A multi-regional adaptation of the ASSA2008 AIDS and Demographic Model for population projections

Greaves Thabo Mofokeng
MFKGRE001
University of Cape Town

Dissertation submitted in partial fulfilment of the requirements for the degree of Master of Philosophy in Demography to the Centre for Actuarial Research

November 2013

PLAGIARISM DECLARATION

I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own. I have used the Harvard-UCT 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 dissertation 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:

Greaves Thabo Mofokeng
Date:

ABSTRACT

The traditional models used in population projections rely on the net migration method. The ASSA2008 AIDS and Demographic model is one such model.

In this research, the nine provinces and the four population groups are aggregated to give rise to three regions. Using STATA12, the directional migration tables for the years covering the period 1996-2007 between the three regions, by age and sex, based on a 10% sample of the 2001 Census and a 2.5% sample of the 2007 Community Survey, are produced.

Using MATLAB *2011a* with built-in Levenberg-Marquardt algorithm with nonlinear least squares methods, Rogers-Castro multi-exponential age schedules are fitted to the census/survey migration data in order to obtain parameters used to estimate migration rates in the model for the period 1996-2007. After 2007, migration rates are extrapolated roughly linearly, assuming that migration will trend towards zero over a fixed number of years.

The multi-regional adaptation of the ASSA model is tested and found to work, with a minor re-calibration to the HIV data for 2008.

The projected regional population age structure and size implied by the model for 1996-2025 are consistent with the same estimates implied by the net migration model, and so are the projected net migration rates per 1,000. The level of the migration rates assumed in the multi-regional model accounts for an average of 89% of the change in the estimates of the population size relative to those generated by the net migration model, and the use of multi-regional modelling itself accounts for 11% of these changes.

The proportions of the changes attributable to the level of migration rates assumed in the multi-regional model, and the use of the multi-regional modelling, show that the choice of the method by which population projections are done is important.

Finally, the three-region model can be extended to a nine-province model that recognises that each province has unique demographic dynamics, but the construction of such a model requires a significant amount of extra work due to its size and complexity.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Prof. Rob Dorrington, for guiding my research and providing the expert supervision throughout this study. The project was difficult and technical, but Prof. Dorrington patiently guided me through the intricacies of the research and helped me build much needed skills.

My sincere thanks also go to Dr. Jani Morrissey-Little of the University of Colorado at Boulder for her unrelenting assistance in helping me fully understand the mathematical properties associated with this research project.

I also wish to extend my heartfelt thanks to the Centre for Actuarial Research for their role in providing the formal training aimed at developing a competent research demographer.

I also thank the Hewlett Foundation and the University of Cape Town Postgraduate Funding Office for their financial support throughout my studies with the Centre for Actuarial Research and my stay in Cape Town.

Very special thanks go to my dear wife Charity, and beloved daughter Mamello, for their patience and unselfish support in seeing me complete this research project and the entire programme. These two have especially been my pillars throughout the programme, and I will always be appreciative of their personal sacrifices in this instance. In addition to my family, my prayerful thanks go to my dear God, Jehovah, for lifting me up when my future in the programme looked uncertain. God directed matters for me and I will always live to tell others about this experience.

Greaves Thabo Mofokeng
Cape Town, November 2013

1 INTRODUCTION

1.1 Background

The ASSA2008 AIDS and Demographic Model (Dorrington, Johnson and Budlender 2010), henceforth referred to as the ASSA (Actuarial Society of South Africa) model, uses the cohort-component projection method for population projection and incorporates the demographic impact of HIV/AIDS. The model includes a table of the net numbers of migrants for each sex at individual ages, for each year from 1985 to 2025. The model projects the population into the future, taking into account these values. The problem with the net migration approach is that only the differences between inward and outward migration are taken into account when estimating the numbers of migrants moving between provinces, without regard for the population age distribution of the sending region.

This weakness is described more clearly in Section 2.1.4 where the demographic accounting equation that underlies it is compared to one that relies on the multi-regional projection methodology.

That section demonstrates the fundamental differences in the treatment of internal migration as one projects future population numbers.

The main reason for the extensive use of the net migration approach is that the method does not require detailed census migration data. The multi-regional modelling approach, on the other hand, requires evidence of directional migration streams from which age-sex-specific out-migration rates can be estimated.

1.2 Statement of the problem

Despite efforts made to carry out population projections as accurately as possible, the impact of migration on the projected population size and age structure is not always known, neither is the impact of the method used to perform population projections. This dissertation attempts to determine if allowing for multi-regional migration in the projection of regional populations in the context of South Africa makes a substantial difference to population estimates relative to the use of net migration modelling.

1.3 Geographic make-up of South Africa

South Africa is divided into nine provinces and these are, in no particular order, Gauteng, Western Cape, KwaZulu-Natal, Free State, Limpopo, Mpumalanga, Northern Cape, Eastern Cape and the North-West. These provinces didn't exist prior to 1991. For the purposes of this research, the nine provinces are collapsed into three regions, namely Gauteng, Western Cape and rest of South Africa. There are sound reasons for collapsing the provinces this way. First, the 'rest of South Africa' is predominantly rural and relatively under-developed. Second, Gauteng and the Western Cape are economic hubs of South Africa and hence attract migrants from the other seven provinces for reasons that will be explained later in Section 4.1.

Combining the seven provinces into one region required the aggregation of their demographic, epidemiological and behavioural parameters as model input variables. The model needed as input a set of mortality rates and fertility rates from 1985 to 2007, and a set of net numbers of migrants from 1985 to 2025, and the base population for each of the three regions as at the middle of 1985.

In setting up the 'rest of South Africa' region, these input parameters were aggregated. For example, the seven provincial base populations by single ages, and the provincial total populations, were summed in order to derive the base population numbers for the region. The same calculation was done for the annual net numbers of migrants for this region.

However, the mortality and fertility assumptions were treated differently. The mortality and fertility rates for the seven provinces were combined as weighted averages for the region. These parameters were weighted by race and by province. Additional parameters that were summed and those that were derived as weighted averages are discussed in Section 3.2.

1.4 Objectives of the research

This dissertation answers the research question above by adapting the ASSA model in order to project the national population multi-regionally, using available data and assumptions to derive the necessary origin-destination migration flows. The demographic assumptions of the nine provinces and the four race groups are collapsed into three regional populations and then used as input parameters for the model.

1.5 Significance of the research

The major advantage of the multi-regional projection methodology is that the approach takes into account the age distribution of the population from which the migrants arise.

This research illustrates an important property of the multi-regional projection methodology, namely that the approach allows for greater information because it treats the population (that gives rise to the migrants) transparently, and is thus sensitive to the changes in the age distribution of that population. The research also makes use of the consistency in the age distribution of migrants over time, and this facilitates plausible projections.

Further, the dissertation uses the ASSA model, a population projection model that allows for the impact of HIV/AIDS, which is necessary because South Africa experiences high HIV incidence and prevalence levels as well as AIDS-related deaths. The model is also well-documented.

Finally, since seven of the nine provinces are predominantly rural and thus grouped together as 'rest of South Africa', the research adds value by assisting in comparing migration patterns between rural parts of the country and urban parts. This, in turn, informs policy planning on migration and development.

Multi-regional modelling has its uses in many scenarios. For instance, the National Planning Commission would require detailed and reasonably accurate population projections in order to plan better for the country's service delivery needs. The Commission, as well as the provinces, also require these types of projections in order to generate internal migration scenarios that inform policy on service delivery, transport infrastructure and other basic services. This is especially true with the Commission's migration scenarios as shown in the National Development Plan[1]. One of the migration scenarios in the document includes constant numbers of migrants moving into Gauteng. This assumption is not plausible since it would require migration levels that would sustain this scenario of migration. Multi-regional models are therefore important in informing such scenarios

1.6 Organisation of the dissertation

The dissertation comprises five chapters. The first chapter presents the background, objectives and relevance of the study. Chapter 2 reviews the demographic literature on multi-regional modelling, the ASSA model and period migration rates and patterns in South Africa. Chapter 3 presents the type and sources of data, describes the method for adapting the ASSA model and how the input parameters for the model are set up, and the estimation and projection of age-sex-specific migration rates in the model. The chapter also outlines the assumptions that underlie the estimation and projection of migration rates. Chapter 4 presents the results of the research. Chapter 5 concludes the research with a discussion of the results and indicates ideas for further research.

[1] The National Development Plan can be accessed on http://www.npconline.co.za/MediaLib/Downloads/Downloads/NDP

This chapter reviews the literature on the multi-regional population projection methodology and the spatial view of the population. Section 2.1 describes the mathematics of multi-regional population modelling. This section starts with developing the foundational mathematics of the uniregional projection model, where the population is projected without regard for internal migration flows. The mathematics of the uniregional model, also called the Leslie matrix in demographic literature, is then extended to a bi-regional projection model. The three-region model is easily developed from the bi-regional model by simply extending the mathematics of this system to a three-region system. This section then concludes by looking at empirical examples in which results were obtained by applying the multi-regional models to specific populations.

In Section 2.2 literature on internal migration in South Africa is reviewed. Particular attention is paid to levels and age patterns of inter-provincial migration and data issues. In Section 2.3, literature on the ASSA model is reviewed.

2.1 Multi-regional projection models

Multi-state demography is a branch of formal demography which includes the field of multi-regional modelling. Multi-state demography is the study of transition patterns between multiple states. For example, the study of patterns between the states of being single to that of being married, married to divorced, and divorced to re-married. Another example is labour force transitions from the 'employed' state to the 'unemployed' state. Thus, in demography one studies transition patterns using life tables, and such life tables include multiple decrement life tables, work life tables, nuptiality tables, tables of educational life, and multi-regional life tables (Rogers and Willekens 1986b). Life tables provide the estimates of transition probabilities. In a mortality table, these can be single decrement events such as death, multiple-decrement events such as death, out-migration, etc. There can also be increment-decrement probabilities, such as those recording entries such as births or in-migration as well as exits such as out-migration or death. These life table models in fact form part of the general class of increment-decrement life tables referred to as multi-state life tables. It is further noted that populations stratified by states of existence can also be projected using multi-state population projections (Rogers and Willekens 1986a).

Multi-state demography can be illustrated by making reference to multi-state life table models available to measure survivorship probabilities. Rogers and Willekens

(1986b) describe two examples in this regard, and these are associated with life tables of working life and married life.

Regarding the labour force example, the authors demonstrate the three assumptions used in estimating labour force expectancies. First, entry into the labour force occurs prior to the peak age of active life. The second assumption in building the labour force life table is that retirement occurs after the peak age of active life in the force, and the third assumption is that all persons in the labour force after a particular age experience the same mortality patterns. These assumptions contrast with standard life tables of working life that are constructed with stocks, rather than flows, as their primary focus. Multi-state life table models estimate survivorship proportions that will be used in devising a multi-state projection model that will yield estimates of future labour force totals. This way, one can handle the flows of persons into, and through, the labour force as they age and then exit the force.

The second example is the application of multi-state life tables to marital states. While the traditional form of these tables do not permit the formerly married portion of the population to re-marry and re-enter the married population, multi-state life tables allow for this event. The result is that influences of current married statuses on future statuses are incorporated, and thus more accurate measures of life expectancy in the state of being married, as well as being divorced or widowed, are derived. The regional dimension can also be incorporated to study the association between marital statuses and interregional migration (Rogers and Willekens 1986b).

Multi-regional demography is considered part of the wider field of multi-state demography. Formal multi-regional demography is defined as a branch in demography that describes the evolution of human populations over time and space mathematically (Rogers 1995c).

One of the more important tools in multi-regional demography is the multi-regional life table. It incorporates multiple modes of entry into and exit out of the population, making it a multiple increment-decrement life table. One becomes a member of a population by being born into it or migrating into it, and one exits from a population by dying or migrating out of it.

Multi-regional life tables have as their point of departure the multiple-radix that recognises that the birth cohort being considered is not homogeneous (Rogers and Willekens 1986b). They recognise that regional populations experience unique mortality and fertility regimes that affect their individual age distributions over time. These multi-regional life tables were developed in order to enable demographers to study internal migration and the extent to which regional population age distributions are affected by the migration, as well as how the regional populations interact within the parent population (Rogers 1986; Rogers and Willekens 1986b).

Multi-regional demography, therefore, provides improved tools to the demographer performing population projections. The field allows one to study the extent to which fertility, mortality and migration propensities interact and how these determine the 'constituent states' of a population and its distribution across all regions (Land and Rogers 1982).

Section 2.1.1 develops the uniregional framework from the foundational literature. Section 2.1.2 builds on the uniregional framework by extending the uniregional projection methodology to a bi-regional framework, describing how a bi-regional model can be opened to migration and how this system can be represented by multiple uniregional models that are open to migration. This is the net migration approach. Section 2.1.3 develops the methodology further by extending the bi-regional techniques to cover multi-regional projection methods. This will be illustrated by simple extensions to the bi-regional population systems and adapting these to three-region systems. Sections 2.1.4 and 2.1.5 demonstrate differences seen between the two sets of models by looking at empirical results seen in specific countries around the world.

2.1.1 The uniregional system

A multi-regional projection model may best be understood if one starts off by setting up a uniregional model that is closed to migration and represented by a matrix Equation. Ignoring migration, demographic projection models traditionally use the cohort-component projection method, where the projected population vector $\mathbf{P}^{t+h}$ at time $t+h$ is a function of the product of the base population vector $\mathbf{P}^{t}$ at time t and the growth matrix $\mathbf{G}$, which consists of a set of birth and survivorship elements.

Applications of demographic projection models are generally carried out by disaggregating the population of interest by age or age group and sex. A uniregional projection model projects the population forward with age and sex as the disaggregation factors and a multi-regional projection model further disaggregates the population by region.

For simplicity, consider first a uniregional population projection model that is closed to migration, with the two sexes combined. The population is disaggregated by 18 five-year age groups at time t and these numbers are contained in the vector $\mathbf{P}^{t}$. The projection horizon is five years and the population aged x to $x+4$ that would have survived the unit time interval of five years into the $x+5$ to $x+9$ age group would then be expressed as follows:

$$\mathbf{P}^{t+5} = \mathbf{S}.\mathbf{P}^{t} \ , \qquad (1)$$

where $\mathbf{P}^{t+5}$ is the vector of the projected cohort of survivors aged $x+5$ to $x+9$ at time $t+5$ and $\mathbf{S}$ is a square matrix with dimension 18. It contains all zeros in the first row, ignoring births for the moment, and the estimates of survivorship proportions are on the off-diagonal. These survivorship proportions are derived from a life table describing the mortality pattern experienced by the population aged x to $x+4$ at time t having survived five years.

Equation 1 above excludes births, for the moment, and can be described as the survivorship matrix $\mathbf{S}$ or a growth matrix consisting of survivorship proportions and zero birth elements. The survivorship matrix is defined as

$$\mathbf{S} = \begin{bmatrix} 0 & 0 & 0 & \cdots\cdots & 0 \\ \frac{L_5}{L_0} & 0 & 0 & \cdots\cdots & 0 \\ 0 & \frac{L_{10}}{L_5} & 0 & \cdots\cdots & 0 \\ \vdots & \vdots & \ddots & \cdots\cdots & \vdots \\ 0 & 0 & \cdots & \frac{T_{85}}{T_{80}} & \frac{T_{85}}{T_{80}} \end{bmatrix} .$$

For a particular group aged x to $x+4$, L_x represents the number of people in a stationary population based on the life table aged x to x+4 and L_{x+5} denotes the survivors to age x+5 to x+9 of those previously included in L_x. So the fraction of the population that survives to the next age group is defined as $\frac{L_{x+5}}{L_x}$.

Based on this specification of $\mathbf{S}$ the last age group would be projected as

$$P_{85+}^{t+5} = \frac{T_{85}}{T_{80}} P_{80}^{t} + \frac{T_{85}}{T_{80}} P_{85+}^{t} .$$

The **S** matrix may be specified with the precise elements as in the equation above as follows:

$$\mathbf{S} = \begin{bmatrix} 0 & 0 & 0 & \cdots\cdots & 0 \\ s_0 & 0 & 0 & \cdots\cdots & 0 \\ 0 & s_5 & 0 & \cdots\cdots & 0 \\ \vdots & \vdots & \ddots & \cdots\cdots & \vdots \\ 0 & 0 & \cdots & s_{80} & s_{85} \end{bmatrix} ,$$

where $s_x = \dfrac{L_{x+5}}{L_x}$ for each age group x=0 to 75 in increments of 5. The survivorship proportions, s_{80} and s_{85}, are defined as

$$S_{80} = S_{85} = \frac{T_{85}}{T_{80}}. \tag{1a}$$

We start by considering only the female population in the discussion of the estimation of numbers of births.

To project the first age group at time $t+5$, b_x is used to denote the average number of babies born during the time interval $(t, t+5)$ and alive at the end of the interval, per person aged x to $x+4$ living at time t (Rogers 1995a). That age group's contribution to the number in the first age group is $b_x P_x^t$. Age 15 is assumed to be the age at which females in this population start bearing children. It is denoted $\alpha = 15$. If age 49 is the assumed age at which they cease childbearing, then age 50 is the age by which child bearing has ceased. It is denoted $\beta = 50$. Let $\mathbf{B}$ be a square matrix of dimension 18. It contains all zeros except for the b_x elements in the first row.

$$B = \begin{bmatrix} 0 & 0 & b_{10} & b_{15} \cdots b_{45} & 0 \\ 0 & 0 & 0 & \ldots\ldots\ldots\ldots & 0 \\ 0 & 0 & 0 & \ldots\ldots\ldots\ldots & 0 \\ & \vdots & \ddots & \ldots\ldots\ldots\ldots & \vdots \\ 0 & 0 & \cdots & \ldots\ldots\ldots\ldots & 0 \end{bmatrix}$$

To derive the b_x elements we begin by defining B_x as the number of live female babies born during a calendar year to women aged x to $x+4$ at last birthday. The average annual age-specific birth rate for the age group x to x+4 is then defined as F_x. This is multiplied by the arithmetic mean of the initial and final populations aged x to $x+4$

$$\frac{1}{2}(P_x^t + P_x^{t+5}).F_x$$

to obtain the annual number of births. However, because $(t, t+5)$ is a five-year interval, the annual number of births is multiplied by 5 to give

$$\frac{5}{2}(P_x^t + P_x^{t+5}).F_x$$

which is the contribution of women aged x to $x+4$ to the total number of births during the five-year period. Therefore, assuming that the population of women aged x changes linearly over the five-year period, the total number of births in the population is

$$\frac{5}{2}\sum_{x=\alpha}^{\beta-5}(P_x^t + P_x^{t+5})F_x \tag{2}$$

where the summation is done in five-year steps.

Assuming that the births are uniformly distributed over the five years, the number of the babies who survive to the end of the year is:

$$\frac{L_0}{l_0} * \frac{1}{5} * \frac{5}{2}\sum_{x=\alpha}^{\beta-5}(P_x^t + P_x^{t+5}).F_x \text{ which equals } \frac{L_0}{2l_0}\sum_{x=\alpha}^{\beta-5}(P_x^t + P_x^{t+5}).F_x \,.$$

Based on Equations (1) and (1a), P_x^{t+5} is equivalent to $s_{x-5}P_{x-5}^t$ the **B** matrix can be defined explicitly as

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & \frac{L_0}{2l_0}(\frac{L_{15}}{L_{10}}.F_{15}) & \frac{L_0}{2l_0}(F_{15} + \frac{L_{20}}{L_{15}}.F_{20}) & 0 & \cdots \\ 0 & 0 & 0 & \cdots \quad \cdots & & \cdots \\ 0 & 0 & 0 & \cdots \quad \cdots & & \cdots \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ \vdots & \vdots & \cdots & \vdots & & \vdots \end{bmatrix}.$$

The complete growth matrix $\mathbf{G}$ then consists of a set of survivorship and birth elements $\mathbf{S}$ and $\mathbf{B}$, and can be defined as

$$\mathbf{G} = \mathbf{S} + \mathbf{B}.$$

The projection Equation becomes

$$\mathbf{P}^{t+5} = \mathbf{G}.\mathbf{P}^t \,,$$

And the generalized Leslie growth matrix $\mathbf{G}$ is explicitly defined as

$$\mathbf{G} = \begin{bmatrix} 0 & 0 & \frac{L_0}{2l_0}(\frac{L_{15}}{L_{10}}.F_{15}) & \frac{L_0}{2l_0}(F_{15} + \frac{L_{20}}{L_{15}}.F_{20}) & \cdots & 0 \\ \frac{L_5}{L_0} & 0 & 0 & 0 & \cdots & 0 \\ 0 & \frac{L_{10}}{L_5} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & \frac{T_{85}}{T_{80}} & \frac{T_{85}}{T_{80}} \end{bmatrix}$$

Note also that the assumptions that underlie this projection model are such that the mortality and fertility patterns experienced by this population will be stable for the five-year projection horizon, and that the population is closed to migration (Rogers 1985).

The general form of the uniregional projection system is expressed as

$$\begin{bmatrix} P_0^{t+5} \\ P_5^{t+5} \\ \vdots \\ \vdots \\ P_{85+}^{t+5} \end{bmatrix} = \begin{bmatrix} 0 & 0 & b_{\alpha-5} & \cdots b_{\beta-5} & 0 & \cdots & 0 \\ s_0 & 0 & & \ddots & & & \vdots \\ 0 & s_5 & & \ddots & & & \vdots \\ \vdots & \vdots & & \ddots & & & 0 \\ & & & & & & \vdots \\ & & & & & s_{80} & s_{85} \end{bmatrix} \cdot \begin{bmatrix} P_0^t \\ P_5^t \\ \vdots \\ \vdots \\ P_{85+}^t \end{bmatrix} \quad (3)$$

where α and β are the ages at which women start and end childbearing, respectively, and $s_{85} = s_{85} = \frac{T_{85}}{T_{80}}$. At this stage, the vectors $\mathbf{P}$ denote the female population at specific 5-year age group at times t and t+1.

In reconverting the single-sex population system to one in which we consider both sexes, the application of the growth matrix **G** differs in its application to the male population since it does not contain the **B** matrix because men are not expected to bear children in this population system. Note, however, that male births require the fertility of females so that the projection is complete.

2.1.2 The bi-regional system

The uniregional projection model can be extended to a bi-regional population system consisting of regions 1 and 2 to allow for interregional migration flows. Assume that this system is still closed to international migration and again considering the female population alone for representational convenience. The growth process is now a bi-regional system where mortality and out-migration decrement each region's age-specific populations and fertility increments the first age group for each region and in-migration increments all age groups for each region. The bi-regional population projection model is given by

$$\begin{bmatrix} \mathbf{P}_0^{t+5} \\ \mathbf{P}_5^{t+5} \\ \vdots \\ \cdot \\ \cdot \\ \cdot \\ \vdots \\ \mathbf{P}_{85}^{t+5} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{B}_{10} & \cdots \mathbf{B}_{45} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{S}_0 & \mathbf{0} & & \ddots & & & \vdots \\ \mathbf{0} & \mathbf{S}_5 & & \ddots & & & \vdots \\ \vdots & \vdots & & \ddots & & & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & & & \mathbf{S}_{80} & \mathbf{S}_{85} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{P}_0^t \\ \mathbf{P}_5^t \\ \vdots \\ \cdot \\ \cdot \\ \cdot \\ \vdots \\ \mathbf{P}_{85}^t \end{bmatrix}, \quad (4)$$

This looks very similar to the uniregional model set out in Equation 3. However, it is important to note that each of the elements, denoted $\mathbf{P}_x^{t+5}$ and $\mathbf{P}_x^t$, in the population vectors is now a 2-element sub-vector, and each of the elements in the growth matrix, $\mathbf{S}_x$, $\mathbf{B}_x$ and $\mathbf{0}$, are now sub-matrices with dimension 2.

Each sub-vector, $\mathbf{P}_x^{t+5}$ and $\mathbf{P}_x^t$, within the two-sex population vectors for *t+5* and *t*, contains the population numbers for region 1 and region 2 in the five-year age group beginning at age x as specified by

$$\mathbf{P}_x^{t+5} = \begin{bmatrix} P_{1,x}^{t+5} \\ P_{2,x}^{t+5} \end{bmatrix} \text{ and } \mathbf{P}_x^t = \begin{bmatrix} P_{1,x}^t \\ P_{2,x}^t \end{bmatrix}. \quad (4a)$$

The survivorship sub-matrices $\mathbf{S}_x$ are defined as

$$\mathbf{S}_x = \begin{bmatrix} {}_1s_{1x} & {}_2s_{1x} \\ {}_1s_{2x} & {}_2s_{2x} \end{bmatrix}$$

where ${}_is_{jx}$ denotes the proportion of residents of region i aged x to $x+4$ at time t who are alive and aged $x+5$ to $x+9$ in region j at time $t+5$.

Similarly, the birth sub-matrices $\mathbf{B}_x$ give the total number of surviving births during the unit time interval. They are defined as

$$\mathbf{B}_x = \begin{bmatrix} {}_1b_{1x} & {}_2b_{1x} \\ {}_1b_{2x} & {}_2b_{2x} \end{bmatrix} \quad (5)$$

where ${}_ib_{jx}$ is the mean number of babies born during the unit age-time interval, and living in region j at the end of that interval, per person aged x to $x+4$ at last birthday and living in region i at the start of the interval. The off-diagonal elements are measures of the mobility of children aged 0 to 4 years old, who were born to x to $x+4$ year old parents.

The $\mathbf{0}$ sub-matrices are defined simply as $\mathbf{0} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

The survivorship sub-matrices

It is conventional to define the survivorship sub-matrices $\mathbf{S}_x$ in terms of the bi-regional life table and the relationship is set out in the matrix Equation

$$\mathbf{S}_x = \mathbf{L}_{x+5}\mathbf{L}_x^{-1}, \text{ where} \quad (6)$$

$\mathbf{L}_x = \begin{bmatrix} {}_1L_{1x} & {}_2L_{1x} \\ {}_1L_{2x} & {}_2L_{2x} \end{bmatrix}$ and $\mathbf{L}_x^{-1}$ is the matrix inverse. Each element in $\mathbf{L}_x$, ${}_iL_{jx}$, gives the expected number of person-years lived in region *j* between ages x and $x+4$ by an individual born in region *i*. It denotes the duration of residence in region *j* by an *i*-born person and it depends on two components. First, the probability of surviving to age *x* and second, the time spent in region *i* in a five-year interval by a person of age x at the beginning of the interval. The duration of residence component may be measured for persons born in the given region and for persons living in the given region at age x. The linear approximation and the computational formula for $\mathbf{L}_x$ is defined as

$$\mathbf{L}_x = 5/2[\mathbf{l}_x + \mathbf{l}_{x+4}] \tag{7}$$

where

$$\mathbf{l}_x = \begin{bmatrix} {}_1l_{1x} & {}_2l_{1x} \\ {}_1l_{2x} & {}_2l_{2x} \end{bmatrix}$$

and each element ${}_il_{jx}$ gives the expected number of survivors in region *j* at exact age x who were born in region *i*. Note that in Equation (7), the assumption is that, on average, deaths occur half-way through the age group x and $x+4$. This assumption is made for algebraic convenience. . Although the approximation is reasonable enough for age 5 and above, where mortality is low and doesn't change rapidly with age, it is less accurate for the first age interval where mortality declines rapidly with age.

