	0.1304	0.2213	Upper Tail	Upper Tail	
21	13	0.5	0.6	0.1259	0.1769	Neither Tail	Neither Tail	
22	13.5	0.2	0.2	0.1782	0.2052	Lower Tail	Lower Tail	
23	13.5	0.3	0.6	0.1718	0.1603	Lower Tail	Lower Tail	
24	13.5	0.4	0.2	0.1154	0.2262	Upper Tail	Upper Tail	
25	14	0.2	0.2	0.1639	0.2088	Lower Tail	Lower Tail	
26	14	0.3	0.6	0.1539	0.1643	Lower Tail	Lower Tail	
27	14	0.5	0.4	0.0850	0.2152	Upper Tail	Upper Tail	
28	14.5	0.1	0.2	0.1838	0.2036	Lower Tail	Upper Tail	
29	14.5	0.2	0.6	0.1751	0.1586	Lower Tail	Lower Tail	
30	14.5	0.3	0.2	0.1132	0.2241	Upper Tail	Upper Tail	
31	15	0.1	0.4	0.1844	0.1789	Lower Tail	Lower Tail	
32	15	0.2	0.2	0.1314	0.2173	Upper Tail	Upper Tail	
33	15	0.3	0.6	0.1150	0.1734	Neither Tail	Neither Tail	

Column 7 in the table indicates the schedules that qualify for inclusion in the tails based on the criteria set out above. Column 8 reproduces the results from Table 3.4 in Booth (1979, p. 57) and shows the tail to which Booth allocates each schedule - 17 for the lower tail and 16 for the upper tail.

The inclusion of 17 schedules for the lower tail and 16 for the upper tail creates the impression that all 33 schedules are used for the tails. However, as can be seen from column 8 of the table, this is not case. Two schedules (7 and 13 marked with t in column 9) are used for both tails and two schedules (21 and 33 indicated by t) are not used for either tail. In addition, schedule 28 (marked *) is included by Booth for the upper tail although it does not meet the pre-defined criterion since f(35-49) 0.204. Moreover, seeing as f(10-19) 0.184, schedule 28 should be included in the group used to calculate the standard fertility for the age bands 10-14, 15-19 and 20-24 i.e. for the lower tail'. The reasons for these inclusions and omissions are never explained, but the effect is negligible

Once the schedules were selected, the transformations were applied and the $AY_s(x)$ values were calculated. As stated before, the method requires a fixed starting point Y,(x) in order to transform the AY, (x) into Y,(x) and, finally, calculate a useful ASFR. This raises a question about how the selection of the fixed Y,(x) affects the resulting ASFR and how this starting point was determined.

Booth (1979) asserts that changing the value of Y,(x) is the same as changing the origin of the standard. She demonstrates that a change of origin amounts to a vertical movement of the Y-curve such that Equation 3.30 holds at all ages.

$$Y_{new}(x) - Y_s(x) = d$$ **Equation 3.30**

Carrying this through to the Gompertz function implies that the cumulative fertility (under the new origin) is related by a constant power function to the cumulative fertility of the original standard (see Equation 3.31).

$$\begin{aligned} F_{new}(x) &= e^{e^{[-Y_{new}(x)]}} \\ &= e^{e^{[-Y_s(x)-d]}} \\ &= [F_s(x)]^{e^{-d}} \end{aligned}$$ **Equation 3.31**

In order to determine the standard, Booth selects the starting point $Y_s(30) = 0.7$ (Booth, 1979). However, alternative selections of a starting point will give rise to different distributions in

[5] Appendix B shows the impact of correctly using schedule 28 for the lower tail

accordance with Equation 3.31. Such an alternative selection can dramatically change the pattern of standard fertility as shown by Tables 3.12 and 3.13 and Figures 3.6 and 3.7

Table 3.12: Comparative standards arising from alternative starting values

	Standard Schedule if Y(30) is:		
Age Group	Y(30) = 1.2	Y(30) = 0.7	Y(30) = 0.2
10 - 14	0.02811	0.00277	0.00006
15 - 19	0.26984	0.13307	0.03714
20 - 24	0.25572	0.24147	0.16329
25 - 29	0.18626	0.23130	0.24050
30 - 34	0.13095	0.18757	0.24575
35 - 39	0.08617	0.13401	0.20079
40 - 44	0.03801	0.06169	0.09912
45 - 49	0.00493	0.00812	0.01335

Table 3.13: Cumulative fertility arising from alternative starting values

	Standard Cumulative Fertility if Y(30) is:		
Age Group	Y(30) = 1.2	Y(30) = 0.7	Y(30) = 0.2
10 - 14	0.02811	0.00277	0.00006
15 - 19	0.29795	0.13583	0.03720
20 - 24	0.55368	0.37731	0.20049
25 - 29	0.73993	0.60861	0.44099
30 - 34	0.87088	0.79618	0.68675
35 - 39	0.95706	0.93019	0.88753
40 - 44	0.99507	0.99188	0.98665
45 - 49	1	1	1

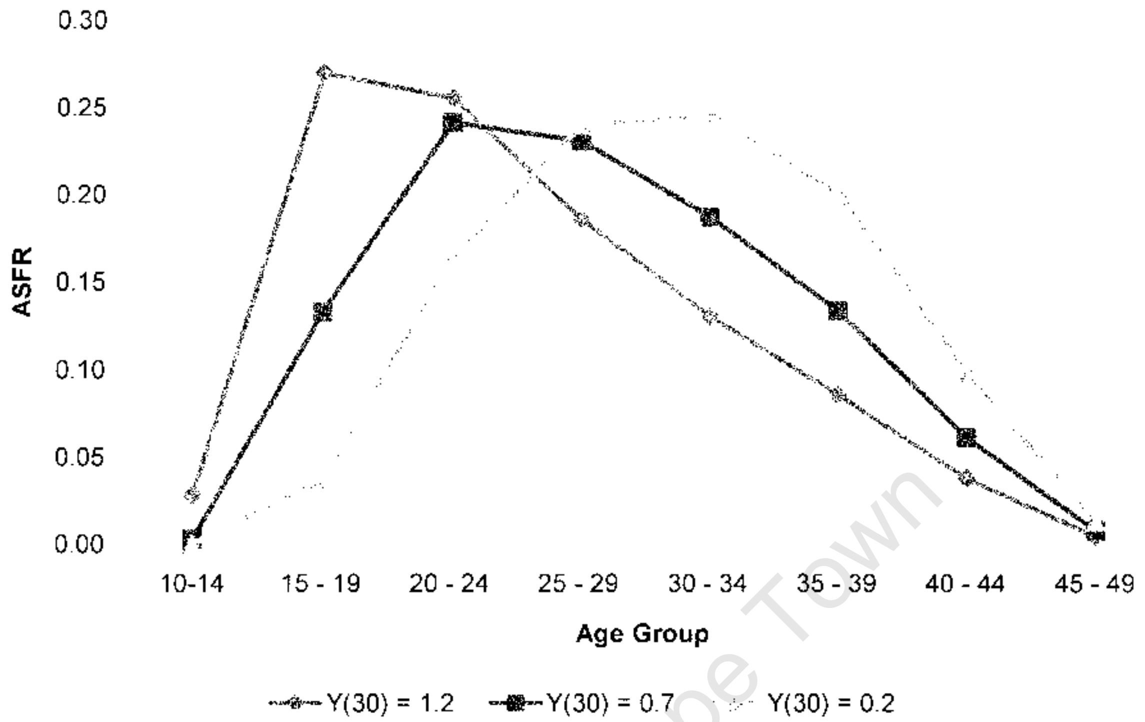

Figure 3.6: Comparative standard schedules arising from different starting values of Y(30)

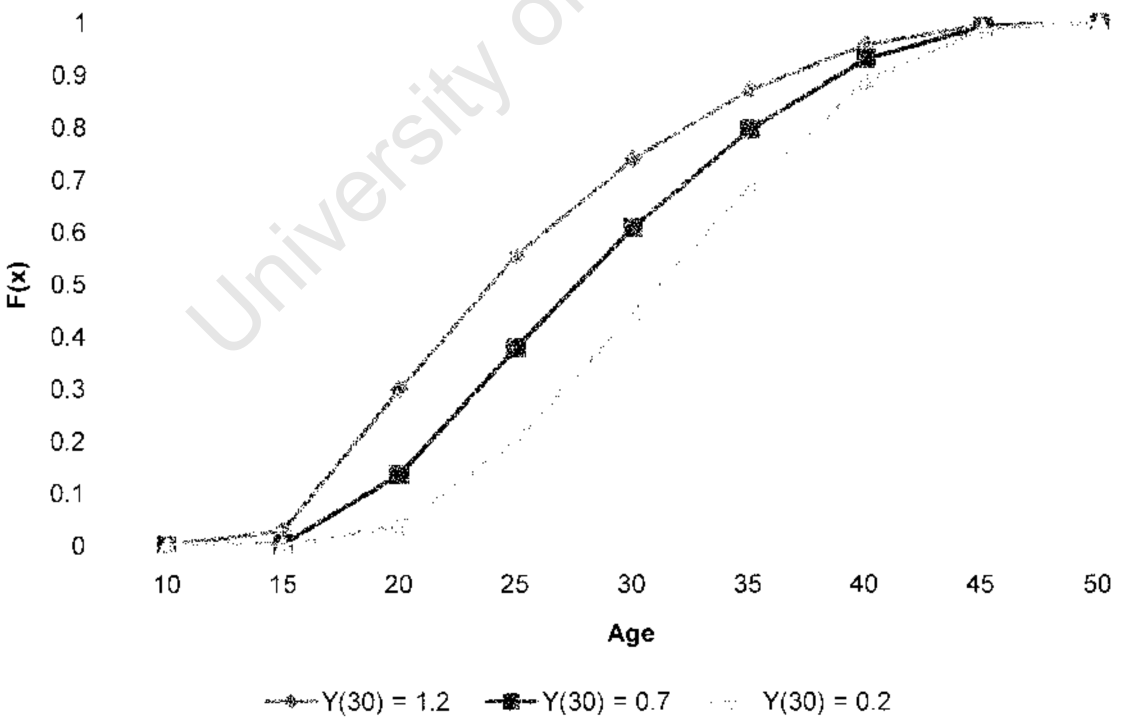

Figure 3.7: Cumulative fertility, F(x), arising from different starting values of Y(30)

The selection of starting point Y,(30) = 0.7 indicates that Booth believes about 60 per cent of fertility will occur by age 30. By contrast, Y,(30) = 0.2 and Y,(30) = 1.2, respectively, imply that 44 and 74 per cent of fertility will occur by age 30 (Figure 3.7).

Once the 5-year standard was determined it was extended to single years and a third apparent irregularity is noticed. Booth (1979, p. 65) states that the "same adjustment factors" are used when calculating the single-year standard as were used for the five-year standard. The factors derived by Booth are k_1 = 0.99135 and k_2 = 1.02287. As discussed in section 3.3, Booth multiplies the AY values - for all ages below 25 - with the adjustment factor k_1.

Consequently, one would expect that the standard AY's (in column 3 of Table 3.14) divided by the average AY's (in column 2) will give implied k_1 values equal to the k_1 value calculated by Booth.

Table 3.14: Average and Standard ΔY values and k_1 implied by these values

Age	Average ΔY	Standard ΔY	Implied k_1
10-11	∞	∞	
11-12	0.4927	0.48844	0.99135
12-13	0.32998	0.32713	0.99136
13-14	0.30295	0.30033	0.99135
14-15	0.30217	0.29956	0.99136
15-16	0.28611	0.28020	0.97934
16-17	0.24783	0.24225	0.97748
17-18	0.21108	0.20582	0.97508
18-19	0.19061	0.18552	0.97330
19-20	0.17290	0.16797	0.97149
20-21	0.15944	0.15805	0.99128
21-22	0.14931	0.14801	0.99129
22-23	0.14224	0.14101	0.99135
23-24	0.13759	0.13640	0.99135
24-25	0.13463	0.13347	0.99138

Source: Booth (1984, p. 501)

From the table it is clear that the implied k_1 values (in column 4) for the ages 15 to 19 do not equal the value of k_1 calculated by Booth. Although this may initially be perceived as an irregularity, Booth explains that this inconsistency is necessary. In particular, it ensures that cumulating the single-year standard (in five-year age groups) gives the same result as the derived five-year standard (Booth 1979, p. 65).

Finally, after calculating both the 5-year and single-year standards, Booth calculates the average parity values corresponding "to the ages at which average and actual parities are equal" and not to the exact midpoint of the age intervals (Booth 1979, p.80).

$$\overline{F}_s(x, x+4) = F_s(x) + \frac{1}{5}[4.5 f_s(x) + 3.5 f_s(x+1) + 2.5 f_s(x+2) + 1.5 f_s(x+3) + 0.5 f_s(x+4)]$$

Equation 3.32

The above equation for P,(x to x+4) is derived through a linear approximation of the integral over each age range. Although the equation is mathematically correct, the coefficients derived using this linear approximation have the effect of weighting parity-estimates towards the first age in the age group. Although this may be reasonable for the middle and older ages, it seems inappropriate for the 10-14 and 15-19 age bands where most of the fertility occurs towards the end of the age category. As a result, the parities for the lower age groups are too low.

3.5 Problems of Misfit

According to Booth (1979; 1984) and Brass and Airey (1988), the standard was designed for use with the relational Gompertz model in order to correct data problems commonly found in high fertility populations. As such, the standard should be particularly useful in Africa and especially sub-Saharan Africa, since this is the region with the highest total fertility rates and arguably the poorest quality data in the world. According to Guengant and May (2001), about a third of African countries were vet to experience large fertility declines by the 1990's.

Evidence from the sub-Saharan Demographic and Health Surveys (DHS) show that, although the TFR is still high by developed countries' standards, fertility in the region has indeed declined and is still declining, supporting the findings by Cohen (1993; 1998), Garenne and Joseph (2002) and Caldwell and Caldwell (2002). In particular, the data exhibits widespread declines in total fertility rates for the region starting two decades ago.

Irrespective of the cause of these declines the result remains that fertility rates can no longer be termed high by 1980 standards. Given this decline one must consider that Booth (1979; 1984) warns against using the standard if the high fertility assumption is not met. She and others (notably, Zaba (1981) and Brass and Airey (1988)) argue that differences in the shapes of fertility will result in poor estimates if the standard is incorrectly used. The reason is that the Relation Gompertz Model is based on the premise that the standard fertility pattern adequately represents the fertility pattern of the population being modelled.

By contrast to Booth, Brass and Zaba, some authors and analysts (e.g. Udjo 2003) consistently apply the relational Gompertz model, using the Booth standard, with disregard for these specified restrictions. By using the relational Gompertz model for analysis despite these warnings, it is argued implicitly that the magnitude of the location parameter a and spread parameter ρ does not adversely affect the accuracy of the analysis. This would be true if the age-specific fertility patterns were the same in high and low fertility settings, since the problem will reduce merely to an estimation of the differences in level. In addition, since Zaba's restatement already eliminates the effect of differences in level, the Booth standard may be used if the patterns of fertility are consistent among all populations. However, Figure 3.8 demonstrates that fertility patterns are not the same for all populations.

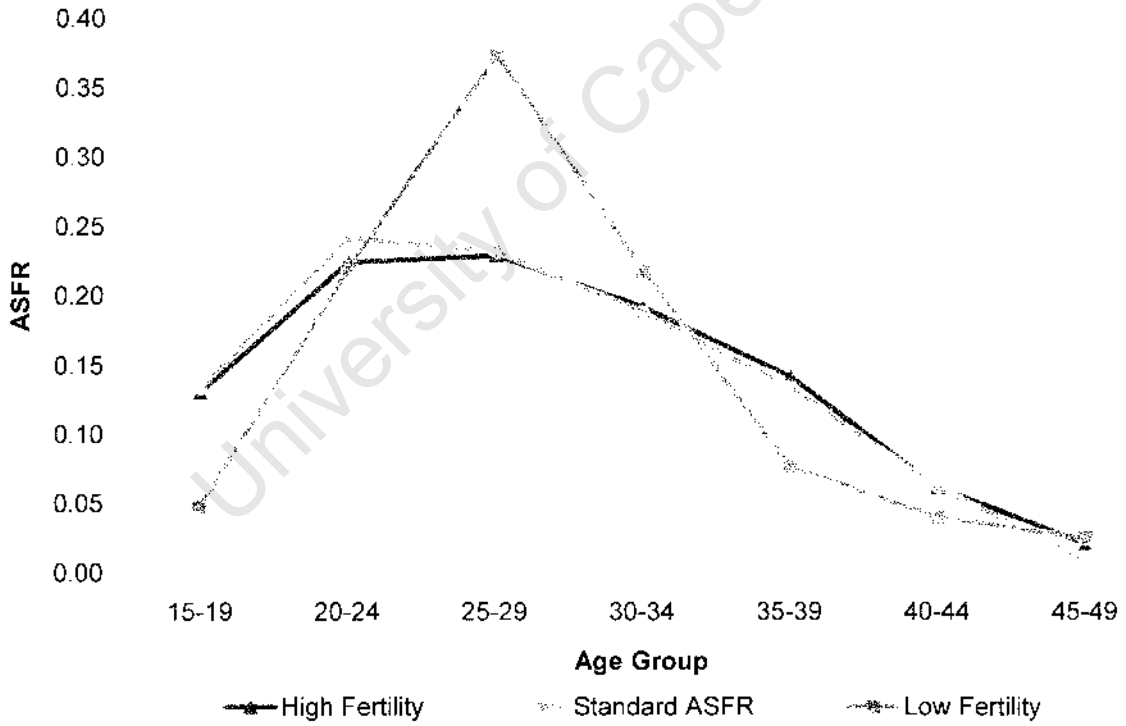

Figure 3.8: Standardised ASFR for the Booth standard, low and high fertility populations[6]

[6] The high- and low fertility populations are, respectively, the African and White populations of South Africa. The 1996 Census data are used for the white population. This population is used because there were sample size problems for this group in the 1998 DHS (Moultrie and Timæus, 2002, p. 19).

As a result, the appropriateness of the Booth standard for different populations is more completely investigated in sections 3.5.1 to 3.5.3.

3.5.1 The Booth standard and low fertility populations

The degree of deviation observed in Figure 3.8 between the standard and the low fertility population is strongly suggestive that the Booth standard cannot be used to model this population. The reason is that the misfit between the schedules will impose bias upon the fertility estimates, since an inappropriate pattern of fertility is forced upon the analysis.

In particular, as already discussed, Booth and others warn that estimates derived using the relational Gompertz model with the Booth standard when the underlying pattern of fertility differs will result in poor and biased estimates. If the parameters fall outside the ranges specified by Zaba (1981) then the assumption that limits the Taylor expansion to two difference terms is violated and the third and subsequent terms of the expansion become significant.

An example of this occurs in the handling of the South African fertility data by Udjo (2003). He applies the relational Gompertz model to both the high and low fertility populations. However, Udjo's results (p. 422-423) show that 5 of the 16 current estimates and 9 of the 24 period estimates have the p-parameter outside the critical range specified by Zaba. Significantly, all these estimates are for the two low fertility populations — Indian and White. In fact, for the two low fertility populations 5 of the 8 current estimates and 9 of the 12 period estimates fall outside Zaba's range.

Undoubtedly, it is the lack of statistical correspondence observed in the patterns of fertility - illustrated by Figure 3.8 - that results in these significant ß-estimates . This strongly suggests that the relational Gompertz model cannot be applied to such dissimilar distributions since the conclusions drawn and estimates derived must be flawed even if the results seem reasonable.

3.5.2 The Booth standard and high fertility

By contrast, there is fair correspondence between the South African high fertility population and the standard. The similarity in shape immediately suggests that the Booth standard may be appropriate when modelling this population.

However, the overstatement of 20-24 fertility and severe underestimate for 45-49 fertility raises concerns. These discrepancies may come from differences in timing, location and tilt between the high fertility population and the standard and this could result in significant a-values in the relational Gompertz model. Although significant a-values still affect the shape of the fertility distribution the problem is less serious, since the simplification applied to the Taylor expansion remains valid.

Despite the large difference in the upper tail there is a much closer fit between the standard and the high fertility population than for the low fertility population discussed in section 3.5.1. This close tit; ß-estimates within an acceptable range; and the fact that Booth (1979; 1984), Zaba (1981) and Brass and Airey (1988) all state the model should be used to correct errors in high fertility populations create the impression that the standard is applicable to all high fertility settings. In addition, the standard has been used for high fertility populations - and, on occasion, inappropriately for low fertility populations - for more than two decades. This supports the belief that the standard is appropriate for analysis in high fertility settings.

Investigations were undertaken on 61 sub-Saharan Demographic and Health Surveys in order to establish if the standard can be employed in these high fertility populations. Figure 3.9 compares the standardised age-specific fertility data from the 61 DHS schedules to the Booth standard.

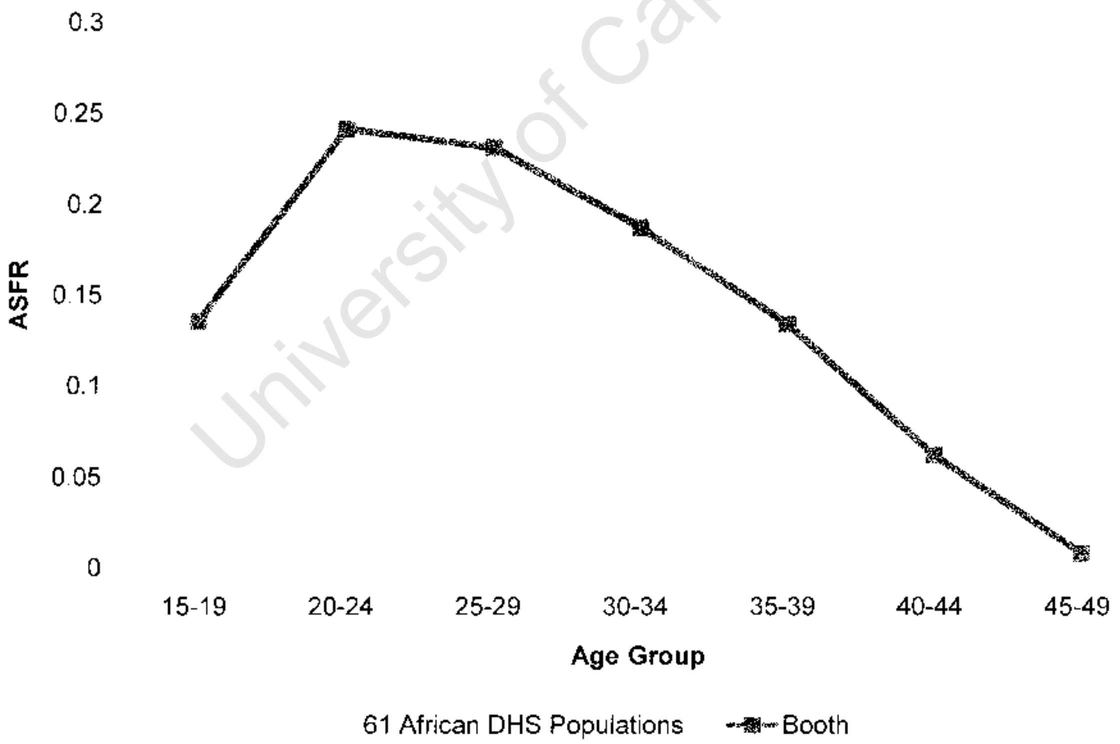

Figure 3.9: Standardised ASFR of the Booth Standard compared to 61 African DHS

From Figure 3.9 one could deduce that the standard is indeed applicable to all high fertility settings. The large number of schedules plotted in Figure 3.9, however, obscures actual differences. Also, looking at the 45-49 age group it appears that the understatement in old age fertility persists for the sub-Saharan DHS.

