CENTRE of POLICY STUDIES and the IMPACT PROJECT

Eleventh Floor, Menzies Building
Monash University, Wellington Road
CLAYTON Vic 3800 AUSTRALIA

Telephone:
(03) 9905 2398, (03) 9905 5112

from overseas:
61 3 9905 2398 or
61 3 9905 5112

Fax:
(03) 9905 2426

61 3 9905 2426

e-mail:

impact@buseco.monash.edu.au

Internet home page:

http//www.monash.edu.au/policy/

Labour-Market Database for South Africa with HIV/AIDS Detail

by

LOUISE ROOS

Centre of Policy Studies
Monash University

General Paper No. G-235 May 2013

ISSN 1 031 9034

ISBN 978 1 921654 43 5

The Centre of Policy Studies (COPS) is a research centre at Monash University devoted to economy-wide modelling of economic policy issues.

Labour-market database for South Africa with HIV/AIDS detail

Louise Roos

Centre of Policy Studies, Monash University, Australia, May 2013

This paper describes the construction of a database that underlies the labour supply module developed for South Africa, with a specific focus on HIV/AIDS. The labour supply theory imposes a stock/flow dynamic mechanism on labour market groups distinguished by labour market activity, age, gender, race, and HIV status/stage. Broadly, the theory specifies that at the start of year t, people aged 15-65 (the working age population, hereafter the WAP) are divided into categories based on common characteristics. These characteristics are age, gender, race, HIV status/stage and labour-market activity undertaken in year t-1. People in categories offer their labour services to activities performed during year t. At the end of year t, people still part of the WAP progress one year in age and may change their HIV status/stage. Some people leave the WAP due to retirement or death. After this transition, people are again grouped into categories, based on common characteristics. The process of labour supply from a category to an activity is then repeated. For the implementation of this theory, we need to create a database that contains matrices that form the initial solution of the model. Three characteristics of this database are noted: (1) it contains detailed information regarding the structure of the WAP in the base year (2002); (2) it includes a transition matrix that allows adults to change their age and HIV stage between year t-1 and year t and (3) it includes matrices describing the flow of adults from categories to activities.

This paper is organised in three parts. The first part describes the construction of the activities matrix in the base year. The activities matrix describes the number of people in each labour-market activity by age, gender, race and HIV stage. The second part of this paper explains the construction of the categories matrix and the flow matrices. The categories matrix shows the number of people in each labour-market activity by age, gender, race and HIV stage at the start of the year. The flow matrices show the number of people by age, gender, race and HIV stage, moving from a labour-market category to an activity. The third part of the paper describes the construction of the transition matrix. This matrix allows people in each labour-market activity, given their gender and race, to change their age from aa to a and change their HIV stage from hh to h.

Keywords: Africa, HIV/AIDS
JEL codes: I190, O55

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AIDS	Acquired immunodeficiency syndrome
AIDS	Acquired immunodeficiency syndrome
ASSA	Actuarial Society of South Africa
BER	Bureau for Economic Research
HIV	Human immunodeficiency virus
ILO	International Labour Organisation
IMF	International Monetary Fund
KZN	KwaZulu-Natal
LFS	Labour Force Survey
PDL	Permanently departed from the labour force
RHS	Right hand side
SAGE-H	South African General Equilibrium model including the labour-market specification
SAM	Social Accounting Matrix
SAQB	South African Reserve Bank Quarterly Bulletin
SARB	South African Reserve Bank
StatsSA	Statistics South Africa
UN	United Nations
WAP	Working age population
WHO	World Health Organisation