For birth cohorts of 100,000 in each region,

$$\mathbf{l}_0 = \begin{bmatrix} 100{,}000 & 0 \\ 0 & 100{,}000 \end{bmatrix}.$$

The computation of the multi-regional life table begins with the estimation of age-specific death and out-migration probabilities. These are the inputs for calculating the life table statistics and they will be denoted simply as $\mathbf{P}_x$ (without a *t* superscript) to distinguish them from the population sub-vectors.

In the two dimensional system $\mathbf{P}_x = \begin{bmatrix} {}_1p_{1x} & {}_2p_{1x} \\ {}_1p_{2x} & {}_2p_{2x} \end{bmatrix}$ and each element ${}_jp_{ix}$ represents the probability that a person in region *j* at exact age x will reside in region *i* at exact age $x+4$. These probabilities can be derived from the observed mortality and migration rates based on the population in the region of residence at the beginning of the interval. The procedures for estimating the age-specific death and out-migration probabilities will be described in a later section.

Given a matrix $\mathbf{P}_x$ for each age x, computation of the successive $\mathbf{l}_x$ matrices begins with the 100,000 births in each region which are the elements of the diagonal matrix $\mathbf{l}_0$, and babies are survived forward as follows:

$$\mathbf{l}_5 = \mathbf{P}_0\,\mathbf{l}_0$$

$$\mathbf{l}_{10} = \mathbf{P}_5\,\mathbf{l}_5$$

.

.

$$\mathbf{l}_{x+5} = \mathbf{P}_x\,\mathbf{l}_x$$

.

.

$\mathbf{l}_z = \mathbf{P}_{z-5}\,\mathbf{l}_{z-5}$

With the expected number of survivors at each age x in the two-region system, Equation 7 can be used to calculate the $\mathbf{L}_x$ matrices, and then from Equation 6 the survivorship sub-matrices are obtained from the life table. The survivorship matrices can also be computed from $\mathbf{P}_x$ with this computational formula:

$$\mathbf{S}_x = [\mathbf{I} + \mathbf{P}_{x+5}]\mathbf{P}_x\left[\mathbf{I} + \mathbf{P}_x\right]^{-1} \qquad (8)$$

The birth sub-matrices

The birth sub-matrices $\mathbf{B}_x$ were defined in Equation 5 and a computational formula can now be derived. If $\mathbf{F_x}$ is a diagonal matrix containing the annual regional birth rates of people aged x to $x+4$ and **I** is the identity matrix the computational formula is

$$\mathbf{B}_x = \frac{5}{4}[\mathbf{P}_0 + \mathbf{I}]\left[\mathbf{F}_x + \mathbf{F}_{x+5}\mathbf{S}_x\right]. \qquad (9)$$

Estimating the probabilities of death and out-migration

There are two ways of estimating the $\mathbf{P}_x$ matrices that give rise to the multi-regional life tables. The first is referred to as Option 1 (Rogers, 1995) and it can be used by countries that have a national population registration system for migration much like that which is common for births and deaths. This is also referred to as the 'movement approach'. The more common approach, known as Option 2, counts migrants from data obtained from the national census and other national surveys. The mathematical exposition of the two approaches is detailed in Rogers (1995a).

Option 1 begins by arranging annual out-migration and death occurrence-exposure rates to define the matrices $\mathbf{M}_x$ for each age x,

$$\mathbf{M}_x = \begin{bmatrix} {}_1M_{xd} + {}_1M_{2x} & -{}_2M_{1x} \\ -{}_1M_{2x} & {}_2M_{xd} + {}_2M_{1x} \end{bmatrix}$$

where ${}_iM_{jx}$ denotes the age-specific out-migration rate from region *i* to region *j* and ${}_iM_{dx}$ denotes the age-specific mortality rate in region *i*.

Estimating the survivorship sub-matrices for the open-ended age group

The two alternatives for closing out the life table depend on the approach for estimating the probability matrices (Rogers 1995a).

For the open-ended age group discussed earlier, the matrix $\mathbf{M}_z$ is used to close off the life table (Rogers 1985) and the Equation $\mathbf{L_z} = \mathbf{M}_z^{-1}\mathbf{l}_z$ corresponds to the uniregional counterpart in closing off the life table.

Option 1

Rogerson (1991) proposed the application of the following identity to find the matrix of survivorship proportions for the open-ended age group, where complete data for the 85+ age group are not available

$$\mathbf{S}_{85} = \mathbf{I} - \frac{\mathbf{5}}{\mathbf{2}}[\mathbf{I} + \mathbf{P}_{85}].\mathbf{M}_{85}$$

(Rogers 1995b).

Option 2

This approach complicates the calculation of the survivorship proportion for the final age interval because the numerator will always contain the open-ended interval. For instance, supposing that the final age group in the population is $z = 85$, the population at the start of the five-year time interval is in the age group (80-84). At the end of the time interval, the data will report not only the survivors in the age interval of 85+ that were aged 80-84, but also those that were older than 85 years at the start of the five-year interval. This way, one has persons in the age group 85+ that were 80-84 plus those that were 85+ at the start of the time interval of five years, in which case the numerator of the survivorship proportion exceeds the denominator. Assuming that this difference is not too great, the approximation of the survivorship matrix $\mathbf{P}$ is thus defined as

$$\mathbf{P}_{80} = \frac{1}{3}\left[2\bar{\mathbf{S}}_{75} + \bar{\mathbf{S}}_{80}\right]\mathbf{P}_{\sigma 80}$$

where $\bar{\mathbf{S}}_x$ is a matrix of conditional survivorship proportions at age x and $\mathbf{P}_{\sigma 80}$ is the diagonal matrix of survival probabilities at age 80 in each region, and reverses the effect of the conditional survival probability matrix at age 80, namely $\bar{\mathbf{P}}_{80}$ (Rogers 1995b).
Note that

$$\bar{\mathbf{P}}_{80} = \frac{1}{2}\left[\bar{\mathbf{S}}_{75} + \bar{\mathbf{S}}_{80}\right]$$ (Rogers 1995b).

2.1.3 The three-region system

The mathematics of the bi-regional population system discussed above can be extended to a three-region system. In this case, the population sub-vectors in Equation 4a, namely P_x^{t+5} and P_x^{t} can be extended to three-element vectors, such that

$$\mathrm{P}_x^{t+5} = \begin{bmatrix} P_{1,x}^{t+5} \\ P_{2,x}^{t+5} \\ P_{3,x}^{t+5} \end{bmatrix} \text{ and } \mathrm{P}_x^{t} = \begin{bmatrix} P_{1,x}^{t} \\ P_{2,x}^{t} \\ P_{3,x}^{t} \end{bmatrix}$$

The associated survivorship and birth sub-matrices $\mathbf{S}_x$ and $\mathbf{B}_x$ for this system, respectively, are defined as

$$\mathbf{S}_x = \begin{bmatrix} {}_1s_{1x} & {}_1s_{2x} & {}_1s_{3x} \\ {}_2s_{1x} & {}_2s_{2x} & {}_2s_{3x} \\ {}_3s_{1x} & {}_3s_{2x} & {}_3s_{3x} \end{bmatrix}$$

and

$$\mathbf{B}_x = \begin{bmatrix} {}_1b_{1x} & {}_1b_{2x} & {}_1b_{3x} \\ {}_2b_{1x} & {}_2b_{2x} & {}_2b_{3x} \\ {}_3b_{1x} & {}_3b_{2x} & {}_3b_{3x} \end{bmatrix} \quad .$$

where ${}_is_{jx}$ and ${}_ib_{jx}$ are elements in these sub-matrices and defined as in the 2-region population system.

The $\mathbf{0}$ sub-matrices are defined simply as $\mathbf{0} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

Survivorship sub-matrices

The survivorship sub-matrices described in terms of the bi-regional life table person-years lived in the age group x to $x+4$ at time t, $\mathbf{L_x}$ can be extended to survivorship sub-matrices in terms of the three-region life tables such that

$$\mathbf{L}_x = \begin{bmatrix} {}_1L_{1x} & {}_2L_{1x} & {}_3L_{1x} \\ {}_1L_{2x} & {}_2L_{2x} & {}_3L_{2x} \\ {}_1L_{3x} & {}_2L_{3x} & {}_3L_{3x} \end{bmatrix}$$ and as with the two-region model developed above, $\mathbf{L}_x^{-1}$ is the matrix inverse. Each element in $\mathbf{L}_x$, ${}_iL_{jx}$, gives the expected number of person-years lived in region j between ages x and $x+4$ by an individual born in region *i*. It denotes the duration of residence in region *j* by an *i*-born person and it depends on two components. First, the probability of surviving to age x and second, the time spent in region *i* in a five-year interval by a person of age x at the beginning of the interval. The duration of residence component may be measured for persons born in the given region

and for persons living in the given region at age x. The linear approximation and the computational formula for $\mathbf{L}_x$ is defined as

$$\mathbf{L}_x = 5/2[\mathbf{l}_x + \mathbf{l}_{x+5}] \qquad (10)$$

where

$$\mathbf{l}_x = \begin{bmatrix} {}_1l_{1x} & {}_2l_{1x} & {}_3l_{1x} \\ {}_1l_{2x} & {}_2l_{2x} & {}_3l_{2x} \\ {}_1l_{3x} & {}_2l_{3x} & {}_3l_{3x} \end{bmatrix}$$

and each element ${}_il_{jx}$ gives the expected number of survivors in region *j* at exact age x who were born in region *i*. For birth cohorts of 100,000 in each region,

$$\mathbf{l}_0 = \begin{bmatrix} 100{,}000 & 0 & 0 \\ 0 & 100{,}000 & 0 \\ 0 & 0 & 100{,}000 \end{bmatrix}.$$

As in the case of the two-region system, the computation of the three-region life tables starts off with the derivation of probabilities of death and out-migration denoted as $\mathbf{P}_x$. These will again be denoted as $\mathbf{P}_x$ (without a *t* superscript) to distinguish them from the population sub-vectors.

In the three dimensional system $\mathbf{P}_x = \begin{bmatrix} {}_1p_{1x} & {}_2p_{1x} & {}_3p_{1x} \\ {}_1p_{2x} & {}_2p_{2x} & {}_3p_{1x} \\ {}_1p_{3x} & {}_2p_{3x} & {}_3p_{3x} \end{bmatrix}$ and each element ${}_jp_{ix}$ represents the probability that a person in region *j* at exact age x will reside in region *i* at exact age $x+4$. The survivorship sub-matrices are therefore computed using Equation 8 above. On the other hand, the birth sub-matrices are computed using Equation 9. The estimates of the probabilities of death and out-migration, as well as the survivorship proportions for the open-ended age group, are simple extensions of the bi-regional system described in Section 2.1.2.

2.1.4 Methodological differences between the multi-regional models and net-migration models

The multi-regional and the net-migration models have differences in presentation. The multi-regional model is conventionally presented using matrix notation. The net-migration models, on the other hand, are presented as sets of individual equations, but could be expressed fairly simply as matrices.

Methodological differences between the traditional uniregional models and their multi-regional counterparts essentially lie in how they treat migration. In the net-migration approach each region is projected separately, hence uniregionally, with the interactions between the regions coming through the net migration mechanism. Net-migration projection models only consider population stocks at a point in time, hence the use of the net migration approach that considers only the differences between in-migration and out-migration. On the other hand, multi-regional models consider migration flows based on changing stocks over time, rather than stocks at a point in time. These models make use of gross migration flows for each region, where the demographic changes in the sending regions govern the expected numbers of migrants for a unit time interval.

Rogers (1990) explains the fundamental differences between the net-migration and multi-regional projection models by looking at an aggregated (ignoring the decomposition by age or age group for the moment) bi-regional projection framework. He describes the model using the following projection Equation, looking at a hypothetical total population consisting of two regions, namely rural and urban, where the national population is closed to international migration:

$$P_u(t+1) = (1 + b_u - d_u - o_u)P_u(t) + o_r P_r(t) \tag{11}$$

Equation 11 above states that the projected urban population $P_u(t+1)$ at time $t+1$ is obtained by adding to the base urban population $P_u(t)$ at time t the increment due to the excess of births over deaths $[(b_u - d_u).P_u(t)]$ in the urban region, and subtracting the decrements resulting from urban dwellers migrating to the rural region over the unit time interval $[o_u.P_u(t)]$, then adding the increment resulting from rural dwellers migrating into the urban region over the unit time interval $[o_r.P_r(t)]$ (Rogers 1990). The same framework can also be applied to obtain the projected rural population at time $t+1$ such that

$$P_r(t+1) = (1 + b_r - d_r - o_r)P_r(t) + o_u P_u(t)$$

In order to understand the inherent weaknesses of net-migration models, consider how the gross migration flow specification is altered when the bi-regional model, and by extension the multi-regional model, is transformed into a net-migration projection model. A net-migration model is obtained by multiplying the second term in Equation 11 by unity expressed as $\left[\dfrac{P_u(t)}{P_u(t)}\right]$ such that the equation now becomes

$$P_u(t+1) = (1 + b_u - d_u - o_u)P_u(t) + o_r P_r(t)\left[\frac{P_u(t)}{P_u(t)}\right].$$

This leads to the equation

$$P_u(t+1) = (1 + b_u - d_u - o_u)P_u(t) + o_r\left[\frac{P_r(t)}{P_u(t)}\right]P_u(t) \tag{12}$$

If we define the rate of in-migration i_u into the urban region such that

$$i_u = o_r\left[\frac{P_r(t)}{P_u(t)}\right] = o_r\left[\frac{1-U(t)}{U(t)}\right] \tag{12a}$$

then the numerator in Equation 12a is the proportion of the national population that is rural at time t whereas $U(t)$ is the fraction that is urban at time t. The net migration rate is then defined as

$$m_u = i_u - o_u.$$

Equation 11 can therefore be rewritten as

$P_u(t+1) = (1 + b_u - d_u - o_u)P_u(t) + i_u P_u(t)$, which finally yields

$P_u(t+1) = (1 + b_u - d_u - o_u + i_u)P_u(t)$, thus

$$P_u(t+1) = (1 + b_u - d_u + m_u)P_u(t). \tag{12c}$$

Thus the natural growth rate for the urban population is

$$r_u = b_u - d_u + m_u$$

From Equation 11 above, note that the model takes into account the demographic changes occurring in the sending population $P_r(t)$, and thus the gross migration flows are indeed treated transparently by the projection model. This is in stark contrast to a projection model that follows the final statement in Equation 12c, where demographic changes in the sending population, the rural region, are completely ignored.

Extending the argument to a comparison between the net-migration model and the three-region model, we start off by defining a three-region population system described by a set of three equations. Ignoring the age decomposition for the moment once again, suppose the projected sub-national population at time $t+1$ for each of the regions 1, 2 and 3, respectively, is defined as

$$P_1(t+1) = (1 + b_1 - d_1 - o_{12} - o_{13})P_1(t) + o_{21}P_2(t) + o_{31}P_3(t),$$

$$P_2(t+1) = (1 + b_2 - d_2 - o_{21} - o_{23})P_2(t) + o_{12}P_1(t) + o_{32}P_3(t)$$

$$P_3(t+1) = (1 + b_3 - d_3 - o_{31} - o_{32})P_3(t) + o_{13}P_1(t) + o_{23}P_2(t) \tag{13}$$

In each of the projections, the periodic numbers of out-migrants in the second and third terms of each equation are a function of the populations of the sending regions since these are at risk of migrating.

Altering the system of equations in 12 above, projections for the three sub-populations can thus be defined analogously as in the bi-regional rural-urban case such that

$$P_1(t+1) = (1 + b_1 - d_1 + m_1)P_1(t)$$
$$P_2(t+1) = (1 + b_2 - d_2 + m_2)P_2(t)$$
$$P_3(t+1) = (1 + b_3 - d_3 + m_3)P_3(t)$$

This system shows that each of the three regions is projected separately, i.e. uniregionally, and without regard for the other two regions. This methodology also views the three-region system through a net-migration perspective since each region is projected forward independently of the rest of the national population.

To illustrate the differences between net-migration models and their multi-regional counterparts, consider an example taken from Rogers and Philipov (1979).

The authors considered this research question by looking at the base 1970 population of 242 million in the Soviet Union. The multi-regional projection method estimated that the population will grow to 265 million in 1980. The net-migration approach they used in this exercise resulted in the projected population size of 269 million in 1980 (Rogers and Philipov 1979).

The illustration above shows that the net-migration model introduces bias and inconsistency when long range (100+ years) projections are performed. Net-migration and bi-regional model projections are consistent with each other in the short term, but net-migration models collapse when long term projections are carried out. The bias and inconsistency seen in this illustration are explained by how the net-migration models treat internal migration. While multi-regional models treat internal migration as gross directional flows over time between regions within a parent population, net-migration models treat these as net flows (Rogers and Philipov 1979). The essential difference identified between the net migration model and the multi-regional model is that while the multi-regional model allows for the transparent treatment of migration, that is, that one knows both in-migration and out-migration rates (i and o, respectively) that should be applied to the correct 'at-risk' population, the net migration model does not allow for that benefit. Instead, all one knows in the net migration model is the number of net migrants, namely m.

2.1.5 Empirical comparisons of multi-regional and uniregional projections

Consider now the empirical results derived in two countries that applied multi-regional models in their population projection exercises, and compare these with known results generated by traditional net-migration projection.

Starting with projections that were done for the states of Florida and Colorado in the United States in 1988, I demonstrate the differences between these models. Rogers and Woodward (1991) illustrated practical weaknesses found in the use of net-migration projection models when they did simulations for some of the states, and used Florida and Colorado as examples. In their work, they noted possible over-projections of the populations they were interested in when the net-migration models were applied to these states. In Florida, the expected population for the year 2000 was first projected at 17.4 million by the U.S. Census Bureau. This compares to the Bureau of Economic Analysis's projection of 14.4 million for the same year. Upon revising the projected estimates five years later, the U.S. Census Bureau adjusted the 2000 projected population for Florida downward by roughly 2 million.

Rogers and Woodward asked whether the revised estimates a result of changing demographic trends with respect to migration into Florida, births and deaths, or just changing models. Their answer was that it was most likely be the latter, as the first estimates were derived from a net-migration model that over-projected the population of Florida (Rogers and Woodward 1991).

The underlying reason for the over- and under-projections by the net-migration models lie in the assumptions about migration and how the internal migration is defined in these projection models.

Raymer, Abel and Rogers (2012) also illustrate the use of the multi-regional population projection method by applying it to England. The country was partitioned into the North, Midlands and South regions. The data were aggregated into these three regions using the nine Government Office Regions of the Office for National Statistics for the period covering 1976 to 2008. Having projected the three regional populations of England to 2008, the authors used both the net-migration and multi-regional models and projected further into 2021. The major difference in the results was that the population of the North in 2021 was projected at 23,000 lower using the net-migration model versus the multi-regional model. The Midlands projection also yielded results of the same order between the two models, and for the South, the net-migration model-generated projections yielded a population that was 29,000 larger than that produced by the multi-regional model (Raymer, Abel and Rogers 2012).

In all, the treatment of internal migration was found to be the major reason for either over-projection or under-projection of the populations being modelled. In defining internal migration as net migration, only the age distribution and demographic profiles of the destination regions were considered, and in each instance, projections were defined by migrants at a point in time rather than flows over time. In order for the multi-regional projection framework to work, it is vital that the inter-regional migration flows are measured, and these can occur either in five-year blocks or annually. Literature on this is reviewed in section 2.2, where we consider internal migration between South Africa's provinces. In this section, we also consider the age patterns of internal migration and how consistent these age patterns have been over time.

2.2 Internal migration in South Africa

In this section, literature on internal migration experienced during the period covering the periods 1991-1996, 1996-2001 and 2001-2007 is reviewed. Two key questions on internal migration are: what levels and age patterns of migration were seen in the periods mentioned above, and did these remain stable over time? These questions are essential to this research because period migration rates that change over time, following a particular trend, as well as stable age patterns, allow for ease of computation and projection of migration into the future.

2.2.1 Levels of inter-provincial migration

There are various reasons for people to move internally. These include a move to find employment, to move back to one's home province to retire there, to stay with a spouse, a start or end of a marital union, or children migrating along with their parents. Migration is caused by considerations such as finding work, schooling for children, or youths moving to study at tertiary institutions (Kok and Collinson 2006). These reasons inform migration decisions by South Africans with respect to choice of destination province. That said, the Eastern Cape and Limpopo have been identified as major net sending provinces and Gauteng and the Western Cape as major receiving provinces (Dorrington and Moultrie 2009). According to the Forced Migration Studies Programme fact sheet, 26 per cent of Gauteng's population growth between 2001 and 2007 is explained by migration into the province (Polzer 2010). From the inter-provincial proportional flow table constructed by Dorrington and Moultrie (2009), Gauteng was the major receiving province, with the Western Cape as second, but to a much lower extent, over the period 1991-2007.

By comparison, all other provinces attracted fewer migrants during this period, with the North-West province attracting the highest proportion of migrants among these seven provinces, at 9 per cent. One can therefore conclude that Gauteng and the Western Cape are the most important industrial and commercial centres in the country and hence we see the internal migration streams dominated by flows into these two provinces.

Another question is whether during the period covering 1991-2007 the inter-provincial period migration rates have changed? Dorrington and Moultrie (2009) show that the change in the age distribution of internal migration was negligible during this period, with insignificant differences between males and females overall and in the proportionate age distribution of the migrants. The period migration rate during the period 1992-1996 (shorter than a 5-year intercensal period) was 3 per cent and during the period 1996-2001 the level increased to 5 per cent. For the period 2001-2007, the rate of migration was 4 per cent. Thus the rates of the total inter-provincial migration during the 1992-2007 period were in the 3%-5% range per five years (Kok and Collinson 2006).

Considering the provincial estimates of intercensal numbers of migrants (StatsSA 1998) for 1993 and 1998 (mid-points of the 1991-1996 and 1996-2001 intercensal periods, respectively), Gauteng and the Western Cape recorded 196,966 and 40,546 out-migrants respectively during the period 1991-1996, and a further 292,992 and 66,193 out-migrants respectively during the period 1996-2001. All other provinces (combined) recorded 811,021 out-migrants between 1991 and 1996 and a further 1.2 million for the subsequent intercensal period (Naidoo, Leibbrandt and Dorrington 2008). Given the mid-year population estimates for 1993 and 1998, respectively (Gauteng: 6.8 million and 7.6 million; Western Cape: 3.7 million and 4.1 million, and all other provinces: 27.2 million and 30.4 million), Gauteng lost 3 per cent of the population to other provinces for each intercensal period. The Western Cape lost 1 per cent during the period 1991-1996 and 2 per cent in the subsequent period. All other provinces, combined, lost 2 per cent to Gauteng and the Western Cape together in the first intercensal period, and 3 per cent in the following period.

The numbers discussed above indicate that the changes in the migration rates over the two periods were low. The period out-migration rates for males and females, respectively, are shown in Table 2.1 and

Table 2.2.

Table 2.1[2]: Male period out-migration rate matrix for 1996-2007 (per cent)

Origin	Destination					
	1996-2001			2001-2007		
	GP	WC	RoSA	GP	WC	RoSA
GP		0.76	3.10		0.55	2.29
WC	0.75		1.50	0.73		1.13
RoSA	2.27	0.74		2.02	0.54	

Table 2.2: Female period out-migration rate matrix for 1996-2007 (per cent)

Origin	Destination					
	1996-2001			2001-2007		
	GP	WC	RoSA	GP	WC	RoSA
GP		0.74	2.75		0.50	2.21
WC	0.76		1.38	0.65		1.09
RoSA	2.06	0.71		1.69	0.46	

From Table 2.1, for example, we see that Gauteng lost nearly 4 per cent of its male population to the Western Cape and the rest of South Africa over the period 1996-2001. Out-migrants as a percentage of the population in Gauteng declined to just less than 3 per cent in the subsequent period covering 2001-2007. The Western Cape lost between 0.73 per cent and 1.50 per cent of its male population to Gauteng and other provinces combined over the period 1996-2007. The other seven provinces combined (excluding Gauteng and the Western Cape) lost a total of 3.01 per cent of their male population over the period 1996-2001 and 2.56 per cent in the subsequent period. The same deductions can be made regarding the female migrants represented in Table 2.2.

The tables above were obtained by creating directional migration tables using STATA12, and the census migration data obtained from a 10 per cent sample of the 2001 South African Census and the 2007 Community Survey (which sampled slightly more than 2% of the population). The data were obtained from the IPUMS (Integrated Public Use Microdata Series) database managed by the Minnesota Population Centre (IPUMS-International 2011).

The percentages were derived by dividing the numbers of out-migrants by the mid-intercensal period population averages for males and females as denominators, separately. Note that the denominators used to derive the period out-migration rates in these tables are regional populations at risk of migrating to other regions.

The overall female out-migration rates fall in the same range as those of males for the two periods.

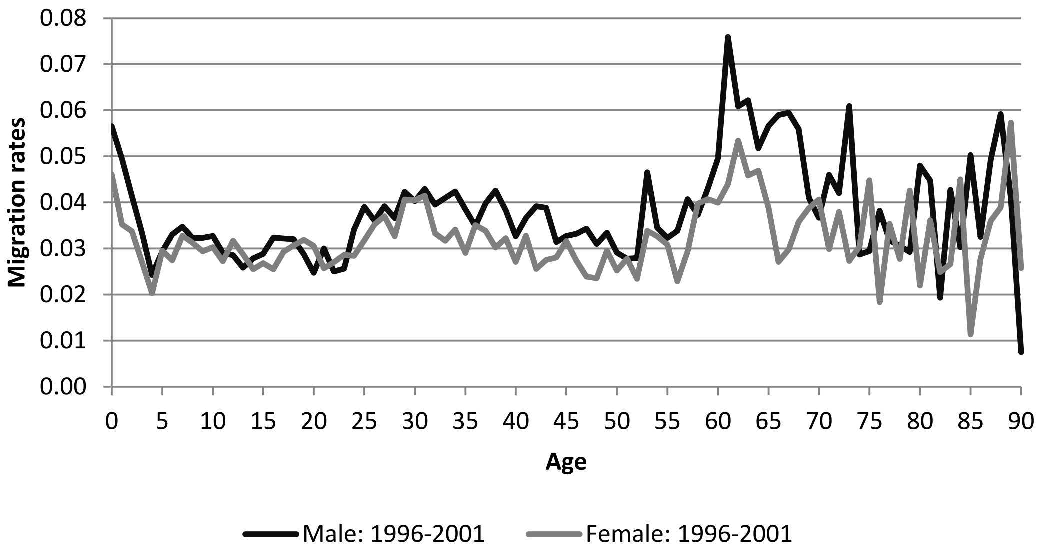
[2] GP stands for Gauteng Province, WC is Western Cape, and RoSA is 'rest of South Africa'

2.2.2 Age patterns of internal migration

Dorrington and Moultrie (2009), in their analysis of the migration patterns experienced over the periods 1991-1996, 1996-2001 and 2001-2007, note that age patterns of migration have the expected shape of a standard migration schedule. This means that the migration curves showed age patterns in which younger migrants moved from most provinces to Gauteng and the Western Cape for study and work opportunities while older migrants moved back to provinces such as the Eastern Cape, Limpopo and others as return migrants at the end of their working careers. Further, the authors note that child migrants (0-4) moved with their parents as confirmed by the 1996 Census results, where child migration showed a peak consistent with migrants aged 17 to 23. Although long birth intervals experienced in South Africa might be expected to spread out the child and parental migration peaks beyond the 17-23 age range, it is important to note that, on the whole, the consistency between the two age bands, namely, the 0-4 and 17-23 age groups, dominates the migration flows experienced during the intercensal periods mentioned above. Also, the authors note that the age patterns of migration were consistent with the data for the period 1975-1980 albeit that the 1980 Census that recorded the migrants for this period excluded the African population of the former TBVC homelands (Transkei, Bophuthatswana, Venda, Ciskei) (Kok and Collinson 2006). These age patterns include the deviations of the age distributions of those from the Eastern Cape and Limpopo as provinces of origin on one hand, and those with Gauteng and the Western Cape as provinces of destination on the other, from the age patterns of all migrants for each of the intercensal periods and the entire 10.35-year period.