3.6 The African Pattern and the Booth Standard

The consistent underestimate of 45-49 fertility and the broadly similar patterns of fertility observed in Figure 3.9 suggest that a distinct African pattern exists among the sub-Sahara African DHS. In addition, deriving an African pattern will ease comparison within the region as well as with the Booth standard. Although the standard will clearly understate 45-49 fertility for the African pattern, since it understates f(45-49) for the individual DHS schedules, it will nevertheless allow a more practical quantification of this disparity.

3.6.1 Developing an African pattern

In order to investigate and develop an African pattern it is important to determine which schedules to include. The age-specific fertility rates were ranked for each of the seven age groups. For each survey a value of one was given to the age group with the highest ASFR and a value of seven to the age group with the lowest ASFR. The results of this ranking process are given in Table 3.16.

Table 3.15: Ranking of ASFR for the 78 Sub-Saharan Africa DHS

	DHS	15-19	20-24	25-29	30-34	35-39	40-44	45-49	Sum 20-24 and 25-29
1	South Africa 1998	4	2	1	3	5	6	7	3
2	Lesotho 2004	5	1	2	3	4	6	7	3
3	Zimbabwe 1999	4	1	2	3	5	6	7	3
4	Gabon 2000	4	1	2	3	5	6	7	3
5	Namibia 2000	5	2	1	3	4	6	7	3
6	Zimbabwe 1994	5	1	2	3	4	6	7	3
7	Ghana 1998	5	2	1	3	4	6	7	3
8	Ghana 2003	4	1	2	3	5	6	7	3
9	Mauritania 2000/01	4	1	2	3	5	6	7	3
10	Comoros 1996	5	1	2	3	4	6	7	3
11	Kenya 1998	5	1	2	3	4	6	7	3
12	Nigeria 1999 (1)	5	1	2	3	4	6	7	3
13	Sudan 1990	4	1	2	3	5	6	7	3

14 Cameroon 1998	4	1	2	3	5	6	7	3
15 Congo 2005	5	2	1	3	4	6	7	3
16 Eritrea 2002	5	2	1	3	4	6	7	3
17 Botswana 1988	4	1	2	3	5	6	7	3
18 Kenya 2003	3	1	2	4	5	6	7	3
19 Cameroon 2004	5	2	1	3	4	6	7	3
20 CAR 1994/95	5	1	2	3	4	6	7	3
21 Cote d'Ivoire 1998/99	5	1	2	3	4	6	7	3
22 Ghana 1993	5	1	2	3	4	6	7	3
23 Madagascar 2003/ 2004	4	2	1	3	5	6	7	3
24 Mozambique 1997	4	1	2	3	5	6	7	3
25 Togo 1998	5	1	2	3	4	6	7	3
26 Cote d'Ivoire 1994	4	1	2	3	4	6	7	3
27 Senegal 2005	5	2	1	3	4	6	7	3
28 Kenya 1993	5	1	2	3	4	6	7	3
29 Zimbabwe 1988	4	1	2	3	5	6	7	3
30 Ethiopia 2005	5	1	2	3	4	6	7	3
31 Namibia 1992	5	1	2	3	4	6	7	3
32 Guinea 1999	5	1	2	3	4	6	7	3
33 Mozambique 2003	5	1	2	3	4	6	7	3
34 Ethiopia 2000	4	1	2	3	5	6	7	3
35 Benin 2001	5	1	2	3	4	6	7	3
36 Tanzania 1999	5	2	1	3	4	6	7	3
37 Guinea 2005	5	2	1	3	4	6	7	3
38 Tanzania 2004	5	2	1	2	4	6	7	3
39 Nigeria 2003	5	2	1	3	4	6	7	3
40 Senegal 1997	5	1	2	3	4	6	7	3
41 Cameroon 1991	5	1	2	3	4	6	7	3
42 Tanzania 1996	4	1	2	3	5	6	7	3
43 Rwanda 2000	4	1	2	3	5	6	7	3
44 Burkina Faso 2003	5	1	2	3	4	6	7	3
45 Zambia 2001/02 (2)	4	1	2	3	5	6	7	3
46 Ondo State 1986	5	2	1	3	4	6	7	3
47 Benin 1996	5	2	1	3	4	6	7	3
48 Madagascar 1997	5	2	1	3	4	6	7	3
49 Malawi 2004	5	2	1	3	4	6	7	3
50 Nigeria 1990	5	1	2	3	4	6	7	3
51 Senegal 1992/93	5	1	2	3	4	6	7	3
52 Eritrea 1995	4	1	2	3	5	6	7	3
53 Madagascar 1992	5	1	2	3	4	6	7	3
54 Zambia 1996	5	1	2	3	4	6	7	3
55 Rwanda 2005	5	1	2	3	4	6	7	3

56 Tanzania 1992	4	1	2	3	5	6	7	3
57 Rwanda 1992	5	1	2	3	4	6	7	3
58 Chad 2004	4	1	2	3	5	6	7	3
59 Malawi 2000	4	1	2	3	5	6	7	3
60 Burkina Faso 1998/99	4	1	2	3	5	6	7	3
61 Chad 1996/97	5	2	1	3	4	6	7	3
62 Ghana 1988	5	3	1	2	4	6	7	4
63 Senegal 1986	6	3	1	2	4	5	7	4
64 Togo 1988	5	3	1	2	4	6	7	4
65 Burkina Faso 1992/93	6	3	1	2	4	5	7	4
66 Zambia 1992	6	3	1	2	4	5	7	4
67 Kenya 1989	5	3	1	2	4	6	7	4
68 Liberia 1986	5	3	1	2	4	6	7	4
69 Malawi 1992	5	3	1	2	4	6	7	4
70 Mali 1995/96	5	3	1	2	4	6	7	4
71 Mali 2001	5	3	1	2	4	6	7	4
72 Uganda 1995	6	3	1	2	4	5	7	4
73 Uganda 2000/01	6	3	1	2	4	5	7	4
74 Burundi 1987	6	3	1	2	4	5	7	4
75 Niger 1992	6	3	1	2	4	5	7	4
76 Mali 1987	7	3	1	2	4	5	6	4
77 Niger 1998	6	3	2	1	4	5	7	5
78 Uganda 1988	5	3	2	1	4	6	7	5

Source: Measure DHS STATCompiler

For 61 of the 78 surveys, the sum of the ranks for these two age groups (given in the last column) is equal to three. That is, fertility was highest between ages 20 and 29 years while exhibiting the same flat distribution over the remaining age ranges. For this reason these 61 schedules (78 per cent) are believed to represent sub-Saharan African fertility. By contrast, the majority of the schedules excluded are West African and, in particular, Sahelian - Ghana, Burkina Faso, Mali, Niger and Senegal.[7]

From the 61 selected schedules a mean and variance are calculated for the standardised fertility rates in each age group. These statistics are then used to identify the 11 schedules containing outliers in one or more age groups. However, only two of these have outliers in more than one age group and the exclusion of these schedules has a negligible effect on the calculated pattern of age-specific

[7] This suggests the presence of a second pattern pertaining specifically to the Sahelian region. However, this thesis focuses only on the 61 surveys that have a similar shape.

fertility. Given that no schedules are excluded the mean fertility rates for each group can be used to obtain an average age-specific fertility schedule (Table 3.17 and Figure 3.10).

Table 3.16: Booth Standard compared to African Pattern

Age Groups	Booth Standard	African Pattern
15-19	0.13584	0.12707
20-24	0.24147	0.22663
25-29	0.23130	0.22020
30-34	0.18757	0.18969
35-39	0.13401	0.13968
40-44	0.06169	0.07053
45-49	0.00812	0.02621

The schedule given in Table 3.17 also minimises the sum of squared deviations and gives the maximum likelihood estimates for each age group. As a result, it is believed to best represent the pattern of African fertility.

3.6.2 Misfit of the Booth standard to the African pattern

As expected, given that the Booth standard understated 45-49 fertility for the individual DE-IS schedules, the Booth standard also underestimates the African pattern (Figure 3.10).

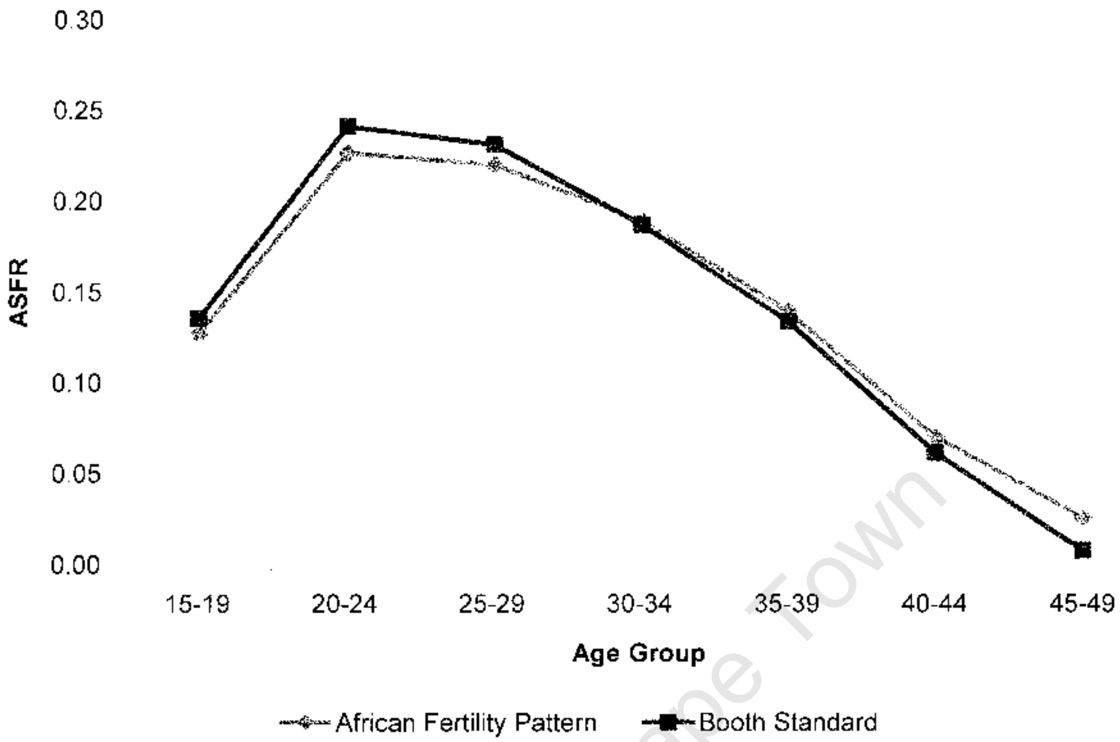

Figure 3.10: Standardised ASFR of the Booth standard compared to the African fertility pattern'

In fact, f(45-49) in the Booth standard is only 31 per cent of the average DHS fertility level. This casts serious doubts over the appropriateness of the Booth standard for sub-Saharan African settings.

Figure 3.10, combined with the results from section 3.5, demonstrates that problems can arise for both high and low fertility populations. It is observed that there are differences between the shapes of the Booth standard and the populations discussed irrespectively of the fertility level. Critically, the conclusion must then be reached that it is the *pattern* of fertility that determines the appropriateness of the standard and not the level of fertility. Although, of course, pattern and level are not separate from each other or independent. As such, it seems inappropriate to refer to a high fertility pattern and reference should instead be made to the African pattern observed for these populations.

Figure 3.9 shows that the Booth standard not only understates f(45-49) but f(40-44) is also low and lies near the bottom of the cluster of DHS schedules. Figure 3.10 shows that the African pattern has a slightly higher f(40-44) than the Booth standard. Although this difference appears small it is significant. In particular, only 13 of the 61 DI IS schedules have f(40-44) smaller than the Booth standard but 32 schedules lie below the African standard.

3.7 Data and Criteria Concerns

As a result of the misfits observed in sections 3.5.2 and 3.6.2, the standard must be further investigated to ascertain the reason for the consistent understatement. It is reasonable to begin by interrogating the data and criteria upon which the standard is based. As discussed in section 3.3, Booth selected 33 Coale-Trussell schedules as basis for the development of the standard, but recognised the need to increase fertility in the tails of the distribution. Two criteria were set to achieve this - f(10-19) > 0.15 and f(35-49) > 0.21. The former ensures that schedules with high early fertility are used to determine fertility at the youngest ages. Similarly, the latter criterion guarantees that schedules with high old age fertility are used to obtain the standard levels for the older age groups.

These two conditions are compared to the age-group specific fertility rates observed for the 61 DI-IS populations (Table 3.18).

Table 3.17: Evaluation of the Booth Criteria for inclusion in the dataset

	33 Coale-Trussell Schedules		61 DHS Populations	
	f(10-19)	f(35-49)	f(10-19)[9]	f(35-49)
Average	0.14818	0.19701	0.12706	0.23692
Minimum	0.08499	0.15864	0.08549	0.18505
Maximum	0.20747	0.23030	0.16845	0.29814

The table results suggest that the criterion f(10-19) > 0.15 is too high. This is evidenced by Booth schedules with a 17 per cent higher average and approximately 25 per cent higher maximum than the DHS schedules. This higher average level of early fertility will suppress 45-49 fertility since fertility is cumulative and must always reach its maximum by age 50 (the accepted end of the fecund period).

By contrast, the second criterion - f(35-49) > 0.21 - would appear to be too low. As can be seen in the table the maximum of the DHS populations is 29.5 per cent above the Booth equivalent. The disparity is further emphasised by the fact that even the average of the DHS schedules is higher than the maximum of the 33 Coale-Trussell schedules. Although the criterion states "greater than" - leaving the upper end open - the starting value of 0.21 clearly allows the inclusion of too many low values and, consequently, 45-49 fertility is further restricted.

[9] The DHS data does not include f(10-14) and it is assumed to be negligible with the result that f(10-14) 0.0. Consequently, f(10-19) = f(15-19) for these populations.

These results suggest that the problem lies not with the methodology Booth employed, but with the dataset upon which the standard is based. As such, it becomes important to assess the dataset used by Booth. She sets the criteria 10 < a_0 < 15, 0.1 < k < 1, 0 < m < 0.6 and SMAM[1] < 21 to select a sub-set of 33 Coale-Trussell schedules believed to capture high fertility patterns (Booth 1979, p. 49).

Exploration of the data in Sections 3.5.2 and 3.6.2 reveal that the disparity between the DHS and Booth data is most evident in the 45-49 age group. Table 3.19 compares the average, minimum and maximum values of fertility for the 45-49 age group.

Table 3.18: Comparison of the f(45-49) values

	33 Coale-Trussell Schedules	61 DHS Populations
Average	0.00810	0.02621
Minimum	0.00580	0.00949
Maximum	0.01021	0.05324

The table shows that the Coale-Trussell values are significantly lower than the equivalent measures for the 61 surveyed populations. The average and maximum of the sub-Saharan populations are respectively 220 per cent and 420 per cent higher than the equivalents for the data used by Booth. Critically, the maximum of the 33 Coale-Trussell schedules - f(45-49) = 0.01021 - is barely higher than the minimum for the DHS populations - f(45-49) = 0.00949.

Figure 3.11, similarly, illustrates the disparity between the fertility rates of the 61 DHS surveys and the Booth standard for the 45-49 age group.

[1] SMANI is a frequently used abbreviation for the Singulate Mean Age at Marriage.

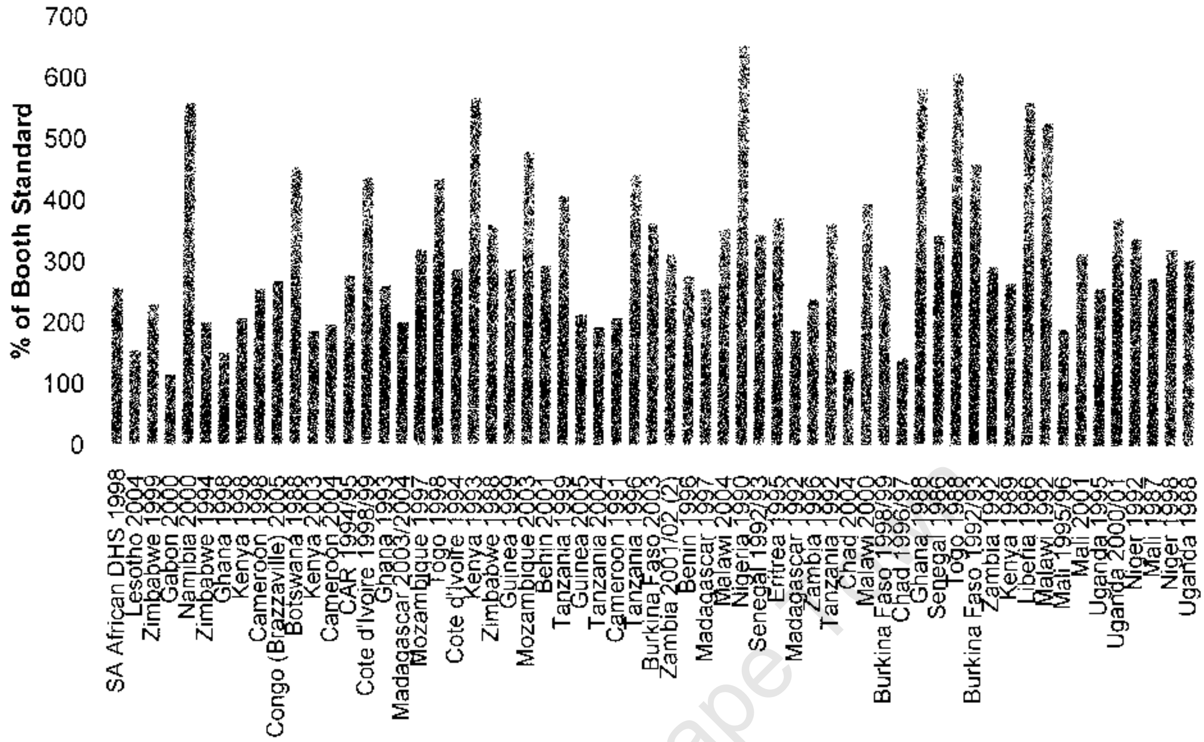

Figure 3.11: Ratio of the 45-49 fertility of the sub-Saharan DHS to the Booth Standard

Having removed the level effect - by standardising the data to a TFR of one - inspection of Figure 3.11 shows that the difference between the Booth standard and sub-Saharan fertility rates for the 45-49 age group ranges from 17 per cent (Gabon 2000) to 555 per cent (Nigeria 1990) with an average difference of 224 per cent.

This supports the finding that it is the *pattern* of fertility - and not the fertility level - that determines if a standard is appropriate. In addition, from Tables 3.18 and 3.19 as well as Figures 3.10 and 3.11, one must conclude that the Booth standard does not and cannot capture the effect of old age fertility for these African populations.

4 ALTERNATIVES TO THE BOOTH STANDARD

Chapter 3 showed that the Booth standard is not appropriate for use in the analysis of sub-Saharan Africa populations and that the data used by Booth cannot be employed to develop an alternative standard for African settings. However, the Coale-Trussell model cannot be summarily rejected as a data source based on evidence from only 33 schedules. As a result, the Coale-Trussell model must be reinvestigated to establish if an alternative data set can be found that will yield a standard fertility pattern more appropriate to the African DHS populations.

4.1 Automating the Coale-Trussell Model

Booth (1979) reworks the original Coale-Trussell model, written in FORTRAN, to include two more ages at the bottom end in order to fully capture high fertility patterns. This meant extending *v(a)* and *n(a)* by two more values and Booth uses the logical (linear) extensions: v(10) = 0.0, v(11) = 0.0, n(10) = 0.005 and n(11) = 0.100 (Booth 1979, p. 232).

In order to assess the appropriateness of the Coale-Trussell model, Booth's extended version was reprogrammed using MS Excel. The problem is that this model only produces one schedule at a time and identifying sufficient schedules to analyse takes considerable time and effort. To avoid this, the coding has been automated using the macro functionality in MS Excel" and the constraints Booth places on the three input parameters (a_0, k and m) are dropped to ensure the capture of a broader, more inclusive range of fertility distributions.

To this end, the value of a_0 is allowed to vary within the range [9.25, 19] in increments of 0.25 years. Similarly, the parameters k and m are both allowed to take values between 0.05 and 2 (inclusive) in intervals of 0.05. This yields 40 categories for each parameter and results in a total of 64000 (40x40x40) schedules.

Each of these schedules includes a 10-14 age category that is not included in the DI IS which starts at the 15-19 age group. In order to accurately compare the model schedules with the observed DHS populations this irregularity must be addressed. There are two possible ways of dealing with this. First, the fertility for the 10-14 age group can be proportionately reallocated to all other age groups. The alternative method - adopted here - is to include the 10-14 fertility in the 15-19 fertility. It is intuitively reasonable, since f(10-14) is part of early fertility, to include it with other early

" The automated MS Excel code is included in Appendix C.

fertility. However, since f(10-14) is small, the effect of its inclusion in f(15-19) is negligible as regards the shape of fertility for these schedules. After this alteration the model schedules and the DHS populations both apply to the same age groups (15-19, 20-24, ..., 45-49) and can be compared sensibly.

4.2 Comparing the Coale-Trussell Model and DHS Populations

The comparison begins by applying the procedure, discussed in section 3.5, for ranking the age-group specific fertility rates. It was observed that for 61 of the 78 DHS for sub-Saharan Africa populations the majority of fertility occurs in the 20-24 and 25-29 age groups. Put differently, if the age-group specific fertility rates are ranked in decreasing order, then the sum of the ranks for f(20-24) and f(25-29) equal three.

This methodology is extended in an analogous manner to the age-specific fertility rates of the 64000 schedules produced using the automated version of the model. The ranks of the 20-24 and 25-29 age groups are added and the totals recorded. Only schedules with a total sum of ranks of three are included for analysis resulting in a subset of 23864 schedules. These schedules form the complete set that can possibly be used to establish a new standard for sub-Saharan populations.