LIST OF SETS

ACT $\in$	OCC, UNEMP and Permanently departed from the labour force (PDL)
AGE $\in$	A15-24, A25-34, A35-44, A45-54, A55-65.
CAT $\in$	OCC, UNEMP, N and Permanently departed from the labour force (PDL).
COM $\in$	Agricultural, Coal, Gold, Other mining, Food, Textiles, Petroleum, Other non-metallic mineral products, Basic iron/steel, Electrical machinery, Radio, Transport equipment, Other manufacturing, Electricity, Water, Construction, Trade, Hotels and restaurants, Transport services, Communications, Financial intermediation, Real estate, Other business activities, General government, Health and social work, Other activities, Owner Dwellings.
EUN $\in$	Employment (OCC), Unemployment (UNEMP), New entrant (N).
EUNP $\in$	Employment (OCC), Unemployment (UNEMP), New entrant (N) and Permanently departed from the labour force (PDL).
EUP $\in$	Employment (OCC), Unemployment (UNEMP), Permanently departed from the labour force (PDL).
EU $\in$	Employment (OCC), Unemployment (UNEMP).
GEN $\in$	Female, Male.
H2 $\in$	HIV negative, HIV positive.
H3 $\in$	HIV negative, Stage 1, Stage 2 and Stage 3.
IND $\in$	Agricultural, Coal, Gold, Other mining, Food, Textiles, Petroleum, Other non-metallic mineral products, Basic iron/steel, Electrical machinery, Radio, Transport equipment, Other manufacturing, Electricity, Water, Construction, Trade, Hotels and restaurants, Transport services, Communications, Financial intermediation, Real estate, Other business activities, General government, Health and social work, Other activities, Owner Dwellings.
MAR $\in$	Trade, Transport services.
N $\in$	New entrant.
OCC $\in$	Legislators, Professionals, Technicians, Clerks, Service workers, Skilled agricultural workers, Craft workers, Plant and machine operators, Elementary occupations, Domestic workers and Occupations not else where specified.
RACE $\in$	African, Other.
SRC $\in$	Domestic, Import.
STG $\in$	HIV negative, Stage 1, Stage 2, Stage 3 and Stage 4.
UNEMP $\in$	Short (S) and long-run (L) unemployment.

LIST OF FIGURES

LIST OF TABLES

1. THE DATABASE DESCRIBING THE LABOUR-MARKET SPECIFICATION

1.1. Coefficients and parameters

The labour-supply module requires separate matrices describing activities undertaken during the base year; categories at the start of the base year; offers and actual flows matrices describing the possible flows between categories and activities; and a transition matrix. These matrices are listed in Table 1. Each matrix is defined by a set of dimensions. These sets are listed in Table 1 under the heading *Sets*. Combinations of labour-market activities/categories yield different sets. For example, the set UN includes short-term (S), long-term (L) and new entrants (N) into the labour force. A complete list of the elements included in different sets is reported in Appendix 1.

The coefficients listed in Table 1 are explicitly specified in the initial database and updated after every simulation year. The following matrices are included in the database:

- The number of adults in *activity* a in year t. This matrix is defined as $ACT_{(o,a,g,r,h)}$ and refers to the number of adults in each labour-market activity o by age a, gender g, race r and HIV stage h.
- The *lagged activities* matrix for year $t-1$. This matrix is defined as $ACT\_L_{(o,a,g,r,h)}$ and is similar in dimensions to $ACT_{(o,a,g,r,h)}$, but defined for the preceding year.
- The number of adults in *category* c at the start of year t. This matrix is defined as $CAT_{(c,a,g,r,h)}$ and shows the number of adults in each labour-market category c by gender g and race r who are allocated to age a and HIV stage h at the start of year t. An additional category, new entrant (N) is exogenously added at the beginning of each year.
- *Planned offers* matrix for year t. This matrix is defined as $OFFER_{(c,a,g,r,h,o)}$ and shows the planned flows from category c to activity o, given a set of age a, gender g, race r, HIV stage h characteristics. This matrix shows the activities adults would like to perform during year t.
- *Actual flows* matrix for year t. This matrix is defined as $H_{(c,a,g,r,h,o)}$ and shows the actual flows from category c to activity o, given a set of age a, gender g, race r, HIV stage h characteristics. This matrix shows who secures employment and what happens to those who do not