Dorrington and Moultrie (2009) further show that Gauteng's ratio of out-migrants to all migrants is higher at older ages (40+) and lower at younger ages (15-35). Figure 2.1 represents the male age patterns of migration from Gauteng to all other provinces except the Western Cape during the period 1996-2001.

Figure 2.1: Male and female age patterns of migration from GP to Rest of SA, 1996-2001



The age distributions of male and female migrants depicted in Figure 2.1 were obtained after creating origin-destination migration tables between three regions in the country prior to adjusting for mortality.

These patterns are also evident for male and female age distributions of migrants moving from Gauteng to the Western Cape and from the Western Cape to the rest of South Africa.

This implies that Gauteng and the Western Cape attracted more labour migrants. Most out-migrants seen at older ages were return migrants who retired to other provinces (presumably the provinces of origin), notably the Eastern Cape and Limpopo. The latter two provinces exhibit peaks in the age-specific ratios of out-migrants to those of all migrants at older ages, and the results at older ages are the reverse of what Gauteng and the Western Cape are showing as provinces of destination.

The predominantly rural provinces have however shown the expected age patterns of migration, and this can be seen in Figure 2.2.

Figure 2.2: Male and female age patterns of migration from Rest of SA to GP and Western Cape, 1996-2001

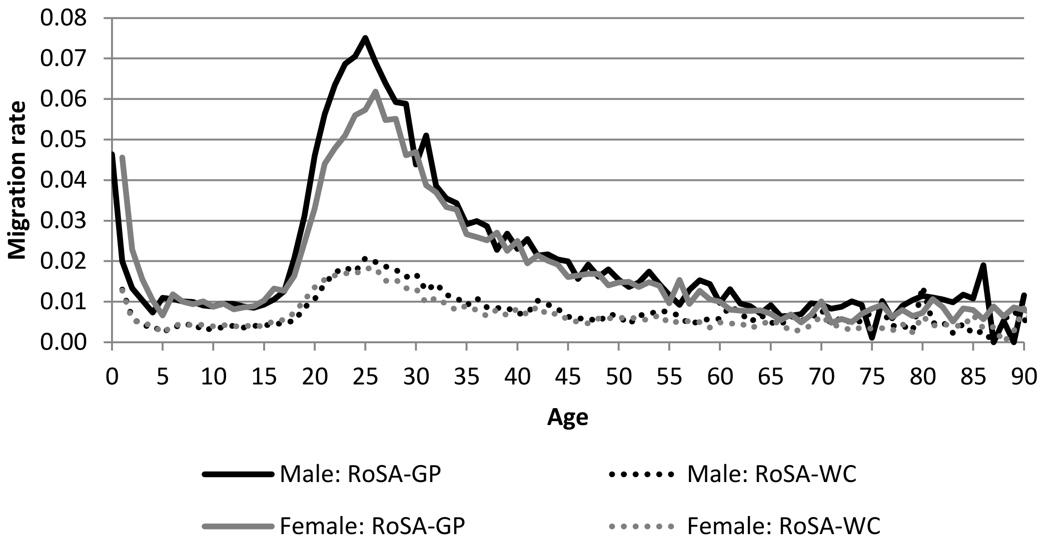


Figure 2.2 shows plots of the migration rates by age from the rest of South Africa to Gauteng and the Western Cape for males and females during the period 1996-2001. Most of the migrants from predominantly rural South Africa to Gauteng and the Western Cape are aged 15-34.

While the overall level of migration may change over time, it is reasonable to expect age patterns of internal migration to be stable into the future on the basis of the evidence just considered.

2.2.3 Data reliability problems

While the literature review above points to stable age patterns of inter-provincial migration and levels, it is important to note that migration data have serious shortcomings for which adjustments would have to be made in order to derive reliable results that capture all migrants as accurately as possible. The data considered thus far were based on the 1996 and 2001 national censuses and the 2007 Community Survey.

Dorrington and Moultrie (2009) made a number of important observations about the data used to measure migration for the period covering 1991 and 2007. First, there is a clear tendency by respondents to report their most recent move as having occurred in the 12 months prior to the census. The migration rates that one derives on an annual basis would thus be too heavily biased towards the more recent move (Dorrington and Moultrie 2009). The second problem emanates from the poor quality of the data on migration of children for the entire 15-year period from 1991. The quality in this regard results *inter alia* from scanning problems, where the scanner could not properly differentiate between province numbers, particularly the province indicators '1s' and '7s'.

A further observation is that when the two censuses and the Community Survey are used together, they are inconsistent with each other with regards to the level of migration (Dorrington and Moultrie 2009). This could be explained by minor variations in data coding and scanning errors as these introduce a disproportionate effect on the estimates of migration (Dorrington and Moultrie 2009).

2.3 Overview of the ASSA2008 AIDS and Demographic Model

The ASSA model was developed by the Actuarial Society of South Africa, and is an Excel-based tool used to model the population and the demographic impact of HIV and AIDS. The cohort-component projection method underlies the model. A description of the model's history, software requirements and logic is provided in the associated manual (Dorrington, Johnson and Budlender 2010). This Section of the literature review will present only the synopsis of the design and operation of the model.

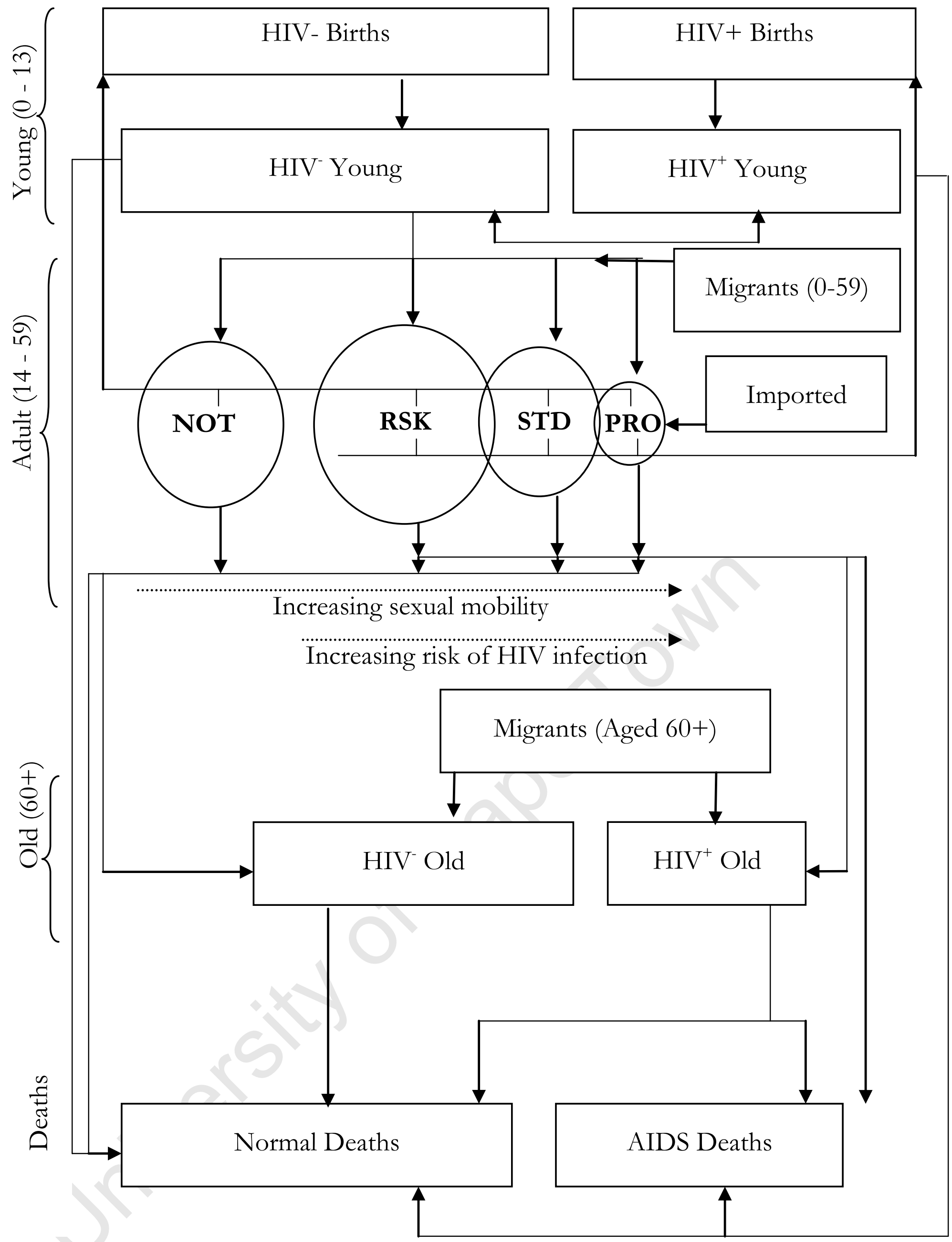
Johnson and Dorrington (2006) show that the ASSA model divides the population into cohorts of individuals by single age and sex, where each cohort consists of individuals having identical characteristics. The model partitions the population into three distinct age groups with respect to the transmission of HIV. The first age group consists of the young population (birth up to, but excluding the exact age of 14), the second age group consists of the adult population (14 to 59 years of age last birthday) and the final age group is that of the old-age population (60+) (Johnson and Dorrington 2006). The adult age group is further partitioned into four risk groups referred to as PRO, STD, RSK and NOT. The risk groups are differentiated by their level of exposure to the risk of contracting the HIV disease. The PRO risk group is a group of individuals whose sexual activity reflects that of commercial sex workers and their regular clients. The STD group is a group of individuals whose HIV prevalence is similar to those regularly infected with STDs. The RSK group is a group of individuals whose sexual activity is assumed to be less risky, but who are still exposed to the HIV infection since they have, on average, one new partner per year and sometimes engage in unprotected sex, but are not regularly infected with STDs. The final risk group, the NOT group, consists of individuals who are not at risk of HIV infection (Johnson and Dorrington 2006).

Modelling of the AIDS epidemic in South Africa started in 1989 with the Doyle-Metropolitan Life model (Doyle and Miller 1990). Since the code of the model was proprietary, the Actuarial Society of South Africa created a non-proprietary model to be accessible to all users and thus in 1996, the first version of the ASSA model was released, the ASSA500. The model has been revised and recalibrated numerous times since 1996 due to the need for the incorporation of new or updated data into it, with the latest model being ASSA2008. The number 2008 at the end of the model name was included to reflect the year in which the most recent, up-to-date data on antenatal HIV prevalence and reported deaths were available to which the model could be calibrated. This naming convention was adopted by the AIDS Committee in 2000 as the model was revised further, and was first used when ASSA2000 was released (Dorrington, Johnson and Budlender 2010).

The ASSA model possesses a feature that sets it apart from all other demographic models. The model integrates behavioural, epidemiological and demographic parameters into the projections (Dorrington, Johnson and Budlender 2010; Nyirenda 2007). Some models are driven purely by demographic parameters, while others integrate demographic, and to a lesser extent, epidemiological parameters. The fact that the model is able to incorporate these three sets of parameters into the projections makes it an ideal tool to use for projecting the population in a sub-Saharan country such as South Africa, where the population experiences one of the highest burdens of HIV/AIDS in the region (Nyirenda 2007). The model is relatively difficult to use though. First, the model has extensive data requirements such as accurate estimates of the number of deaths against which to calibrate the model, and these are often not available. Second, the model only allows for cycles once a year and does not allow for people to change from one risk group to another over their life span.

Figure 2.3 presents a schematic diagram of the model.

Figure 2.3: A schematic representation of the ASSA model



Source: Dorrington *et.al,* 2010; p. 8

Figure 2.3 illustrates the three broad age groups into which the population is partitioned on the far left. These represent different modes of transmission, namely: mother-to-child; via heterosexual activity; and no new infection above age 59.

Note the four circles that denote the four HIV risk groups into which the adult population (14-59) is further partitioned. These represent different categories in terms of the risk of getting infected. The NOT group is notably detached from the other risk groups as individuals in this group are assumed not to be exposed to HIV infection. The three broad age groups are assumed to differ in their exposure to HIV infection. The young population (birth to 13 at last birthday) is assumed to be infected at birth or from breastfeeding. At the age of 14, these individuals are allocated to the risk groups according to the proportions recorded on the *Assumptions* sheet.

The model assumes that, in the absence of prevention and treatment, 20 per cent of the babies born to HIV+ mothers are HIV+, while a further 16 per cent of the remaining HIV- babies contract the virus from their HIV+ mothers through breast-feeding.

The young (birth to 13) and the old-age (60+) populations are divided into only two groups, namely, those that are HIV- and those that are HIV+. At the age of 60, all individuals are allocated to the OLD group. The duration since infection is still modelled for this group, but no further infections occur beyond the age of 60. The OLD worksheets function as a run-off of the population, and the open age interval 90+ estimates those aged 90+ from those aged 89+ a year ago (Dorrington, Johnson and Budlender 2010).

The latest suite of ASSA models, the ASSA2008 suite, was released by the Actuarial Society of South Africa in early 2010, together with the associated manual and results generated by the model at national and provincial level.

The model has four different versions namely, the *full* national model, the *lite* national model and the full *provincial* model. The fourth model, namely the *rural-urban* model, has not been updated in many years and will therefore not be discussed here. The *full* model disaggregates the population into four race groups, namely Asian, Coloured, Black and White. It does this by separately modelling each of the four population groups at national level, and then aggregating the results to derive the national population over a user-defined projection horizon. The *full* model was developed in order to address user demand which was motivated, in part, by the observation that the impact of the epidemic was unique to each population group. These differences are among the reasons for differences between the prevalence in the Western Cape and Northern Cape on one hand and all the other provinces on the other. In addition to this, some users use the results by population group and geographic sub-region as proxy for socio-economic groups, and thus they extrapolate the results of the model to different socio-economic groups (Dorrington, Johnson and Budlender 2010).

The *lite* model treats the national population as one population group without regard for differences in race-specific demographic and epidemiological profiles. The *provincial* model takes into account the geographic differences in the spread of the HIV/AIDS epidemic, and thus models each province separately. The *provincial* models are created using the workbook called AssumptionsProv. This workbook pastes provincial starting assumptions into the ASSA model (*full* version) to create a model for a specific province. The workbook contains tables that will be explained in more detail in

Chapter 3, where the aggregation of the starting assumptions for the regions to be modelled as part of this research will be discussed. The AssumptionsProv workbook also contains starting assumptions, at provincial level, for each of the four population groups, and an additional sheet that records the provincial aggregate of the race-based assumptions.

The *full* model projects the national population to 50.4 million as at the middle of 2010, and projects that the population would rise to 50.86 million by the middle of 2011, implying a yearly growth rate of 0.9 per cent. Presently, the model projects individual provincial populations separately, using the net migration approach to projections.

For purposes of this study, only the *full* and *provincial* versions will be discussed and used to generate results in Chapters 3 and 4, and the *lite* version will be used to check that the multi-regional adaptation of the model works.

The ASSA model starts off with the base population that reflects the actual population as at the middle of 1985. The base population was derived by projecting the 1970 population to 1996 in a way that is most consistent with births, deaths, numbers of migrants and the population counted in 1996. With regards to the *provincial* model, the provincial base populations were reconstructed to yield numbers of people that would be expected to be living within the borders of the current provinces had they existed back in 1985, as the current provinces did not exist then. A process of remapping the 1991 census into the new provincial boundaries and the establishment of the inter-provincial migration patterns between 1985 and 1996 was used to arrive at the provincial 1985 base populations.

As the focus of this research is adapting the ASSA model to allow for multi-regional migration, it is important for us to understand how the model currently handles migration. The model has two sheets, namely *Male Migration* and *Female Migration*. These sheets contain assumed age-specific net numbers of in-migrants (i.e. in-migrants less out-migrants) for single years of age for each year from the middle of 1985 to the middle of 2025. The model uses these numbers by apportioning them to each of the four risk groups according to the *Immigration Risk Group* proportions in the *Assumptions* worksheets. All other migrants are added to the numbers in the *Young*, *MaleOLD* and Fem*OLD* worksheets as at the end of each projection year (i.e. just before the middle of each calendar year.

The model assumes that migration takes place at the end of each projection year and that migrants have the same age-sex infection history (i.e. proportions infected by duration of infection) as that of the risk groups they join. Further, it assumes that if net migration is positive, then after the year 2006, net migration declines from its peak in

2006 and approaches zero over a 30-year period. The opposite is true of net migration if negative. The net migration is then assumed to rise as it trends toward zero over the 30-year period. The model assumes that after 2025, the numbers of migrants remain the same as those in 2025. The model has been calibrated and released publicly for South Africa.

We now consider how the model produces results that appear in the *Results* worksheet. Not all worksheets will be discussed, only those relevant to the research will be described.

2.3.1 Annual population projection in the *full* model, using the VBA code

The model starts off at 1 July 1985 with the sum of the race-specific initial population numbers in each of the four *Assumptions* worksheets as at that date. These sheets also have the starting assumptions that include, among others, the imported infectivity that starts the epidemic in 1985. The starting assumptions then generate the HIV prevalence and incidence rates in future years as the model runs. The total numbers by age and sex are distributed into the risk groups and the sum then appears on the *Population* sheet. The *Population* sheet also records the deaths derived from each of the four risk groups, that is, AIDS and non-AIDS deaths from the PRO, STD, RSK and the NOT group for males and females at individual ages.

Further, the worksheet separates the population into three broad age groups as mentioned in the preceding subsection. Note that the national 'young' population is the sum of the races which are each modelled as per the *lite* model, where each of the races consists of the three broad age groups and the four risk groups for the adult population.

The annual projection calculates the net numbers of migrants in each of the HIV risk groups. To illustrate, consider the female population comprising the four risk groups in the ASSA2008 workbook. The female risk groups are represented by the sheets named *FemPRO*, *FemSTD*, *FemRSK* and *FemNOT*. The first three sheets use the rates in the *SexActivity* worksheet that estimate age-specific sexual activity parameters, the *HIVTable* worksheet that estimates antenatal clinic HIV prevalence and the *MortTable* worksheet that computes age-specific mortality rates, mortality improvement factors and ultimate mortality rates. The model then uses these four sheets to derive annual numbers of uninfected adults by age, numbers of new infections and numbers infected by age and duration of infection, plus numbers of infected and uninfected new-borns. These worksheets also use the mortality rates to compute annual numbers of AIDS and non-AIDS deaths by age, as well as annual net numbers of migrants for inclusion in the risk groups. The computation process in the *FemPRO* worksheet and the other two sheets is completed by the calculation of net numbers of migrants by age using results in the *Female Migration* sheets. The *FemNOT* worksheet computes the annual numbers of babies born to women aged 14 to 59 that are uninfected, and also incorporates into the *NOT* group net numbers of migrants from the *Migration* sheets. The worksheet does not calculate numbers of new infections as, by definition, individuals in the *NOT* group are not at risk of an infection.

The essential feature of the model is that there is a 'Before' table and an 'After' table, recording the numbers of uninfected people by age and the numbers of infected people by age and duration of infection. The 'Before' table presents the numbers as at the middle of the current year, the 'After' table presents the numbers as of the middle of the following year. The 'After' table is calculated from the 'Before' table, where the numbers of migrants by age are added to the 'After' table, and both are used to project the numbers of births and deaths arising from each age group over the projection year. The VBA code then projects the population for the year by pasting the 'After' table numbers into the 'Before' table of the following year.

The calculation process is exactly the same for the male population as well, except that births are confined to the female sheets.

2.3.2 Population numbers and results

Each race-specific *Population* worksheet records annual age-specific sums of the race groups. For the young national population, the worksheet calculates the sum of individuals from birth to age 13 from all four race-specific *Young* worksheets. For the adult national population, the *Population* worksheets derive the sum from each race group, with each of these sums being derived from age-race-specific population numbers as stated in the preceding subsection. For the old population, the *Population* worksheets derive the male and female sums from the *MaleOLD* and *FemOLD* worksheets, respectively, where each of these worksheets also calculates totals by adding population numbers from associated race-specific *MaleOLD* and *FemOLD* worksheets, respectively.

Chapter 3 will describe how the model was adapted to allow for multi-regional projection of migration, particularly regional out-migration rates from 2007 going forward. The chapter will also explain how the VBA code of the ASSA model will be re-configured so that the model is adapted as a multi-regional projection tool to model the national population as a three-region population system, as well as to project internal migration between the three regions.

3 METHODOLOGY

This chapter introduces the methodology followed in answering the research question as stated in Chapter 1. Section 3.1 describes the general approach for changing the ASSA model into a multi-regional model. Section 3.2 describes how the starting assumptions for the model are set up. Section 3.3 discusses the adaptation and testing of the multi-regional model. Sections 3.4 and 3.5 consider the approach applied in deriving annual internal age-specific migration rates for the period 1996-2007, and the incorporation of international migration, respectively. Section 3.6 describes how these rates are projected after 2007 for each region.

The approach starts off by setting the aggregate starting assumptions for the model.

The next step in the process of the adaptation of the model was adding extra worksheets to the model. The extra worksheets were needed in order to generate regional out-migration rates and the resulting numbers of out-migrants. After this, the VBA code of the model was edited and an additional module written to project the migration. The VBA module on migration was necessary as the original model did not require code to allow for migration. The module thus allows the model to calculate directional out-migration rates for each region, and the associated annual numbers of out-migrants. Variables were also edited using the *Name Manager* tool in the ASSA model workbook. These steps were necessary since they would enable the VBA code, including the additional migration module, to enable the model to execute the projections successfully.

Finally, once the multi-regional model was set up, starting demographic and epidemiological assumptions were fed into the model and the model was run iteratively in order to perform a minor re-calibration to antenatal HIV clinic data. Subsequent to this, the model was tested against the original ASSA model to ensure correctness of its results.

This whole process is explained in more detail in Sections 3.2 and 3.3.

3.1 Regional *lite* versions of the ASSA model

The original parameters in the ASSA2008 AssumptionsProv workbook, described in Section 2.3, were aggregated to obtain input parameters for each region. The parameters were obtained since there are no *lite* versions of the model for Gauteng, the Western Cape and other provinces. For purposes of this research, three separate *lite* versions were created, one for Western Cape, one for Gauteng and one for the rest of South Africa. The third region comprises all other seven provinces combined because these are predominantly rural and because Gauteng and the Western Cape mostly receive migrants from the seven provinces. The multi-regional model thus is a set of the three *lite* versions that interact through the multi-regional migration mechanism. In the next two Sections the adaptation of these parameters for the multi-regional model is described.

A brief description of the parameters follows in Section 3.2, with Gauteng as an example, keeping in mind that the same aggregation was done for all other regions.

3.2 Setting input parameters and assumptions for the multi-regional model

The aggregation of the model input parameters was not uniform for all the parameters. Some of these parameters were derived as population-weighted averages, others as simple sums, while some were weighted by age and sex.

The aggregation was necessary because the *provincial* model parameters are all race-specific, that is, each population group has its own set of starting assumptions unique to it. Noting that each region has all four population groups, and that the aggregate model is needed for each region, it was necessary to combine all four race groups in the region in order to model regional populations and thus simplify the modelling methodology. I use Gauteng as an example for illustrative purposes.

3.2.1 Parameters summed

The following parameters were added to derive sums and these are shown in Table 3.1.

Table 3.1 Starting assumptions for Gauteng in 1985: Sums

Race group	Parameter weighting factors			
	Population	Adult female population	Pregnancies	Infectivity
Asian	144,085	44,715	3,501	0.1
Black	3,376,896	857,099	94,977	46.4
Coloured	263,784	82,376	8,383	0.7
White	2,156,915	734,341	42,593	0.9
Totals	5,941,680	1,718,531	149,454	48.0

Table 3.1 contains starting parameters as at 1985 for Gauteng, where these are race-specific numbers for total population, adult female population, 1985 numbers of pregnancies and the imported infectivity. Imported infectivity is a variable that introduces the HIV epidemic into the population. For Gauteng, imported infectivity was set at the starting number of a total of 48 people, where these are not part of the population in 1985 but are used to start the epidemic in the model. This number is also used to determine the time of the start of the epidemic.

In addition to the above variables, the age-specific annual net numbers of migrants were summed across all four population groups to derive the aggregate numbers.

3.2.2 Population-weighted averages

Table 3.2 lists parameters that were derived as population-weighted averages.

Table 3.2: Starting assumptions for Gauteng in 1985: Population-weighted averages

	Parameters weighting factors				
Race Group	Male births	Ratio of condom use to aver	Sex activity Factors	ART take-up rates	ART eff. Assumptions
Asian	0.02	0.03	0.03	0.02	0.02
Black	0.64	0.50	0.50	0.55	0.64
Coloured	0.06	0.05	0.05	0.04	0.06
White	0.28	0.43	0.43	0.39	0.28
Totals	1.00	1.00	1.00	1.00	1.00

The first column in Table 3.2 contains race-specific proportions of the male births. It shows the race-specific sex ratios at birth, i.e. the proportion of all live births that are male per population group. The next column shows weighting factors for the ratio of condom usage within each population group to the national average. The third column contains weighting factors for the sex activity parameters. Sex activity parameters, which will be described more clearly in Section 3.2.4, define the shape and position factors of sexual behaviour in the adult population, and these are shown in the *SexActivity* worksheets of the model. ART take-up rates measure the proportion of the HIV+ population that has access to antiretroviral treatment nationally and provincially for each year starting with the year in which the treatment was introduced, namely, the year 2000. The ART effectiveness assumptions in the final column of Table 3.2 measure the effectiveness of ART treatment in prolonging the survival times of individuals in each disease stage, such as the progression from stage 5 to stage 6 and progression from stage 5 to death.