Investigation of these schedules shows that the understatement of f(45-49) observed for the Booth schedules is not an artefact of the selection criteria. Instead, this underestimate exists for all 23864 Coale-Trussell schedules. In fact, the subset maximum of f(45-49) = 0.0135 is only 52 per cent of the average level observed for the 61 DHS populations - f(45-40) = 0.0262.

Furthermore, the schedule with the highest 45-49 fertility does not reflect the pattern of the African fertility pattern (Figure 4.1). By contrast, the Coale-Trussell schedule defined by the parameters *a,* = 12.5, *k* = 0.45 and m = 0.15 provides the closest fit to the African pattern based on minimising the sum of squared error.

However, as seen in Figure 4.1, despite the remarkably close fit over the ages 15 to 44 a dramatic underestimate is observed for the 45-49 age group. In fact, the best-fitting schedule understates 45-49 fertility by more than 150 per cent.

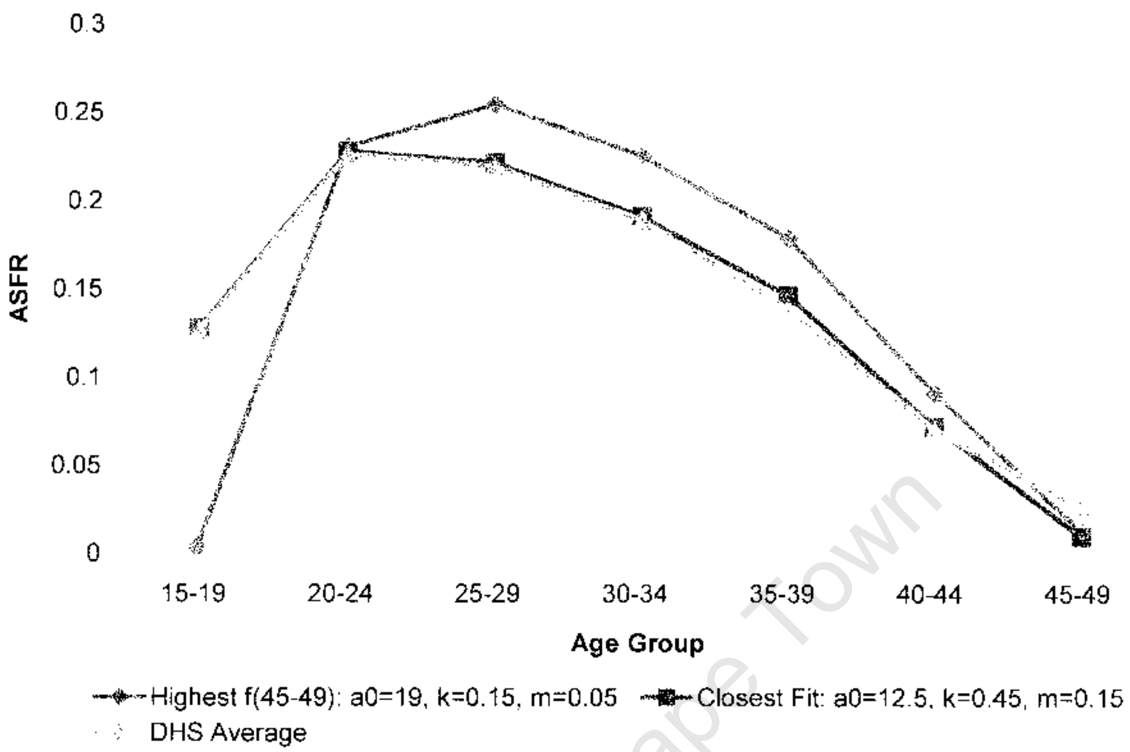

Figure 4.1: Comparison of average DHS fertility with two Coale-Trussell populations

4.3 Coale-Trussell Model Reworked

The differences observed in sections 3.5 and 4.1 necessitate a reassessment of the Coale-Trussell formulation. However, as discussed in section 2.5, Xie (1990) and Xie and Pimentel (1992) already reformulate the Coale-Trussell model in an attempt to maintain its relevance in the face of mounting criticism. Xie (1990) focuses on adjusting the vector *n(a)* whilst Xie and Pimentel (1992) modify *v(a)*.

4.3.1 Coale-Trussell using Xie n(a)

Xie (1990) derived revised values of *n(a)* and these values - for the age groups 20-24, 25-29, …, 40-44 and 45-49 - are given in Table 4.1.

Table 4.1: Standardized n(a) values for the Xie Independence and Coale-Trussell models

Age Group	20-24	25-29	30-34	35-39	40-44	45-49
Xie n(a)	0.460	0.436	0.392	0.333	0.199	0.043
Coale-Trussell n(a)	0.460	0.431	0.395	0.322	0.167	0.024

Obtained from Table 2 by Xie (1990, p 660)

As can be seen, the differences between the n(a)-values derived by Xie and Coale-Trussell values are only large in the last age group. This increase suggests that fertility modelled using the revised *n(a)* will have higher fertility for the 45-49 age group. In particular, it must be tested if this increase will result in a sufficient increase in f(45-49) to adequately model the African pattern.

However, to test this the n(a)-values need to be extended to include the 10-14 and 15-19 age groups. Examination of the sub-Saharan DHS population fertility rates suggests a value for the 15-19 age group of 57 per cent of the ASFR for the 20-24 is reasonable. By contrast, the fertility for the 10-14 age range is small and a value of 0.05 is used.

Since the Coale-Trussell model requires age-specific values of *n(a)* the values derived for the age groups must be separated into single year values. This extension to single years was done using Beers' formula (Shryock and Siegel 1976) and results in the *n(a)* values given in Table 4.2 and figure 4.2.

Table 4.2: Xie single-year n(a) resulting from Beers

Age	Xie n(a) after applying Beers	Age	Xie n(a) after applying Beers
10	0.000	31	0.404
11	0.006	32	0.395
12	0.049	33	0.385
13	0.094	34	0.374
14	0.140	35	0.361
15	0.185	36	0.346
16	0.232	37	0.329
17	0.278	38	0.308
18	0.323	39	0.285
19	0.365	40	0.258
20	0.401	41	0.228
21	0.431	42	0.195
22	0.452	43	0.162
23	0.464	44	0.128
24	0.468	45	0.097
25	0.464	46	0.069
26	0.456	47	0.045
27	0.445	48	0.026
28	0.434	49	0.013
29	0.423	50	0.007
30	0.413		

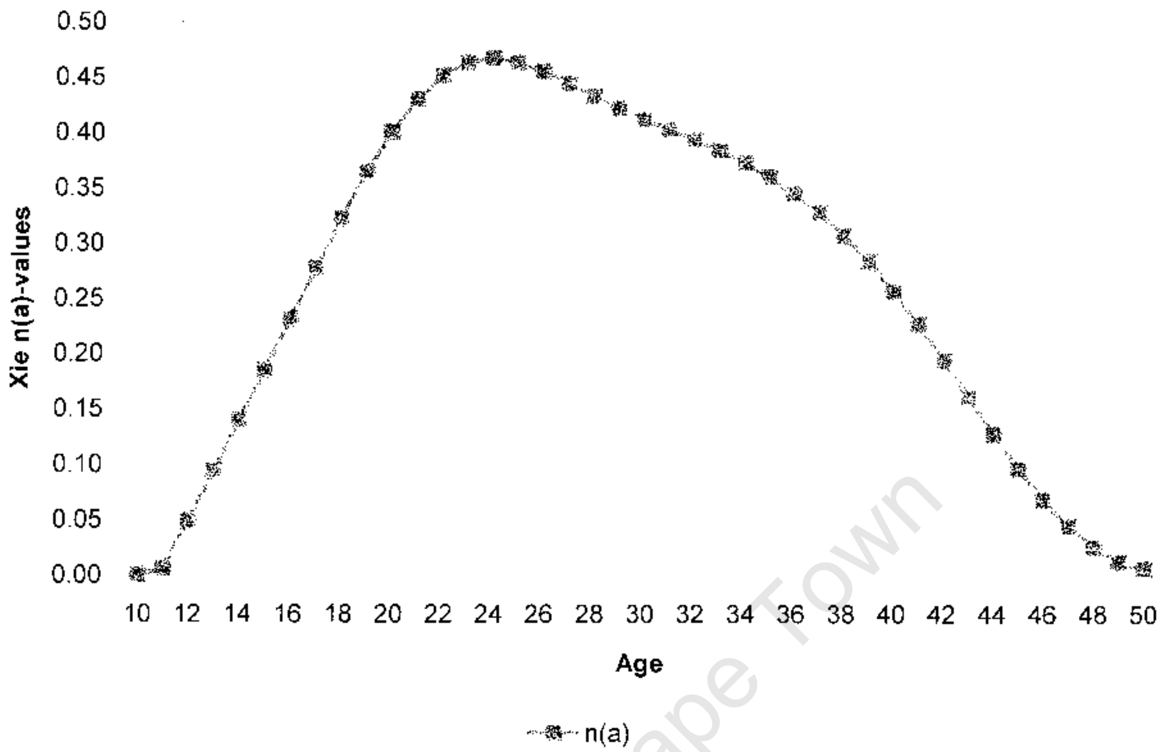

Figure 4.2: Natural fertility schedule, n(a), as derived by Xie (1990)

Next, the new *n(a)* standard - listed in Table 4.2 - is included in the automated Coale-Trussell model. Using the same process as before establishes an alternative dataset of 64000 schedules based on these revised values. Although the resulting schedules are unlikely to be too different the hope is that the schedules will be sufficiently different in the last age group to overcome the misfit observed for the original Coale-Trussell model.

Following the same procedure as before the age-group specific fertility rates are ranked from large to small and the ranks are summed. Of the 64000 schedules 27144 have a sum of three for the ranks of the 20-24 and 25-29 age groups.

Not only do more schedules meet the sum of ranks criterion but, as expected given the higher n(45-49), f(45-49) is also higher for the second model than in the original formulation. Inspection of the new subset shows that the maximum ASFR in the oldest age category has increased by 32 per cent from f(45-49) = 0.0135 to f(45-49) 0.0178 . However, this is still only about 68 per cent of the average African level - f(45-49) = 0.0262.

In addition, as before, the population with the highest 45-49 fertility rate (given by the parameters a_0 = 18.5, k = 0.2 and m = 0.05) does not reflect the African pattern. A better overall fit can be achieved by an alternative parameterisation (a_0 = 10.5, k = 0.55 and m = 0.15) but with a disparity of 90 per cent for the oldest age group (Figure 4.3).

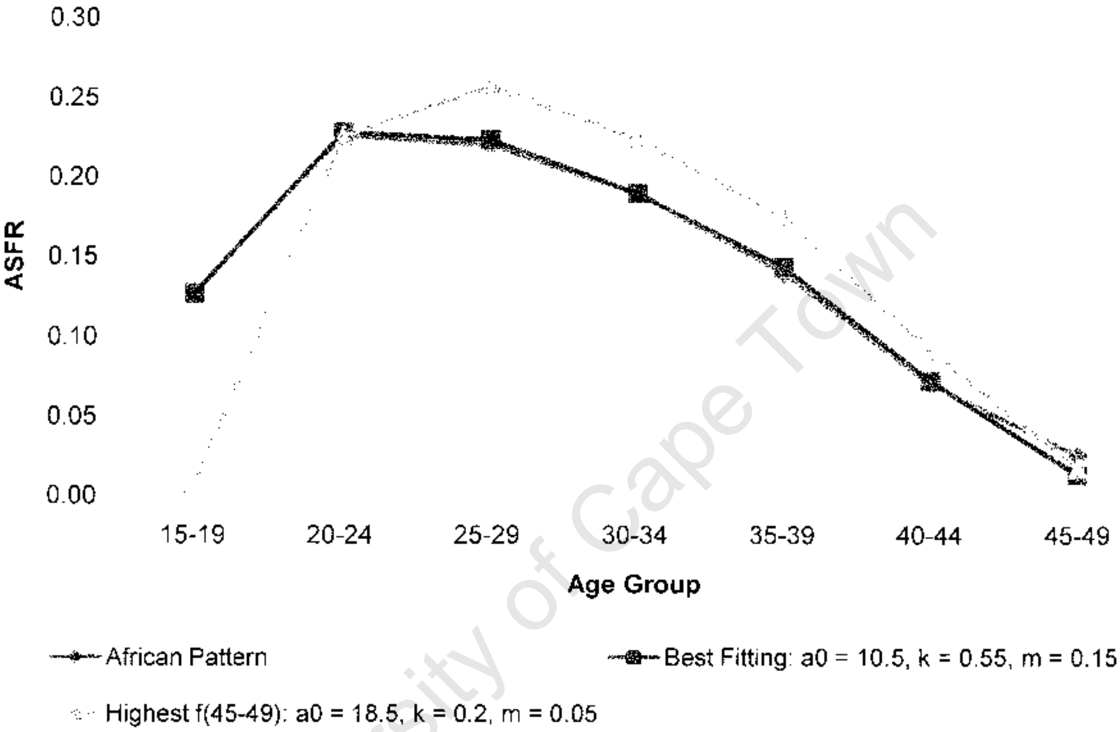

Figure 4.3: Comparison of the African pattern with two schedules using changed n(a)

Consequently, the modified Coale-Trussell model still cannot accurately model the African fertility pattern and the problem clearly persists for the oldest age group despite the increases recorded in f(45-49).

4.3.2 Coale-Trussell with alternative v(a)

The improvement in old age fertility is encouraging since the second rework by Xie and Pimentel (1992) - discussed in section 2.5.2 - is also expected to increase f(45-49). The reformulation by Xie and Pimentel tries to rescue the Coale-Trussell model by restating it as a statistical method and deriving a new series of v(a)-values. Table 4.3 compares the v(a)-values derived by Xie and Pimentel (1992) with the original *v(a)* determined by Coale and Trussell (1974).

Table 4.3: Standardized v(a) values for the Coale-Trussell and Xie-Pimentel models

Age Group	20-24	25-29	30-34	35-39	40-44	45-49
Xie-Pimentel v(a)	0	-0.329	-0.713	-1.194	-1.671	-1.082
Coale-Trussell v(a)	0	-0.279	-0.677	-1.042	-1.414	-1.671

By contrast to the original v(a)-values, the *v(a)* schedule derived by Xie and Pimentel decreases over the age range 25-44, but increases again for 45-49. The upshot of the increase in *v(a)* for the 45-49 age group is that fertility schedules derived using these *v(a)* should show higher 45-49 fertility since the reduction effect - fertility control - is smaller.

To test this, the Coale-Trussell model utilized in section 4.2.1 is extended further to include the new estimates of *v(a)*. This is achieved by, first, converting the v(a)-values listed in Table 4.3 to single year values[12] (Table 4.4 and Figure 4.4) and, second, replacing the original *v(a)* in the Coale-Trussell model with the revised values.

Table 4.4: Xie and Pimentel single-year v(a) resulting from Beers

Age	Xie and Pimentel v(a)	Age	Xie and Pimentel v(a)
10	0	31	-0.629
11	0	32	-0.715
12	0	33	-0.805
13	0	34	-0.898
14	0	35	-1.001
15	0	36	-1.115
16	0	37	-1.235
17	0	38	-1.356
18	0	39	-1.470
19	0	40	-1.556
20	0	41	-1.601
21	0	42	-1.602
22	-0.012	43	-1.557
23	-0.051	44	-1.473
24	-0.103	45	-1.365
25	-0.166	46	-1.246
26	-0.238	47	-1.133
27	-0.313	48	-1.036
28	-0.391	49	-0.968
29	-0.469	50	-0.937
30	-0.548		

[12] This is again done using Beers' formula and setting v(a) to 0.0 for ages 20 and below.

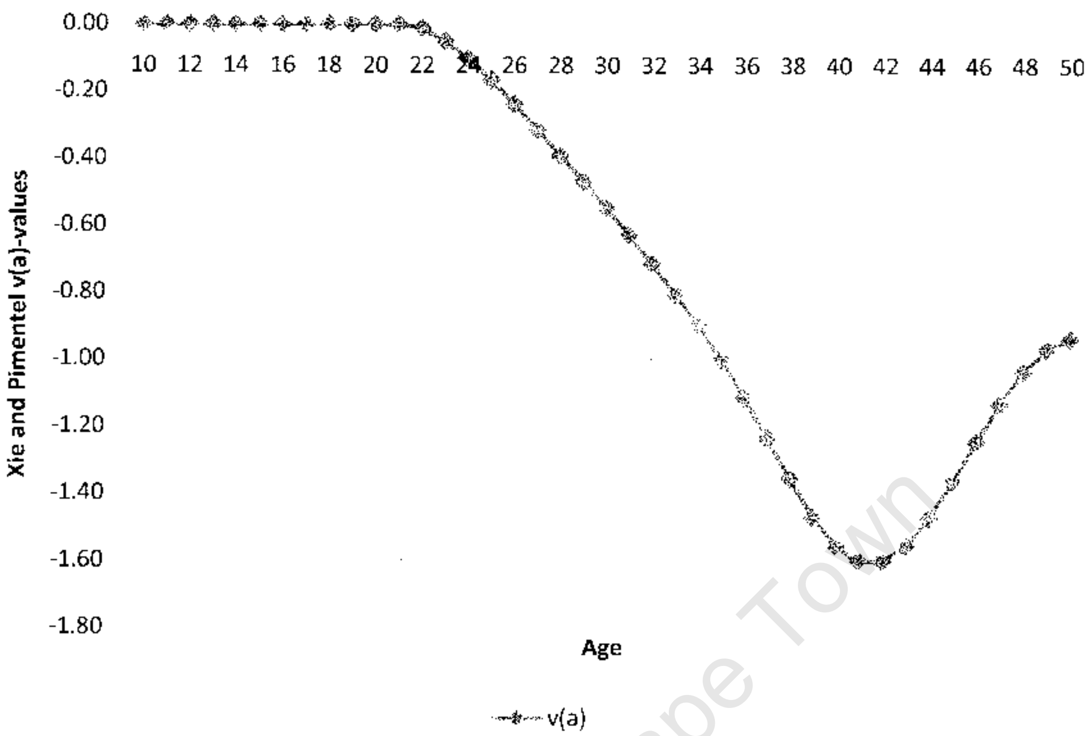

Figure 4.4: Schedule of fertility control, v(a), derived by Xie and Pimentel (1992)

This latest revision incorporates both the revised *n(a)* and *v(a)* and produces another dataset of 64000 schedules. As before, the sum of ranks criterion is used to identify a subset of schedules most likely to be appropriate for African populations. Using the sum of ranks constraint reduces the dataset to 29570 schedules for analysis.

Examination shows that these schedules have a maximum f(45-49) of approximately 0.0195 obtained for the schedule given by $a,= 11$, $k = 2$ and m = 1.3. This is a further increase of 9.6 per cent over the maximum recorded in section 4.2.1. Despite the improvement of almost 44 per cent over the original 45-49 fertility, this maximum is about 35 per cent below the equivalent African level and the problem clearly persists. In addition, this schedule - like the schedules in sections 4.2 and 4.3.1 - must be rejected for the African populations since it does not sufficiently resemble the African fertility pattern (Figure 4.5).

As with the two models discussed in sections 4.1 and 4.2.1, a better overall fit to the African pattern can be found with an alternative parameterisation. The schedule with the best overall fit (i.e. the lowest sum of squared error) is defined by the parameters $a, = 9.5$, $k = 0.65$ and *in* = 0.15.

However, this schedule reports age-group specific fertility for the 45-49 age group of 0.0151 which is a 72 per cent understatement relative to the average sub-Saharan DHS level.

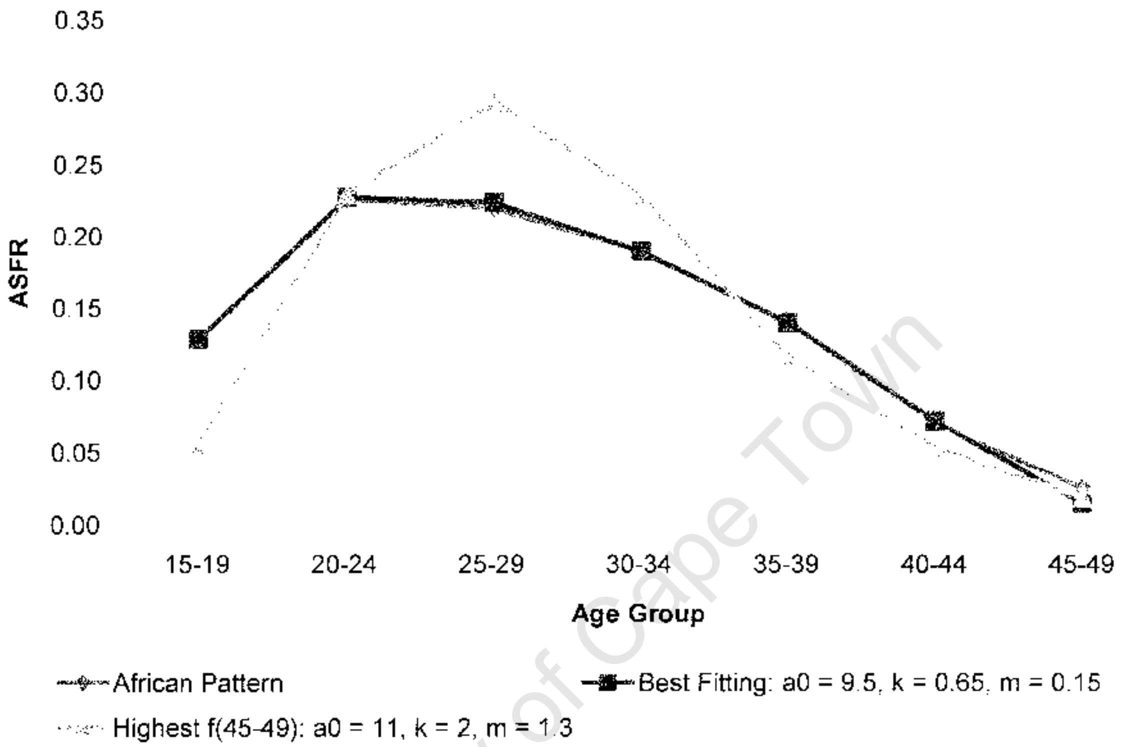

Figure 4.5: Comparison of the African pattern with two schedules using changed v(a)

Despite the improvements by Xie (1990) and Xie and Pimentel (1992) the model continues to understate fertility in the oldest age group. Although successfully adjusting *r(a)* and increasing fertility in the oldest age group, even the best-fitting schedule understates f(45-49) by more than 70 per cent.