secure employment

Table 1. Contents of the labour-market database

Symbol	Name	Dimension
1. Sets		
AGE	Set AGE (age)	5 age groups
EU	Set EU (employed and unemployed)	11 occupations + S + L
EUN	Set EUN	OCC + UN
EUNP	Set EUNP (employed, unemployed, new entrant, PDL)	11 occupations + S + L + N + PDL
GEN	Set GEN (gender)	2 genders
UN	Set UN	S + L + N
NEWENT	Set NEWENT (new entrant)	1 new entrant (N)
EUP	Set EUP	OCC + S + L + PDL
OCC	Set OCC (occupations)	11 occupations
PDL	Set PDL (permanently departed from the labour force)	PDL
RACE	Set RACE (race groups)	2 race groups
STG	Set STG (HIV stages)	5 HIV stages
UNEMP	Set UNEMP (unemployment)	S + L
2. Coefficients in the labour-market database		
ACT	Activities during year t	EUP*AGE*GEN*RACE*STG
ACT_L	Activities during year $t-1$	EUP*AGE*GEN*RACE*STG
ATW	Real post-tax wage	OCC
CAT	Categories at the start of year t	EUNP* AGE*GEN*RACE*STG
H	Actual flows from category c to activity a	EUNP*AGE*GEN*RACE* STG *ACT
OFFER	Desired flows from category c to activity a	EUNP*AGE*GEN*RACE* STG *ACT
TRANS	Proportion of people moving from activity a to category c	EUP*AGE*GEN*RACE*STG*AGE*STG
VAC	Occupation-specific vacancies	OCC
3. Elasticities and parameters		
α	Sensitivity of wages to employment	Scalar
η	Substitution elasticity in labour supply between types of labour	Scalar
μ	Fraction of workers who are fired	Scalar

- Occupation-specific *vacancies* for year t. This matrix is defined as $VAC_{(o)}$ and defined as the difference between the number of employment opportunities (jobs) and incumbents.
- The *transition matrix*. This matrix is written as $T_{(o,aa,g,r,hh,a,h)}$ and shows the probability of a person, in labour-market activity o of gender g and race r, moving from age aa to a and HIV stage hh to h. This matrix

allows adults to move from $ACT\_L_{(o,aa,g,r,hh)}$ to $CAT_{(o,a,g,r,h)}$ at the end of each year.

Parameters included in the database are listed in Table 1, part 3. The α parameter shows the sensitivity of wages to employment during a policy simulation. This parameter is set at 0.6 in the base year. η is the substitution elasticity included in the labour-supply equation and shows the substitution between labour income. μ is the proportion of employed adults who are fired every year.

1.2. Data sources

1.2.1. Labour force surveys

The Labour Force Survey (LFS), published by Statistics South Africa, provides information on the number of adults who are employed, unemployed, and not economically active (NEA).[1] The LFS contains information on the:

- number of employed persons by occupation, gender and race;
- number of short-term unemployed persons by gender and race;
- number of long-term unemployed persons by gender and race; and
- number of not economically active (NEA) persons by gender and race. Note that the number of people who are NEA is part of our PDL activity. PDL further includes those with a Stage 4 status.

1.2.2. Census data

Data from the 2001 Census is used to create an age dimension for the unemployment and PDL activities matrix in 2002. The Census data provides, among other details, information regarding the demographic composition of the South African population.[2] For each unemployment and not economically active activity the gender and race-specific age shares are determined. These shares are then multiplied with the unemployment and PDL activities matrix based on LFS data. The following Census data was used (Statistics South Africa, 2003b):

- number of unemployed persons (official definition) by age, gender and

[1] Statistics South Africa defines "people who are *out of the labour-market* or who are *not economically active* as those who are not available for work. This category includes full-time scholars and students, full-time homemakers, those who are retired, and those who are unable or unwilling to work" (Statistics South Africa, 2003a: xiii). In this thesis, those classified as NEA form part of the PDL activity. PDL includes everybody who is permanently excluded from the labour force and those with Stage 4 HIV.
[2] The Census night was held on 9–10 October 2001.

race;

- number of unemployed persons (expanded definition) by age, gender and race; and
- number of NEA persons by age, gender and race.

1.2.3. HIV data

Various surveys, health reports and the ASSA model database were used to create the HIV status dimension of HIV negative and Stage 1 to Stage 4. The main sources of HIV data are the South African National HIV Survey, conducted in 2002, and the ASSA model developed by Actuarial Sciences South Africa (ASSA).