3.2.3 Age-group-weighted averages for pregnant females attending private clinics

Table 3.3 lists weighting factors for the only parameter for which aggregate averages were derived with respect to each quinquennial age group, but it shows how such averages were derived for pregnant women attending private clinics between the ages of 15 and 49.

Table 3.3: Starting assumptions for Gauteng in 1985: Averages weighted by pregnant females

	Distribution of pregnant women by race group					
Age group	Asian	Black	Coloured	White	Prop.	All
15-19	0.03	0.49	0.06	0.42	1.00	264,172
20-24	0.03	0.56	0.05	0.36	1.00	288,859
25-29	0.03	0.55	0.05	0.38	1.00	267,516
30-34	0.03	0.53	0.05	0.40	1.00	223,084
35-39	0.03	0.47	0.04	0.46	1.00	176,451
40-44	0.03	0.45	0.04	0.48	1.00	142,734
45-49	0.03	0.45	0.04	0.49	1.00	121,770

The first four columns after the column of the age groups in Table 3.3 show the distribution of the five-year age groups by population group for pregnant women. The fifth column shows the sum total of the proportions. For example, 0.55 in the third row, third column is simply 55 per cent of women attending private clinics that are African aged 25-29 in 1985. The final column in Table 3.3 contains the total numbers of pregnancies per five-year age group.

Note that the numbers derived in the table are numbers of pregnancies.

3.2.4 Age-sex-weighted averages

In order to complete the *lite* model for the region of Gauteng, age-sex-weighted averages were derived for the following parameters:

- Condom usage for the RSK risk group, where the female age-specific weighted averages were used
- Male and female mortality improvement indices
- Male and female ultimate mortality rates
- Male and female annual non-AIDS mortality rates from 1985 to 2007
- Female annual non-HIV fertility rates from 1985 to 2006 (aged 15-49)
- Fertility improvement factors (aged 15-49)
- Ultimate fertility rates (aged 15-49)
- HIV fertility factors, and these are the start ratios, initial impact and the reduction factors (note: the start ratios are ratios of HIV+ to HIV- fertility rates holding immediately prior to contracting the disease; the initial inpact factor is the initial impact on the fertility ratio on contracting HIV; the reduction factor is the rate at which the fertility ratio drops per year infected).
- Sex activity parameters for $f(y|x)$ for the sexually-active population (aged 14-59), where $f(y|x)$ is defined as the percentage of partners of females aged x

who are aged y (Dorrington, Johnson and Budlender 2010). The parameters for which weighted averages were derived are the mean and variance of $f(y|x)$ at ages 17, 22, 27, 32, 37, 42 and 47.

The same process of deriving the weighted averages and sums was used for the Western Cape and the 'rest of South Africa'. However, for the 'rest of South Africa', the same process was repeated, where the averages were province-weighted while others were simple sums.

3.2.5 Proportions

The starting population age distribution and the age-specific male proportions were derived directly from the population numbers as at the middle of 1985. These factors were pasted onto the Gauteng *Population* worksheet in the model so that the Gauteng base population could be reproduced accurately. The same calculation was carried out for the Western Cape and 'rest of South Africa'.

3.3 Model adaptation and testing

3.3.1 Additional worksheets and tables

The new regional totals were used as base population totals and pasted in the *Assumptions* worksheets for each region. The provincial population totals were added together to determine the national estimate as at the middle of 1985 and this is illustrated in Table 3.4.

Table 3.4: Per cent distribution of the 1985 regional base populations

ASSA2008		Multi-regional model	
Region	Population size	Region	Population size
National	32 306 335	Gauteng	5 941 680
		Western Cape	2 985 796
		Rest of SA	23 378 858
		National	32 306 335

The *full* version of the original ASSA model has four projections, one for each of the four population groups. However, for our purposes the three suffice to test the adaptation of the ASSA model; hence the fourth projection was surplus to need. In order to avoid having to rewrite code but also to avoid 'divide by zero' results that arise as a consequence of a fourth projection that is not needed, the numbers in the fourth projection were made positive but as small as possible (a fictitious population of 1) so that a three-region system is possible. The migration numbers for the fourth projection were set to zero to avoid generating 'negative' numbers of people.

Additional worksheets were included in the ASSA multi-regional model for each region, and these are *Results MaleOut* and *Results FemOut* worksheets. For example, Gauteng had the sheets *Results MaleOutGP* and *Results FemOutGP* added for its projection of migration. Each of these sheets record age-specific annual rates of migration from Gauteng. The directional age-specific out-migration rates are derived from a set of migration parameters obtained by means of a curve-fitting process, as will be explained in Section 3.4.3.

The multi-regional model then calculates the annual net numbers of in-migrants in the *Male Internal Migration* and *Female Internal Migration* worksheets at individual ages for each region, and then adds to each of these, numbers of immigrants moving into each region as a proportion of total numbers of immigrants. Net numbers of immigrants are found on the *Male Immigration* and *Female Immigration* worksheets, and these were obtained from the original ASSA model.

The *Results MaleOutGP* and *Results FemOutGP* worksheets have three additional tables that are intended to inform model assumptions for migration. The first of these is titled *Fitted Rogers-Castro multi-exponential migration schedules*. This table contains migration parameters that result in smooth out-migration age patterns to the other two regions during the periods 1996-2001 and 2001-2007, for males and females, separately. The second table is titled *% change in migration rates*. This table contains percentage changes to out-migration rates applicable to the two intercensal periods. The third table is titled *Scalars applied to migration rates 2007+*. This table assumes that migration rates will decline roughly linearly over time after 2007, and the values in this table are equivalent to $1-\Delta$, where Δ denotes the percentage change to migration rates contained in the preceding table.

3.3.2 The VBA code

The VBA code in the ASSA model allows for the projection of four population groups simultaneously, without any interaction between these. The task therefore was to adapt the model so that it projects three sub-populations simultaneously, taking into account interregional flows between these sub-populations.

The first part of this phase involved editing the names originally assigned to the four population group projections and renaming three of them. In place of 'Asian Group', 'Black Group' and 'Coloured Group' as these names appear in the buttons on the *Assumptions All* sheet of the original model, the model buttons are now renamed 'Gauteng', 'Western Cape' and 'Rest of South Africa', respectively. In place of 'White Group', the button is blank since the fourth projection is not used and thus it is not part of the three-region system for purposes of this research. The VBA macros were also edited in line with these changes so that the code can recognise each of the names assigned to the regions. Four additional buttons were created and placed on the *Assumptions All* sheet, and these are named *Male Outmigr* and *Female Outmigr* as well as *Male Inmigr* and *Female Inmigr* (shortened names *Outmigr* and *Inmigr* denote out-migration and in-migration, respectively). These buttons yield charts that plot separate male and female age distributions of out-migrants and in-migrants in each region.

In order to complete the charts function of the model, all twelve chart buttons were updated so that each of these displays plots only for the three regions, thus the fourth button in each chart, where this button would allow the user to access the plots for the fourth projection, was deleted to allow for a neat presentation of the charts showing 'GP', 'WC' and 'Rest' only.

For the adapted VBA code to work, together with the *Assumptions* buttons in the model, the 'Name Manager' facility was used to edit the worksheet variable names and associated worksheet formulae so that these can recognise the regional acronyms, namely 'GP', 'WC' and, 'Rest'.

An additional VBA module, shown in Appendix 2, was written for the model so that the model not only calculates the out-migration rates annually, but also updates the *Results MaleOut* and *Results FemOut* worksheets for each region. Thus the new VBA module contains macros that perform the projection of migration in the adapted multi-regional model.

3.3.3 Testing the adapted multi-regional model

Model-testing was performed in two phases. The first involved testing the *lite* models for each region. This was not the crucial part of the model-testing phase. The aim of this phase was merely to ensure that the assumptions that underlie the *lite* models for the regions, and the results generated by these, are sensible.

The second phase was to check the correctness of the aggregation of input parameters for the multi-regional model as well as to ensure that the model works. In other words, every effort was made to ensure that, given the same input parameters, the adapted multi-regional model gives the same results as the net-migration model. This phase involved obtaining age-specific annual net numbers of in-migrants from the multi-regional model between 1985 and 2025 and pasting these onto the *Male Migration* and *Female Migration* worksheets of the *lite* versions for Gauteng, the Western Cape and 'rest of South Africa'. The same starting assumptions used in the multi-regional model were also used in the three *lite* versions. The purpose of the test was to ensure that the multi-regional model reproduces the current model for the period 1985 to 2025.

The check for the correctness of the model was carried out using the ASSA2008 ProvOutput_110216[3] workbook, adapted in order to record the results from the ASSA multi-regional model instead of provincial results and the *lite* versions mentioned above. The purpose of the ASSA2008 ProvOutput_110216 workbook is to display the projected provincial and national demographic and epidemiological results derived from the ASSA model. The original workbook was adapted to cater for the purpose of reporting results derived from the multi-regional model. The adapted workbook allowed the results from the multi-regional model to be compared to those obtained from the original ASSA model for the three regions.

The ProvOutput workbook allows for a visual inspection of the results from the two models, by age and sex, since the workbook records aggregate results from the ASSA model and displays these using tables and charts. The workbook was therefore reconfigured such that each region had the *lite* worksheet and the *m.r.* worksheet. For example, the workbook for Gauteng has worksheets *GP - lite* and *GP - m.r.* The two projections would thus be compared directly for each variable as the user keys in relevant worksheet row numbers describing variables for which comparisons are to be made.

3.4 Estimation of historical regional out-migration rates

3.4.1 Data and methods

The census migration data were obtained from a 10 per cent sample of the 2001 Census and the 2007 Community Survey, an approximately 2.5 per cent sample of national population. Both national inquiries were conducted by Statistics South Africa and the two data sets were provided by IPUMS.

The internal migration rates for the periods 1996-2001 and 2001-2007 were derived using the IPUMS data sets while the immigration (international migration) numbers for the same intercensal periods were obtained from the *full* version of the ASSA model.

The 2001 Census and the 2007 Community Survey migration data sets did not capture age patterns of child migrants because the place of residence five years prior to the census for the population aged 4 and younger is undefined. As a result, the assumption made was that children aged 4 and younger move at most once during the intercensal period (Dorrington and Moultrie 2009). That is, that children aged 4 would have migrated between their region of birth and region of enumeration at exact age 2.25

[3] ASSA2008 ProvOutput_110216 can be found on the Actuarial Society's website http://www.actuarialsociety.org.za/Societyactivities/CommitteeActivities/AidsCommittee/Models

(i.e. half-way between the middle of the year of birth (4 to 5 years before the census) and the time of the census). The age location of 2.25 for those aged 4 is assumed as one compensates for mortality as a result of the enumerated population reflecting only the survivors. Dorrington and Moultrie (2009) demonstrate the process of estimating numbers of child migrants, where the authors state that children aged 3 would migrate at exact age 1.75, 2-year olds at exact age 1.25, 1-year olds and infants at exact ages 0.75 and 0.25, respectively (Dorrington and Moultrie 2009). For purposes of this exercise, the same algorithm was applied to the census migration data.

3.4.2 The 1996-2001 intercensal out-migration rates

The age patterns of migration yield rates at which people at specific age groups or individual ages migrate between regions. In order to derive rates of migration between the three regions in the country, it was necessary to consider age-specific numbers migrating internally during each intercensal period.

Tables of age-specific population numbers were created for the periods 1996-2001 and 2001-2007 separately. The results for the former period are discussed since the same process was followed in estimating migration numbers for the latter period.

The census population and migration tables based on the IPUMS-International data sets were created for the period 1996-2001 using the statistical software package STATA12 (StataCorp 2011). In order to obtain plausible estimates of numbers of migrants, international migrants were isolated from all other migrants in order to avoid over-estimating internal migrants in deriving tables of numbers of the latter (refer to Appendix 1 for the STATA12 code used to produce the tables of numbers of migrants.

Once the regional age-specific population numbers and migration tables were obtained for the period 1996-2001, the average population of the region of origin (*i*) midway through this period was derived in order to obtain the population at risk of migrating from region *i* to region *j*.

The intercensal numbers of out-migrants per region were adjusted for mortality, using mortality estimates obtained from the ASSA model, and assuming the following: first, migrants experience the same level and age pattern of mortality as the population from which they are migrating; second, the migrants would have been exposed to the risk of dying for half of the intercensal period and third, that mortality rates are constant within each intercensal period of five years (Dorrington and Moultrie 2009). The resulting numbers of migrants were divided by the average population numbers mid-way through the intercensal periods 1996-2001 and 2001-2007 at each age to obtain out-migration probabilities for each region.

Each of the intercensal migration streams, again expressed in individual ages 0 to 90+, were then divided by the length of the period (exactly 5 years between 1996 and 2001 and 5.35 years between 2001 and 2007) to convert these to average annual out-migration rates - this simple conversion method was used since census and survey results between 1996 and 2007 showed that respondents were more likely to report the most recent move as having occurred more recently over the intercensal period than it actually occurred (Dorrington and Moultrie 2009), hence annual migration estimates would be too heavily biased towards the twelve months preceding the national census or survey concerned.

3.4.3 Fitting multi-exponential curves to observed patterns

A curve-fitting process to census migration data was carried out in order to derive smooth annual migration rates. The Rogers-Castro *n-parameter* models (Appendix 6) that describe the standard age patterns of migration (Rogers and Raymer 2006) were fitted to the census migration rates in order to estimate parameters defining migration volumes at specific age bands.

A mathematical modelling software package MATLAB 2011a was used to fit Rogers-Castro multi-exponential curves to the regional census migration rates, where the total number of curves fitted was 24, implying eight curves per region, four curves for males and females, separately. For example, migration curves for males moving from Gauteng to the Western Cape and rest of South Africa were fitted for each of the two intercensal periods covering the years 1996-2001 and 2001-2007. So, one fitted Rogers-Castro age schedules to the two migration streams for each period for males. The same fitting process applied to female migration streams. The Levenberg-Marquardt algorithm with nonlinear least squares methods was used in this exercise.

For the Western Cape, the curve-fitting was carried out using the simple 7-parameter age schedule, except for female out-migrants leaving the Western Cape for the 'rest of South Africa' during the period 1996-2001.

This age pattern of migration exhibited age patterns similar to the *11-parameter* schedule since it showed a retirement hump that could not be ignored and could well be an indication of female return migrants leaving for the Eastern Cape and the Northern Cape at older ages.

For the 'rest of South Africa', the simple *7-parameter* schedule was fitted to the migration rates since most of the migrants moving into Gauteng and the Western Cape are younger people, plus a portion of the middle-aged people. The only exception for this region, for which the *11-parameter* model was fitted to the census migration rates, is the age pattern of migration for males leaving the RoSA or the Western Cape during the period covering 2001-2007. The largest proportion of all the migrants is aged between, and including, 16 and 23.

For Gauteng, all eight migration age patterns (2 intercensal periods x 2 destination regions x 2 sexes) consistently exhibited a retirement hump; hence the *11-parameter* model was fitted to the census migration rates here.

3.5 Incorporating immigration (international migration)

Age-specific annual net numbers of immigrants were obtained from the *full* version of the ASSA model. These were apportioned between the three regions such that Gauteng received 46.8 per cent of the male and female net immigration, the Western Cape received 13.4 per cent and the 'rest of South Africa' received 39.8 per cent (StatsSA 2007).

The net numbers of immigrants were recorded into the *Male Immigration* and *Female Immigration* worksheets of the multi-regional model.

3.6 Multi-regional methodology for migration estimates

Two of the additional worksheets (described in Section 3.3.1) created in the multi-regional adaptation of the ASSA model, namely *Results MaleOut* and *Results FemOut*, were used as tools for calculating annual migration for each year of the projection. These rely on fitted model migration schedules described above and included in the *Assumptions* worksheet for each region.

Based on these fitted schedules, the model estimates directional out-migration rates and uses these to estimate annual numbers of migrants from 1997 onwards. The model then calculates total out-migrants and total in-migrants for each region, then the net numbers of in-migrants (leaving region *i* and out-migrants leaving regions *j* and *k* moving into region *i* at exact age *x*) such that the net numbers of migrants, numbers of in-migrants and numbers of out-migrants at exact age *x* at time *t* for males and females in each region are defined as nm_x, im_x and om_x, respectively, where, for region *i*,

$$nm_x = im_x - om_x$$

Each of the worksheets *Results MaleOut* and *Results FemOut* has a column that updates with each cycle as the model runs. The migration rates on that column are recorded on the *ProjSpace* Section of the same worksheet as projected migration rates per age are generated by the model. The migration rates used as input to the multi-regional model are obtained by extrapolating each of the level parameters one year ahead in each cycle, using the formula

$$a_r^{t+1} = a_r^t . k^{(t+1)-2007} \qquad\qquad r = 0,1,2,3 \text{ and } k \geq 0$$

where a_r^{t+1} represents the level parameters in the Rogers-Castro curve describing the level of migration in the various age ranges over the period t to $t+1$, k is the scalar by which the level of migration is multiplied each year. The level parameter subscript r in a model migration schedule identifies the component of the labour force. Thus the schedule fitted to the census data contains the migration level parameters a_1, a_2, a_3, a_0, where these denote levels of pre-labour migration, labour migration, post-labour migration and the constant component. a_0 is also written c in other demographic literature.

In order to obtain the scaling factors k, separate male and female intercensal and annual migration rates were derived as age-specific proportions of the average population as at the middle of the intercensal periods for 1996-2001 and 2001-2007. The total equivalents of these rates were also obtained this way. In order to determine the percentage change in the level of migration between 1996 and 2007, the difference in the total intercensal migration rates, expressed as a proportion of the total rates in the 1996-2001 period, were calculated for males and females, separately, using the formula:

$$\Delta = k = \frac{r^{2001-2007} - r^{1996-2001}}{r^{1996-2001}} . \frac{1}{5.35} \qquad (20)$$

where $r^{1996-2001}$ denotes the aggregate migration rate expressed as the proportion of the population receiving the migrants. The percentage change described by Equation 20 was divided by 5.35 years (period between the night of 9/10 October 2001 and mid-February 2007) in order to obtain the annual percentage change k.

Chapter 4 looks at results generated by the multi-regional model in relation to the projected values derived from the full version of the ASSA model.

4 DATA ANALYSIS AND RESULTS

This chapter discusses the results obtained by means of the multi-regional approach and compares these to results obtained from the net migration version of the ASSA model.

Section 4.1 discusses trends in internal migration for the period 1996-2007, age patterns of internal migration over the same period, and how these have changed over time. The section also considers the age-specific rates of migration obtained by means of curve-fitting, how well the curves fit the census migration data and how plausible these rates are. Section 4.2 looks at the multi-regional projection of internal migration rates and how well the projected rates retain their age patterns going forward. Finally, Section 4.3 discusses the results in two stages. First, model-testing results and, second, the impact of the level of the assumed migration rates on the projected population size and age structure, relative to the impact of the use of multi-regional modelling itself.

4.1 Internal migration

Multi-regional models need extensive census migration data. These were derived from the 10 per cent sample of the 1996 Census and the 2007 Community Survey. However, it must be borne in mind that the census migration data are often incomplete or deficient. The Community Survey is also not entirely accurate in capturing period migration rates.

The data issues associated with the two national enquiries will be discussed in more detail in Chapter 5. In short, however, the two major issues with the data are that, first, the 1996 and 2001 censuses undercounted the population by 10.7 per cent and 17.6 per cent, respectively, and particularly children aged 0-4. Second, period reference errors were evident in these national enquiries, and these were in the form of a general preference for reporting most recent moves as having occurred in the twelve months before the censuses. Finally, the quality of the 2001 census results was affected by the scanning problems, where province identifiers such ('1' through '9') were misread by the scanner, particularly 1's and 7's. These shortcomings will be discussed in more detail in the next chapter.

4.1.1 Overall trends in migration

The 2001 South African census and the 2007 Community Survey estimated a total of 1,836,411 and 1,638,449 inter-provincial migrants (male and female combined) in the population, respectively. These numbers translate into a total of 1,360,771 migrants and 1,192,654 migrants for the respective periods, when one allows for the combination of seven provinces (barring Gauteng and the Western Cape) into the RoSA. Gauteng had a total of 317,541 migrants who moved to the Western Cape and the rest of South Africa during the period 1996-2001, and a further 299,257 during the period 2001-2007, a drop of 6 per cent in the level of out-migration for the region. The Western Cape had 102,334 migrants who left during 1996-2001, and another 98,626 in the subsequent intercensal period, and recorded a drop of 4 per cent in the level of migration for the province. All other seven provinces, combined (thus referred to as RoSA), had a total of 940,896 people migrating to Gauteng and the Western Cape during 1996-2001, and another 794,772 during the period 2001-2007. This is a decline of 16 per cent in the numbers of out-migrants leaving for the Western Cape and Gauteng.

The numbers above were obtained before adjusting both the census/survey migration data and population numbers for mortality, and they translate into the overall period migration rates as shown in Table 4.1, ignoring international migration for the moment.

Table 4.1: Out-migration per region as a proportion of the population at risk of migrating to other regions, 1996-2001 and 2001-2007

	Male			Female		
Period	GP	WC	RoSA	GP	WC	RoSA
1996-2001	0.045	0.026	0.033	0.040	0.024	0.030
2001-2007	0.033	0.022	0.029	0.031	0.020	0.024

It can be seen from Table 4.1 that the general level of migration within the country has been in the 2%-4% range per five-year period over the fifteen years.

Note also the directional strength of the flows between the three regions in terms of their importance as sources of migrants and receiving regions, and this is shown in Table 4.2, representing the combined male and female numbers of migrants flowing between the three regions.

Table 4.2: Combined male and female origin-destination flows between regions, 1996-2001 and 2001-2007

Period	Origin	Destination region			
		GP	WC	RoSA	Total
1996-2001	GP		62,263	255,278	317,451
	WC	34,717		67,617	102,334
	RoSA	705,774	235,123		940,896
		740,491	297,386	322,984	1,360,771
2001-2007	GP		50,267	248,989	299,257
	WC	38,056		60,570	98,626
	RoSA	640,656	154,116		794,772
	Total	678,712	204,383	309,559	1,192,654

Recall that the RoSA is a combination of seven provinces that exclude Gauteng and the Western Cape. As a result, migrants moving between these provinces are deemed as changing residence within the region RoSA, thus a large number of them is ignored for the purpose of this study. Thus, Gauteng is the largest source of migrants just as it is the most important as a destination region.

Table 4.2 shows that migration numbers for the period 2001-2007 are lower than those of the preceding intercensal period. Thus the general level of internal migration (for males and females combined) in the 1996-2001 period was 12 per cent higher than that of the second period.

Consider the combined male and female annual origin-destination migration rates for the two periods, as shown in Table 4.3, and calculated as described in Section 3.4.2.

Table 4.3: Annual origin-destination period migration rates: 1996-2007 (per cent)

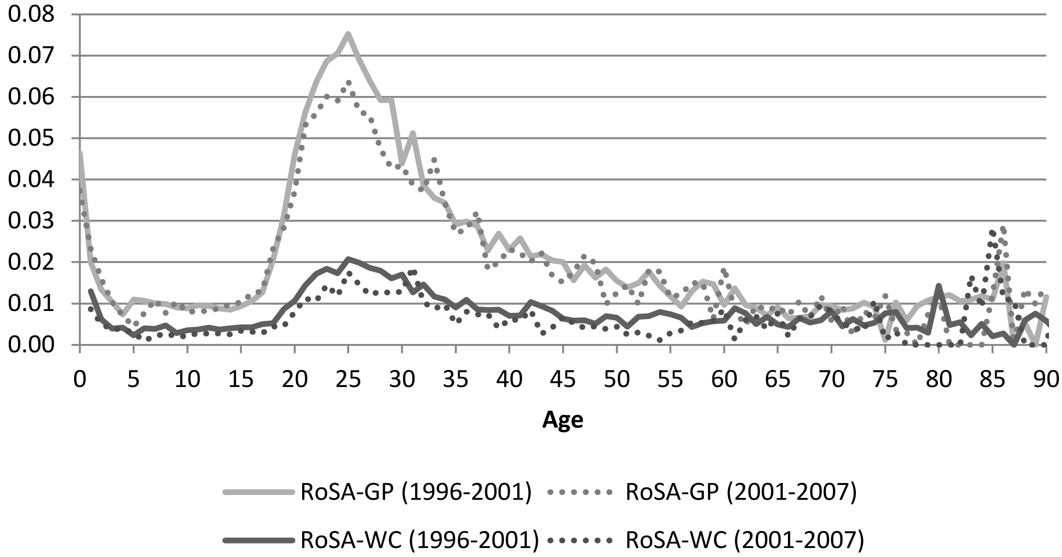
Period	Origin	Destination		
		GP	WC	RoSA
1996-2001	GP		0.17	0.62
	WC	0.17		0.31
	RoSA	0.47	0.16	
2001-2007	GP		0.10	0.47
	WC	0.14		0.23
	RoSA	0.36	0.10	

Based on Table 4.3, annual rates for the period 2001-2007 were, on average, 28 per cent lower than those of the preceding period.

4.1.2 Age patterns of migration

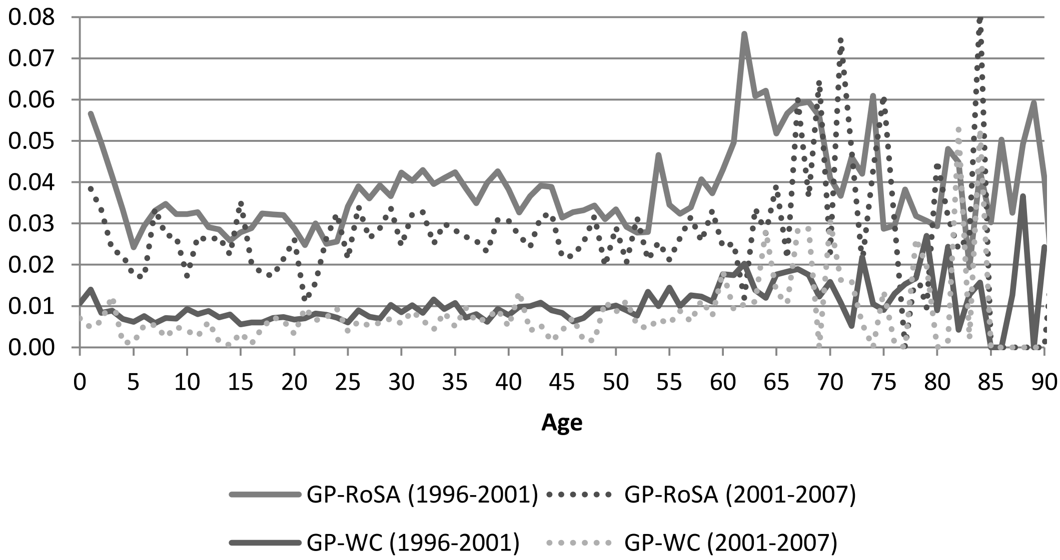
Figure 4.1 shows the plot of period age-specific out-migration rates for males from the RoSA to Gauteng and the Western Cape over the period 1996-2007. Note that age patterns of female migration exhibit the same features.