4.4 Problems with the Coale-Trussell Model

The systematic underestimation of the 45-49 fertility observed in all three formulations of Coale-Trussell model indicates problems underlying the model. As discussed in section 2.4, the model calculates age-specific fertility rates using Equation 4.1.

$$f(a) = G(a) * r(a) \qquad \textbf{Equation 4.1}$$

That is, fertility at age a is the product of the marital fertility, $r(a)$, and the proportion of women married by that age, $G(a)$. However, this formulation assumes that fertility occurs mostly within marriage and that the proportion ever-married is high. Critically, Coale and Trussell state that problems can be expected "when nuptiality is changing rapidly" suggesting that the method is highly dependent on nuptiality (Coale and Trussell 1974, p. 193).

4.4.1 Emphasis on marital fertility

The initial emphasis on marital fertility was appropriate because marriage acted as a proxy for entrance into sexual activity and exposure to fertility. However, since the 1960s there has been an increase in non-marital fertility. In particular, Anderson and Silver (1992) and Hinde (1998) report high non-marital fertility in the United Kingdom (30 per cent reported by Hinde) and the rest of Europe. In addition, Tomasson finds that 54 per cent of Swedish fertility occurred out-of-marriage in 1996 (Tomasson 1998). These results suggest that marriage is no longer a good indicator of fertility (Anderson and Silver 1998; Tomasson 1998; Department of Health 2007).

Hinde argues that the legal definition of marriage is too narrow and that fertility analysis should rather consider cohabitation (Hinde 1998, p. 122). This reformulation appears to be strengthened by the results obtained by Tomasson (1998) which shows that single-parenting is much lower (below 16 per cent) than non-marital fertility demonstrating a shift in attitudes towards marriage and the timing of marriage. Regarding Africa, Budlender, Chobokoane and Simelane (2004) show that marriage in South Africa is not clear owing to differences in definition between cultures and religions. Similar differences can be expected in many other African countries since the same problems - as described by Budlender, Chobokoane and Simelane (2004) - exist for these populations.

In terms of the Coale-Trussell method, the observed increase in illegitimate births, difficulty in defining marriage and changing attitudes towards marriage violate the core assumptions of the method. This necessitates the re-evaluation of the model formulation and the functions $G(a)$ and $r(a)$.

One alternative is to build on Hinde's reasoning and possibly recast $G(a)$ and $r(a)$ in terms of sexual exposure rather than marriage. Although outside the scope of the current project - which emphasises the use of alternative methods to the Coale-Trussell model for analysis of African data - this will require a reassessment of both $r(a)$ and the Coale-McNeil function $G(a)$.

Conversely, it can be argued that people are still marrying but that this occurs later - as implied by the results from Tomasson (1998) - yet they are still exposed to the same risk of fertility through sexual activity (possibly while cohabiting or in stable sexual relationships). Using this reasoning the progression into "stable sexual union" or "sexual activity" could conceivably follow the same pattern as was previously used for marriage - the Coale-McNeil *G(a)*.

With such a large portion of fertility occurring out of marriage one can, further, reason that fertility occurs at the desired time. That is, the timing of births are no different to what it would have been had the persons been married. Again, although outside the scope of this project, this hypothesis should be tested. However, if the hypothesis is accepted, then one can conclude that the same control as before, *v(a)*, is exercised to restrict fertility below natural levels, *n(a)*. As a result, the Coale-Trussell model can be used unchanged except for the interpretation of the functions *G(a)* and *r(a)* and vectors *n(a)* and *v(a)*. This hypothesis, then, would imply that the Coale-Trussell method can be used to model fertility and develop a fertility standard.

However, the fundamental problem with the continued use of the Coale-Trussell method is that it does not address the misfit observed between the model and empirical data. In particular, *r(a)* has already been altered by Xie (1990) and Xie and Pimentel (1992), through changes to *n(a)* and *v(a)*, with little impact on the underestimate of f(45-49). This necessitates an investigation of the function *G(a)*.

4.4.2 Small values of k

According to Coale and Trussell (1974) the function *G(a)* depends on two variables a_0 and k. As explained in section 2.4, a_0 is the starting age and k represents the deviation in the pace of first marriage from that of the 19th century Swedish population used as a standard.

In particular, a value of k = 0.5 is equivalent to stating that 50 per cent - of those who will eventually marry - do so within 5 years after the starting age. Similarly, k = 0.3 and k = 0.1 imply 50 per cent are married within 3 years and 1 year, respectively.

This is of concern since, to obtain a reasonable fit over the majority of the age intervals, the value of k must be small. For example 23 of the 33 schedules identified by Booth as having high fertility patterns report k .5 0.5. Of these, 12 have k 5 0.3 with two schedules giving k = 0.1. Likewise, the best-fitting schedules from the three models discussed in sections 4.1 and 4.2 give two k-values of 0.45 and one of 0.65.

The problem is illustrated graphically in Figure 4.2 where the small k-values result in a sharp increase of marriage rates. In particular, the figure shows a rapid progression to marriage as demonstrated by schedule 28 from Booth (defined by parameters a_0 = 14.5, k = 0.1 and in = 0.2). This schedule goes from 0 per cent married at age 14 to more than 90 per cent married at age 16. The majority of the other schedules Booth identified experience similarly quick marital rates.

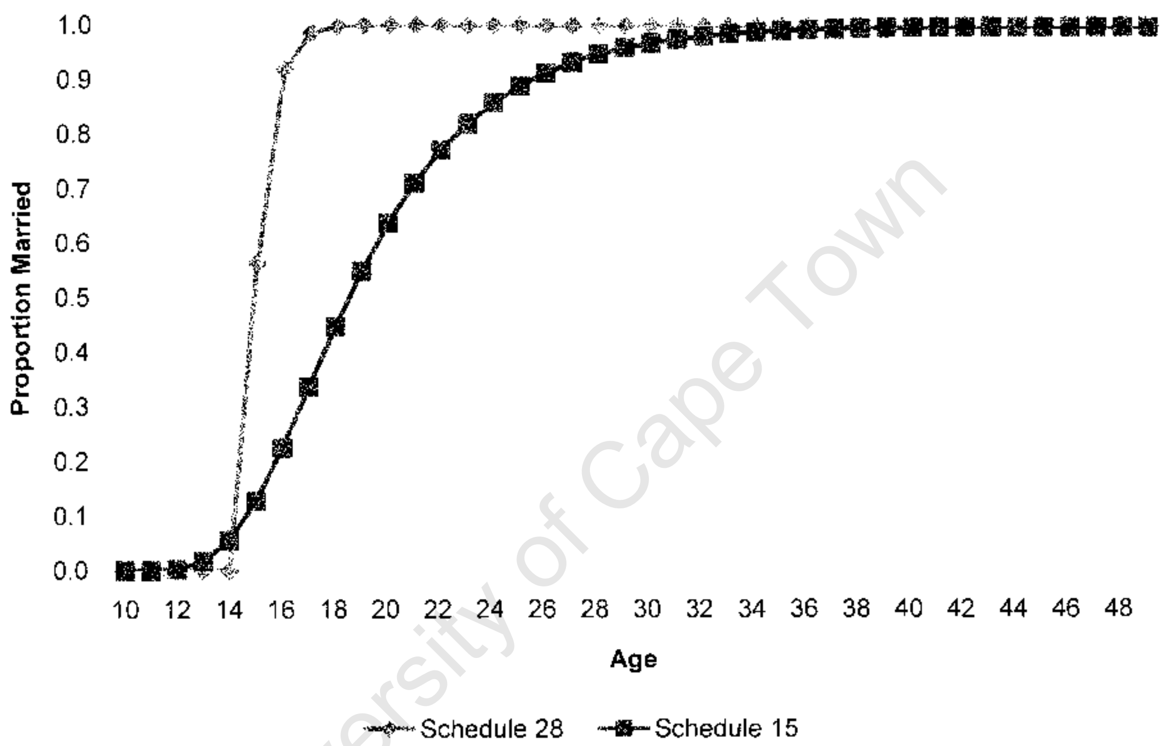

Figure 4.6: Proportion married by age, G(a), for schedules 15 and 28 identified by Booth

4.4.3 Shape of *G(a)*

Figure 4.6 shows the high early marriage rates that result from small values of k, but also reflects the S-shape of *G(a)*. Figure 4.7 further demonstrates the differences in shape of the nuptiality curve, *G(a)*, for different levels of k for a_0 = 10.

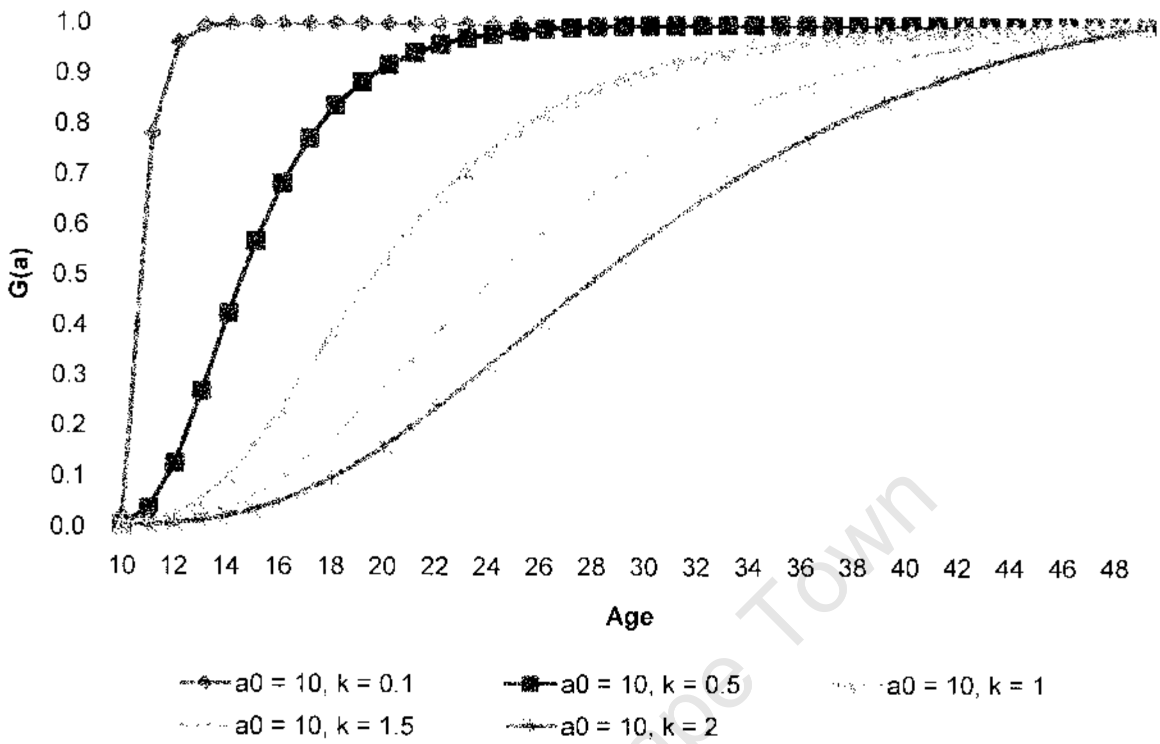

Figure 4.7: Proportion married by age, G(a), for different levels of k holding a_0 constant

The S-shape of *G(a)* is clear even for the highest values of *k* and this can be explained using Equations 4.2 and 4.3.

$$g(a) = \frac{0.19465}{k} \exp[-0.174W - \exp(-0.2881W)]$$ **Equation 4.2**

where

$$W = \frac{a - a_0 - 6.06k}{k}$$ **Equation 4.3**

Since *k* and a_0 are positive constants, the value of *g(a)* increases up to a maximum at age $a = 7.8103 * k + a_0$ (Appendix D). This translates into an increase in *G(a)* at an increasing rate until *g(a)* reaches its maximum. After *g(a)* reaches its maximum it begins to decrease and the rate of increase in *G(a)* slows down giving rise to the observed S-shape.

The asymptotic approach to 50 resulting from the S-shape, combined with the rapid progression to marriage, makes it virtually impossible to achieve the appropriate level of 45-49 fertility. The reason is that given the cumulative nature of fertility the inflated early fertility will suppress later fertility. In addition, the gradual approach to the maximum occurs in small, decreasing increments further suppressing 45-49 fertility.

4.5 Alternatives to the Coale-Trussell Model

The misfit of 45-49 fertility observed in section 4.2 and the problems leading to these underestimates (discussed in section 4.3) necessitate an analysis of alternative methods of representing sub-Saharan Africa data, particularly if a different standard is to be developed.

Hoem et al. (1981) analyse Danish fertility data using a number of different fertility distributions. They conclude that the cubic spline provides the best fit and that the Brass polynomial is less accurate. They also find that the Hadwiger function, Gamma density and Coale-Trussell model produce equally good results for human populations and deem these functions joint second best (Hoem, et al. 1981). As a result, the Gamma density is dismissed for the purposes of analysing the sub-Saharan populations since it provides no additional benefit over the Coale-Trussell model and Hadwiger function.

However, section 4.2 has already shown that the Coale-Trussell model is not appropriate for the sub-Saharan populations being considered. Gage also dismisses cubic splines for mammalian populations since it "requires good underlying empirical data" (Gage 2001, p.490). Hence, by the same argument cubic splines are dismissed for the sub-Saharan populations due to the enduring problems around data quality.

Gage further shows that the Brass polynomial cannot be rejected in favour of other, more complex, methods (Gage 2001). Consequently, both the Brass polynomial and Hadwiger functions are investigated for the usefulness in measuring and capturing African fertility patterns. These functions will then be evaluated against each other using graduation methods.

4.5.1 Brass polynomial

As shown in Section 2.6.1 and Appendix E, the Brass polynomial can be generalised and simplified to express the age-specific cumulative fertility rates in terms of the starting age, s, and the length of the fertility period, iv.

In order to determine the appropriateness of the Brass polynomial to an African pattern this generalised equation was applied to the DHS data. As before, a macro in MS Excel 2007 was used to automate the production of these schedules. The parameters were allowed to vary such that 9 < s < 21 and 22 5_ w < 52. This produced 36000 schedules for analysis and the resultant age-specific fertility rates are ranked using the same procedure as before. Investigation shows that 11055 meet the criterion that the sum of the ranks for the 20-24 and 25-29 age groups equals three. The average DHS fertility rates are then compared to the approximately 11000 remaining schedules by minimising the sum of squared differences (SSE).

The best fit is achieved by the schedule with parameters s = 13.2 and w = 38.9 (SSE = 0.00034). However, like the Booth standard and the Coale-Trussell models discussed above, this schedule still understates 45-49 fertility by almost 27 per cent. Although this is an improvement, a closer match to f(45-49) is achieved with the parameters s = 13 and n) = 39.8 while maintaining good overall fit (SSE = 0.00040). Table 4.5 compares the African pattern to the latter model schedule.

Table 4.5: Model schedule using the Brass polynomial

x	African f(x,x+4)	Model Schedule f(x,x+4)	Percentage Error
15-19	0.12707	0.13285	4.5%
20-24	0.22663	0.20925	-7.7%
25-29	0.22020	0.22360	1.5%
30-34	0.18969	0.19410	2.3%
35-39	0.13968	0.13899	-0.5%
40-44	0.07053	0.07646	8.4%
45-49	0.02621	0.02475	-5.6%

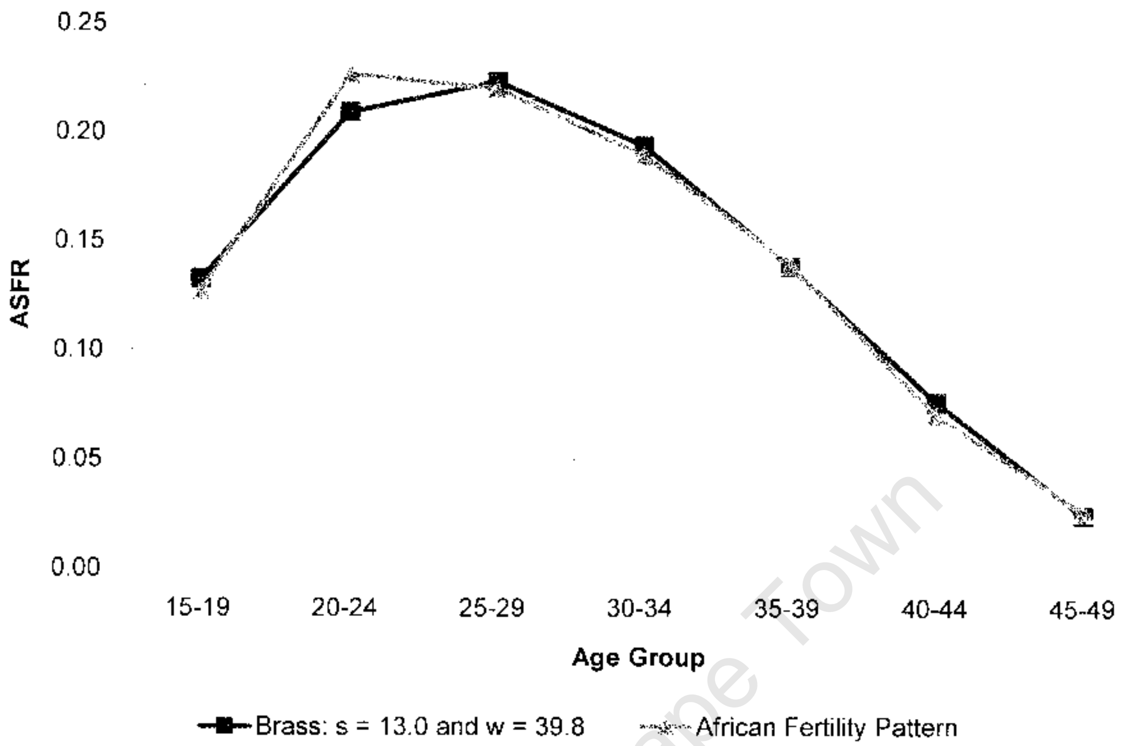

Figure 4.8: Comparison of ASFR for the African pattern and Brass model schedule

Using the Brass polynomial has reduced the error in the fertility for the oldest age group to 5.6 per cent. This is a significant improvement over the Coale-Trussell schedules which showed under-estimates of 151, 90 and 72 per cent, respectively, for the original, Xie and Xie and Pimentel formulations discussed earlier.

4.5.2 Hadwiger function

As discussed in section 2.6.2, Hoem et al. (1981) and Gage (2001) indicate that a reasonable alternative to the Coale-Trussell model may be the Hadwiger function. Despite having two parameters, *b* and *c*, with no clear demographic interpretation and being more complex than the Brass polynomial, the Hadwiger function must be considered. The reason for this is that both Hoem et al. and Gage show that the Hadwiger function consistently provides higher estimates of 45-49 fertility than the Coale-Trussell model.

As with the Brass polynomial and Coale-Trussell models a procedure was set up in MS Excel to automate the production of fertility schedules using the Hadwiger function. Since the investigation is concerned with pattern rather than level the total fertility parameter, *a*, is given a value of one. The remaining parameters — *b, c, s* and *u* — were allowed to take values within predefined ranges: 1.7 *b* **5** 3.4 with step-size 0.1, 23 *c* 5- 34 in steps of 0.25, 47 **5** *u* **5** 53 where u takes only integer values and 11 **s 5** 17 at half-year ages. However, analysis showed that no additional benefit is gained by including half-year ages for s or incrementing *c* by 0.25.

As a result, the parameter *c* was set to increase by steps of 0.5 and starting age, s, takes on integer ages. This process results in 20286 schedules to be compared with the sub-Saharan Africa experience and the average fertility derived from the 61 DHS populations. Of these schedules, 10731 satisfied the ranking criterion that fertility must be highest between ages 20 and 30.

A number of Hadwiger schedules may be used to describe the average DHS fertility pattern. Some of these schedules slightly overstate and some understate 45-49 fertility. However, the best fitting schedule is defined by the parameters *b* = 1.9, *c* = 31 and 17c x 5_ 47. That is, x starts at age *s* = 17 and ends at the maximum *u* = 47. Table 4.6 and Figure 4.5 compare this model schedule to the DHS average age-group specific fertility rates.

Table 4.6: Model schedule using the Hadwiger function

x	DHS Average f(x,x+4)	Model Schedule f(x,x+4)	Percentage Error
15-19	0.12707	0.13020	2.5%
20-24	0.22663	0.22570	-0.4%
25-29	0.22020	0.22538	2.4%
30-34	0.18969	0.18143	-4.4%
35-39	0.13968	0.12872	-7.8%
40-44	0.07053	0.08440	19.7%
45-49	0.02621	0.02417	-7.8%

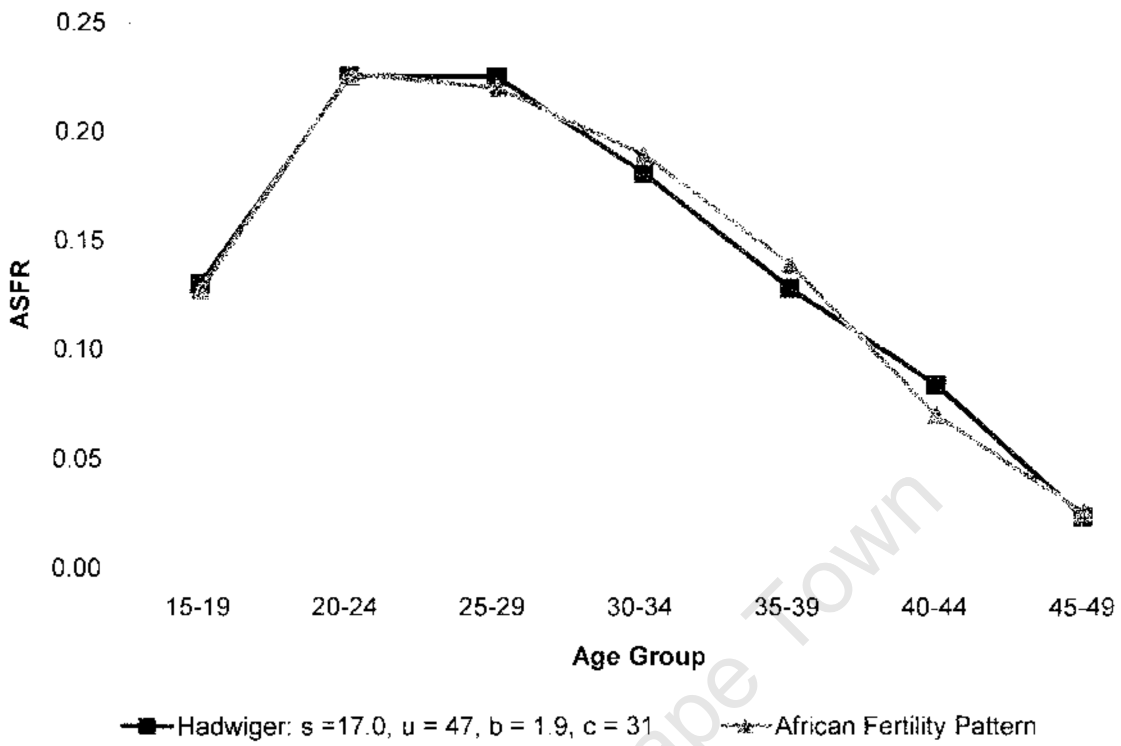

Figure 4.9: ASFR of African pattern compared to Hadwiger model schedule

As with the Brass polynomial, there is an improvement in the estimates of f(45-49) on those obtained from the Coale-Trussell schedules discussed in sections 4.1 and 4.2. The Hadwiger-based standard understates 45-49 fertility by 7.8 per cent and has SSE 0.00042.