Figure 4.1: Male out-migration rates from the RoSA to GP and WC, by age, 1996-2001 and 2001-2007



The feature that stands out in Figure 4.1 is that the shapes of the plots show standard age profiles of a migration curve. This is expected because most of the migrants leaving the RoSA for Gauteng and the Western Cape do so for work reasons. The second feature is that in each migration stream, namely RoSA to Gauteng and RoSA to the Western Cape, the 2001-2007 migration rates are lower than the 1996-2001 level. This is evident in the young adult (ages 17-23) peaks. The same can be said of other migration rates as shown in Figure 4.2 with the Gauteng to RoSA and the Western Cape flows.

Figure 4.2: Male out-migration rates from GP to WC and RoSA, by age, 1996-2001 and 2001-2007



Note that the age distribution of the migration rates in Gauteng is not standard. This is to be expected because the prospects of finding employment are greater in

Gauteng than the RoSA. Out-migration rates for Gauteng are in the older ages because at these ages migrants are generally retirees.

Plots of the rates of out-migration from the Western Cape to RoSA are also not standard migration age patterns due to lack of work opportunities in the RoSA. The second feature noted in Figure 4.2 is the decline in the level of migration over the 10.35-year period.

Comparing the migration rates obtained above to those derived by Dorrington and Moultrie (2009), note that the two sets of estimates are not entirely consistent with each other. This is due to combining seven of the nine provinces to form one region for the purposes of this research, where migrants moving between the seven provinces are now deemed as 'non-migrants' since these are still in this region. This adjustment thus lowered the general level and volume of migration, but the change was negligible. The aggregate period migration rate of 4 per cent for 1996-2001 declined to 3 per cent for that period, and similarly, the rate of 3 per cent for 2001-2007 declined to 2 per cent when migrants moving between the RoSA provinces are excluded. Therefore the change in the aggregate period migration rates stated above did not have a significant impact on the overall results.

4.1.3 Fitting multi-exponential schedules to migration data

Rogers-Castro multi-exponential curves were fitted to the annual age-specific migration rates, using the curve-fitting software package MATLAB 2011, in order to derive shape and level parameters for each year between 1996 and 2007, inclusive (The MathWorks 2011).

In order to fit the Rogers-Castro age schedules to census migration rates, it was necessary to derive annual migration rates on the basis of a set of assumptions as stated in the Dorrington and Moultrie (2009) paper. Mortality and migration assumptions, as well as the method used to derive annual migration rates, are stated in Section 3.4.2. Additional assumptions needed for this process to be undertaken are that, first, there is no selection effect of duration of residence for onward migration - regardless of the duration of residence; all migrants experience the same migration propensity. Second, that in any one year, the force of migration at any age is constant with respect to time over that year (Dorrington and Moultrie 2009).

Figure 4.3 and Figure 4.4 show the *7-parameter* Rogers-Castro age schedules fitted to annual census and survey migration rates from the RoSA to Gauteng for the periods 1996-2001 and 2001-2007, and compares these to observed migration rates.

Figure 4.3: Age schedules fitted to observed male migration rates from RoSA to GP, 1996-2001 and 2001-2007

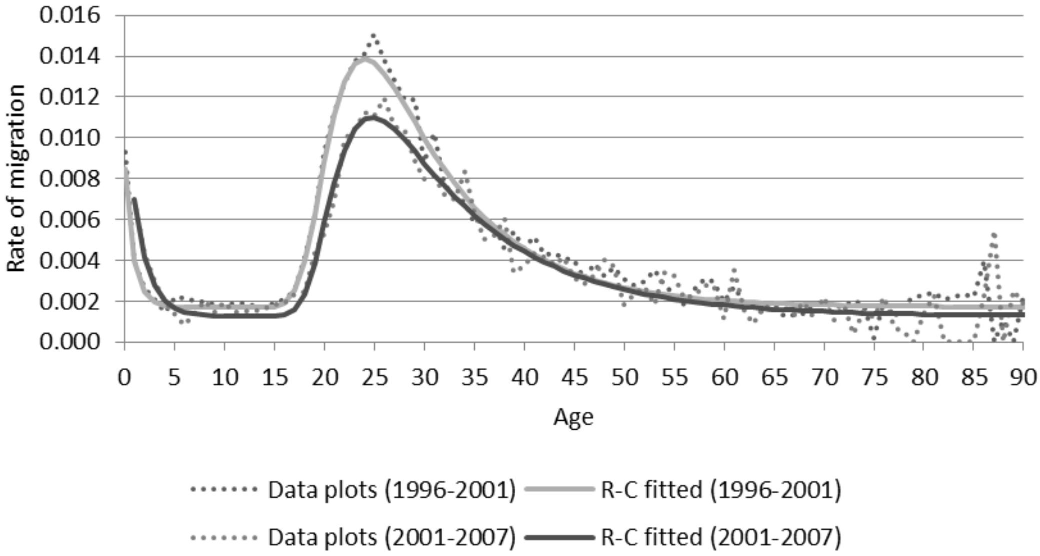
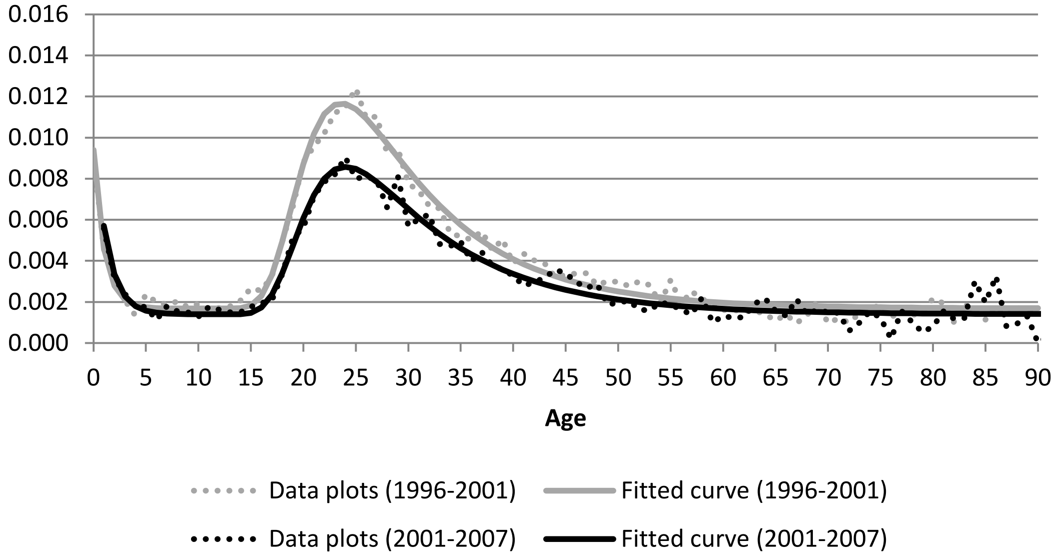


Figure 4.4: Age schedules fitted to observed female migration rates from RoSA to GP, 1996-2001 and 2001-2007



Note that the standard age schedules fit the census/survey migration data well for each sex since the observed migration age profile is regular and the young adult peaks are pronounced in each case. The goodness-of-fit (R^2) for the two migration curves for each sex is close to 1. Plots of the equivalent female migration streams and the RoSA-Western Cape (for the same periods) are not shown here, but the same scenario holds for those flows as well. The associated parameters are given in Appendix 5.

It is important also to note that the rates in Figure 4.3 and Figure 4.5 are roughly one-fifth of those in Figure 4.1 and Figure 4.2 because the period migration rates were converted into annual migration rates for curve-fitting purposes

Figure 4.5: Age schedules fitted to observed male migration rates from GP to RoSA, 1996-2001 and 2001-2007

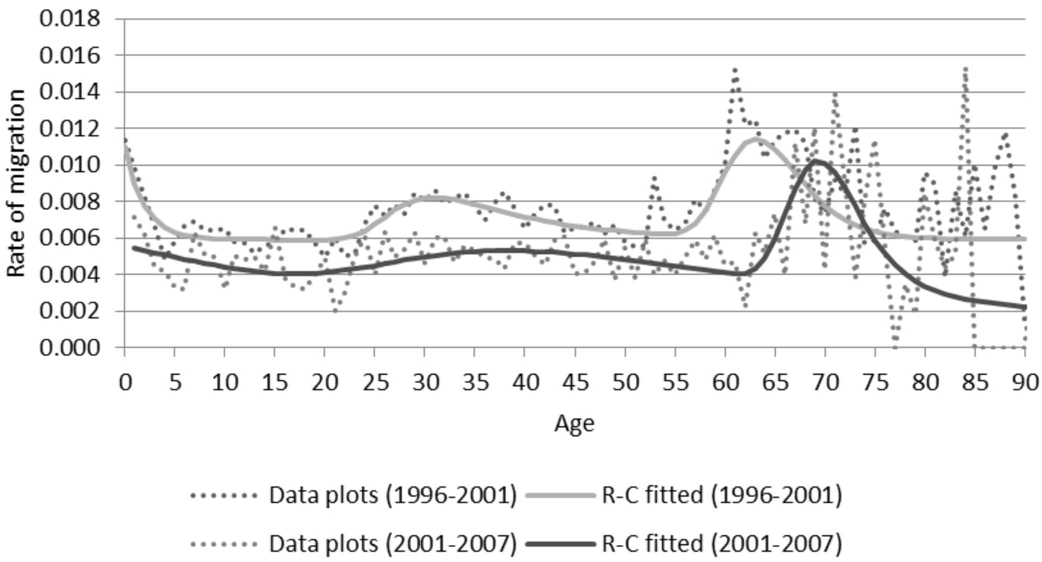


Figure 4.6: Age schedules fitted to observed female migration rates from GP to RoSA, 1996-2001 and 2001-2007

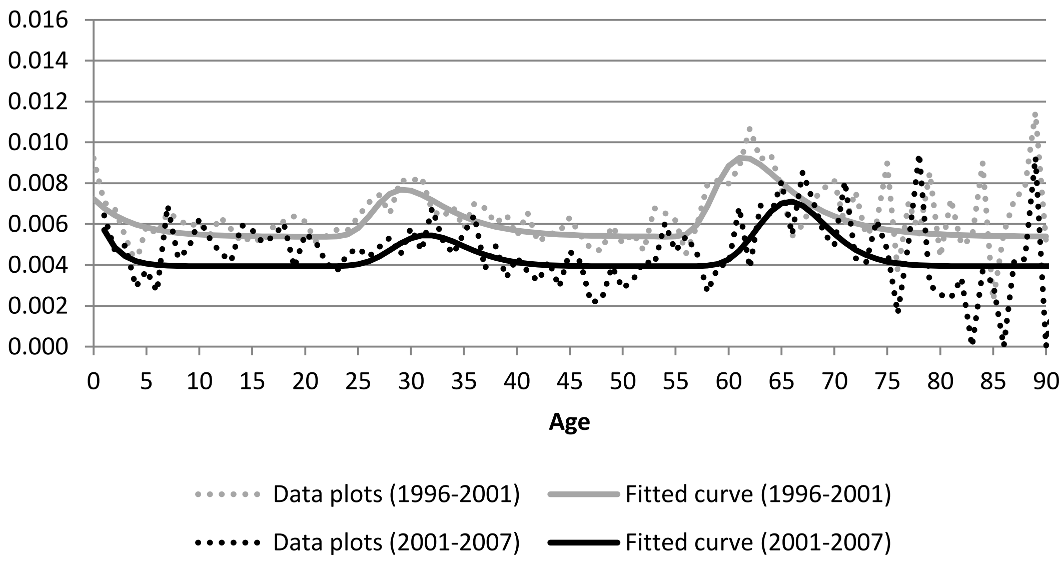


Figure 4.5 and Figure 4.7 compare the plots of the fitted *11-parameter* Rogers-Castro curves to the 2001 census and 2007 survey migration data plots for males and females for the Gauteng-RoSA flow. In both cases, the fitted curves confirm the earlier observation that the migration rates are highest for migrants leaving Gauteng for other regions at older ages (50+). Note the right-ward shift in the retirement peak in the 2001-2007

intercensal period. It is not clear why there is a rightward shift – this will need further investigation. The male and female plots for the Western Cape are not shown, but the same scenario holds for these.

All three parameter tables are shown in the Appendices on pages 83-85.

The drop in migration, as seen in the diagrams above, is somewhat difficult to explain and hence also requires further investigation.

4.2 Multi-regional annual migration estimates

The multi-regional adaptation of the ASSA model requires that the level parameters, namely, a_0, a_1, a_2 and a_3, be projected forward, applying a common scalar to all of these without altering the shape of the age patterns of migration (Gullickson 2001). The extrapolated migration rates were allowed to decline roughly linearly over time after 2007. The scaling factors used in the calculations of internal migration rates after 2007 were obtained by means of the method described in Section 3.6, and these are shown in Table 4.4.

Table 4.4: Male and female migration scaling factors for the period 2007+

Origin	Destination					
	GP		WC		RoSA	
	Male	Female	Male	Female	Male	Female
Gauteng			0.95	0.93	0.95	0.96
Western Cape	0.99	0.98			0.96	0.96
Rest of SA	0.98	0.97	0.95	0.93		

Based on the scaling factors in Table 4.4, male and female migration rates from Gauteng to other regions are expected to decline into the future at average annual rates of 5 per cent and 7 per cent; from the Western Cape to other regions, by 1 per cent to 4 per cent and from 'rest of South Africa' to other regions, by 2 per cent to 7 per cent.

4.3 Multi-regional population projection

This section consists of two sub-sections. Section 4.3.1 demonstrates that the multi-regional adaptation of the ASSA model works. This is done by first running the multi-regional model, after which the annual age-specific net numbers of migrants implied by the model for each region are copied and pasted onto the migration sheets of the *lite* version for each region, with the same starting demographic, behavioural and epidemiological assumptions as in the multi-regional model. This is explained in detail in Chapter 3 of this dissertation. Section 4.3.2 demonstrates the impact of using the multi-regional migration model in projecting migration on the results versus the net migration model. This section briefly compares the two approaches by looking at the impact on the population size, age structure and net migration rates per 1,000.

4.3.1 Testing the multi-regional adaptation of the ASSA model

The multi-regional model was tested in order to ensure that it works. Further, model-testing was carried out in order to ensure that any divergence between the multi-regional and net migration models is not due to the modelling of migration. The multi-regional model was set to reproduce the net migration numbers in the original model.

Comparisons done with respect to all components measured by both the ASSA net migration and multi-regional migration models show that the two models match exactly. This is illustrated by the comparisons of the models in terms of the annual net migration rates per 1,000 and annual numbers of patients on antiretroviral treatment.

Consider the visual comparisons shown in the following diagrams.

Figure 4.7: Comparison of the projected net migration rate per 1,000 for Gauteng, Western Cape, and rest of South Africa, 1985-2025

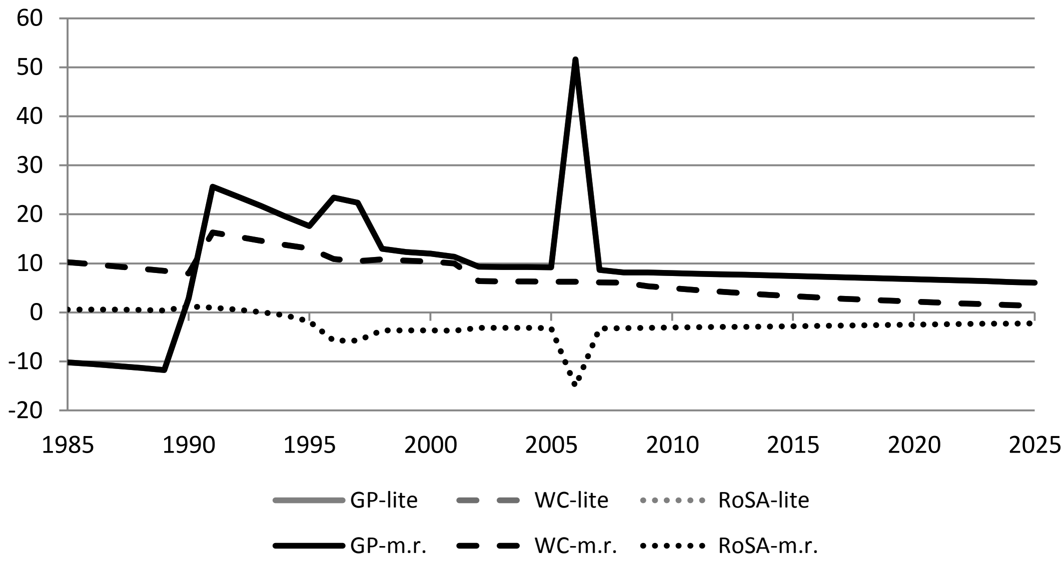


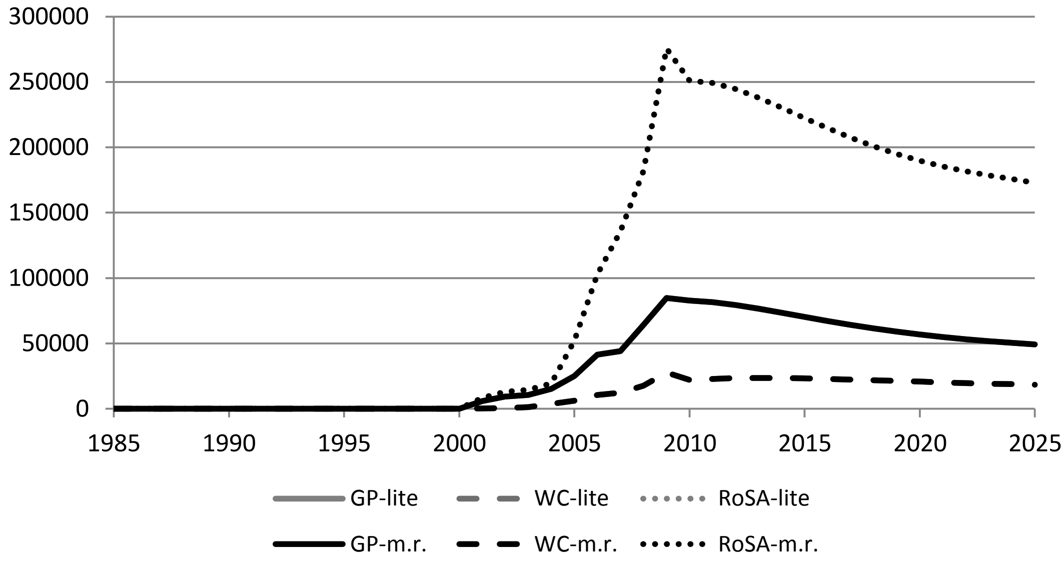
Figure 4.7 compares projected annual net migration rate per 1,000 obtained from the multi-regional model to that obtained from the *lite* versions of the model[4] over the projection period 1985-2025. Note that both the *lite* and multi-regional versions of the model produce the same results over the entire projection period since the net migration curves for each region coincide exactly, for all years - spike and trough for Gauteng and 'rest of South Africa' in 2006 are a result of the boundary changes for Gauteng and the neighbouring provinces such as the North-West in 2006. This change meant that some municipalities were moved between provinces. This is discussed in more detail in section 4.3.2.

[4] File name: AssumptionsProv_110207, modified on 2011/02/17

Recall that the comparability of the two models was ensured by first running the multi-regional model for the period 1985-2025 in order to derive the annual numbers of migrants. These numbers were then copied and pasted into the male and female migration sheets of the *lite* versions – for each of the three regions – that were subsequently run in order to generate equivalent projections for the period 1985-2025, for each region, for comparison. Thus the migration component in the multi-regional model works.

In order to establish further that the multi-regional model works, the projected annual numbers of new patients on antiretroviral treatment were compared, and this is shown in Figure 4.6.

Figure 4.8: Comparison of the projected annual numbers of new patients on ART in Gauteng, Western Cape and rest of South Africa, 1985-2025



Note that, as with the projected net migration numbers in Figure 4.7, the curves plotting projected annual numbers of new patients on antiretroviral therapy coincide exactly for the whole projection period.

Additional visual comparisons were made with respect to deaths, births, growth rates per 1,000 and total population, and these are shown in Appendix 7.

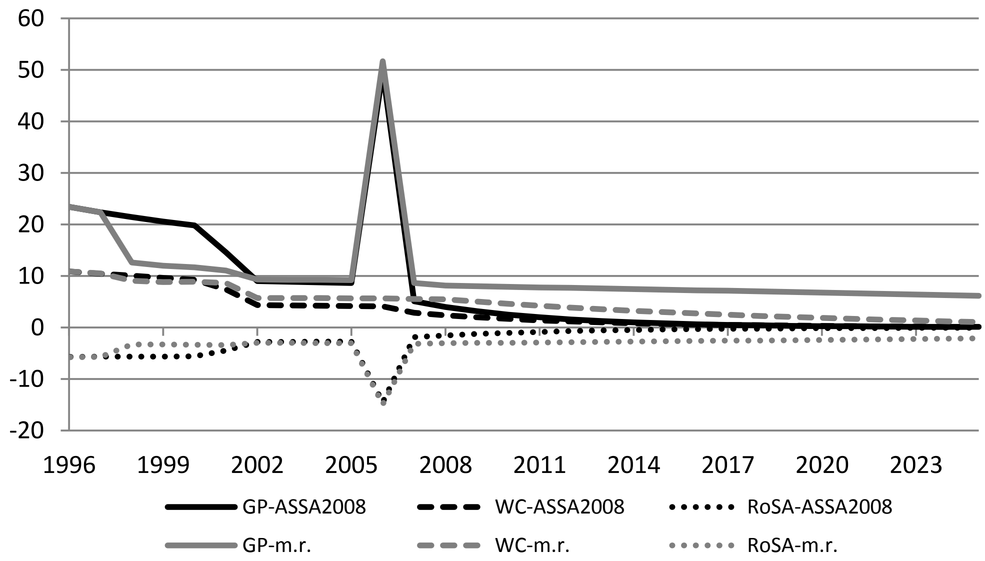
In conclusion, the multi-regional model works as it should in terms of all the demographic, epidemiological and behavioural estimates, though not without some calibration that incorporated the fact that this is now a multi-regional model. Once the new net numbers of migrants were derived from the multi-regional model, these were pasted into the *lite* three versions for Gauteng, the Western Cape and 'rest of South Africa'. Each of the regional *lite* versions was also re-calibrated to antenatal HIV clinic data. Thus the projections derived from the multi-regional model coincide with the estimates derived from the ASSA *lite* versions for Gauteng, the Western Cape and 'rest of South Africa'.

4.3.2 Multi-regional adapted projection modelling versus ASSA net-migration modelling

Having established that the multi-regional model works, we now turn our attention to considering the effect of projecting the population using the multi-regional approach. The two models are first compared in terms of projected annual net migration rates per 1,000 in each region over the period 1997-2025.

Projected net migration rates per 1,000 for Gauteng, Western Cape and the RoSA, after 1997, obtained from the multi-regional and ASSA models, are shown in Figure 4.9.

Figure 4.9: Projected annual net migration rates per 1,000 for Gauteng, Western Cape, and rest of South Africa 1996- 2025



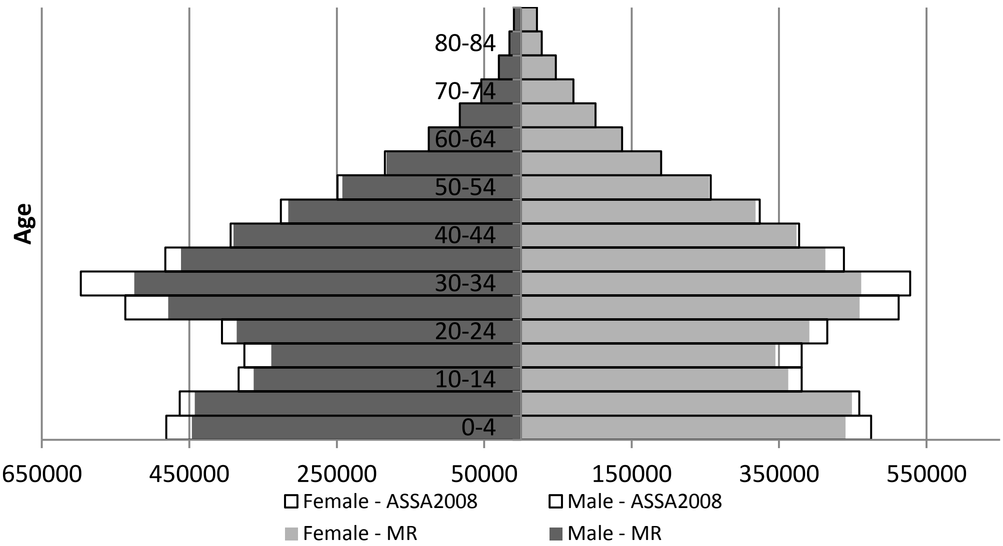
Note that, for Gauteng, the ASSA model exaggerates the annual rates for the period 1998-2001 and under-estimates the rates after 2007. A particular feature noted in Figure 4.9 is that the projected net migration rate for 2006 for Gauteng, as obtained from the net migration model, increases by 41 per 1,000 from the 2005 levels. This is connected to the projected rate for the RoSA for the same year declining by 12 per 1,000 from the levels in the preceding year. The peak and the trough in the net migration rates projected for Gauteng and the RoSA, respectively, for 2006, are a result of boundary changes that occurred in 2006. The changes were accounted for in the calculation of the rates in the multi-regional model and thus the higher net numbers of migrants for 2006 are not the result of this hiatus.

For the Western Cape, the projected rates, as obtained from the ASSA model, are consistently higher than those derived from the multi-regional model for the period after 1997. The Western Cape does not have any peak or trough in its projected rates because the boundaries of the province did not change in 2006.

The results obtained above have a direct bearing on the projected population age pyramids for all three regions, but Gauteng and the RoSA deserve particular attention. Consider the impact of modelling the regional population using the multi-regional projection model that contains directional out-migration rates for each region and compare the results obtained from the two models in terms of the regional population age structure, as shown in Figure 4.9.

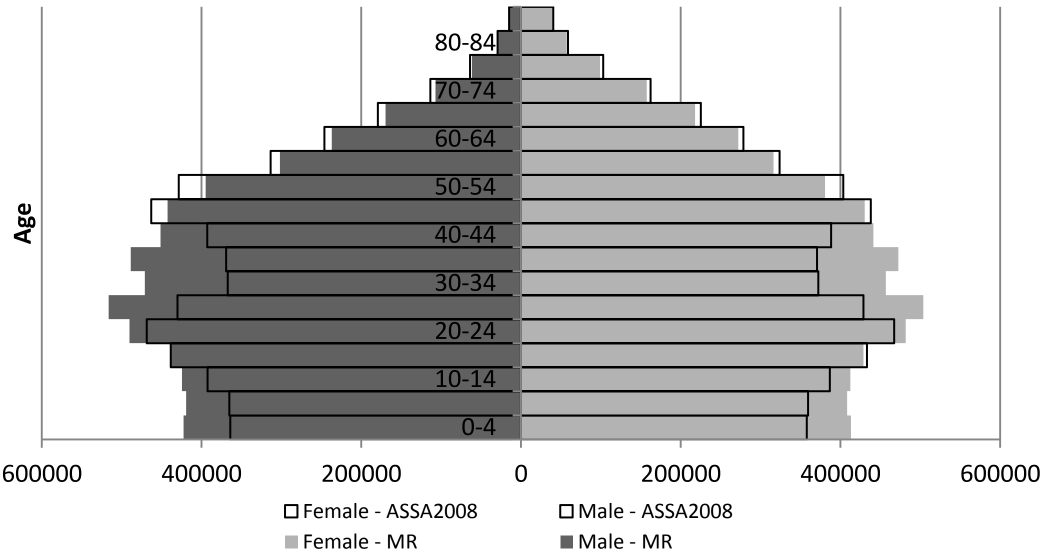
Figure 4.10: Comparison of the projected age structure of the population in Gauteng, 2007 and 2025

(i)



Panel (i) shows the projected age structure for 2007.

(ii)



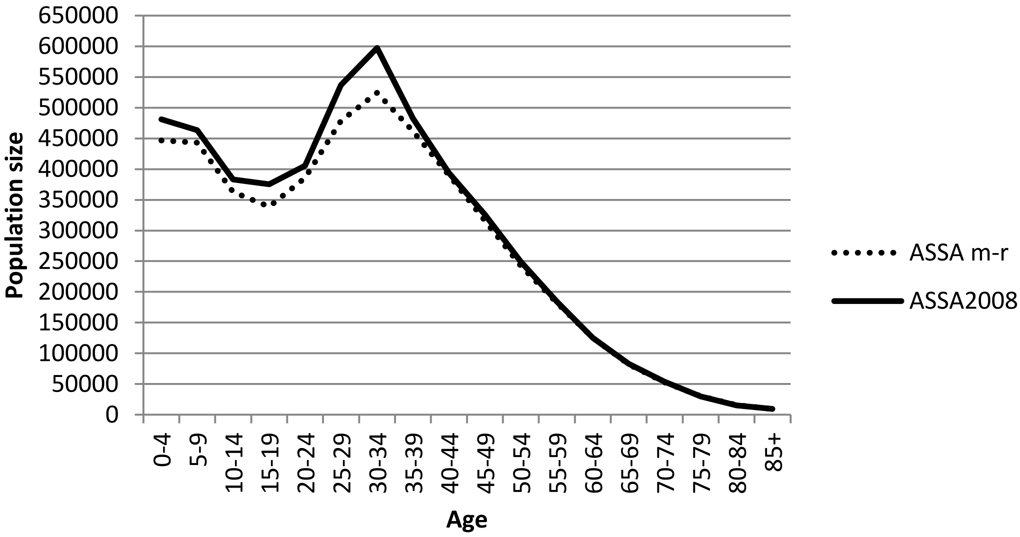
Panel (ii) shows the projected age structure for 2025.

For 2007, we see a relatively wider population pyramid for ages 0-14 and 24 -44, than obtained from the ASSA model for males and females, separately. Figure 4.10 also shows the demographic consequences of under-estimating the annual net numbers of migrants, and this is evident in the later years up to 2025. Consider the projected population age structure in Gauteng for 2025, for example. The age pyramid for Gauteng, derived from the multi-regional model is expected to be relatively wider than that obtained from the ASSA model, particularly for males in the 0-44 age range.

The impact of modelling the population multi-regionally was presented in the preceding population pyramids in aggregate for any particular year, in this case 2007 and 2025. Consider the age-specific effects introduced by the methodology, and we look at these using the male population age distribution for all three regions as shown in Figure 4.11. The female distributions are shown in Appendix 8.

Figure 4.11: Comparison of the projected population age distributions for males in Gauteng, 2007 and 2025

(i) 2007



(ii) 2025

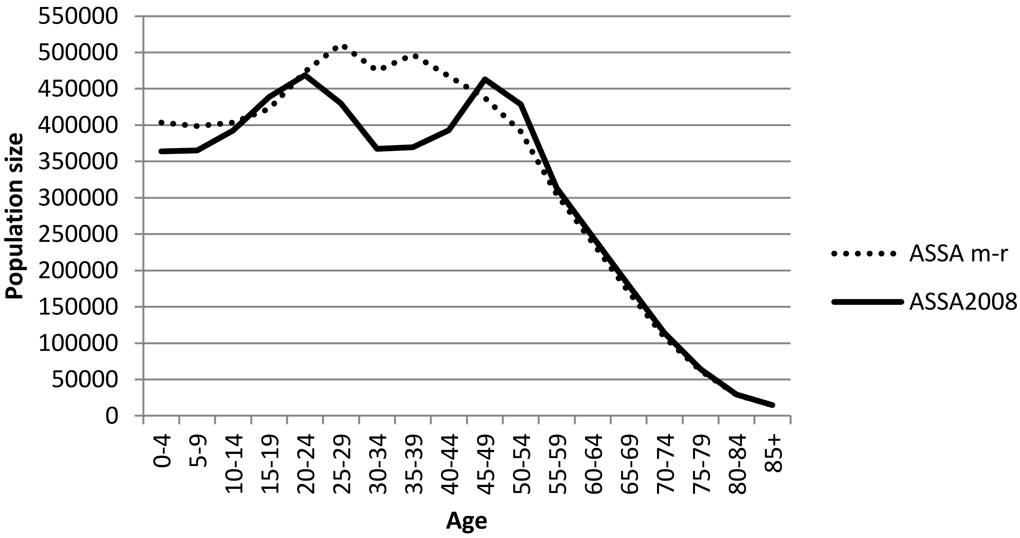


Figure 4.11 shows the projected population distribution of males by age in Gauteng for 2007 and 2025, and compares the age distributions obtained from the two models. The age axis is expressed in quinquennial age groups.

Note the effects of the application of age-specific migration rates on the regional populations at risk of migrating, using the multi-regional model. The age-specific effects of the methodology are significant for Gauteng, and this is seen with the substantial differences between the two models for age groups in 2007, where ASSA2008 projects the age distribution that is higher than that projected by the ASSA multi-regional model by 4 per cent to 12 per cent around these ages.

These differences are reversed between the two models in 2025. First, the ASSA2008 model over-projects the population aged 15-19 by 4 per cent, and underprojects the population aged 0-4 and 5-9 by 11 per cent and 9 per cent, respectively. The inconsistency is also significant for ages 25-29 through to 40-44, where the ASSA2008 model under-estimates the population in these age groups by 19 per cent to 35 per cent.

These differences are explained by the fact that numbers of migrants generated in the multi-regional model are a function of the underlying 'at-risk' age distributions, that is, populations that give rise to the migrants.

Consider the results obtained by projecting the population using out-migration rates – and doing so multi-regionally – relative to results obtained from the ASSA model.

For Gauteng, we see a decrease of 3 per cent in the population estimate for 2007 from the estimate obtained from the ASSA model, and an increase of 9 per cent for 2025 in the projected population as a result of projecting the out-migration rates multi-regionally. The multi-regional population modelling for Gauteng, without the assumed level of the migration rates, causes declines of 0.1 per cent for 2007 and 0.4 for 2025 in the mid-year population estimates relative to the ASSA model. This impact is better illustrated in Table 4.5.

Table 4.5: Average proportions of changes in population size attributable to multi-regional modelling and migration rates: 2007, 2010 and 2025

Region	GP	WC	RoSA
Migration rates	94%	95%	78%
M.R. modelling	6%	5%	22%

Based on this, the level of the assumed migration rates in the multi-regional model has a greater impact on projected population numbers than does the use of the multi-regional modelling approach on its own. Consider projections for Gauteng, for example. For this region, the multi-regional projection of migration accounts for an average of 94 per cent of the changes in population estimates as the two models are compared, while 6 per cent of these changes are attributable to the multi-regional projection approach itself.

An important point to note in Table 4.5 is that the proportions for the 'rest of South Africa' are different from those of the other two regions. This is due, in part, to the fact that the region is a combination of seven provinces that differ from one another in their demographic and epidemiological profiles. For instance, the region includes KwaZulu-Natal, a province with the highest HIV incidence and prevalence rates, and thus the highest AIDS mortality rates. Further, the seven provinces do not grow at the same rate.

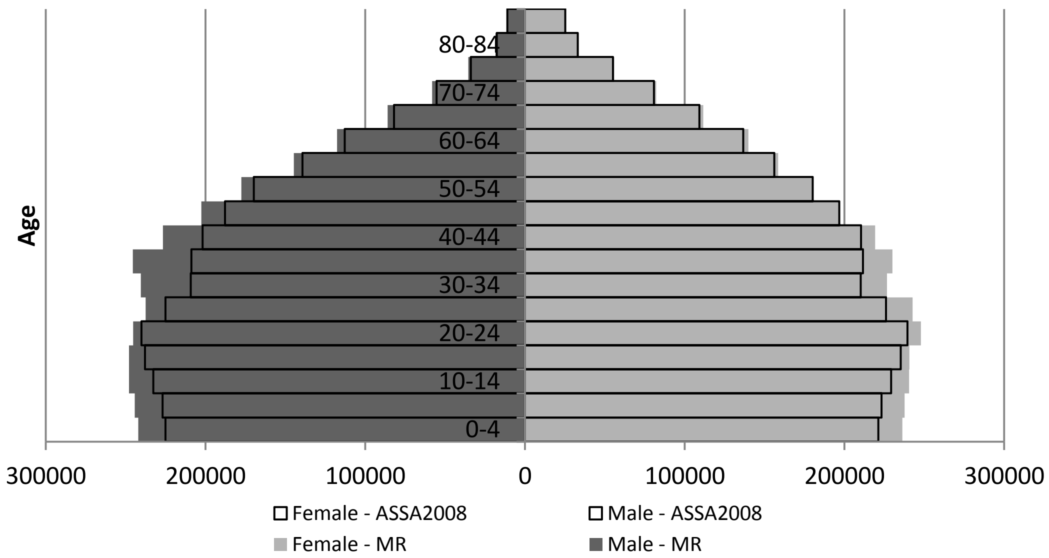
These proportions were derived by calculating the ratio of the percentage change in the multi-regional adaptation of the ASSA model when only the net numbers of migrants in the original ASSA model are used, to the total percentage change between the two models for each of the years 2007, 2010 and 2025, where the total percentage change is introduced as a result of the use of the assumed migration rates in the multi-regional model. Thus for each of the years 2007, 2010 and 2025, the proportions of changes in population size attributable to multi-regional modelling and migration rates, are given as

$$proportion = \frac{\%\Delta^t_{net}}{\%\Delta^t_{mr}}$$

where the numerator denotes the percentage change in population estimates as at the middle of year t obtained from the multi-regional model that incorporates the original net numbers of migrants in the ASSA model, and the denominator denotes the percentage change in population estimates as at the middle of year t obtained from the same adaptation when the migration rates are used in the model.

For the Western Cape, the two models do not differ significantly for 2007 since the population estimate for 2007 rises by 0.5 per cent from the estimates obtained from the ASSA model, but we see some divergence by 2025, with the estimate for this year increasing by 5 per cent relative to the estimate obtained from the ASSA model applicable to 2025. This is shown in Figure 4.12.

Figure 4.12: Comparison of the projected age structure of the population in Western Cape, 2025



The projected age pyramid for the Western Cape for 2025, shown in Figure 4.12, shows that the multi-regional population projections are consistently higher than the ASSA model projections. This is, once again, a result of modelling the population by incorporating the migration rates in the multi-regional model and treating these transparently.

The age-specific effects of modelling the Western Cape population multi-regionally, once again, are evident in Figure 4.13

Figure 4.13: Comparison of the projected population age distributions for males in the Western Cape, 2025

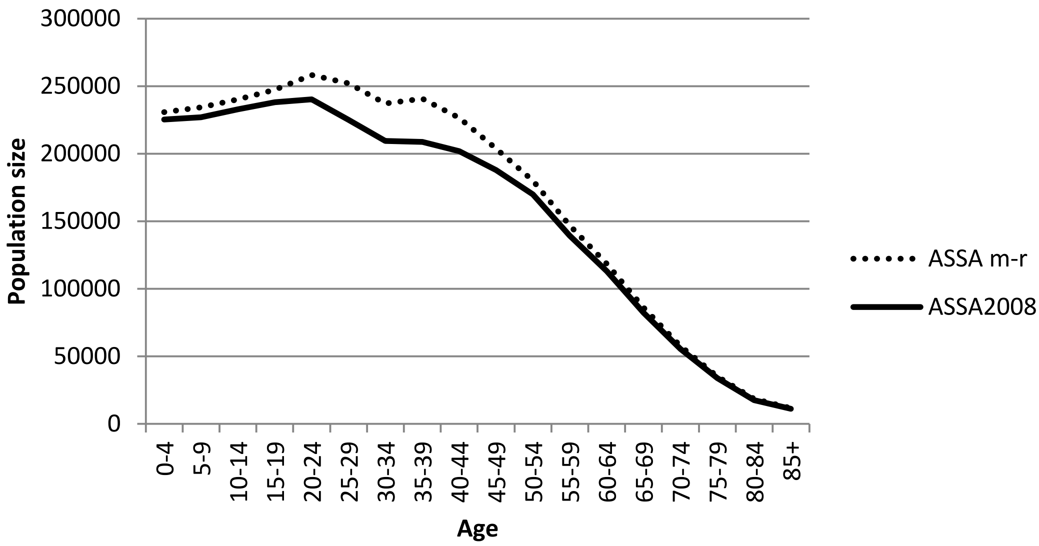


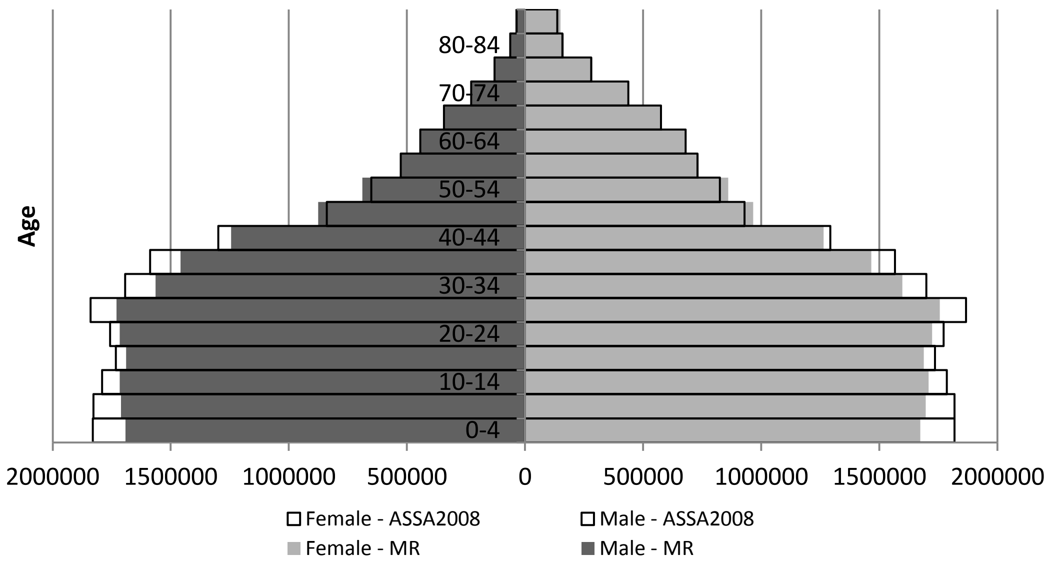
Figure 4.13 plots the age distribution of the male population in the Western Cape as at 2025. As with the population age distribution for males in Gauteng in 2025, the same can be said of the comparisons of the male age distributions in the Western Cape for the same year. The multi-regional model takes into account the 'at-risk' populations giving rise to the migrants, as opposed to the original ASSA model, and this is more evident in the age groups 0-4 through to 50-54. The differences are smaller for older ages 55+.

The age-specific effects noted in the diagram are consistent with the net migration rates per 1,000 noted in Figure 4.9, and this is also true of Gauteng. Note that, in the Western Cape, the differences range between 8 per cent and 15 per cent if one considers the age groups 20-24 through to 45-49, with the smallest difference of 8 per cent recorded at ages 20-24 and 45-49. The largest difference of 15 per cent is recorded at ages 35-39. These differences are consistent with the fact that most of the migrants moving into the Western Cape are those younger than the age of 50.

Looking at the relative impact of modelling the migration rates multi-regionally versus the multi-regional approach itself on the projected population of the Western Cape for 2025, one notes that the migration rates assumed in the model account for 95 per cent of the increase in the population size of the Western Cape as compared to the estimate obtained from the ASSA model, while the multi-regional modelling accounts for only 5 per cent.

For the RoSA, the opposite of what is seen for Gauteng is true for 2007 and 2025. The male and female population age structure in this region for 2025 is shown in Figure 4.14.

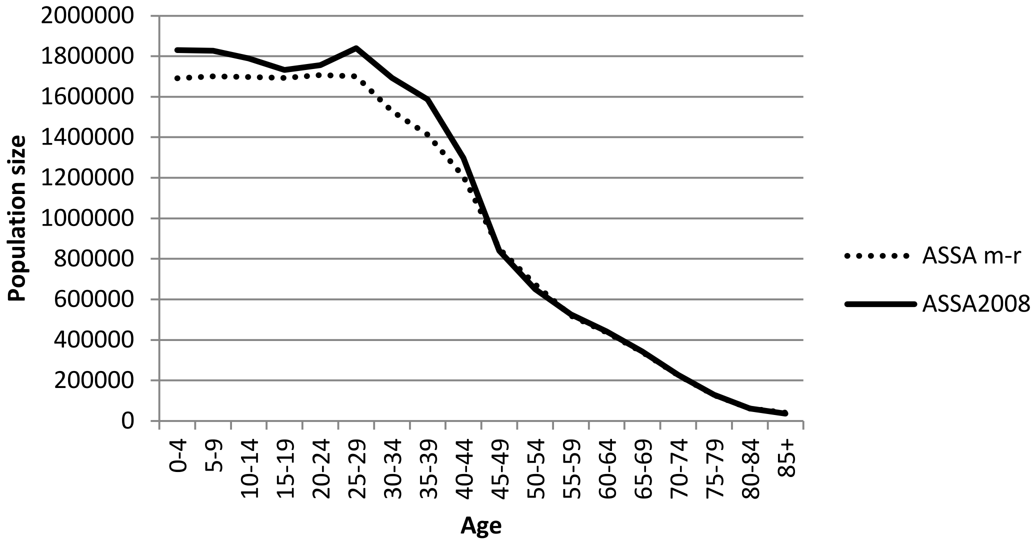
Figure 4.14: Comparison of the projected age structure of the population in the rest of South Africa, 2025



The demographic impact of using the multi-regional approach to projections on the population age structure for the RoSA is not noticeably different for the year 2007, but this impact is evident in the later years. The impact is more pronounced in the 0-44 age range for males and females since the majority of migrants in this segment of the population migrates to Gauteng and the Western Cape for work and study opportunities.

This impact can also be demonstrated by means of the population age distribution for 2025, as shown in Figure 4.15.

Figure 4.15: Comparison of the projected population age distributions for males in the rest of South Africa, 2025



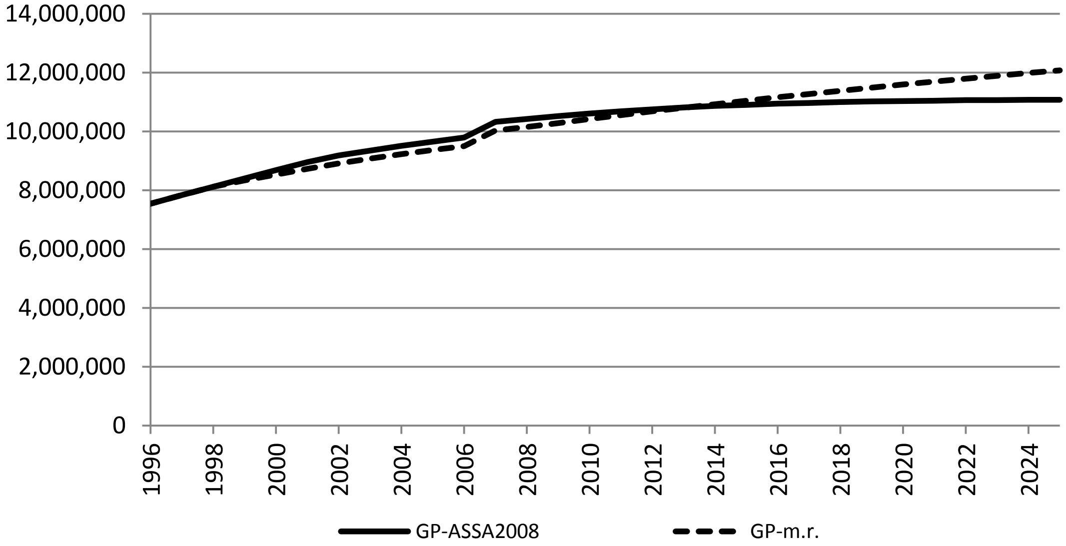
Note that, for this region, the original version of the ASSA model projects a higher male age distribution for 2025 than does the multi-regional model at ages 0-4 through to 40-44. This is, once again, a result of the migrants leaving this region for Gauteng and the Western Cape, with most migrants from the Eastern Cape moving to the Western Cape and those from the North-West, KwaZulu-Natal , Mpumalanga, Limpopo and the Free State moving to Gauteng.

The age differences between the two models are negligible for 2007, ranging between 0 per cent and 2 per cent across all age groups.

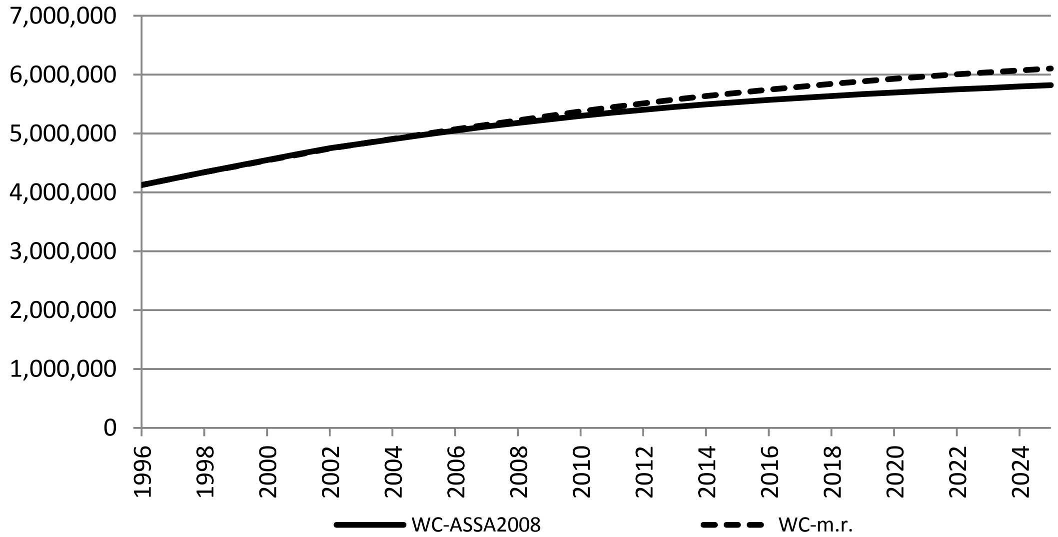
The relative average impact of using migration rates in projecting the population size for the RoSA accounts for 78 per cent of the decline from the estimate obtained from the ASSA model while the multi-regional model itself explains only 22 per cent of the decline. The demographic impact of the multi-regional projection is evident for all regions. The observations above are best captured visually, as shown in, where we plot the projected population size for Gauteng, the Western Cape and the rest of South Africa, respectively, over the period 1996-2025.

Figure 4.16: Projected population size, 1996-2025

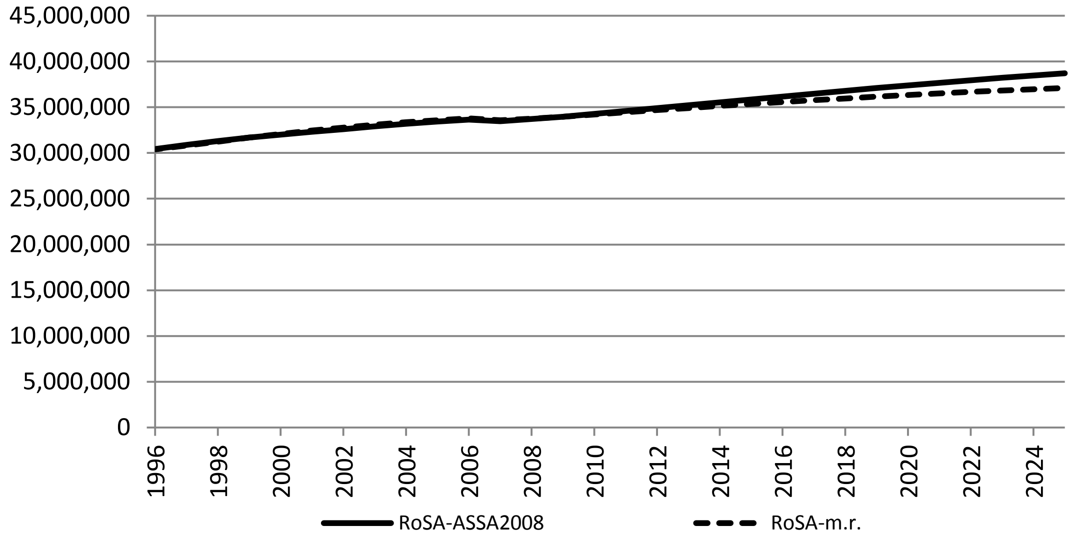
(a) Gauteng



(b) Western Cape



(c) Rest of South Africa



Note that the population size generated by the multi-regional model for the RoSA for the period 2013-2025 is consistently lower than that projected by the ASSA model. This is consistent with the fact that young adults are expected to migrate to Gauteng and the Western Cape over the period, though the rates of migration are expected to decline roughly linearly over time. For Gauteng, the multi-regional model yields a relatively smaller population size over the period 1998-2015, as compared to that obtained from the ASSA model. After 2015, the comparison is reversed. For the Western Cape, the projected population size obtained from the multi-regional model diverges from that obtained from the ASSA model for the period 2005-2025 and is consistently higher than that obtained from the ASSA model.