4.6 Comparing the Brass Polynomial and Hadwiger Function

Both the Brass polynomial and the Hadwiger function show improvements over the Coale-Trussell based models. Seeing as both methods yield reasonable results a decision must be made on which method to use.

The Brass polynomial has two mathematical advantages over the Hadwiger function. First, the polynomial requires one fewer parameter and the parameters have clear demographic interpretations - cis a measure of TFR whilst s is the starting age and w the length of the fertility period.

Second, the Brass cumulative fertility function can be simplified using the properties of integrals. This allows the direct calculation of cumulative and age-specific fertility rates once s and w are set. By contrast, the Hadwiger function cannot be explicitly evaluated and numerical methods or statistical packages are required to obtain the cumulative and age-specific fertility rates.

Although both these arguments favour the Brass polynomial neither is sufficient to dismiss the Hadwiger function. As such, graduation tests must be conducted on both standards to consider both smoothness and goodness-of-fit to the African fertility pattern (Appendix F). The results of the goodness-of-fit tests are summarised in Table 4.7.

Table 4.7: Comparison of Brass and Hadwiger standards with the African fertility pattern

Age Group	African ASFR	Hadwiger standard	Signs of Differences	Brass standard	Signs of Differences
15-19	0.12707	0.13020	+	0.13285	+
20-24	0.22663	0.22570	-	0.20925	-
25-29	0.22020	0.22538	+	0.22360	+
30-34	0.18969	0.18143	-	0.19410	+
35-39	0.13968	0.12872	-	0.13899	-
40-44	0.07053	0.08440	+	0.07646	+
45-49	0.02621	0.02417	-	0.02475	-
SSE (15-49)		0.00042		0.00040	
SSE (35-49)		0.00032		0.00004	

First, both tests show good overall fit to the data as a result of minimising the squared difference terms. The Brass polynomial has lower SSE than the Hadwiger function over the entire age range. The second test is for consistent over- or underestimation identified by excessive numbers of deviations with the same sign. For both standards there is no evidence of a consistent under- or overestimation since both have three deviations of one sign and four with the other sign. The third check tests for correlation between deviations and looks for runs of the same sign. Again, both fertility schedules give no evidence to indicate correlation between differences.

As shown in Appendix F, despite meeting the goodness-of-fit criterion both schedules must still be tested for smoothness. Smoothness is defined as smooth third differences. Based on this measure of smoothness the standard derived using the Brass polynomial is smooth whereas the Hadwiger-based standard is not. As such, the Hadwiger schedule is dismissed based on lack of smoothness.

In addition, although both series have good overall fit, it is the fit in the older age groups that has consistently been the problem (Sections 3.5 to 4.3). Critically, the Brass polynomial fits the older age groups better than the Hadwiger function as illustrated by the lower SSE for the 35-49 age group.

So, in terms of simplicity, goodness-of-fit and smoothness the Brass polynomial with parameters s = 13 and iv =39.8) yields a better standard for the African pattern than the Hadwiger function.

4.7 The African Standard and the Relational Gompertz Model Coefficients

The chosen standard was derived to assist in the analysis of African fertility data and, in particular, when using the relational Gompertz model. The relational Gompertz model requires cumulative fertility rates without a half-year shift, cumulative fertility with a half-year shift and average parities (Table 4.8). In addition, the single-year standardised age-specific fertility rates of this standard are presented in Table 4.9.

Table 4.8: P(i), f(x, x+4) and F(x) for standard excluding the 10-14 age group

Age	F(x) without shift	F(x) with shift	i	Age Group	f(x,x+4)	P(i)
20	0.13285	0.11535	1	15-19	0.13285	0.07548
25	0.34210	0.31974	2	20-24	0.20925	0.24084
30	0.56570	0.54395	3	25-29	0.22360	0.45302
35	0.75980	0.74232	4	30-34	0.19410	0.66640
40	0.89879	0.88740	5	35-39	0.13899	0.83427
45	0.97525	0.96998	6	40-44	0.07646	0.94186
50	1.00000	0.99903	7	45-49	0.02475	0.99087

Table 4.9: Single-year ASFR for standard excluding the 10-14 age group

Age	f(x)	Age	f(x)	Age	f(x)
15	0.01687	27	0.04507	39	0.02277
16	0.02237	28	0.04445	40	0.02021
17	0.02721	29	0.04350	41	0.01768
18	0.03141	30	0.04226	42	0.01521
19	0.03499	31	0.04076	43	0.01282
20	0.03799	32	0.03903	44	0.01055
21	0.04043	33	0.03709	45	0.00842
22	0.04235	34	0.03497	46	0.00647
23	0.04377	35	0.03271	47	0.00472
24	0.04472	36	0.03033	48	0.00320
25	0.04524	37	0.02786	49	0.00194
26	0.04534	38	0.02533		

However, the standard rates in Table 4.8 sand 4.9 are applicable to populations, like the 61 sub-Saharan DHS, where no data are available for the 10-14 age group. By contrast to the DHS data, some fertility data includes the 10-14 age group. The inclusion of this age group means that the standard rates in Tables 4.8 and 4.9 cannot be used without the loss of potentially valuable information about early fertility.

Tables 4.10 and 4.11 give the cumulative fertility rates, age-specific fertility rates and parity for the standard when data are available for the 10-14 age group. As before, the Brass polynomial with parameters s = 13 and *w* = 39.8 is used and ensures consistency between the standards.

Table 4.10: P(i), f(x, x+4) and F(x) for standard including the 10-14 age group

Age x	F(x) without shift	F(x) with shift	i	Age Group	f(x,x+4)	P(i)
15	0.01417	0.00892	0	10-14	0.01417	0.01202
20	0.14514	0.12789	1	15-19	0.13097	0.08858
25	0.35142	0.32938	2	20-24	0.20628	0.25160
30	0.57185	0.55041	3	25-29	0.22043	0.46250
35	0.76321	0.74597	4	30-34	0.19135	0.67113
40	0.90022	0.88900	5	35-39	0.13702	0.83662
45	0.97560	0.97040	6	40-44	0.07538	0.94268
50	1.00000	0.99904	7	45-49	0.02440	0.99100

Table 4.11: Single-year ASFR for standard including the 10-14 age group

Age	f(x)	Age	f(x)	Age	f(x)
10	0	24	0.04409	37	0.02746
11	0	25	0.0446	38	0.02497
12	0	26	0.0447	39	0.02245
13	0.00367	27	0.04443	40	0.01992
14	0.01051	28	0.04382	41	0.01743
15	0.01663	29	0.04288	42	0.01499
16	0.02206	30	0.04166	43	0.01264
17	0.02683	31	0.04018	44	0.01040
18	0.03096	32	0.03847	45	0.00830
19	0.03449	33	0.03656	46	0.00638
20	0.03745	34	0.03447	47	0.00465
21	0.03986	35	0.03224	48	0.00315
22	0.04175	36	0.0299	49	0.00192
23	0.04315				

The relational Gompertz model requires the calculation of the coefficients e(x) and g(x) from the cumulative fertility rates (with and without 1/2-year shift). Similarly, the average parities require that e(i) and g(i) be calculated. Table 4.12 lists the coefficients of the relational Gompertz model associated for the African pattern where the 10-14 age group is excluded.

Table 4.12: Standard relational Gompertz model parameters (10-14 age group excluded)

	No Shift			IA-year Shift			Parity	
Age	e(x)	g(x)	Age	e(x)	g(x)	Age Group	e(i)	g(i)
20	1.4750	-1.4193	19 %	1.4651	-1.4844	15-19	1.5321	-1.6807
25	1.4741	-0.7869	24 ½	1.4825	-0.8502	20-24	1.5475	-1.0884
30	1.3374	-0.1166	29 ½	1.3570	-0.1888	25-29	1.4214	-0.4694
35	1.0924	0.6915	34 Y2	1.1239	0.5991	30-34	1.2418	0.2513
40	0.7123	1.7930	39 1/2	0.7625	1.6569	35-39	0.9480	1.1615
45	0.0000	3.6865	44 1A	0.1459	3.3772	40-44	0.5054	2.4758
						45-49	0	4.6917

For the standard including the 10-14 age group, the model coefficients are recalculated and given in Table 4.13.

Table 4.13: Standard relational Gompertz model parameters (10-14 age group included)

	No Shift			1/2-year Shift			Parity	
Age	e(x)	g(x)	Age	e(x)	g(x)	Age Group	e(i)	g(i)
15	1.2603	-2.1046	14 1/2	1.2138	-2.1932	10-14	1.0628	-2.645
20	1.5052	-1.3822	19 1/2	1.4999	-1.4444	15-19	1.2897	-1.7438
25	1.4837	-0.764	24 1/2	1.4931	-0.8265	20-24	1.4252	-1.0157
30	1.341	-0.0985	29 1/2	1.361	-0.1703	25-29	1.3726	-0.3355
35	1.0937	0.7074	34 1/	1.1254	0.6152	30-34	1.1421	0.4391
40	0.7127	1.8079	39 1/2	0.7629	1.6719	35-39	0.7061	1.5117
45	0.0001	3.701	44 1/2	0.1459	3.3917	40-44	0.2765	3.2104
						45-49	0	6.0547

4.8 Comparison of Relational Gompertz Results from the Booth and African Standards

The final test of the revised African standard is to apply the relational Gompertz model - with the African standard - to actual data from sub-Saharan populations. The results can then be compared to those obtained using the Booth standard. Two populations were selected from the available sub-

Saharan censuses and assessed (Appendix G. In particular, it was decided not use Demographic and Health Survey data since: a) There is little need to use relational Gompertz models on DHS data and b) this may lead to incorrect conclusions about the appropriateness of the African standard given that the standard was derived using this data. Furthermore, the Kenya Census 1979 was chosen to represent old surveys (before the onset of HIV) whereas the Botswana Census 2001 represents more recent censuses and surveys.

Table 4.14 shows the original census data for Kenya 1979 and the standardised results obtained from the Booth and African standards. The last column of the table shows the percentage difference between the Booth estimates and the African estimates.

Table 4.14: Standardised relational Gompertz model results for Kenya Census 1979

Age Group	Original	Booth	African	% Difference between Booth and African standards
15-19	0.08416	0.10957	0.11865	8%
20-24	0.22580	0.22991	0.21251	-8%
25-29	0.23589	0.23629	0.23417	-1%
30-34	0.19329	0.19960	0.20139	1%
35-39	0.14771	0.14656	0.13942	-5%
40-44	0.07757	0.06885	0.07253	5%
45-49	0.03558	0.00923	0.02133	131%

As can be seen from Table 4.14 the Kenya census exhibits the classic, African uptick in 45-49 fertility. A look at the estimates derived using the Booth and African standard show that the Booth standard dramatically understates f(45-49) by almost 75 per cent. By contrast, the estimate based on the African standard is much higher at 60 per cent of the observed level. This constitutes a 35 per cent reduction in the understatement and an increase of 131 per cent over the Booth estimate for the oldest age group.

Table 4.15, in an analogous fashion to Table 4.14, presents the census data for Botswana 2001 as well as the relational Gompertz estimates derived using the Booth and African standards.

Table 4.15: Standardised relational Gompertz model results for Botswana Census 2001

Age Group	Original	Booth	African	% Difference between Booth and African standards
15-19	0.07812	0.09800	0.11080	13%
20-24	0.22759	0.21085	0.19101	-9%
25-29	0.20620	0.22887	0.21944	-4%
30-34	0.19616	0.20532	0.20289	-1%
35-39	0.16087	0.16147	0.15410	-5%
40-44	0.09459	0.08284	0.09010	9%
45-49	0.03647	0.01266	0.03164	150%

As with the Kenyan data, the Botswana census exhibits high fertility in the 40-44 and, particularly, 45-49 age groups. The African standard again yields estimates of 45-49 fertility that are markedly higher than those obtained when using the Booth standard while still maintaining comparable levels over the remainder of the age range. In fact, the estimates of f(45-49) based on the Booth standard are only about 35 per cent of the observed levels in the Botswana population. By contrast, the estimates derived from the African standard are 150 per cent higher than those derived using the Booth standard and at about 87 per cent of the observed rate.

For both censuses the relational Gompertz model using the Booth standard understates fertility in the oldest age group. As explained in previous chapters this results not from Booth's methodology, but from the data upon which the standard is based.

Nevertheless, both the models using the Booth standard are found to provide slightly better overall fit to the data than the models using the African standard (0.0015 against 0.0017 for Kenya and 0.0020 against 0.0027 for Botswana). The reason for this lies in the relative magnitude of the age-specific fertility rates. In particular, a 10 per cent error in the 15-24 age group has a much larger effect of the sum of squared error than a 10 per cent deviation in the 40-49 age group.

The difference in the overall error is, however, small and the estimates obtained using the African standard are more representative of the African fertility pattern.

5 CONCLUSION

In all demographic enquiries, it is preferable that direct methods of analysis be used to obtain fertility rates. Unfortunately, the data collected in sub-Saharan Africa censuses are generally of such poor quality that direct techniques cannot be used. One might question why censuses in this region are not abandoned completely and substituted by, for example, the demographic and health surveys. The reason is that demographic and health surveys have relatively small sample sizes and are subject all the statistical problems related to small samples. Censuses, by contrast, provide full enumeration and large sample sizes (with all the associated benefits). In addition, the DHS sometimes do not work well, for example South Africa 2003.

Hence indirect techniques are required to obtain reasonable estimates of fertility and the foremost among these is the relational Gompertz model. As discussed in Chapter 2, the main drawback to this model is that it requires the use of a standard fertility pattern appropriate to the population under investigation and, since 1979, the Booth standard has been the benchmark most widely (and sometimes incorrectly) used.

However, the preceding chapters have described how changes in the demographic landscape since the development of the standard affect the way it is used. In particular, it necessitated an investigation into the applicability of the Booth standard - for sub-Saharan populations - in order to maintain the relevance of the relational Gompertz model as analysis tool.

To this end the DF-IS data for sub-Saharan Africa was analysed and an African fertility pattern was identified to ease the testing of possible standards for use in African settings. The findings in Chapter 3 showed that the Booth standard does not produce good fits to the individual DHS populations nor the identified African pattern. In particular, the standard understates the fertility in the 45-49 age group and fails to capture the observed uptick at the oldest ages. As a consequence, estimates derived using this standard may be biased with the result that vital statistics and population growth rates are incorrect. In addition, decisions relating to reproductive health, intervention programmes, economic and social development are all based on biased results.

5.1 Methodology

In order to overcome the observed underestimation of fertility at older ages and its associated problems, a number of alternative fertility distributions were considered to develop an African fertility standard for use with the relational Gompertz model: Three versions of the Coale-Trussell model, the Brass polynomial, the Hadwiger function, cubic splines and the Gamma density. In addition, the possibility that *G(a)* and *r(a)* in the Coale-Trussell model be reformulated in terms of sexual exposure rather than marriage was investigated.

However, three of these alternatives were dismissed. First, the method of cubic splines was rejected because it requires good quality data and introduces unnecessary complexity into the model. The second method dismissed is the Gamma function, since both the Coale-Trussell model and the Hadwiger function has been found to deliver equally good results for human populations. Finally, the recasting of the Coale-Trussell *r(a)* and *G(a)* in terms of sexual exposure was set aside for two reasons: a) It falls outside the scope of the current thesis which focuses on methods other than the Coale-Trussell model and b) the data are currently not available in terms of sexual exposure and exposure to marriage is the generally accepted approximation.

The remaining five alternatives were extensively investigated and for each of these models a minimum of 20000 model schedules were produced. This means that every model had a large sample from which an African standard can be developed if an appropriate schedule (or schedules) can be identified. The sum of squared error method - which is statistically appropriate - was used to analyse these schedules and test overall goodness-of-fit.

Once suitable schedules were identified they had to be compared and assessed to determine which schedule best represents African fertility. The comparison was done using graduation tests to check for adherence-to-data as well as smoothness of progression of standard rates. Finally, the selected African standard was tested on empirical data from two DHS to ascertain its appropriateness by compare estimates derived using the relational Gompertz model based on the new standard with estimates based on the Booth standard.

This methodology guarantees the appropriateness of the standard and the Brass polynomial, with parameters s = 13 and w = 39.8, was identified as the most suitable schedule. In particular, it has good overall adherence to the data and it reduces the understatement in the 45-49 age group from almost 70 per cent to less than 6 per cent. The closer match in the oldest age group and the superior overall fit is indicative of a standard more appropriate to the sub-Saharan populations.

5.2 Uses and Benefits of African Standard

Using the identified schedule as the African standard goes some way to addressing the misfit observed for the Booth standard and mitigating the problems that it causes. In particular, the use of the Booth standard for sub-Saharan populations can impose an incorrect shape on the analysis and introduce error and bias into the estimates derived. By contrast, the African standard was developed for use in these populations and has the appropriate shape. As such, it does not introduce the bias associated with the Booth estimates, but produces improved rates - particularly for the oldest age group.

The elimination of this bias does not merely lead to more accurate statistics, but generally more accurate information about fertility and population changes. Having access to accurate, unbiased fertility estimates allows decision makers to make informed choices regarding reproductive health programs, schools, healthcare and social development. The improved estimates also allow researchers to better understand trends in fertility, the factors affecting it and the social dynamics surrounding childbearing in a particular population.

In addition, the development of a distinctly African fertility standard is the first step towards Booth's vision of region specific standards. It is a move away from a single standard that is indiscriminately applied to all situations and towards multiple standards that are suitable for the populations being analysed. The existence of a region-specific standard, such as the African pattern, eases comparison of fertility rates within the region and provides a benchmark for comparison in the absence of more information.

The development of this standard thus ensures the continued relevance of the relational Gompertz method as an important tool in fertility analysis. In particular, since data in the sub-Saharan region is of such poor quality.

5.3 Limitations

Despite these improvements the African standard developed in Chapter 4 has a number of limitations. Foremost among these is the fact that the standard does not explicitly address the issue of HIV. This is of specific concern since HIV is likely to be the greatest single determinant of the future demographic composition of populations in Sub-Saharan Africa, in particular, since these countries are currently experiencing the most severe HIV epidemics.

Figure 5.1 reproduces the distinctive distribution of age-specific fertility observed by Zaba and Gregson (1998) for the HIV-positive population and contrasts this with the African standard.

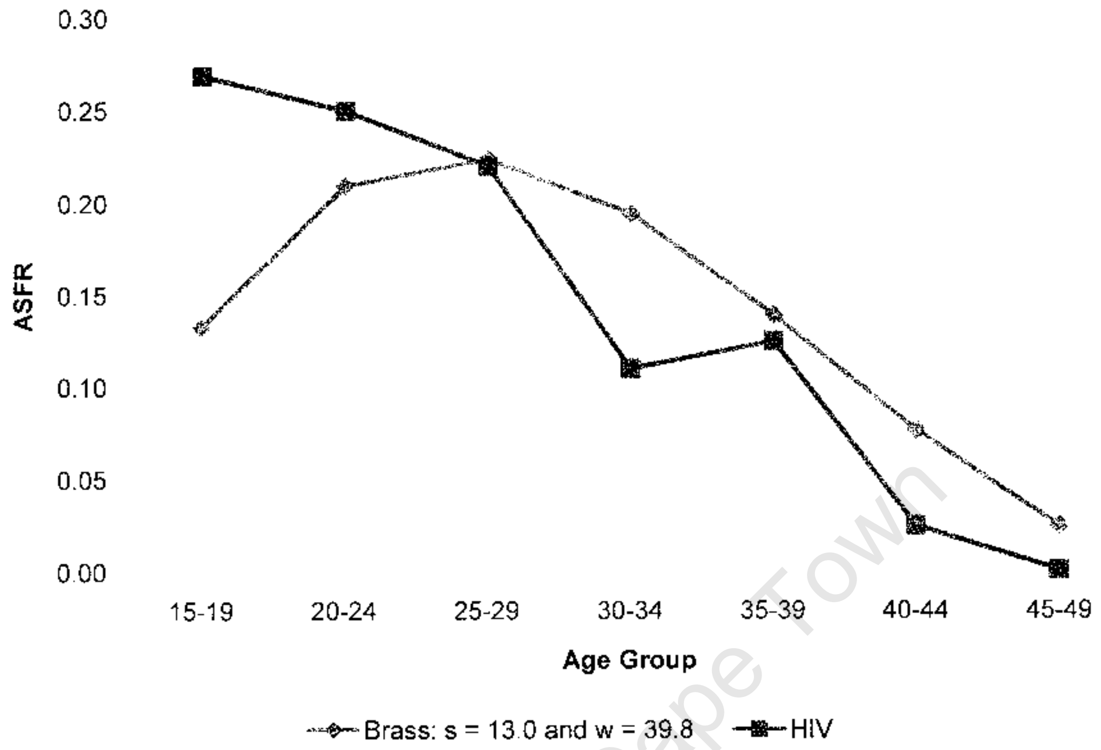

Figure 5.1: ASFR of the African standard and the HIV-positive population

The figure shows that at the youngest ages, the fertility of the sero-positive population can be seen to be higher than the African standard. By contrast, for the oldest the HIV-positive population has lower fertility rates than the standard.

In particular, the difference for the youngest age group results from a selection effect at these ages (Lewis, Ronsman, Ezeh and Gregson 2004). That is, the women falling pregnant in the youngest age group are also those exposed to HIV at an early age due to their sexual activities. For this group the risk behaviour - resulting in exposure to pregnancy and HIV - dominates the debilitating biological effects of the disease. In particular, the impact of the biological factors is limited because duration since infection is relatively short and, consequently, the burden that the disease places on the body is still minor.

Given that standardised fertility rates are cumulative, the inflated early fertility - in conjunction with the debilitating effects of the illness - results in the low ASFR observed for the oldest age group. The dissimilarity in the shape of fertility, as shown in Figure 5.1, means that the standard cannot be used to model the HIV population and derive reasonable, unbiased fertility estimates.

A major drawback of not being able to model the fertility for the HIV-positive population separately is that the impact of interventions or changes in the disease prevalence cannot be catered for. However, despite this restriction the changes at population level can be observed and modelled using the developed African fertility pattern since it models the total population including data for the HIV-positive sub-population.