In conclusion, virtually all changes in the population estimates, as noted above, are attributable to the transparent multi-regional projection of migration rates to a significant extent, while the use of the multi-regional modelling approach, on its own, explains a small proportion of the differences.

5 DISCUSSION AND CONCLUSION

This chapter discusses five points that were considered in this study. It sums up the objective of the research and the need thereof. The chapter goes on to summarise the research questions that were raised in Chapter 1, with particular attention given to the adaptation of the ASSA model and the associated research objectives. The chapter then considers whether the research objectives were met by considering the population size for each region, net migration rates per 1,000 and the population age structure. It concludes by stating the limitations of the research and identifying possible areas for future research.

The aim of the research was to determine the impact of modelling migration using the multi-regional modelling approach on the projected population age structure, size and net numbers of migrants implied by a multi-regional model, as opposed to using the net migration projections incorporated in the ASSA model. In investigating this question, the research set out to adapt the ASSA model in order to carry out multi-regional population projections, allowing for the estimation and projection of rates of origin-destination migration.

The main research question contains four sub-problems. First, given a three-region system, can the ASSA model be adapted into a multi-regional population projection tool? Second, what historical migration rates were experienced between 1997 and 2007 inclusive, and once inserted into the model in the form of migration parameters, what effect do these have on the projected population numbers between 1997 and 2007? Third, what projected rates of age-specific origin-destination migration does one obtain from the multi-regional model going forward, and are these plausible? Finally, what demographic impact does the application of the migration rates to the 'at-risk' populations in the multi-regional model have on the projected population estimates for each region?

While the numbers of migrants changed between the two intercensal/survey periods, the change in the age patterns of migration was not substantial. The assumption underlying the projected rates is that the migration rates will trend towards zero over a period of thirty years. This simplifying assumption ensures that the projected rates are consistent with past rates and thus within acceptable bounds.

The age patterns of migration are expected to be stable over a long period after 2007 given what has been observed in the past in South Africa (Dorrington and Moultrie 2009). This means that while the net migration may decline over time, and trend to zero over a thirty-year period, the age profiles of migration will remain stable over this period. One expects to see higher migration rates among the young adult population migrating (possibly with their children) from the RoSA to Gauteng and the Western Cape each year up to 2025, as well as return migrants from Gauteng and the Western Cape to the RoSA at older ages.

The treatment of migration in the multi-regional adaptation of the ASSA model yields results that diverge from those produced by the original ASSA model that uses the net migration approach for each region. The projected net migration rates per 1,000 obtained from the multi-regional model are greater than those obtained from the net migration model for all regions after 2007, except for the peak and trough for Gauteng and the Western Cape in 2006. This affects the projected population age structure and the population size for all regions, particularly after 2007.

The level of the migration rates assumed in the multi-regional adaptation of the ASSA model has a greater impact on the projected population numbers than does the use of the multi-regional modelling itself. For all three regions, an average of 89 per cent of the difference in the estimate of the population size – estimate obtained from the multi-regional model as opposed to one obtained from the ASSA net migration model – is attributable to the trajectory of the migration rates assumed, whereas the use of the multi-regional modelling accounts for an average of 11 per cent. For instance, consider the projected population for Gauteng as at 2025. The projected population, as obtained from the multi-regional model, is 12.07 million. The net migration model projects the population at 11.08 million for the region. The multi-regional model estimate is therefore 9 per cent higher than the net migration model estimate for that year. Hence, 96 per cent of this difference in the two estimates is attributable to the migration rates assumed in the multi-regional projections, whereas 4 per cent is attributable to the multi-regional modelling approach.

We now turn our attention to the practicality of creating a larger model, namely a nine-province model. Having created a three-region multi-regional adaptation of the ASSA model, one notes that the model is 36 per cent[5] larger than the *full* version of the original ASSA model because the former incorporates the multi-regional component in its projections. Further, while the *full* version of the original ASSA model contains a total of 147 (29 x 4 races + 31 at national level) worksheets (including charts), the multi-regional version contains 160 (31 x 3 regions + 1 'empty' projection + 36 at national level) of these.

A nine-province model, on the other hand, would take longer to deliver results with each cycle due to three reasons. First, the model would incorporate a total of 315 (31 x 9

[5] The size of the *full* model is 22.9 megabytes while that of the 3-region model is 31.2 megabytes. The percentage difference is thus [(31.2-22.9)/22.9] = 58%. The file sizes were obtained from the EXCEL properties page

provinces + 36 at national level) worksheets, including charts. This then translates into a file size 1.6 times (49.4/31.2) as large as that of the three-region model and twice (49.4/22.9) as large as that of the *full* (net migration) model. This, in turn, implies that the run-time of the model increases by 60 per cent[6] relative to the three-region model. This is expected since provincial projections require that in-migrants from, – and out-migrants to – all other eight provinces be incorporated in the calculation of demographic and epidemiological variables.

Additional worksheets necessary for a nine-province model imply an increase in the number of variables on which the model runs. The three-region adaptation of the ASSA model (including the 'empty' projection) incorporates a total of 682 variables edited through the *Name Manager* tool on the Excel workbook. With the addition of more worksheets to cater for all nine provinces individually, an additional 1,364 variables would have to be incorporated, a task that further exacerbates the extension of the multi-regional adaptation of the ASSA model to a nine-province system. If one were to look at adapting the ASSA model to a 9-province model using Excel, ignoring alternative software development platforms for the moment, the value of the 9-province model is likely to be greater because in such a model, each province is treated separately and thus retains its unique demographic and epidemiological profile. This treatment of the provinces lends itself to more accurate projected mid-year population estimates.

Multi-regional population projections are relevant for different end-users since they introduce greater accuracy in the estimates.

A few users that could use multi-regional population models include, but may not be limited to, the national and provincial departments of health, social security, the National Planning Commission, etc. These users would especially be interested in the population age distributions derived from such models. For example, a point was made earlier regarding the Planning Commission's interest in generating scenarios with respect to internal and international migration, where these scenarios give rise to projected demographic trends a few years into the future.

The limits on computing power are imposed on this research because the ASSA model is based entirely on Excel. It is true that these limits are probably minor when one considers computing alternatives, one of which is C++, a software development language that would make the incorporation of computation-intensive calculations into multi-regional demographic models easier, albeit at the cost of transparency. Excel, as a software package, is probably not a desirable development platform on which multi-

[6]The 3-region model runs for approximately 5 minutes over 40 cycles with 31.2 megabytes, thus a nine-region model is expected to run for approximately 8 minutes over the 40 cycles. Hence the change in run time = [(8-5)/5] = 60%

regional models such as this can be created. Although Excel provides transparency in this regard, it does so at the cost of ease of model development.

The research had a number of limitations. First, the research assumed that internal migration is independent of primary components of demographic change which are mortality, fertility and in the ASSA model, the HIV epidemic. This assumption is not necessarily onerous since the impact of mortality on a population age distribution is evident only after a sufficiently long period. However, one should also note that mortality may affect migration decisions made by individuals, for example, some of the migrants leave Gauteng and move back to home provinces to die there.

The second limitation of the research is related to the census migration data. Both the multi-regional adaptation of the ASSA model and the original ASSA model suffer this limitation. The 1996 and 2001 censuses undercounted the national population by approximately 10.7 per cent and 17.6 per cent, respectively (Masiteng and Kekovole 2011). In addition to this, Dorrington and Moultrie (2009) identified two problems associated with the 2001 census and 2007 community survey migration data. First, upon estimating annual migration rates, the authors noted that there were period reference errors in the data, with a preference for reporting most recent moves as having occurred twelve months prior to the national enquiries. Second, in 2001 the scanner had problems differentiating between province numbers such as '1s' and '7s', particularly in the 2001 census (Dorrington and Moultrie 2009). This affected the migration estimates of children born during the period within which migration is being measured. The 2007 Community Survey was not without errors either. Census migration data are therefore incomplete and unreliable.

Based on these shortcomings in the census/survey migration data, an assumption had to be made regarding annual migration. In order to complete the research, a simplifying assumption was introduced that the level of migration during each of the two intercensal/survey periods was level. This way, an attempt was made at addressing the period reference errors in the migration data.

Two experimental changes were made to the model to investigate the sensitivity of the model results to changes in some of the underlying assumptions. First, the assumption mentioned in the preceding paragraph was altered to allow for migration rates to decline roughly linearly each year during the two intercensal periods, and continue the trend after 2007. Note that, of the three regions in the multi-regional model, results for Gauteng and 'rest of South Africa' for 1996-2001 were the most sensitive to this change since the difference in population estimates between the two migration regimes ranged from 0 per cent to -4.8 per cent for Gauteng, and from 0 per cent to 1.3 per cent for 'rest of South Africa'. Results for the Western Cape were not as sensitive to the change in the migration assumption as those of the other two regions. For 2001-2007, changes in population estimates for Gauteng ranged from -6.3 per cent to -3.4 per cent, and the magnitude was lower for the Western Cape (-0.2 per cent to 0.4 per cent) and 'rest of South Africa' (1.8 per cent to 1.1 per cent). By 2025, the differences between the two migration regimes are negligible – Gauteng (2.4 per cent), Western Cape (-0.5 per cent) and 'rest of South Africa' (-0.6 per cent).

Reverting back to the initial assumptions on migration, we now turn our attention to assumptions associated with the HIV epidemic. The original ASSA model assumes that migrants arrive with the proportions infected by duration of infection that are equal to those of the destination population. If one relaxes this assumption, that is, if the allowance is made in the multi-regional model for migrants to move in proportions that represents the epidemic in the provinces from which they arise, we note that the mid-year population estimates for Gauteng decline by an average of 1.5 per cent between 1997 and 2025. The effects for the Western Cape (0.1 per cent) and 'rest of South Africa' (0.2 per cent) are much lower by comparison. This implies that the HIV prevalence and incidence rates in Gauteng and the Western Cape increase as a result of migrants moving into these provinces, while the opposite is true of the RoSA.

The two above experiments demonstrate the importance of ensuring that assumptions that underlie the models one builds must be plausible as model-based results rely heavily on these.

A third limitation of the research is that the ASSA model was not designed to represent the classic multi-regional population projection model, but was designed to carry out projections without disaggregating the national population into regions that interact in the form of directional migration flows. Thus the major difference between the multi-regional adaptation of the ASSA model and the classic multi-regional model is that while the latter expresses components of demographic change in the form of matrices, particularly the growth matrix, the former follows the logic of the original ASSA model very closely.

Ultimately, the goal is eventually to implement the multi-regional model for South Africa, applying the classic matrix equation after adapting it to work for a single-year, single-age 9-province system. However, this was not the goal for this research. Instead, the goal was to make an experimental adaptation of the existing, AIDS software model, namely the ASSA model.

The final limitation of the research is the assumption about international migration. Just as the two models suffer the deficiency of the census migration data, the two models also suffer this limitation. The research used net immigration numbers obtained from the original ASSA model as international migration. Regarding immigration, the research has two major problems. The first is that undocumented migration, a sizeable proportion of which may be illegal, is difficult to estimate. This results from the fact that census data on immigration are either unavailable or defective. Second, international migration in the ASSA model is net and this entails a problem with estimating emigration, and this exercise is not entirely impossible. It is merely complicated by defective and incomplete data.

The choice of regions is also a limitation because the provinces in their current form did not exist prior to 1991. For purposes of this research, Gauteng and the Western Cape were isolated as regions because these hold the highest population proportions that are urban (96 per cent and 90 per cent, respectively) (Kok and Collinson 2006) whereas all other provinces were combined into a region that is predominantly rural, with the exception of KwaZulu-Natal. Also, the choice of three regions, while allowing for the least work in producing regions of all possibilities, ignores the fact that each province has demographic, epidemiological and behavioural features unique to itself.

A number of areas for future research have been identified. First, while a three-region model provided the least amount of work to test the multi-regional adaptation of the ASSA model, the extension of the model to a nine-province model is well worth the effort due to reasons already stated earlier in this chapter. Such a model would capture the demographic features unique to each province as much as creating one for each population group would afford us the same benefit.

Second, the analysis of the 2011 Census data had not been completed at the time of preparing this research dissertation. With the availability of the latest census migration data, the multi-regional adaptation of the ASSA model can be updated with additional parameters that underlie the most recent trends and levels in internal migration for more accurate projections of national and regional demographic indices going forward, with particular focus placed on the regional population size, net migration rates per 1,000 and population age structures of the regions. The benefit associated with additional migration data is that the demographic projections for the period 2007-2011 can be improved. This implies that assumptions about rates of migration after 2011 can be updated to reflect recent trends and levels in migration. This is also true of the data on reported deaths.

A third area for further research is that a multi-regional adaptation of the ASSA model lends itself to extending the model to incorporate the interaction of migration and the HIV epidemic. One could allow for migrants to move with the HIV profile other than that of the province they are entering. A particular example of where this might be useful is the case of Gauteng, where the number of deaths registered as occurring in Gauteng is markedly less than expected because, it is speculated, HIV-infected people return to their province of origin to die there.

We have seen how multi-regional projection models relate, perhaps at a basic level, to the wider field of multi-state demography. These models have been shown to rely on multi-regional life tables, a variation of the multi-state increment-decrement life tables. We have also seen how multi-dimensional population analysis relates to multi-regional demography, and have established that multi-dimensional population analysis considers multiple dimensions such as age, sex and region in carrying out demographic analysis. We also know now that multi-dimensional population analysis requires adequate data in order to carry out demographic projections, and where such data are inadequate or defective, expected details of such are filled in along reasonable assumptions about those. The multi-regional adaptation of the ASSA model demonstrated these features successfully.

The need to seek a population projection methodology that treats internal migration transparently emerges as advisable, and the multi-regional projection methodology already demonstrated by this dissertation addresses this need. However, one should note two key issues that were identified by the research. First, the multi-regional modelling methodology alone has little impact on the provincial projections over the time horizon considered. The assumed level of migration rates, when applied to a population 'at risk' of migrating, together with the multi-regional projection methodology, have a greater impact on projected demographic results arising from the projection exercise. Second, the extension of the three-region model to a nine-province model is indeed possible, but would require quite a bit of work in its construction. Third, given that both methods – multi-regional modelling and net migration modelling – give rise to projections that are based on questionable (although necessary) assumptions, perhaps the added accuracy is in some respects spurious. For the accuracy to be trusted and genuine, the assumptions upon which these methods are based should be revisited and perhaps improved so that they may be plausible. The nine-province model has been shown to be more complex in its requirements of additional variables, would take longer to run and its size would be significantly larger due to the additional worksheets required for it.

ASSA2008. 2010. *Actuarial Society of South Africa AIDS and Demographic model: Beta Version*. Cape Town: Actuarial Society of South Africa. www.actuarialsociety.org.za.

Dorrington, R., L. Johnson and D. Budlender. 2010. *ASSA2008 AIDS and Demographic Models User Guide (beta Version)*. Cape Town: Actuarial Society of South Africa.

Dorrington, R. and T. A. Moultrie. 2009. *Making use of the consistency of patterns to estimate age-specific rates of inter-provincial migration in South Africa*. Cape Town: Centre for Actuarial Research.

Doyle, P. R. and D. B. Miller (1990) In *Transactions of the Actuarial Society of South Africa*.

Gullickson, A. 2001. *"Multiregional probabilistic forecasting"*. California: University of California, Berkeley. www.demo.berkeley.edu/~aarong/PAPERS/gullick_iiasa_stochmig.pdf

IPUMS-International (ed). 2011. *Integrated Public Use Microdata Series, International: Version 6.1 [Machine-readable database]*. Minneapolis: University of Minnesota:

Johnson, L. and R. E. Dorrington. 2006. "Modelling the demographic impact of HIV/AIDS in South Africa and the likely impact of interventions", *Demographic Research* **14**(22):541-574.

Kok, P. and M. Collinson. 2006. *Migration and Urbanisation in South Africa*. Report 04-03-02. Pretoria, South Africa: Statistics South Africa.

Land, K. C. and A. Rogers. 1982. "Multidimensional mathematical demography: An overview," in *Multidimensional Mathematical Demography*. University of Maryland, College Park, Maryland: Academic Press, pp. pp. 1-35.

Masiteng, K. and J. Kekovole. 2011. *Planning of Census 2011: Approach and Implementation Strategies*. Statistics South Africa.

Naidoo, N., M. Leibbrandt and R. Dorrington. 2008. "Magnitudes, personal characteristics and activities of Eastern Cape migrants: A comparison with other migrants and with non-migrants using data from the 1996 and 2001 censuses", *South African Journal of Demography* **11**(1):3-38.

Nyirenda, M. 2007. *Small area projections: Modelling the demographic and epidemiological dynamics of a rural area in South Africa*. Cape Town: University of Cape Town.

Polzer, T. 2010. *Population Movements in and to South Africa*. Johannesburg: Forced Migration Studies Programme.

Raymer, J., G. J. Abel and A. Rogers. 2012. "Does specification matter? Experiments with simple multiregional probabilistic population projections", *Environment and Planning A* **44**(11):1-34.

Rogers, A. 1980. "Essays in multistate mathematical demography: special IIASA issue", *Environ Plan A* **12**(5):485-622.

Rogers, A. 1985. "Regional population projection models", *Scientific Geography Series* **4**:63-88.

Rogers, A. 1986. "Parameterized multistate population dynamics and projections", *Journal of the American Statistical Association* **81**(393):48-61.

Rogers, A. 1990. "Requiem for the net migrant", *Geographical Analysis* **22**(4):282-300.

Rogers, A. 1995a. "Multiregional demography : principles, methods and extensions," in Chichester New York: J. Wiley, pp. pp. xi, 236.

Rogers, A. 1995b. "The multiregional life table," in *Multiregional Demography: Principles, Methods and Extensions* John Wiley & Sons, Inc., pp. pp. 81-110.

Rogers, A. 1995c. "Multiregional perspective and multiregional models," in *Multiregional Demography: Principles, Methods and Extensions*. John Wiley & Sons, Inc., pp. pp. 1-8.

Rogers, A. and D. Philipov. 1979. *Multiregional Methods for Subnational Population Projections*. Laxenburg: International Institute for Applied Systems Analysis.

Rogers, A. and J. Raymer. 2006. *Applying model migration schedules to represent age-specific migration flows*. Colorado: Research Program on Population Processes.
Rogers, A. and F. Willekens. 1986a. "Conclusion," in *Migration and Settlement: A Multiregional Comparative Study*. D Reidel Publishing Company, pp. pp. 457-473.
Rogers, A. and F. Willekens. 1986b. "A short course on multiregional mathematical demography," in *Migration and Settlement: A Multiregional Comparative Study*. D Reidel Publishing Company, pp. pp. 355-385.
Rogers, A. and J. Woodward. 1991. "Assessing state population projections with transparent multiregional demographic models", *Population Research and Policy Review* **10**(1):1-26.
StataCorp. 2011. *Stata Statistics/Data Analysis*. College Station, Texas: http://www.stata.com.
StatsSA. 1998. *Mid-Year Population Estimates*. P0302. Pretoria,: Statistics South Africa.
The MathWorks, I. 2011. *MATLAB*. Massachussetts:

Appendix 1 Code for generating the 3-way multiplicative log-linear-based migration tables for 1996-2001 and 2001-2007 in STATA12 (StataCorp 2011)

2001 Census migration do-file

```
*CREATING 3 REGIONS OUT OF THE 9 PROVINCES

use "C:\StataMigration\2001census.dta", clear

*CURRENT REGION OF RESIDENCE
gen regza=1 if provza==7
replace regza=1 if provza==7
replace regza=2 if provza==1
replace regza=3 if provza==5
replace regza=3 if provza==2
replace regza=3 if provza==3
replace regza=3 if provza==4
replace regza=3 if provza==6
replace regza=3 if provza==8
replace regza=3 if provza==9
tab regza
tab regza [iw=wtper]

*REGION OF RESIDENCE 5 YEARS AGO
gen reg5yr=1 if migza2==7
replace reg5yr=1 if migza2==7
replace reg5yr=2 if migza2==1
replace reg5yr=3 if migza2==5
replace reg5yr=3 if migza2==2
replace reg5yr=3 if migza2==3
replace reg5yr=3 if migza2==4
replace reg5yr=3 if migza2==6
replace reg5yr=3 if migza2==8
replace reg5yr=3 if migza2==9
tab reg5yr
tab reg5yr [iw=wtper]

*GENERATING INTER-REGIONAL MIGRANTS
gen migrant=1 if reg5yr!=regza
replace migrant=2 if mgctry!=.
sort sex

*RENAMING VARIABLES
rename regza region
rename reg5yr region5yrs

*TABULATING MULTIREGIONAL MIGRATION FLOWS
tab region5yrs region [iw=wtper]
```

```
*RELABELLING REGIONS
label define region 1 "Gauteng" 2 "Western Cape" 3 "Rest of SA"
label val region region
label define region5yrs 1 "Gauteng" 2 "Western Cape" 3 "Rest of SA"
label val region5yrs region5yrs

*EXTENDING THE NUMBER OF AGE GROUPS TO 90+
gen age3=age
recode age3 min/4=1 5/9=2 10/14=3 15/19=4 20/24=5 25/29=6 30/34=7 35/39=8
40/44=9 45/49=10 50/54=11 55/59=12 60/64=13 65/69=14 70/74=15 75/79=16
80/84=17 85/89=18 90/max=19

label define age3 1 "0-4" 2 "5-9" 3 "10-14" 4 "15-19" 5 "20-24" 6 "25-29" 7 "30-34" 8
"35-39" 9 "40-44" 10 "45-49" 11 "50-54" 12 "55-59" 13 "60-64" 14 "65-69" 15 "70-74" 16
"75-79" 17 "80-84" 18 "85-89" 19 "90+"
label val age3 age3

br age age3

tab age3

*TABULATING REGIONAL OUTMIGRANTS

*AGE-SPECIFIC NUMBERS OF MALE OUTMIGRANTS (in single ages)
tab age region5yrs [iw=wtper] if region==1 & sex==1
tab age region5yrs [iw=wtper] if region==2 & sex==1
tab age region5yrs [iw=wtper] if region==3 & sex==1

*AGE-SPECIFIC NUMBERS OF FEMALE OUTMIGRANTS (in single ages)
tab age region5yrs [iw=wtper] if region==1 & sex==2
tab age region5yrs [iw=wtper] if region==2 & sex==2
tab age region5yrs [iw=wtper] if region==3 & sex==2

*ESTIMATING MIGRANT NEW-BORNS
gen regborn=1 if bplza==7
replace regborn=1 if bplza==7
replace regborn=2 if bplza==1
replace regborn=3 if bplza==5
replace regborn=3 if bplza==2
replace regborn=3 if bplza==3
replace regborn=3 if bplza==4
replace regborn=3 if bplza==6
replace regborn=3 if bplza==8
replace regborn=3 if bplza==9
label define regborn 1 "Gauteng" 2 "Western Cape" 3 " Rest of SA"
label val regborn regborn
```

*AGE-SPECIFIC NUMBERS OF MALE OUTMIGRANTS (in single ages)
tab regborn region [iw=wtper] if sex==1

*AGE-SPECIFIC NUMBERS OF FEMALE OUTMIGRANTS (in single ages)
tab regborn region [iw=wtper] if sex==2

*CREATING MIGRATION FLOW TABLES
tab region5yrs region [iw=wtper]
tab region5yrs region [iw=wtper] if sex==1
tab region5yrs region [iw=wtper] if sex==2

*TABULATING AGE-SPECIFIC CENSUS SUB-POPULATIONS
tab age region [iw=wtper] if sex==1
tab age region [iw=wtper] if sex==2

*CREATING MIGRATION FLOW TABLES

tab region5yrs region [iw=wtper]
tab region5yrs region [iw=wtper] if sex==1
tab region5yrs region [iw=wtper] if sex==2

*TABULATING AGE-SPECIFIC CENSUS SUB-POPULATIONS

tab age region [iw=wtper] if sex==1
tab age region [iw=wtper] if sex==2

2007 Community Survey migration do-file

*CREATING 3 REGIONS OUT OF THE 9 PROVINCES

use "C:\StataMigration\2007communitysurvey.dta", clear

*CURRENT REGION OF RESIDENCE
gen regza=1 if provza==7
replace regza=1 if provza==7
replace regza=2 if provza==1
replace regza=3 if provza==2
replace regza=3 if provza==3
replace regza=3 if provza==4
replace regza=3 if provza==5
replace regza=3 if provza==6
replace regza=3 if provza==8
replace regza=3 if provza==9