The second limitation is that one standard can never be appropriate to all populations. Although the African standard developed in Chapter 4 is a step towards Booth's vision of region specific standards, as a single standard it cannot capture all possible fertility patterns. The caveat remains that the standard used in the relational Gompertz model must be appropriate to the population under investigation. If this is not the case, then the same limitations and problems apply to this standard as was observed for the Booth standard when the latter was incorrectly used. So, for example, the African standard cannot be used for the low fertility sub-populations of South Africa despite being suitable for the high fertility group. Similarly, the standard is also not appropriate for the Sahelian countries that demonstrate a different fertility pattern reflecting differences in attitudes towards fertility.

Finally, some critics may argue that the uptick observed for the African populations stems from age misreporting and that the standard - being based on empirical (DHS) data - incorrectly tracks and models an error in the data. In a sense, these critics are correct as there, surely, is an element of age misreporting present. In particular, it may be the case that 45-49 fertility is overstated as a result of the grandmother effect whereby the child is incorrectly matched to the grandmother (United Nations 1983, p. 183). This is particularly likely where mothers are absent due to migration or the effect of AIDS.

However, the uptick is too consistent among the African populations to be explained away by one factor (age misreporting). The belief that the uptick disappears the moment better data becomes available also appears flawed on two counts. First, the sub-Saharan DHS data does not seem to show this. For example, the South African fertility data from the 1996 Census and 1998 DHS also exhibits this fertility pattern despite relatively reliable data. Second, the DHS is designed to minimise age reporting errors and, as such, if the uptick is present it is likely to be an actual feature of the fertility pattern and should be reflected by the standard. Finally, if the uptick disappears when better quality data are available this may be the result of other factors. Although it is true that improved education and socio-economic circumstances will eliminate age misreporting, these factors will also affect fertility through the proximate determinants. As a result, this may lead to a change in fertility preferences and, consequently, the fertility pattern which could - just as easily as the elimination of age misreporting - account for the disappearance of the uptick.

In addition, the standard developed in Chapter 4 slightly understates 45-49 fertility relative to the average African pattern and slightly overstates the 40-44 fertility. As a result, the standard allows for both an element of age misreporting as well as reflecting the consistent uptick in 45-49 fertility.

5.4 Looking to the Future

Despite these limitations the standard provides a means of improving fertility estimation using the relational Gompertz model in sub-Saharan Africa, whilst the limitations indicate possible areas of future research. For example, some evidence exists to suggest that there is more than one fertility pattern present in sub-Saharan African. Evidence from the sub-Saharan DHS suggests that the pattern of fertility may be linked to geographic location and the level of fertility. As already discussed, the Sahelian countries exhibit a different pattern to other sub-Saharan populations and a standard needs to be developed for these West African populations.

Apart from the Sahelian countries, the DHS data also hints at other potential patterns related to region and total fertility. In particular, this means that future research may possibly result in a family of sub-Saharan fertility standards being developed (for example consisting of North, South, East and West standards) in much the same way as the Coale-Demeny mortality patterns.

The second significant area for future research is HIV. Given the debilitating effects of HIV and its importance when looking into the future, further research is imperative to gain information on the factors driving the disease and the impacts of intervention campaigns and ART (anti-retroviral therapy). Specifically, a method is required to explicitly model fertility of the HIV-positive population and develop a sub-Saharan HIV standard. This will enable better understanding, planning and analysis of the disease and its effects on fertility.

A blending function is then envisioned to obtain population age-specific fertility rates by using the age-specific HIV prevalence rates in combination with the HIV standard and a non-HIV standard (Equation 5.1).

$$f_T(x, x+4) = (1-p) * f_N(x, x+4) + p * f_H(x, x+4) \qquad \textbf{Equation 5.1}$$

Where p is the HIV prevalence for age group (x, x+4) and f_1 (x, x+4), f, (x, x+4) and f,,(x, x+4) are the ASFRs for the total, HIV-negative and HIV-positive populations, respectively. The standard for the HIV-positive population and this blending function can then be extended to incorporate the duration since infection and the availability of anti-retroviral drugs as more data becomes available.

Finally, and critically, attention must be given and emphasis placed on improving data collection methods. It is clear that the enduring problem in demographic analysis remains the quality of the data available and - no matter what other future research is conducted to improve estimation procedures - the results will remain constrained by the data quality.

However, until such time as better quality data are collected in the sub-Saharan countries the only alternative remains estimating fertility rates using indirect techniques like the relational Gompertz model, but ensuring that an appropriate standard fertility pattern - like the African standard for sub-Saharan populations - is used.

APPENDIX A

Table A 1: Single-year values of n(a) and v(a)

Age	n(a)	v(a)	Age	n(a)	v(a)
0	0.000	0.000	21	0.420	-0.520
1	0.005	0.000	22	0.410	-0.600
2	0.100	0.000	23	0.400	-0.680
3	0.175	0.000	24	0.389	-0.760
4	0.225	0.000	25	0.375	-0.830
5	0.275	0.000	26	0.360	-0.900
6	0.325	0.000	27	0.343	-0.970
7	0.375	0.000	28	0.325	-1.040
8	0.421	0.000	29	0.305	-1.110
9	0.460	0.000	30	0.280	-1.180
10	0.475	0.000	31	0.247	-1.250
11	0.477	-0.004	32	0.207	-1.320
12	0.475	-0.030	33	0.167	-1.390
13	0.470	-0.060	34	0.126	-1.460
14	0.465	-0.100	35	0.087	-1.530
15	0.460	-0.150	36	0.055	-1.590
16	0.455	-0.200	37	0.035	-1.640
17	0.449	-0.250	38	0.021	-1.670
18	0.442	-0.310	39	0.011	-1.690
19	0.435	-0.370	40	0.003	-1.700
20	0.428	-0.440			

Source: Coale and Trussell 1979

Table A 2: Standard e(x) and g(x) values (no shift)

Age	e(x)	g(x)
15	0.9866	-2.3138
20	1.3539	-1.3753
25	1.4127	-0.6748
30	1.2750	0.0393
35	0.9157	0.9450
40	0.3966	2.3489
45	--	4.8097

Table 3 in Brass (1988) and Table 3A in /.aba (1981)

Table A 3: Standard e(x) and g(x) values (1/2-year shift)

Age	e(x)	g(x)
14.5	0.9760	-2.4020
19.5	1.3364	-1.4501
24.5	1.4184	-0.7430
29.5	1.2978	-0.0382
34.5	0.9670	0.8356
39.5	0.4509	2.1649
44.5	0.0462	4.4564
49.5		--

Source: Brass (1988) and Zaba (1981)

Table A 4: Standard e(i) and g(i) values

Age	e(i)	g(i)
10-14	1.0632	-2.6447
15-19	1.2897	-1.7438
20-24	1.4252	-1.0157
25-29	1.3725	-0.3353
30-34	1.1421	0.4391
35-39	0.7061	1.5117
40-44	0.2763	3.2105
45-49		--

Source: Brass (1988) and Zaba (1981)

APPENDIX B

As discussed in section 3.5.1 Booth incorrectly uses schedule 28 in Table 18 for the upper tail when it should have been used for the lower tail. Tables B 1 to B 6 below illustrates the steps in the procedure used by Booth as well as showing the impact of correctly classifying schedule 28 in the lower tail. In each of the tables the first column is the age group, the second column the values as derived by Booth and the third column the calculated after including schedule 28 in the lower tail and excluding it from the upper tail. Table B 1 compares the AY(x) values derived by Booth with the AY(x) based on the criteria. Table B 2 gives the adjustment factors calculated by Booth and those based on the criteria.

Table B 1: Comparison of the Booth and Criteria AY(x)

Age Group	Booth	Criteria
10-14	00	
15-19	1.09120	1.11470
20-24	0.72320	0.71944
25-29	0.67436	0.67435
30-34	0.77872	0.77903
35-39	1.14730	1.14731
40-44	2.13486	2.13598
45-49		

Table B 2: The Adjustment Factors used by Booth and based on the Criteria

Adjustment Factors	Booth	Criteria
k1	0.99135	0.99494
k2	1.02287	1.02074

Table B 3 shows the Y(x) values after back-transforming the AY(x) values and B 4 shows the Y(x) values after applying the adjustment factors in Table B 2.

Table B 3: Y(x) for Booth and Criteria

Age Group	Booth	Criteria
10-14	∞	∞
15-19	1.09120	1.11470
20-24	0.72320	0.71944
25-29	0.67436	0.67435
30-34	0.77872	0.77903
35-39	1.14730	1.14731
40-44	2.13486	2.13598
45-49	∞	∞

Table B 4: Adjusted Y(x) for Booth and Criteria

Age Group	Booth	Criteria
10-14		
15-19	1.08176	1.10906
20-24	0.71694	0.71580
25-29	0.67436	0.67435
30-34	0.77872	0.77903
35-39	1.14730	1.14731
40-44	2.18368	2.18029
45-49	∞	∞

Finally, the cumulative fertility, F(x), values are derived by reversing the Gompertz transform (Table B 5) and these are differenced to obtain the age-group specific fertility rates, f(x), in Table B 6.

Table B 5: Derived cumulative fertility, F(x)

Age Group	Booth	Criteria
10	0	0
15	0.00277	0.00237
20	0.13584	0.13615
25	0.37731	0.37731
30	0.60861	0.60861
35	0.79618	0.79624
40	0.93019	0.93021
45	0.99188	0.99186
50	1	1

Table B 6: Derived standard age-group specific fertility, f(x)

Age Group	Booth	Criteria
10 - 14	0.00277	0.00237
15 - 19	0.13307	0.13378
20 - 24	0.24147	0.24117
25 - 29	0.23130	0.23129
30 - 34	0.18757	0.18763
35 - 39	0.13401	0.13398
40 - 44	0.06169	0.06165
45 - 49	0.00812	0.00814

APPENDIX C

Included in this appendix is the VBA code used to generate the three Coale-Trussell models (section C.1), the Brass polynomial data (section C.2) and the Hadwiger function schedules (section C.3) in MS Excel 2007. The code for the Coale-Trussell model utilizes the original $n(a)$ and $v(a)$ values given by Booth. To obtain the other two models the values of $n(a)$ and $v(a)$ must be changed.

C.1 Coale-Trussell Model

```
Sub CTSchedules()
'
' CTSchedules Macro
'
'

Application.ScreenUpdating = False

' Name sheets
    Sheets("Sheet1").Select
    Sheets("Sheet1").Name = "Model"
    Sheets("Sheet2").Select
    Sheets("Sheet2").Name = "Results"

' Label columns and generate output
   Application.Run "Labels"
   Application.Run "Multiple"

' Saves workbook in a new folder called Results as CoaleTrussell
    MkDir "Results"
    ChDir "Results"
    ActiveWorkbook.SaveAs Filename:= _
        "CoaleTrussell.xlsm", FileFormat:= _
        xlOpenXMLWorkbookMacroEnabled, CreateBackup:=False
End Sub

Sub Labels()
'
' Labels Macro
'
'

' Set Labels
    Sheets("Results").Select
```

```
    Range("A1").Select
    ActiveCell.FormulaR1C1 = "A0"
    Range("B1").Select
    ActiveCell.FormulaR1C1 = "K"
    Range("C1").Select
    ActiveCell.FormulaR1C1 = "M"
    Range("D1").Select
    ActiveCell.FormulaR1C1 = "'-"
    Range("E1").Select
    ActiveCell.FormulaR1C1 = "'-"
    Range("F1").Select
    ActiveCell.FormulaR1C1 = "Mean"
    Range("G1").Select
    ActiveCell.FormulaR1C1 = "Std Dev"
    Range("H1").Select
    ActiveCell.FormulaR1C1 = "'R1"
    Range("I1").Select
    ActiveCell.FormulaR1C1 = "PTY1"
    Range("J1").Select
    ActiveCell.FormulaR1C1 = "PTY2"
    Range("K1").Select
    ActiveCell.FormulaR1C1 = "PTY3"
    Range("L1").Select
    ActiveCell.FormulaR1C1 = "PAR1"
    Range("M1").Select
    ActiveCell.FormulaR1C1 = "PAR2"
    Range("N1").Select
    ActiveCell.FormulaR1C1 = "'-"
    For i = 10 To 49 Step 1
        Sheets("Results").Cells(1, i + 5).Select
        ActiveCell.FormulaR1C1 = i
    Next i
    Range("BC1").Select
    ActiveCell.FormulaR1C1 = "-"
    For i = 0 To 7 Step 1
        Sheets("Results").Cells(1, 56 + i).Select
        ActiveCell.FormulaR1C1 = "'" & (10 + 5 * i) & "-" & (14 + 5 * i)
    Next i
    Range("BL1").Select
    ActiveCell.FormulaR1C1 = "Sum Rank"
End Sub

Public Sub GenOutput(aaa As Double, kkk As Double, mmm As Double)
'
' GenOutput Macro
'
'
  Dim BB(7), ZS(423), ZSS(501), EM2(41), F(41), T(8), V(41), N(41), RR(6), ZU(423) As Double
```

```
Dim Sumf, w, SS, FR, TT, last, first, A, cons, sum, Sumsq As Double
Dim sigma, smean, R1, Q1, Q2, Q3, PAR1, PAR2 As Double
Dim i, J As Integer

Application.ScreenUpdating = False

' Initialize v(a) and n(a)
  V(0) = "0"
  V(1) = "0"
  V(2) = "0"
  V(3) = "0"
  V(4) = "0"
  V(5) = "0"
  V(6) = "0"
  V(7) = "0"
  V(8) = "0"
  V(9) = "0"
  V(10) = "0"
  V(11) = "-0.004"
  V(12) = "-0.03"
  V(13) = "-0.06"
  V(14) = "-0.1"
  V(15) = "-0.15"
  V(16) = "-0.2"
  V(17) = "-0.25"
  V(18) = "-0.31"
  V(19) = "-0.37"
  V(20) = "-0.44"
  V(21) = "-0.52"
  V(22) = "-0.6"
  V(23) = "-0.68"
  V(24) = "-0.76"
  V(25) = "-0.83"
  V(26) = "-0.9"
  V(27) = "-0.97"
  V(28) = "-1.04"
  V(29) = "-1.11"
  V(30) = "-1.18"
  V(31) = "-1.25"
  V(32) = "-1.32"
  V(33) = "-1.39"
  V(34) = "-1.46"
  V(35) = "-1.53"
  V(36) = "-1.59"
  V(37) = "-1.64"
  V(38) = "-1.67"
  V(39) = "-1.69"
  V(40) = "-1.7"
```

```
    N(0) = "0"
    N(1) = "0.005"
    N(2) = "0.1"
    N(3) = "0.175"
    N(4) = "0.225"
    N(5) = "0.275"
    N(6) = "0.325"
    N(7) = "0.375"
    N(8) = "0.421"
    N(9) = "0.46"
    N(10) = "0.475"
    N(11) = "0.477"
    N(12) = "0.475"
    N(13) = "0.47"
    N(14) = "0.465"
    N(15) = "0.46"
    N(16) = "0.455"
    N(17) = "0.449"
    N(18) = "0.442"
    N(19) = "0.435"
    N(20) = "0.428"
    N(21) = "0.42"
    N(22) = "0.41"
    N(23) = "0.4"
    N(24) = "0.389"
    N(25) = "0.375"
    N(26) = "0.36"
    N(27) = "0.343"
    N(28) = "0.325"
    N(29) = "0.305"
    N(30) = "0.28"
    N(31) = "0.247"
    N(32) = "0.207"
    N(33) = "0.167"
    N(34) = "0.126"
    N(35) = "0.087"
    N(36) = "0.055"
    N(37) = "0.035"
    N(38) = "0.021"
    N(39) = "0.011"
    N(40) = "0.003"

' Establishes the cumulative of the cumulative ever married
' schedule in 0.1 year intervals with age 0 as origin
        ZS(0) = 0
        ZU(0) = (0.19465 / kkk) * Exp((-0.174 / kkk) * ( 6.06 * kkk) - Exp((-0.2881 / kkk) * ( 6.06 *
kkk)))
```

```
    For i = 1 To 422
       w = i / 10
       ZU(i) = (0.19465 / kkk) * Exp((-0.174 / kkk) * (w - 6.06 * kkk) - Exp((-0.2881 / kkk) * (w -
6.06 * kkk)))
       ZS(i) = ZS(i - 1) + (0.05 * (ZU(i) + ZU(i - 1)))
    Next i

' Shift the Origin of the cumulative ever married schedule to a0
    J = Int(10 * aaa)
    last = 500 - J

    For i = 1 To last Step 1
       J = J + 1
       ZSS(J) = ZS(i)
    Next i

' Establish the average for each year of age by averaging the
' cumulative of the cumulative ever married schedule for the 100
' values in each year of age
    For i = 1 To 40 Step 1
       k = 100 + 10 * i
       w = 0
       For l = 1 To 10 Step 1
          w = w + 0.5 * (ZSS(k - l + 1) + ZSS(k - l))
       Next l
       EM2(i) = w / 10
    Next i

' The following section fits an exponential to the 15-19 section of
' the ASF schedule such that the area under the curve is the same
' before and after transformation
    For i = 1 To 40 Step 1
       F(i) = EM2(i) * N(i) * Exp(mmm * V(i))
    Next i

    For k = 1 To 7 Step 1
       BB(k) = 0
       For l = 1 To 5 Step 1
          s = l + 5 * (k - 1) + 5
          BB(k) = BB(k) + F(s)
       Next l
       T(k) = BB(k) / 5
    Next k

    first = (F(1) + F(2) + F(3) + F(4) + F(5)) / 5

    If aaa >= 15 Then
```

```
        TT = T(1) * 5
        FR = 0.476 * ZSS(200)
        SS = (FR * 5 / TT) - 1
        cons = FR / (5 ^ SS)
        For i = 1 To 5 Step 1
            RR(i) = cons * (i ^ (SS + 1)) / (SS + 1)
        Next i

        F(6) = RR(1)

        For k = 2 To 5 Step 1
            l = k + 5
            F(l) = RR(k) - RR(k - 1)
        Next k
    End If

' Calculate the mean, standard deviation, parities and R1
    For i = 1 To 40 Step 1
        Sumf = F(i) + Sumf
    Next i

    For i = 1 To 8 Step 1
        T(i) = T(i) / Sumf
    Next i

    first = first / Sumf
    A = 10.5

    For i = 1 To 40 Step 1
        F(i) = F(i) / Sumf
        sum = sum + A * F(i)
        Sumsq = Sumsq + A * A * F(i)
        A = A + 1
    Next i

    sigma = (Sumsq - sum * sum)   1 / 12
    smean = sum
    Q1 = (4.5 * F(6) + 3.5 * F(7) + 2.5 * F(8) + 1.5 * F(9) + 0.5 * F(10)) / 5 + (5 * first)
    Q2 = (4.5 * F(11) + 3.5 * F(12) + 2.5 * F(13) + 1.5 * F(14) + 0.5 * F(15)) / 5 + 5 * (T(1) +
first)
    Q3 = (4.5 * F(16) + 3.5 * F(17) + 2.5 * F(18) + 1.5 * F(19) + 0.5 * F(20)) / 5 + 5 * (T(2) -
T(1) + first)
    PAR1 = Q1 / Q2
    PAR2 = Q2 / Q3
    R1 = T(1) / T(2)

' Output data
    Sheets("Model").Select
```

```
        Range("A4").Select
        ActiveCell.FormulaR1C1 = "-"
        Range("A5").Select
        ActiveCell.FormulaR1C1 = "-"
        Range("A6").Select
        ActiveCell.FormulaR1C1 = smean
        Range("A7").Select
        ActiveCell.FormulaR1C1 = sigma
        Range("A8").Select
        ActiveCell.FormulaR1C1 = R1
        Range("A9").Select
        ActiveCell.FormulaR1C1 = Q1
        Range("A10").Select
        ActiveCell.FormulaR1C1 = Q2
        Range("A11").Select
        ActiveCell.FormulaR1C1 = Q3
        Range("A12").Select
        ActiveCell.FormulaR1C1 = PAR1
        Range("A13").Select
        ActiveCell.FormulaR1C1 = PAR2
        Range("A14").Select
        ActiveCell.FormulaR1C1 = "-"

        For i = 0 To 7 Step 1
           For k = 0 To 4 Step 1
              Sheets("Model").Cells(5 * i + k + 15, 1).Select
              ActiveCell.FormulaR1C1 = 1000000 * F(5 * i + k + 1)
           Next k
        Next i

        Range("A55").Select
        ActiveCell.FormulaR1C1 = "-"

        For i = 0 To 7 Step 1
           Sheets("Model").Cells(56 + i, 1).Select
           If i = 0 Then
           ActiveCell.FormulaR1C1 = 1000000 * first
           Else
           ActiveCell.FormulaR1C1 = T(i) * 1000000
           End If
        Next i

        Range("A64").Select
        ActiveCell.FormulaR1C1 = _
        "=RANK(R[-6]C,R[-8]C:R[-1]C)+RANK(R[-5]C,R[-8]C:R[-1]C)"
End Sub

Sub Multiple()
```

```
'
' Multiple Macro
'
'
  Dim x As Double, y As Double, z As Double
  Dim l_aaa As Double, l_kkk As Double, l_mmm As Double
  Dim c As Double, d As Double, e As Double

  c = 2

  Application.ScreenUpdating = False

' Create schedule output
  For x = 1 To 40 Step 1
   l_aaa = 9 + (x / 4)
   Sheets("Model").Select
   Range("A1").Select
   ActiveCell.FormulaR1C1 = l_aaa

    For y = 1 To 40 Step 1
     l_kkk = y / 20
     Sheets("Model").Select
     Range("A2").Select
     ActiveCell.FormulaR1C1 = l_kkk

      For z = 1 To 40 Step 1
       l_mmm = z / 20
       Sheets("Model").Select
       Range("A3").Select
       ActiveCell.FormulaR1C1 = l_mmm

' Generate output
   Call GenOutput(l_aaa, l_kkk, l_mmm)

' Copy output to Results Sheet
       Range("A1").Select
       Range(Selection, Selection.End(xlDown)).Select
       Selection.Copy
       b = c + d + e
       Worksheets("Results").Select
       Sheets("Results").Cells(b, 1).Select
       Selection.PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
        :=False, Transpose:=True
       e = z + 2
      Next z

    l_mmm = 0
    d = 40 * y
```

```
  c = 2
  Next y

 l_kkk = 0
 c = 1600 * x
 d = 0
Next x
End Sub
```

C.2 Brass Polynomial

```
Sub Brass()
'
' Brass Macro
'
'
    Application.ScreenUpdating = False

' Name worksheets
    Sheets("Sheet1").Select
    Sheets("Sheet1").Name = "BrassCalc"
    Sheets("Sheet2").Select
    Sheets("Sheet2").Name = "Results"

' Label columns and generate output
    Application.Run "Labels"
    Application.Run "Multiple"

'Number Formatting
    Range("D2:BM2").Select
    Range(Selection, Selection.End(xlDown)).Select
    Selection.NumberFormat = "0.00000"

' Saves workbook in a new folder called Results as Brass.xlsm
    MkDir "Results"
    ChDir "Results"
    ActiveWorkbook.SaveAs Filename:= _
        "Brass.xlsm", FileFormat:= _
        xlOpenXMLWorkbookMacroEnabled, CreateBackup:=False
End Sub

Sub Labels()
'
' Labels Macro
'
'
    Dim j As Integer, l As Integer
```

```
    Application.ScreenUpdating = False

' Set column headings on Results sheet
    Sheets("Results").Select
    Range("A1").Select
    ActiveCell.FormulaR1C1 = "s"
    Range("B1").Select
    ActiveCell.FormulaR1C1 = "w"
    Range("C1").Select
    ActiveCell.FormulaR1C1 = "-"

' f(10) to f(53)
        For j = 0 To 8 Step 1
            For l = 0 To 4 Step 1
                Sheets("Results").Select
                Sheets("Results").Cells(1, 5 * j + l + 4).Select
                ActiveCell.FormulaR1C1 = "f(" & 5 * j + l + 9 & ")"
            Next l
        Next j

    Range("D1").Select
    ActiveCell.FormulaR1C1 = "-"
    Range("AS1").Select
    ActiveCell.FormulaR1C1 = "-"
    Range("AV1").Select
    ActiveCell.FormulaR1C1 = "-"

' f(10-14) to f(45-49)
        For j = 1 To 8 Step 1
            Sheets("Results").Cells(1, 45 + j).Select
            ActiveCell.FormulaR1C1 = "f(" & 5 * j + 5 & "," & 5 * j + 9 & ")"
        Next j

    Sheets("Results").Select
    Range("BB1").Select
    ActiveCell.FormulaR1C1 = "-"

' f(15-19) to f(45-49)
        For j = 2 To 8 Step 1
            Sheets("Results").Cells(1, 53 + j).Select
            ActiveCell.FormulaR1C1 = "f(" & 5 * j + 5 & "," & 5 * j + 9 & ")"
        Next j

    Range("BJ1").Select
    ActiveCell.FormulaR1C1 = "-"
    Range("BK1").Select
    ActiveCell.FormulaR1C1 = "Rank"
```

```
End Sub

Public Sub GenOutput(sss As Double, www As Double)
'
' GenOutput Macro
'
'
  Dim x As Integer, i As Integer, k As Integer
  Dim asfr(54) As Double, cumf(54) As Double, FF(8) As Double
  Dim T As Double
  Application.ScreenUpdating = False

' Produce 36000 schedules as output taking values 9 <= s <= 20.9 and 22 <= w <= 51.9
      Sheets("BrassCalc").Select
      Range("A3").Select
      ActiveCell.FormulaR1C1 = "-"

' Loop through age x

      cumf(0) = 0
      asfr(0) = 0

      For x = 1 To 54
         If x < sss Then
            cumf(x) = 0
         Else
            If x <= (sss + www) Then
               cumf(x) = 3 * ((sss + www - x) / www) ^ 4 - 4 * ((sss + www - x) / www) ^ 3 + 1
            Else
               cumf(x) = 1
            End If
         End If
         asfr(x - 1) = cumf(x) - cumf(x - 1)
      Next x

      T = cumf(50) - cumf(10)

      For i = 1 To 8 Step 1
         FF(i) = cumf(5 * i + 10) - cumf(5 * i + 5)
      Next i

'Output results
      For i = 0 To 8 Step 1
         For k = 0 To 4 Step 1
            Sheets("BrassCalc").Cells(5 * i + k + 4, 1).Select
            ActiveCell.FormulaR1C1 = asfr(5 * i + k + 9) / T
         Next k
      Next i
```

```
        Range("A45").Select
        ActiveCell.FormulaR1C1 = "-"

        For i = 1 To 8 Step 1
            Sheets("BrassCalc").Cells(45 + i, 1).Select
            ActiveCell.FormulaR1C1 = FF(i) / T
            Sheets("BrassCalc").Cells(53 + i, 1).Select
'           If i = 2 Then
'               ActiveCell.FormulaR1C1 = (FF(1) + FF(2)) / T
'           Else
'               ActiveCell.FormulaR1C1 = FF(i) / T
'           End If
            ActiveCell.FormulaR1C1 = FF(i) / (T - FF(1))
        Next i

        Range("A4").Select
        ActiveCell.FormulaR1C1 = "-"
        Range("A54").Select
        ActiveCell.FormulaR1C1 = "-"
        Range("A62").Select
        ActiveCell.FormulaR1C1 = "-"
        Range("A63").Select
        ActiveCell.FormulaR1C1 = _
        "=RANK(R[-7]C,R[-8]C:R[-2]C)+RANK(R[-6]C,R[-8]C:R[-2]C)"
End Sub

Sub Multiple()
'
' Multiple Macro
'
'
  Dim y As Integer, z As Integer, b As Double
  Dim s As Double, w As Double
  Application.ScreenUpdating = False
  b = 2

' Create schedule output
  For y = 0 To 119 Step 1
    s = 9 + (y / 10)
    Sheets("BrassCalc").Select
    Range("A1").Select
    ActiveCell.FormulaR1C1 = s

    For z = 0 To 299 Step 1
        w = 22 + z / 10
        Sheets("BrassCalc").Select
        Range("A2").Select
```

```
        ActiveCell.FormulaR1C1 = w

' Generate output
        Call GenOutput(s, w)

' Copy output to Results Sheet
        Sheets("BrassCalc").Select
        Range("A1").Select
        Range(Selection, Selection.End(xlDown)).Select
        Selection.Copy
        Worksheets("Results").Select
        Sheets("Results").Cells(b, 1).Select
        Selection.PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
          :=False, Transpose:=True
         b = b + 1
    Next z
  Next y
End Sub
```

C.3 Hadwiger Function

```
Sub Hadwiger()
'
' Hadwiger Macro
'
'
    Application.ScreenUpdating = False

' Name worksheets
    Sheets("Sheet1").Select
    Sheets("Sheet1").Name = "HadwigerCalc"
    Sheets("Sheet2").Select
    Sheets("Sheet2").Name = "Results"

' Label columns and generate output
    Application.Run "Labels"
    Application.Run "Multiple"

'Number Formatting
    Range("D2:BM2").Select
    Range(Selection, Selection.End(xlDown)).Select
    Selection.NumberFormat = "0.00000"

' Saves workbook in a new folder called Results as Hadwiger.xlsm
     MkDir "Results"
    ChDir "Results"
    ActiveWorkbook.SaveAs Filename:= _
        "Hadwiger.xlsm", FileFormat:= _
```

```
        xlOpenXMLWorkbookMacroEnabled, CreateBackup:=False
End Sub

Sub Labels()
'
' Labels Macro
'
'
    Dim j As Integer, l As Integer
    Application.ScreenUpdating = False

' Set column headings on Results sheet
    Sheets("Results").Select
    Range("A1").Select
    ActiveCell.FormulaR1C1 = "s"
    Range("B1").Select
    ActiveCell.FormulaR1C1 = "u"
    Range("C1").Select
    ActiveCell.FormulaR1C1 = "b"
    Range("D1").Select
    ActiveCell.FormulaR1C1 = "c"
    Range("E1").Select
    ActiveCell.FormulaR1C1 = "-"

' f(10) to f(53)
        For j = 0 To 8 Step 1
            For l = 0 To 4 Step 1
                Sheets("Results").Select
                Sheets("Results").Cells(1, 5 * j + l + 6).Select
                ActiveCell.FormulaR1C1 = "f(" & 5 * j + l + 9 & ")"
            Next l
        Next j

    Range("F1").Select
    ActiveCell.FormulaR1C1 = "-"
    Range("AU1").Select
    ActiveCell.FormulaR1C1 = "-"

' f(10-14) to f(45-49)
        For j = 1 To 8 Step 1
            Sheets("Results").Cells(1, 47 + j).Select
            ActiveCell.FormulaR1C1 = "f(" & 5 * j + 5 & "," & 5 * j + 9 & ")"
        Next j

    Sheets("Results").Select
    Range("BD1").Select
    ActiveCell.FormulaR1C1 = "-"
```

```
' f(15-19) to f(45-49)
        For j = 2 To 8 Step 1
            Sheets("Results").Cells(1, 55 + j).Select
            ActiveCell.FormulaR1C1 = "f(" & 5 * j + 5 & "," & 5 * j + 9 & ")"
        Next j

    Range("BL1").Select
    ActiveCell.FormulaR1C1 = "-"
    Range("BM1").Select
    ActiveCell.FormulaR1C1 = "Rank"
End Sub

Public Sub GenOutput(b As Double, c As Double, s As Double, u As Double)
'
' GenOutput Macro
'
'
  Dim x As Integer, i As Integer, k As Integer
  Dim asfr(54) As Double, F(54) As Double, F2(54) As Double, m(54) As Double, FF(8) As Double
  Dim T As Double, TT As Double
' Dim asfr2(54) As Double
  Application.ScreenUpdating = False

' Produce 36000 schedules as output taking values 9 <= s <= 20.9 and 22 <= u <= 51.9
        Sheets("HadwigerCalc").Select
        Range("A5").Select
        ActiveCell.FormulaR1C1 = "-"

' Loop through age x
        T = 0
        TT = 0
        m(0) = 0
        asfr(0) = 0
        F(0) = 0

        For x = 1 To 53 Step 1
            If x < s Or x > u Then
                m(x) = 0
            Else
                m(x) = (b / c) * (c / x) ^ (3 / 2) * Exp(-(b ^ 2) * (c / x + x / c - 2))
            End If
        Next x

        For x = 1 To 53 Step 1
            F(x) = m(x) + F(x - 1)
            T = T + m(x)
        Next x
```

```
        For x = 1 To 53 Step 1
            F2(x) = F(x) / T
        Next x

        For x = 1 To 52 Step 1
            asfr(x) = F2(x + 1) - F2(x)
        Next x
            asfr(53) = 1 - F2(53)

        T = F2(50) - F2(10)

        For i = 1 To 8 Step 1
            FF(i) = F2(5 * i + 10) - F2(5 * i + 5)
        Next i

'Output results
        For i = 0 To 8 Step 1
            For k = 0 To 4 Step 1
                Sheets("HadwigerCalc").Cells(5 * i + k + 6, 1).Select
                ActiveCell.FormulaR1C1 = asfr(5 * i + k + 9) / T
            Next k
        Next i

        Range("A47").Select
        ActiveCell.FormulaR1C1 = "-"

        For i = 1 To 8 Step 1
            Sheets("HadwigerCalc").Cells(47 + i, 1).Select
            ActiveCell.FormulaR1C1 = FF(i) / T
            Sheets("HadwigerCalc").Cells(55 + i, 1).Select
            ActiveCell.FormulaR1C1 = FF(i) / (T - FF(1))
        Next i

        Sheets("HadwigerCalc").Select
        Range("A6").Select
        ActiveCell.FormulaR1C1 = "-"
        Range("A56").Select
        ActiveCell.FormulaR1C1 = "-"
        Range("A64").Select
        ActiveCell.FormulaR1C1 = "-"
        Range("A65").Select
        ActiveCell.FormulaR1C1 = _
        "=RANK(R[-7]C,R[-8]C:R[-2]C)+RANK(R[-6]C,R[-8]C:R[-2]C)"
End Sub

Sub Multiple()
'
' Multiple Macro
```

```
'
'
 Dim rowi As Double, bbb As Double, ccc As Double, sss As Double, uuu As Double
 Application.ScreenUpdating = False
 rowi = 2

' Create schedule output
   For bbb = 1.7 To 3.4 Step 0.1
      Sheets("HadwigerCalc").Select
      Range("A3").Select
      ActiveCell.FormulaR1C1 = bbb

      For ccc = 23 To 34 Step 0.5
        Sheets("HadwigerCalc").Select
        Range("A4").Select
        ActiveCell.FormulaR1C1 = ccc

          For sss = 9 To 19 Step 1
           Sheets("HadwigerCalc").Select
           Range("A1").Select
           ActiveCell.FormulaR1C1 = sss

             For uuu = 47 To 53 Step 1
              Sheets("HadwigerCalc").Select
              Range("A2").Select
              ActiveCell.FormulaR1C1 = uuu

' Generate output
              Call GenOutput(bbb, ccc, sss, uuu)

' Copy output to Results Sheet
              Sheets("HadwigerCalc").Select
              Range("A1").Select
              Range(Selection, Selection.End(xlDown)).Select
              Selection.Copy
              Worksheets("Results").Select
              Sheets("Results").Cells(rowi, 1).Select
              Selection.PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
                :=False, Transpose:=True
              rowi = rowi + 1
             Next uuu
          Next sss
      Next ccc
   Next bbb
End Sub
```

APPENDIX D

Equations D.1 and D.2 are the identifying equations of g*(a)* in the Coale-Trussell model.

$$g(a) = \frac{0.19465}{k} \exp[-0.174W - \exp(-0.2881W)]$$ **Equation D. 1**

where

$$W = \frac{a - a_0 - 6.06k}{k}$$ **Equation D. 2**

Figure D.1 illustrates g(a) and *G(a)* for an example population where $a_0 = 10$ and $k = 0.8$.

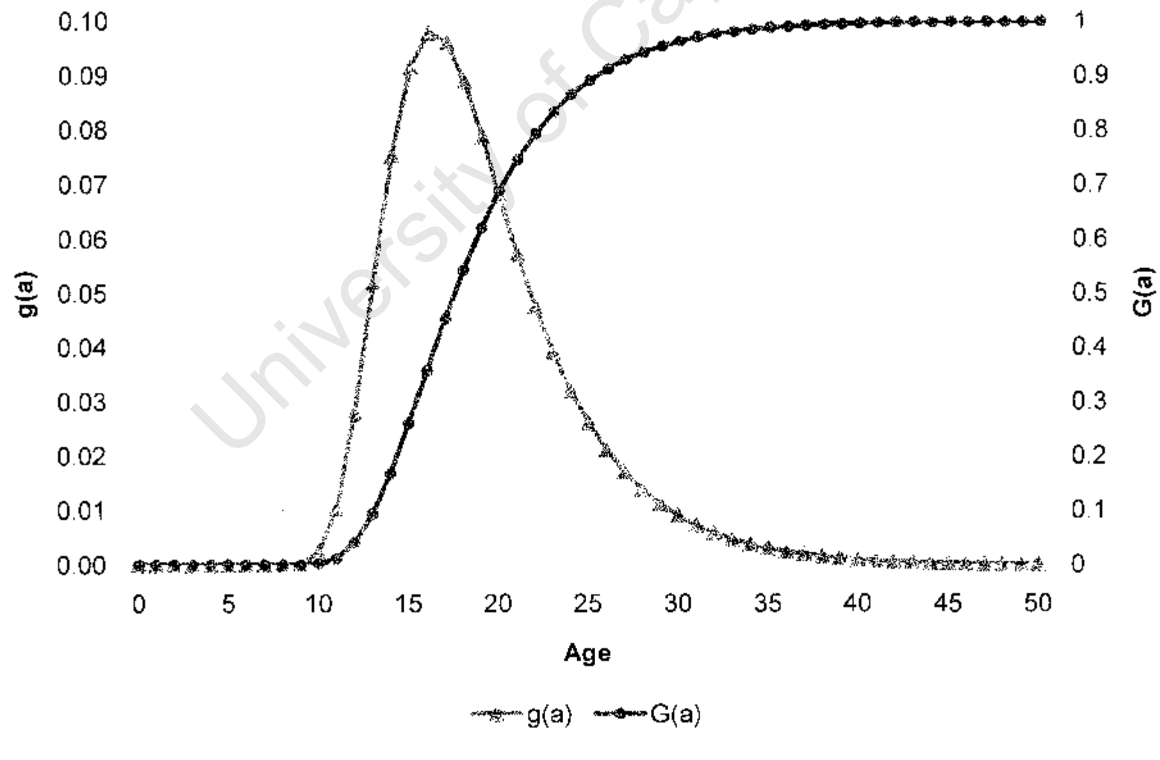

Figure D. 1: A plot of g(a) and G(a) on the same x-axis

Since *G(a)* is found by integrating *g(a)*, the value of *G(a)* increases at an increasing rate as *g(a)* increases to its maximum. After *g(a)* reaches its peak it decreases and the rate of increase for *G(a)* slows down giving rise to the characteristic S-shape (Figure D.1). The maximum point of *g(a)* can also be seen to be the inflection point of *G(a)*.

For further analysis it is useful to define a function, f(a), such that Equation D.3 holds.

$$f(a) = -0.174W - \exp(-0.2881W)$$ **Equation D. 3**

where W is defined as in Equation D.2. Defining f(a) in such a way changes Equation D.1 to Equation D.4.

$$g(a) = \frac{0.19465}{k}\exp[f(a)]$$ **Equation D. 4**

Figure D.2 shows plots f(a) on the same x-axis as *g(a)*. It can be seen that *g(a)* increases as f(a) increases and decreases when f(a) decreases. In order to find the inflection point of *G(a)* the maximum of *g(a)* must be found. However, inspection of Figure D.2 shows that this can be achieved by finding the maximum of f(a).

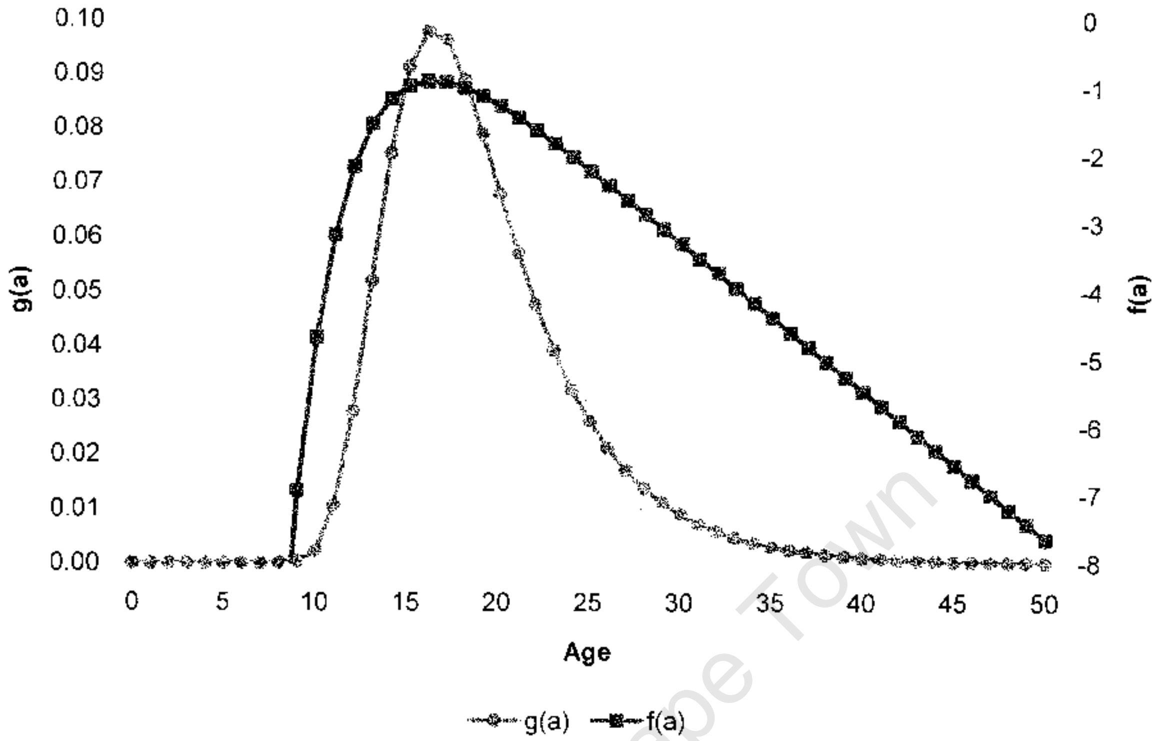

Figure D. 2: f(a) and g(a) plotted on the same x-axis

The maximum of f(a) can be found by setting the derivative, f′(a), equal to zero and solving for the age, a (Equation D.5 and D.6).

$$f'(a) = \frac{\partial}{\partial a}\left\{-0.174*\left(\frac{a-a_0-6.06k}{k}\right) - e^{\left[-0.2881*\left(\frac{a-a_0-6.06k}{k}\right)\right]}\right\} \qquad \textbf{Equation D. 5}$$

$$= -0.174*\left(\frac{a}{k}\right) + 0.2881*\left(\frac{a}{k}\right)e^{\left[-0.2881*\left(\frac{a-a_0-6.06k}{k}\right)\right]} \qquad \textbf{Equation D. 6}$$

By setting Equation D.6 equal to zero and solving for *a* gives the age for the inflection point of *G(a)* (Equation D.7).

$$a = a_0 + 7.8103k$$ **Equation D. 7**

For the example population, used in Figures D.1 and D.2, where $a_0 = 10$ and $k = 0.8$ the maximum of f(a) and $g(a)$ - the inflection point of $G(a)$ - is $a = 16.248$ years.

APPENDIX E

Equation E.1 is the formula for the Brass polynomial as given by Gage (2001).

$$F(z) = c\int_{s}^{z} (x - s)(s + w - x)^2 \, dx \qquad \textbf{Equation E. 1}$$

If this integral can be evaluated directly to produce estimates of F(z) and F(z) without numerical procedures it would speed calculation and provide a useful alternative to the Coale-Trussell schedules. Consequently, the integral was manipulated using the properties of definite integrals in an effort to simplify the equation.

First, Equation E.2 is established by adding and subtracting w in the first term of the integral.

$$F(z) = c\int_{s}^{z} (x - s - w + w)(s + w - x)^2 \, dx \qquad \textbf{Equation E. 2}$$

Using the properties of integrals and rearranging terms results in Equation E.3 and simplifying to give Equation E.4. The integral can now be solved to give Equation E.5 and evaluated between s and z to yield Equation E.6.

$$F(z) = c[\int_{s}^{z} (x - s - w)(s + w - x)^2 \, dx + w\int_{s}^{z} (s + w - x)^2 \, dx] \qquad \textbf{Equation E. 3}$$

$$F(z) = c[\int_{s}^{z} -(s + w - x)^3 \, dx + w\int_{s}^{z} (s + w - x)^2 \, dx] \qquad \textbf{Equation E. 4}$$

$$F(z) = c\left[\frac{1}{4}(s + w - x)^4 \Big|_{s}^{z} - \frac{1}{3}w(s + w - x)^3 \Big|_{s}^{z}\right] \qquad \textbf{Equation E. 5}$$

$$F(z) = c\left[\frac{1}{4}(s+w-z)^4 - \frac{1}{3}w(s+w-z)^3 + \frac{1}{12}w^4\right]$$ **Equation E. 6**

By dividing through by the total fertility rate (TFR), the cumulative fertility, F(z), takes on a value of one at the end of the reproductive life. Standardising the cumulative fertility in this way removes the effect of level while leaving the pattern of fertility unchanged. Since the end of reproductive life by definition occurs at z = s + w then F(z) = F(s + w) = 1. This constraint allows the value of the constant, c, to be calculated as $c = \frac{12}{w^4}$.