*REGION OF PREVIOUS RESIDENCE
gen regprev=1 if migza2==7
replace regprev=1 if migza1==7
replace regprev=2 if migza1==1
replace regprev=3 if migza1==2
replace regprev=3 if migza1==3

```
replace regprev=3 if migza1==4
replace regprev=3 if migza1==5
replace regprev=3 if migza1==6
replace regprev=3 if migza1==8
replace regprev=3 if migza1==9

*GENERATING MIGRANTS/OUT-MIGRANTS PER REGION
gen migrant=1 if regprev!=regza
replace migrant=2 if mgctry!=.
sort sex

*RENAMING VARIABLES
rename regza region
rename regprev regionprev

*TABULATING MULTIREGIONAL MIGRATION FLOWS
tab regionprev region

*RELABELLING REGIONS
label define region 1 "Gauteng" 2 "Western Cape" 3 "Rest of SA"
label val region region
label define regionprev 1 "Gauteng" 2 "Western Cape" 3 "Rest of SA"
label val regionprev regionprev

*EXTENDING THE NUMBER OF AGE GROUPS TO 90+
gen age3=age
recode age3 min/4=1 5/9=2 10/14=3 15/19=4 20/24=5 25/29=6 30/34=7 35/39=8
40/44=9 45/49=10 50/54=11 55/59=12 60/64=13 65/69=14 70/74=15 75/79=16
80/84=17 85/89=18 90/max=19

label define age3 1 "0-4" 2 "5-9" 3 "10-14" 4 "15-19" 5 "20-24" 6 "25-29" 7 "30-34" 8
"35-39" 9 "40-44" 10 "45-49" 11 "50-54" 12 "55-59" 13 "60-64" 14 "65-69" 15 "70-74" 16
"75-79" 17 "80-84" 18 "85-89" 19 "90+"
label val age3 age3
br age age3
tab age3

*TABULATING REGIONAL OUT-MIGRANTS

*AGE-SPECIFIC NUMBERS OF MALE OUT-MIGRANTS
tab age3 regionprev if region==1 & sex==1
tab age3 regionprev if region==2 & sex==1
tab age3 regionprev if region==3 & sex==1

*AGE-SPECIFIC NUMBERS OF FEMALE OUT-MIGRANTS
tab age3 regionprev if region==1 & sex==2
tab age3 regionprev if region==2 & sex==2
tab age3 regionprev if region==3 & sex==2

*AGE-SPECIFIC NUMBERS OF MALE OUT-MIGRANTS (in single ages)
tab age regionprev [iw=wtper] if region==1 & sex==1
tab age regionprev [iw=wtper] if region==2 & sex==1
```

```
tab age regionprev [iw=wtper] if region==3 & sex==1

*AGE-SPECIFIC NUMBERS OF FEMALE OUT-MIGRANTS (in single ages)
tab age regionprev [iw=wtper] if region==1 & sex==2
tab age regionprev [iw=wtper] if region==2 & sex==2
tab age regionprev [iw=wtper] if region==3 & sex==2

*ESTIMATING MIGRANT NEW-BORNS
gen regborn=1 if bplza==7
replace regborn=1 if bplza==7
replace regborn=2 if bplza==1
replace regborn=3 if bplza==2
replace regborn=3 if bplza==3
replace regborn=3 if bplza==4
replace regborn=3 if bplza==5
replace regborn=3 if bplza==6
replace regborn=3 if bplza==8
replace regborn=3 if bplza==9
label define regborn 1 "Gauteng" 2 "Western Cape" 3 "Rest of SA"
label val regborn regborn

*CREATING MIGRATION FLOW TABLES
tab regionprev region [iw=wtper]
tab regionprev region [iw=wtper] if sex==1
tab regionprev region [iw=wtper] if sex==2

*TABULATING AGE-SPECIFIC CENSUS SUB-POPULATIONS
tab age region [iw=wtper] if sex==1
tab age region [iw=wtper] if sex==2
```

```
        ActiveCell.Offset(0, [ProjOffset] + 1).Range("A1").Select
    End If
    Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, _
        SkipBlanks:=False, Transpose:=False
    Application.Goto Reference:="RESULTSET_MALEWC"
    Calculate
    Application.CutCopyMode = False
    Selection.Copy
    If [CurrYear] < 2025 Then
        ActiveCell.Offset(0, [ProjOffset] + 1 - 12).Range("A1").Select
    Else
        ActiveCell.Offset(0, [ProjOffset] + 1).Range("A1").Select
    End If
    Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, _
       SkipBlanks:=False, Transpose:=False
    Application.Goto Reference:="RESULTSET_MALERest"
    Calculate
    Application.CutCopyMode = False
    Selection.Copy
    If [CurrYear] < 2025 Then
        ActiveCell.Offset(0, [ProjOffset] + 1 - 12).Range("A1").Select
    Else
        ActiveCell.Offset(0, [ProjOffset] + 1).Range("A1").Select
    End If
    Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, _
        SkipBlanks:=False, Transpose:=False
End Sub

'  StepResult Female Out-migration() Macro
'  Steps through the female migration results array and updates it one more year
Private Sub StepResult_FemOut()
    Application.Goto Reference:="RESULTSET_FEMGP"
    Calculate
    Application.CutCopyMode = False
    Selection.Copy
    If [CurrYear] < 2025 Then
        ActiveCell.Offset(0, [ProjOffset] + 1 - 12).Range("A1").Select
    Else
        ActiveCell.Offset(0, [ProjOffset] + 1).Range("A1").Select
    End If
Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, _
SkipBlanks:=False, Transpose:=False
    Application.Goto Reference:="RESULTSET_FEMWC"
    Calculate
        Application.CutCopyMode = False
    Selection.Copy
    If [CurrYear] < 2025 Then
        ActiveCell.Offset(0, [ProjOffset] + 1 - 12).Range("A1").Select
    Else
        ActiveCell.Offset(0, [ProjOffset] + 1).Range("A1").Select
    End If
```

```
   Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, _

      SkipBlanks:=False, Transpose:=False
   Application.Goto Reference:="RESULTSET_FEMRest"
   Calculate
      Application.CutCopyMode = False
Selection.Copy
   If [CurrYear] < 2025 Then
      ActiveCell.Offset(0, [ProjOffset] + 1 - 12).Range("A1").Select
   Else
      ActiveCell.Offset(0, [ProjOffset] + 1).Range("A1").Select
   End If
   Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, _
      SkipBlanks:=False, Transpose:=False
End Sub

'  StartResult Migration
'  Resets the Results MaleOut Pages
Private Sub StartResult_MaleOutRegion(Year As Integer, aRegion As String)
  Dim aSheetName As String
  If Len(aRegion) > 0 Then
     aSheetName = "Results MaleOutGP" & aRegion
     aSheetName = "Results MaleOutWC" & aRegion
     aSheetName = "Results MaleOutRest" & aRegion
   Else
      aSheetName = "Results MaleOut"
   End If
   Sheets("Results MaleOutGP").Range("ProjSheet_GPm").ClearContents
   Sheets("Results MaleOutWC").Range("ProjSheet_WCm").ClearContents
   Sheets("Results MaleOutRest").Range("ProjSheet_Restm").ClearContents
   Sheets("Results MaleOutGP").Range("ProjOffset_GPm").Formula = 0
   Sheets("Results MaleOutWC").Range("ProjOffset_WCm").Formula = 0
   Sheets("Results MaleOutRest").Range("ProjOffset_Restm").Formula = 0
End Sub
Private Sub StartResult_FemOutRegion(Year As Integer, aRegion As String)
   Dim aSheetName As String
   If Len(aRegion) > 0 Then
      aSheetName = "Results FemOutGP" & aRegion
      aSheetName = "Results FemOutWC" & aRegion
      aSheetName = "Results FemOutRest" & aRegion
Else
      aSheetName = "Results FemOut"

End If
   Sheets("Results FemOutGP").Range("ProjSheet_GPf").ClearContents
   Sheets("Results FemOutWC").Range("ProjSheet_WCf").ClearContents
   Sheets("Results FemOutRest").Range("ProjSheet_Restf").ClearContents
   Sheets("Results FemOutGP").Range("ProjOffset_GPf").Formula = 0
   Sheets("Results FemOutWC").Range("ProjOffset_WCf").Formula = 0
   Sheets("Results FemOutRest").Range("ProjOffset_Restf").Formula = 0
 End Sub
Private Sub StartResult_MaleOut(Year As Integer)
   [ProjYear] = Year
```

```
   StartResult_MaleOutRegion Year, "GP"
   StartResult_MaleOutRegion Year, "WC"
   StartResult_MaleOutRegion Year, "Rest"
End Sub
Private Sub StartResult_FemOut(Year As Integer)
   [ProjYear] = Year
   StartResult_FemOutRegion Year, "GP"
   StartResult_FemOutRegion Year, "WC"
   StartResult_FemOutRegion Year, "Rest"
End Sub
```

Appendix 3 ***11-parameter* schedules fitted to male and female census out-migration data, Gauteng, 1996-2007**

	1996-2001		2001-2007		1996-2001		2001-2007	
	To Western Cape		To Western Cape		To Rest of SA		To Rest of SA	
	Male	Female	Male	Female	Male	Female	Male	Female
a1	0.0006	0.0017	0.0011	0.0000	0.0050	0.0032	0.0038	0.0028
α1	0.3453	0.8210	0.1412	0.8500	0.4778	0.8592	0.5539	0.1003
a2	0.0004	0.0005	0.0021	0.0027	0.0048	0.0038	0.0022	0.0077
μ2	23.78	17.91	23.01	21.75	27.39	30.78	21.81	16.00
α2	0.0000	0.0088	0.0371	0.6683	0.1042	0.3652	0.0386	0.0618
λ2	0.2635	0.2653	0.1000	1.8500	0.2819	0.2329	1.5000	0.1000
a3	0.0039	0.0020	0.0024	0.0044	0.0037	0.0064	0.0053	0.0042
μ3	61.03	54.71	75.13	64.32	71.02	63.03	73.00	75.00
α3	0.1316	0.0642	0.3775	0.2580	0.4536	0.2237	0.6740	0.1876
λ3	0.2117	0.3005	0.1246	0.1731	0.1478	0.2060	0.2184	0.0744
c	0.0014	0.0013	0.0005	0.0010	0.0059	0.0058	0.0045	0.0029

Appendix 4 ***7-paramter*** **and** ***11-parameter*** **schedules fitted to male and female census out-migration data, Western Cape, 1996-2007**

	1996-2001		2001-2007		1996-2001		2001-2007	
	To Gauteng		To Gauteng		To Rest of SA		To Rest of SA	
	Male	Female	Male	Female	Male	Female	Male	Female
a1	0.0020	0.0021	0.0016	0.0024	0.1390	0.0027	0.0032	0.0014
α1	0.3227	0.7638	0.0839	0.1150	2.2360	0.0677	0.0514	1.1250
a2	0.0048	0.0040	0.0064	0.0062	0.0028	0.0058	0.0067	0.0067
μ2	24.38	24.92	25.00	26.78	23.36	25.00	27.00	26.00
α2	0.0968	0.1394	0.0610	0.0712	0.0879	0.0469	0.0470	0.0325
λ2	0.3119	0.2107	0.1026	0.1092	0.3118	0.0885	0.0966	0.0333
a3					0.0022	0.0015	0.0018	
μ3					64.51	57.75	72.08	
α3					0.3024	0.0288	0.6441	
λ3					0.1760	0.6326	0.2087	
c	0.0009	0.0011	0.0000	0.0000	0.0028	0.0009	0.0001	0.0005

Appendix 5 *7-paramter* and *11-parameter* schedules fitted to male and female census out-migration data, rest of South Africa, 1996-2007

	1996-2001		2001-2007		1996-2001		2001-2007	
	To Gauteng		To Gauteng		To Western Cape		To Western Cape	
	Male	Female	Male	Female	Male	Female	Male	Female
a1	0.0069	0.0077	0.0038	0.0042	0.0015	0.0018	0.0012	0.0012
α1	0.7987	0.9730	0.6891	0.6787	1.5880	1.9470	0.8341	1.1250
a2	0.0186	0.0200	0.0121	0.0126	0.0048	0.0051	0.0048	0.0028
μ2	19.83	20.19	18.56	19.57	18.81	19.02	20.49	18.26
α2	0.0966	0.1068	0.0822	0.0893	0.1006	0.1096	0.1040	0.0790
λ2	0.3411	0.3270	0.3861	0.3060	0.3974	0.4063	0.3098	0.3958
a3							0.0020	
μ3							65.48	
α3							0.2744	
λ3							0.1555	
c	0.0015	0.0017	0.0014	0.0014	0.0008	0.0009	0.0005	0.0004

Appendix 6 Rogers-Castro Equations for migration curve-fitting

7-parameter model

$$m_x = a_1 e^{(-\alpha_1 x)} + a_2 e^{\{-\alpha_2 (x-\mu_2) - e^{[-\lambda_2 (x-\mu_2)]}\}} + a_0 \tag{1}$$

11-parameter model

$$m_x = a_1 e^{(-\alpha_1 x)} + a_2 e^{\{-\alpha_2 (x-\mu_2) - e^{[-\lambda_2 (x-\mu_2)]}\}} + a_3 e^{\{-\alpha_3 (x-\mu_3) - e^{[-\lambda_3 (x-\mu_3)]}\}} + a_0 \tag{2}$$

13-parameter model

$$m_x = a_1 e^{(-\alpha_1 x)} + a_2 e^{\{-\alpha_2 (x-\mu_2) - e^{[-\lambda_2 (x-\mu_2)]}\}} + a_3 e^{\{-\alpha_3 (x-\mu_3) - e^{[-\lambda_3 (x-\mu_3)]}\}} + a_4 e^{(\alpha_4 x)} + a_0$$

Description of the models

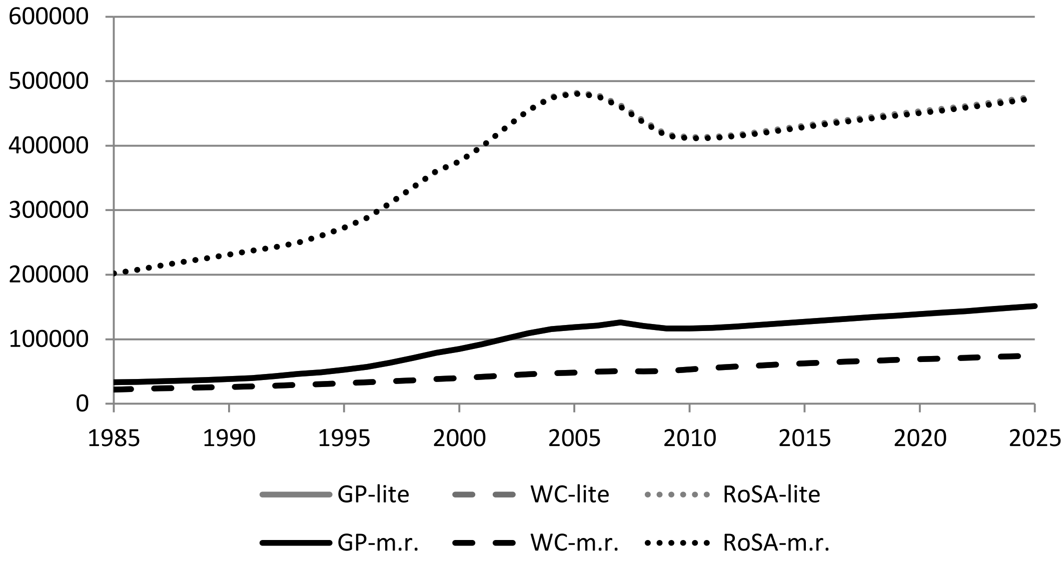
The model age migration schedules above contain, at most, thirteen parameters. Consider Equation 2, for example. The Equation contains eleven parameters. The *level* of migration is described by four parameters, namely a_0, a_1, a_2 and a_3.

The other seven parameters describe the migration *age profile*, and these are denoted α_1, α_2, α_3, μ_2, μ_3, λ_2 and λ_3.

Thus m_x denotes the rate of migration from region *i* to region *j* at exact age *x* between time *t* and *t*+h.

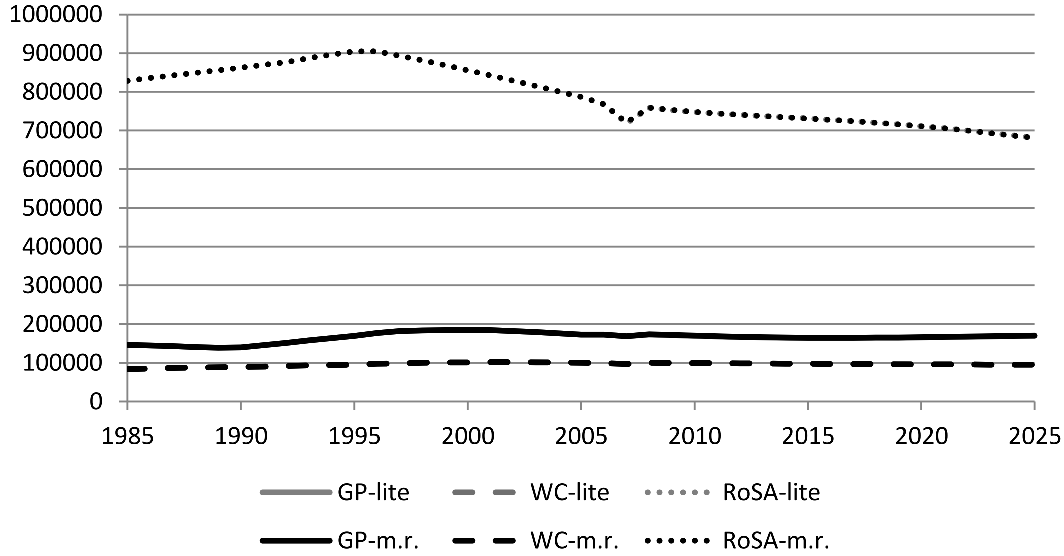
(i)

Comparison of the projected annual deaths (AIDS and non-AIDS) for Gauteng, Western Cape and rest of South Africa, 1985-2025



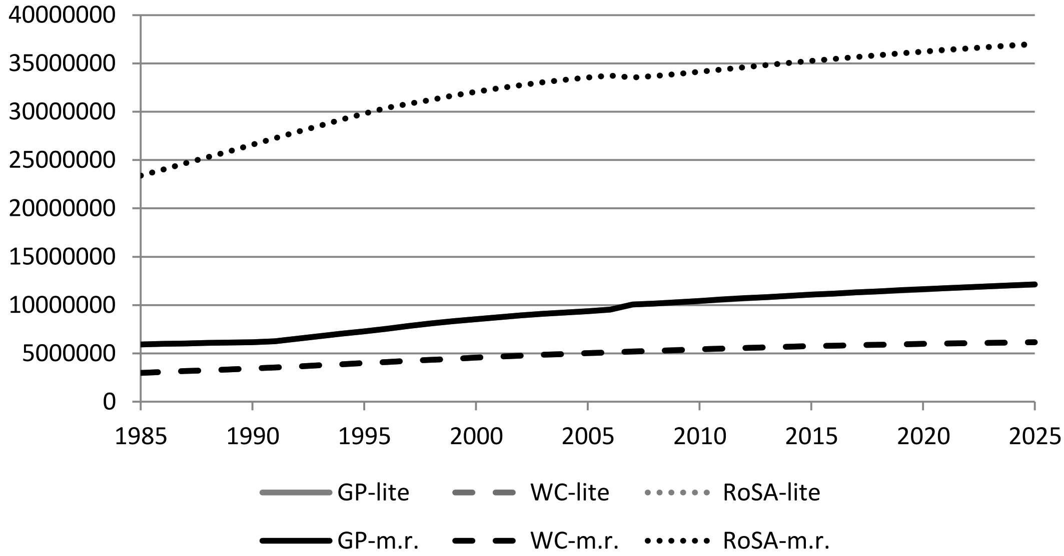
(ii)

Comparison of the projected annual births (HIV+ and HIV-) for Gauteng, Western Cape and rest of South Africa, 1985-2025



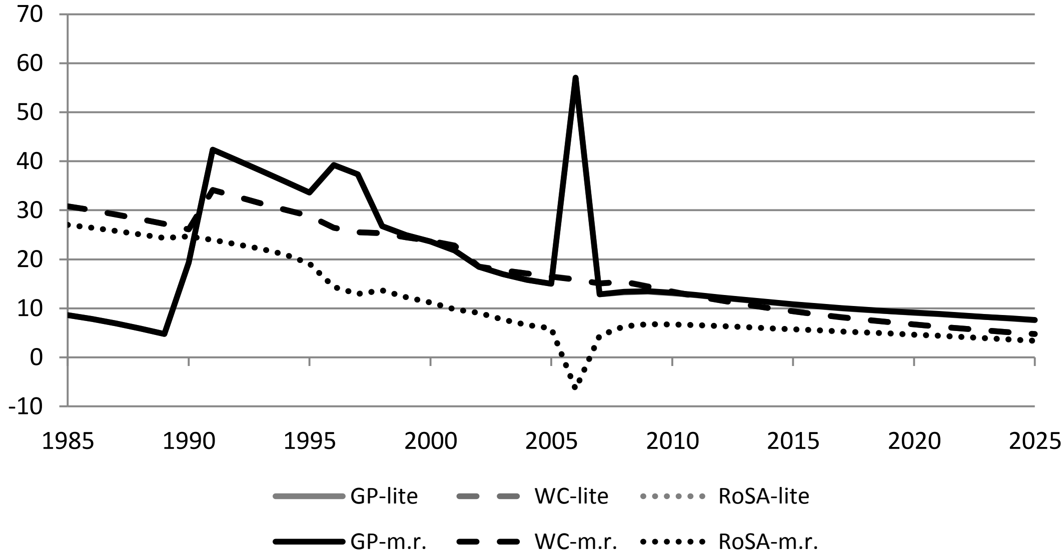
(iii)

Comparison of the projected total population for Gauteng, Western Cape and rest of South Africa, 1985-2025



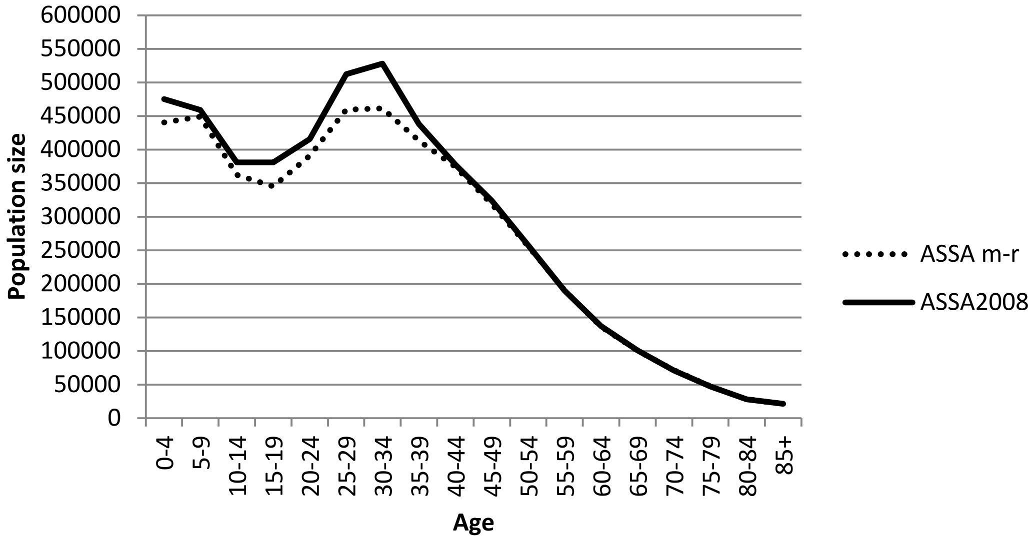
(iv)

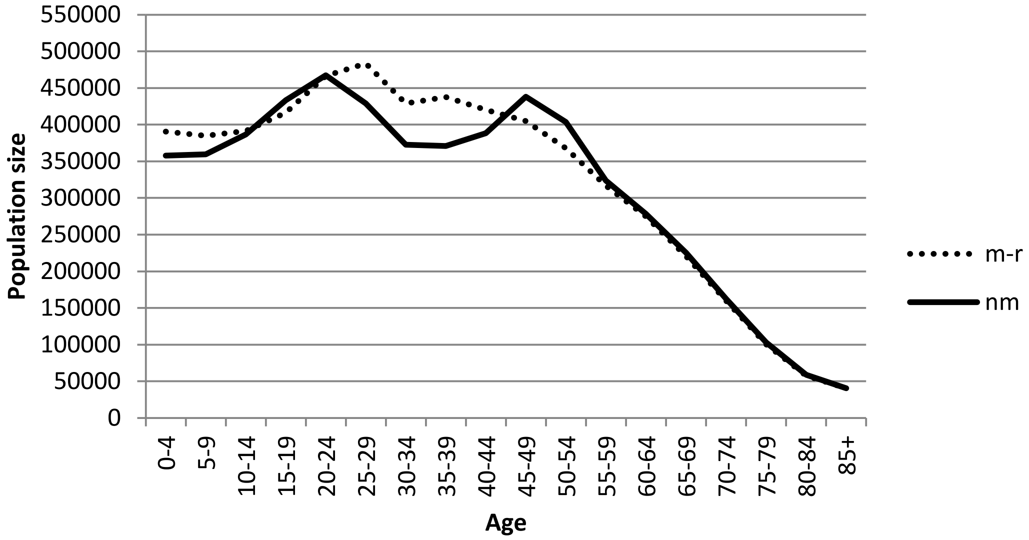
Comparison of the projected annual growth rate per 1,000 for Gauteng, Western Cape and rest of South Africa, 1985-2025



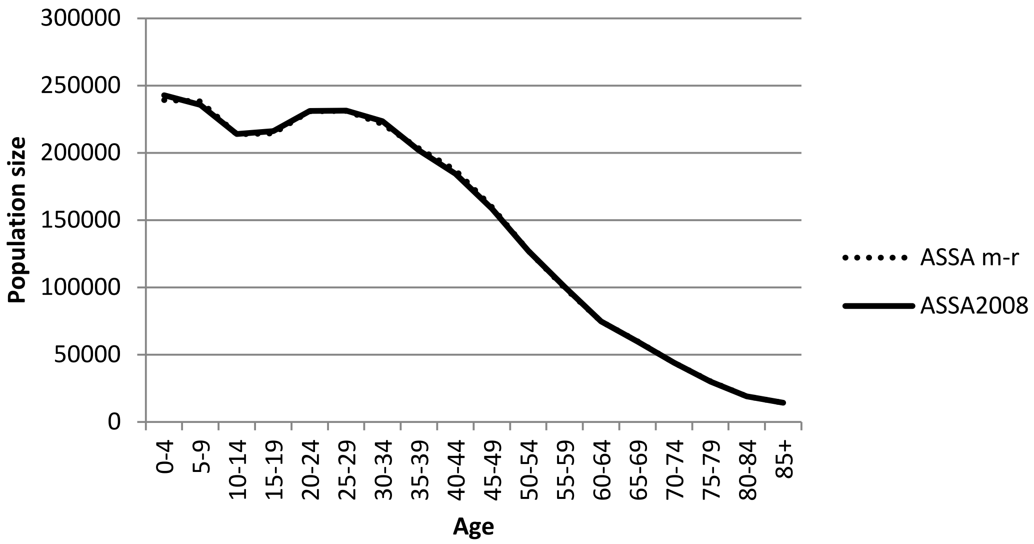
Appendix 8 Population age distributions for females as in 2007 and 2025, respectively

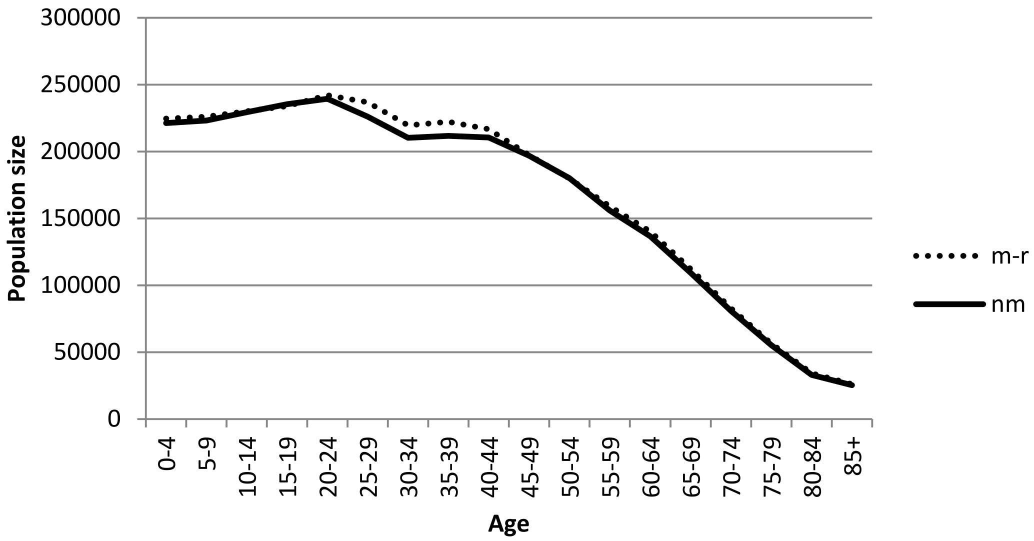
(i) Gauteng





(ii) Western Cape





(iii) Rest of South Africa