Replacing c in Equation E.6 establishes Equation E.7 and by multiplying through by 12 inside the bracket and dividing by 12 outside gives Equation E.8.

$$F(z) = \frac{12}{w^4}\left[\frac{1}{4}(s+w-z)^4 - \frac{1}{3}w(s+w-z)^3 + \frac{1}{12}w^4\right]$$ **Equation E. 7**

$$F(z) = \frac{1}{w^4}\left[3(s+w-z)^4 - 4w(s+w-z)^3 + w^4\right]$$ **Equation E. 8**

Set t = s + w – z and simplify to yield Equation E.9.

$$F(z) = 3\frac{t^4}{w^4} - 4\frac{wt^3}{w^4} + 1$$ **Equation E. 9**

Equation E.10 is established by subtracting one both sides and multiplying by -1.

$$1 - F(z) = 4\frac{wt^3}{w^4} - 3\frac{t^4}{w^4}$$ **Equation E. 10**

The first term on the left-hand side of Equation E.10 can be simplified to give Equation E.11.

$$1 - F(z) = 4\frac{t^3}{w^3} - 3\frac{t^4}{w^4}$$ **Equation E. 11**

Let $p = \frac{t}{w} = \frac{s + w - z}{w}$ be the portion of the fertility period remaining, then Equation E.12 results.

$$1 - F(z) = 4p^3 - 3p^4$$ **Equation E. 12**

The equation is a function of the portion, *p*, of the fertility period remaining at age *z* and provides a simple and convenient way of determining cumulative fertility.

Retherford (1979) highlights additional properties of the original formulation of the Brass polynomial. In particular, he derives formulae for the mean age at childbearing, mean parities, median age at childbearing and the time distribution of children ever born. These formulae can easily be extended to apply to the generalised polynomial derived above and, as a consequence, any specific case once *s* and *w* have been selected.

APPENDIX F

The two standards developed in Chapter 4 can be interpreted as graduated rates. In order for a graduation (in this case the standards) to be appropriate a balance must be found between two often conflicting ideals: 1) adherence-to-data and 2) smoothness.

On the one hand 100 per cent adherence-to-data implies following the data exactly. This may lead to fluctuating rates and erratic progression. By contrast, smoothing observed rates may remove vital information. For example, if mortality rates are smoothed too much then the "accident hump" may be flattened out as a data anomaly.

In fact, the consistent understatement of f(45-49) for African populations arises from the mistaken belief that the observed uptick in 45-49 fertility is a data problem. In particular, poor data quality resulting from illiteracy and enumerator error, amongst others, has been blamed. However, evidence from the sub-Saharan Africa DI-IS indicates that this upwards trend is not an anomaly to be ignored and smoothed away.

A good standard must satisfactorily meet both of these contrasting goals. As such, to formalise the comparison of the standards with the average DHS fertility rates a number of graduation tests are used.

F.1 Goodness-of-Fit

Normally, four tests are required to check goodness-of-fit: a) A test for overall adherence to data, b) a test checking for a few large discrepancies offset by many small ones, c) the signs or cumulative deviations test for overall bias and d) a check for runs of over- or underestimation (Benjamin and Pollard 1992, p. 222 — p. 227)

The Chi-square statistic, test a), is normally used to check adherence-to-data, that is, to compare how closely the graduation fits the observed data. Similarly, the individual standardised deviations test is used to expose problem b). However, it came to light during the comparison of the standards that tests for a) and b) cannot be used in this analysis. A caveat of the Chi-square test - used in both a) and b) above - is that it does not work for small values (Benjamin and Pollard 1992).

Fortunately, the method used to obtain the two standards being investigated minimises the sum of squared error (abbreviated SSE) thus identifying the best-fitting schedules. These schedules are then investigated and a small observed error indicates a relatively close overall adherence to the data (fable F.1). If the observed squared-error is sufficiently small then this methodology approximates the Chi-square test.

The second test looks for a few large differences offset by many small differences. In such a case, the overall goodness-of-fit test may well be passed when, in fact, there is a significant misfit in one (or more) ages or age groups. Since the individual standardised deviations test cannot be used with these small values an alternative method must be found. Given that the problem with the Booth standard and the Coale-Trussell model is an understatement of fertility in the oldest age group, it is particularly important to check adherence in the 35-39, 40-44 and 45-49 age group.

As a result, the same procedure is followed as for overall goodness-of-fit and the sum of squared differences for the last three age groups are compared (Table F.1).

Table F. 1: Brass and Hadwiger differences used in the Signs and Grouping of Signs tests

Age Group	African ASFR	Hadwiger standard	Squared Error	% Difference	Brass standard	Squared Error	% Difference
15-19	0.12707	0.13020	0.00001	2.5%	0.13285	0.00003	4.5%
20-24	0.22663	0.22570	0.00000	-0.4%	0.20925	0.00030	-7.7%
25-29	0.22020	0.22538	0.00003	2.4%	0.22360	0.00001	1.5%
30-34	0.18969	0.18143	0.00007	-4.4%	0.19410	0.00002	2.3%
35-39	0.13968	0.12872	0.00012	-7.8%	0.13899	0.00000	-0.5%
40-44	0.07053	0.08440	0.00019	19.7%	0.07646	0.00004	8.4%
45-49	0.02621	0.02417	0.00000	-7.8%	0.02475	0.00000	-5.6%
SSE (15-49)			0.00042			0.00040	
SSE (35-49)			0.00032			0.00004	

The results in Table F.1 show that the Brass-based standard has better overall adherence-to-data and superior fit to the 35-49 age groups.

Since the least squares methodology uses squared error it ignores the direction of the differences. As such, it is possible that the majority of the age groups could have errors is the same direction - all positive or all negative - resulting in problem b). Two tests can be used to identify overall bias - the signs test and the cumulative deviations test. However, the cumulative deviations test has a zero value by design, since the standards and the DHS average population rates accumulate to one.

As a result, the signs test must be used. Table F.2 shows the differences between the derived standards and the DHS average population (columns 5 and 6 respectively).

Table F. 2: Brass and Hadwiger differences used in the Signs and Grouping of Signs tests

Age Group	DHS Average	Brass based Standard	Hadwiger based Standard	Brass Difference Values	Hadwiger Difference Values	Brass Signs	Hadwiger Signs
15-19	0.12707	0.13285	0.13020	0.00578	0.00313	+	+
20-24	0.22663	0.20925	0.22570	-0.01738	-0.00093	-	-
25-29	0.22020	0.22360	0.22538	0.00340	0.00518	+	+
30-34	0.18969	0.19410	0.18143	0.00441	-0.00826	+	-
35-39	0.13968	0.13899	0.12872	-0.00069	-0.01096	-	-
40-44	0.07053	0.07646	0.08440	0.00593	0.01387	+	+
45-49	0.02621	0.02475	0.02417	-0.00146	-0.00204	-	-

Table F.2 shows that there are four positive signs and three negatives signs for the Brass-based standard. Conversely, there are four negative signs and three positives for the Hadwiger-based standard. In both cases logic dictates that three values of a particular sign and four of the other sign demonstrates no consistent bias above or below the DHS population. In addition, the p-value calculated using a binomial test confirms these results. In both cases the p-value of 0.5 far exceeds the critical level of 0.05. As such the null hypothesis that there is no overall bias cannot be rejected.

Table F.2 would, usually, also be used for the grouping of signs test for runs of the same sign, but the low number of age groups makes the test unreliable. Similarly, the alternative test for runs or clumps - serial correlations test - cannot be used either. However, no clear evidence exists in columns 7 and 8 of Table F.2 to suggest that runs of the same sign exist.

Based on the above results, both standards appear to adhere well to the data. However, the standard derived using the Brass polynomial is superior in both the overall adherence and the adherence to the data for the oldest age groups. As such, the Brass based standard is favoured based on goodness-of-fit.

F.2 Third Differences

The second requirement of a good graduation is that the rates should progress smoothly. The test commonly used for smoothness is to analyse third differences. So, the standard rates calculated using the Brass polynomial and the Hadwiger function should show a regular progression in third differences. Tables F.3 and F.4 show the Brass polynomial and Hadwiger results respectively.

Table F. 3: First, second and third differences for Brass-based standard

Age Group	f(x,x+4)	1st Differences	2nd Differences	3rd Differences
15-19	0.13285			
		-0.07640		
20-24	0.20925		-0.06206	
		-0.01435		-0.01822
25-29	0.22360		-0.04384	
		0.02949		-0.01822
30-34	0.19410		-0.02562	
		0.05512		-0.01822
35-39	0.13899		-0.00741	
		0.06252		-0.01822
40-44	0.07646		0.01081	
		0.05172		
45-49	0.02475			

Table F. 4: First, second and third differences for Hadwiger-based standard

Age Group	f(x,x+4)	1st Differences	2nd Differences	3rd Differences
15-19	0.13020			
		-0.09551		
20-24	0.22570		-0.09583	
		0.00032		-0.05219
25-29	0.22538		-0.04364	
		0.04396		-0.03489
30-34	0.18143		-0.00875	
		0.05270		-0.01713
35-39	0.12872		0.00838	
		0.04432		0.02429
40-44	0.08440		-0.01591	
		0.06023		
45-49	0.02417			

The standard derived using the Brass polynomial shows smooth third differences (Table F.3). By contrast, the standard derived using the Hadwiger function exhibits an irregular progression in third differences (Table F.4).

As with the adherence to data, the standard derived using the Brass polynomial meets the criterion. By contrast the Hadwiger-based standard fails the smoothness test. Consequently, the standard based on the Hadwiger function can be rejected for the average African population because it fails both the smoothness and the goodness-of-fit criteria (greater misfit to 35-49 fertility than the Brass polynomial). By contrast, the Brass-based standard is accepted as the most appropriate for use in sub-Saharan Africa settings.

APPENDIX G

In order to test the suitability of the African standard in sub-Saharan settings, two census populations - Kenya 1979 and Botswana 2001 - were analysed. The relational Gompertz model was used without half-year shift for both populations.

Table G.1 gives the original data for the Kenya population.

Table G. 1: Original data from the 1979 Kenya Census

Age Group	P(i)	f(x)
15-19	0.3206	0.0984
20-24	1.8529	0.2640
25-29	3.6521	0.2758
30-34	5.3881	0.2260
35-39	6.4703	0.1727
40-44	7.0215	0.0907
45-49	7.1735	0.0416

Applying the standard relational Gompertz model as set out by Zaba (1981) and using the Booth standard gives the results in Table G.2.

Table G. 2: Summary results using the Booth Standard

	ASFRs		Parities
Age Group	Shift	No Shift	
10-14	0.0016	0.0029	0.0016
15-19	0.1482	0.1767	0.3201
20-24	0.3615	0.3709	1.7917
25-29	0.3837	0.3812	3.7150
30-34	0.3294	0.3220	5.4916
35-39	0.2459	0.2364	6.8966
40-44	0.1253	0.1111	7.7858
45-49	0.0202	0.0149	8.0584
TFR	8.0712	8.0653	

Applying the African standard developed in Chapter 4 and using the recalculated e(x), z(x), e(i) and z(i) values yields the results in Table G.3 for the 1979 Kenya Census.

Table G. 3: Summary results using the African Standard

	ASFRs		Parities
Age Group	Shift	No Shift	
10-14	0.0081	0.0139	0.0574
15-19	0.1671	0.1881	0.5779
20-24	0.3269	0.3369	1.8664
25-29	0.3720	0.3713	3.6287
30-34	0.3270	0.3193	5.3839
35-39	0.2318	0.2210	6.7457
40-44	0.1250	0.1150	7.5814
45-49	0.0403	0.0338	7.9372
TFR	7.9504	7.9268	

Similarly, the 2001 Botswana Census data (Table G.4) can be analysed using the relational Gompertz model with the Booth standard (Table G.5) and the African standard (Table G.6).

Table G. 4: Original data from the 2001 Botswana Census

Age Group	P(i)	f(x)
15-19	0.13407	0.04220
20-24	0.85889	0.12293
25-29	1.68648	0.11138
30-34	2.66135	0.10595
35-39	3.61718	0.08689
40-44	4.57589	0.05109
45-49	5.27185	0.01970

Table G. 5: Summary results using the Booth Standard

	ASFRs		Parities
Age Group	Shift	No Shift	
10-14	0.0008	0.0015	0.0009
15-19	0.0663	0.0790	0.1451
20-24	0.1649	0.1699	0.8063
25-29	0.1846	0.1844	1.7109
30-34	0.1682	0.1655	2.5958
35-39	0.1343	0.1301	3.3411
40-44	0.0745	0.0668	3.8493
45-49	0.0136	0.0102	4.0211
TFR	4.0320	4.0294	

Table G. 6: Summary results using the African Standard

	ASFRs		Parities
Age Group	Shift	No Shift	
10-14	0.0054	0.0088	0.0370
15-19	0.0806	0.0895	0.2924
20-24	0.1496	0.1544	0.8821
25-29	0.1766	0.1773	1.7027
30-34	0.1666	0.1640	2.5693
35-39	0.1292	0.1245	3.2990
40-44	0.0781	0.0728	3.7948
45-49	0.0296	0.0256	4.0356
TFR	4.0517	4.0409	

REFERENCES

Anderson, Barbara A. and Brian D. Silver. 1992. "A Simple Measure of Fertility Control", *Demography* **29**(3):343-356. http://www.jstor.org/stable/2061822 Accessed: 12 March 2009.

Benjamin, B. and J.H. Pollard. 1992. *The Analysis of Mortality and Other Actuarial Statistic.* Oxford: Butterworth-Heinemann Ltd.

Blake, J. 1985. "The Fertility Transition: Continuity or Discontinuity with the Past?", International Population Conference 4. Liège, IUSSP, pp. 393-405.

Bongaarts, John. 1994. "Population Policy Options in the Developing World", *Science* **263**(5148):771-776.

Booth, Heather. 1979. "The Estimation of Fertility from Incomplete Cohort Data by means of the Transformed Gompertz Model", Unpublished Ph.D., University of London.

Booth, Heather. 1984. "Transforming Gompertz's Function for Fertility Analysis: The Development of a Standard for the Relational Gompertz Function", *Population Studies* **38**:495-506.

Brass, W. and P. Airey. 1988. *The Relational Gompertz Model of Fertility.* London: Centre for Population Studies, London School of Hygiene and Tropical Medicine.

Brass, William. 1968. "Methods of Analysis and Estimation", In Brass, William, Ansley Coale, Paul Demeny, et al. (ed). *The Demography of Tropical Africa.* Princeton: Princeton University Press.

Brass, William. 1974. "Perspectives in Population Prediction: Illustrated by Statistics of England and Wales", *Journal of the Royal Statistical Society,* A 137(4):532-583.

Brass, William. 1975. "Methods for Estimating Fertility and Mortality from Limited and Defective Data", Laboratories for Population Studies An Occasional Publication, Carolina Population Center.

Brass, William. 1996. "Demographic Data Analysis in Less Developed Countries: 1946-1996", *Population Studies* **50**(3):451-467.

Budlender, Debbie, Ntebaleng Chobokoane and Sandile Simelane. 2004. "Marriage Patterns in South Africa: Methodological and Substantive Issues", *Southern African Journal of Demography* **9**(1):1-26.

Caldwell, John C. and Pat Caldwell. 2002. "The Fertility Transition is Sub-Saharan Africa", Paper presented at the Conference Fertility and the Current South African Issues of Poverty, HIV/AIDS and Youth. Pretoria, 24 October 2002.

Cleland, John. 1996. "Demographic Data Collection in Less Developed Countries: 1946-1996", *Population Studies* **50**(3):433-450.

Coale, Ansley J. and Paul Demeny. 1983. *Regional Model Life Tables and Stable Populations.* New York: Academic Press, Inc.

Coale, Ansley J. and Donald R. McNeil. 1972. "The Distribution by Age of the Frequency of First Marriage in a Female Cohort", *Journal of the American Statistical Association* 67(340): 743-749.

Coale, A. and J. Trussell. 1974. "Model Fertility Schedules: Variations in the Age Structure of Childbearing in Human Populations", *Population Index* **40**(2):185-258.

Cohen, Barney. 1993. "Fertility Levels, Differentials, and Trends," In Foote, Karen, A., Kenneth H. Hill and Linda G. Martin (ed). *Demographic Change in Sub-Saharan Africa.* Washington D.C.: National Academy Press, pp. 8-67.

Cohen, Barney. 1998. "The Emerging Fertility Transition in Sub Saharan Africa", *World Development* **26**(8):1431-1461.

Demeny, Paul. 2003. "Population Policy: A Concise Summary", *Policy Research Division Working Paper* no. 173. New York: Population Council. http://www.popcouncil.org/pdfs/wp/173.pdf Accessed: 02 March 2009.

Department of Health, Medical Research Council, OrcMacro. 2007. *South Africa Demographic and Health Survey 2003* Pretoria: Department of Health.

El-Badry, M.A. 1961. "Failure of Enumerators to Make Entries of Zero: Error in Recording Childless Cases in Population Censuses", *Journal of the American Statistical Association* **56**(296):909-924.

Feeney, Griffith. 1998. "A New Interpretation of Brass' P/F Ratio Method Applicable When Fertility is Declining", electronic resource. http://www.gfeeney.com/notes/pfnote/pfnote.htm Accessed: 02 March 2009.

Gage, Timothy B. 2001. "Age-Specific Fecundity of Mammalian Populations: A Test of Three Mathematical Models", *Zoo Biology* **20**:487-499.

Garenne, M. and V. Joseph. 2002. "The Timing of the Fertility Transition in Sub-Saharan Africa", *World Development* **30**(10):1835-1843.

Guengant, Jean-Pierre and John F. May. 2001. *Impact of the Proximate Determinants on the Future Course of Fertility in sub-Saharan Africa*, Workshop on Prospects for Fertility Decline in High Fertility Countries. United Nations Population Division. New York, 9-11 July 2001.

Heaton, L.M. and K.A. Stanecki. 2003. "A Descriptive Analysis of Dependency Ratios and Other Demographic Indicators in Populations Affected by HIV/AIDS," Paper presented at a Scientific Meeting on the Empirical Evidence for the Demographic and Socio-Economic Impact of AIDS. Durban, South Africa. US Census Bureau Population Division.Data on Natural Fertility"

Henry, Louis. 1961."Some Data on Natural Fertility", *Eugenics Quarterly* **8:**81-91.

Hinde, Andrew. 1998. *Demographic Methods: Chapter 10 The Determinants of Fertility.* New York: Arnold/Oxford.

Hobcraft, J.N., N. Goldman and V.C. Chidambaram. 1982. "Advances in the P/F Ratio Method for the Analysis of Birth History", *Population Studies* **36**(2):291-316.

Hoem, Jan M., Dan Madsen, Jorgen Lovgreen Nielsen, Else-Marie Ohlsen, Hans Oluf Hansen and Bo Rennermalm. 1981. "Experiments in Modelling Recent Danish Fertility Curves", *Demography* **18**(2):231-244. http://www.jstor.org/stable/2061095 Accessed: 12 March 2009.

Lam, David. 2005. "How the World Survived the Population Bomb: An Economic Perspective," In Asefa, Sisay (ed). *The Economics of Sustainable Development.* Kalamazoo: Upjohn Institute for Employment Research, pp. 99-132.

Lewis, James J.C., Carine Ronsman, Alex Ezeh and Simon Gregson. 2004. "The Population Impact of HIV on Fertility in Sub-Saharan Africa", *AIDS* **18**(suppl 2):S35-S43.

Livi-Bacci, Massimo. 1968. "Fertility and Population Growth in Spain in the Eighteenth and Nineteenth Centuries", *Daedalus* **97**(2):523-535.

Merrick, Thomas W. 2002. "Population and Poverty: New Views on an Old Controversy", *International Family Planning Perspectives* **28**(1):41-46.

MeasureDHS. 2009. Statcompiler. http://www.statcompiler.com/ Accessed: 01 June 2009.

Moultrie, Tom A. and Rob Dorrington. 2008. "Sources of Error and Bias in Methods of Fertility Estimation Contingent on the P/F Ratio in a Time of Declining Fertility and Rising Mortality", *Demographic Research* **19**:1635-1662. http://www.demographic-research.org/volumes/vol19/46/ Accessed: 02 March 2009.

Moultrie, T.A. and Timæus, I.M. 2002. *Trends in South African Fertility between 1970 and 1998: An Analysis of the 1996 Census and the 1998 Demographic and Health Survey.* Cape Town: Medical Research Council. http://www.mrc.ac.za/bod/trends.pdf Accessed: 13 March 2009.

Pevistera, P. and A. Kostaki. 2007. "Modeling fertility in modern populations", *Demographic Research* **16**:141-194

Potter, J.E. 1977. "Problems in using Birth-History Analysis to Estimate Trends in Fertility", *Population Studies* **31**(2):335-364.

Retherford, R.D. 1979. "The Brass Fertility Polynomial", *Asian and pacific Census Forum* **5**(4):15-19, 22

Shryock, Henry S. and Jacob S. Siegel. 1976. *The Methods and Materials of Demography.* San Diego: Academic Press.

Sloggett, Andy (ed). 1994. *Estimation of Demographic Parameters from Census Data.* Tokyo, Japan: Statistical Institute for Asia and the Pacific.

Statistical Office of the United Nations. 1966. *Demographic Yearbook 1965.* New York: United Nations.

Tomasson, Richard F. 1998. "*Modern Sweden: The Declining Importance of Marriage*", Scandinavian Review **86**(2)83-89.

Udjo, Eric O. 2003. "A Re-examination of Levels and Differential in Fertility in South Africa from Recent Evidence", *Journal of Biosocial Science* **35**:413-431.

United Nations, Population Division. 1983. "Manual X: Indirect Techniques for Demographic Estimation", *Population Studies* **81**

Wilson, C., J. Oeppen and M. Pardoe. 1988. "What is Natural Fertility? The Modelling of a Concept", *Population Index* **54**(1): 4-20. Stable URL: http://www.jstor.org/stable/3644106 Accessed: 27 May 2009.

Xie , Yu. 1990. "What Is Natural Fertility? Remodeling of a Concept." *Population Index* **56**(4):656-663. http://www.jstor.org/stable/3645028
Accessed: 02 March 2009.

Xie, Yu and Pimentel, Ellen E. 1992. "Age Patterns of Marital Fertility: Revising the Coale-Trussell Method", *Journal of the American Statistical Association* **87**(420):977-984.

Zaba, Basia. 1981. *Use of the Relational Gompertz Model in Analysing Fertility Data Collected in Retrospective Surveys.* London: Centre for Population Studies, London School of Hygiene and Tropical Medicine.

Zaba, Basia and Simon Gregson. 1998. "Measuring the Impact of HIV on Fertility in Africa", *AIDS* **12**(suppl 1):S41-S50.