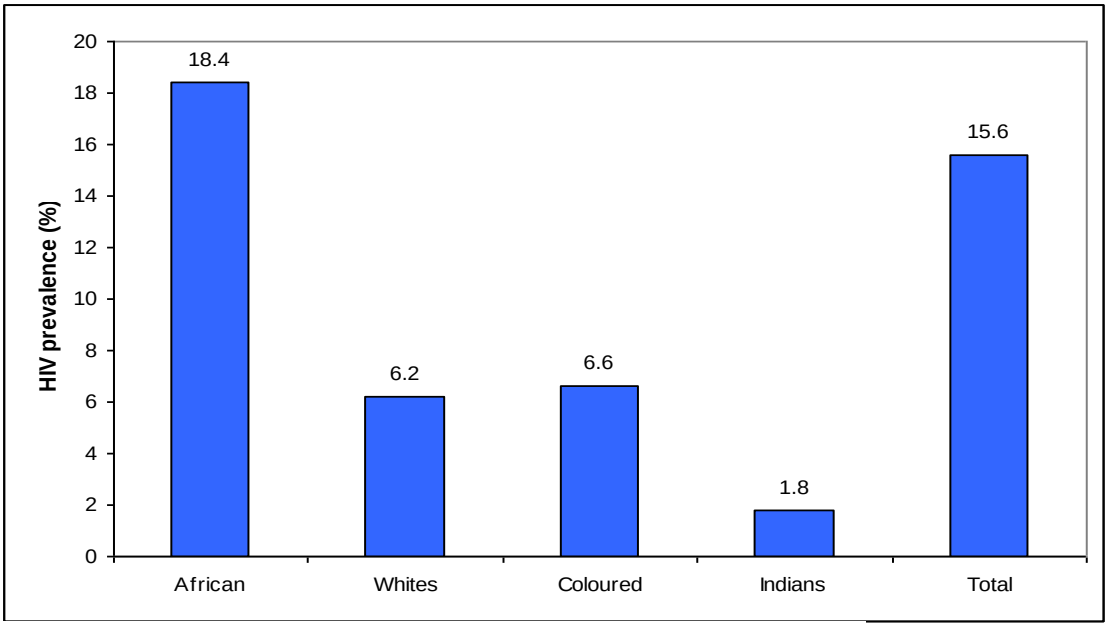
The ASSA2003 model is one that represents the HIV/AIDS epidemic and its demographic impact on the population of South Africa (ASSA, 2005; Dorrington *et al.*, 2005). The model is calibrated to data from the population census, antenatal survey and registered deaths. For the purposes of this thesis, the default settings of the ASSA model are used to derive selected outputs for 2002 (Dorrington *et al.*, 2005: 1–5).

We use the *Full* version of the ASSA model because it generates results for each of the four population groups by age and gender (African, Coloured, Indian and White). The outputs of each population group are aggregated to produce results for the population as a whole (Dorrington *et al.*, 2005: 1–5).

In our specification we do not require HIV-specific results for the entire population or for all four population groups. Instead, we only require information on the number of people between the age of 15 and 65 and two population groups, African and "Other"[3]. We decided on only two population groups because HIV is more prevalent among Africans compared to any other race in South Africa. The remaining races in South Africa report significantly lower levels of HIV. The prevalence rates among the four races are illustrated in Figure 1.

[3] *Other* includes Coloured, Indian and White population groups.

Figure 1. HIV prevalence among adults aged 15–49 years by race (2002) (percentage)



Source: Adapted from Shisana and Simbayi, 2002: 8.

The following data is adopted from the ASSA model:

- Age, gender and race-specific HIV prevalence rates. These prevalence rates are used to determine the age, gender and race-specific number of HIV infections in the base year. The sum across age, gender and race yields the total number of HIV-positive people in the base year. These prevalence rates are introduced when we create the HIV dimension of the activities matrix in the base year.
- Transition rates out of each HIV stage to the next stage. The ASSA model specifies rates of transition out of each HIV-positive stage based on the duration of the time spent in each stage. In the SAGE-H model, transition rates are used to regulate the number of people moving between stages by age, gender, race and all labour-market activities. However, the ASSA model specifies transition rates by age, gender and race but not by labour-market activity.
- Incidence rates. The incidence rates determine the annual number of new HIV infections. As with the above-mentioned transition rates, these incidence rates are not specified by labour-market activity. We use the incidence rates implied by the ASSA model to distribute the number of new HIV infections across all labour-market activities.
- Mortality rates. SAGE-H includes a transition matrix that allows a person, with a unique set of activity o, gender g and race r-specific

characteristics, to survive from year $t-1$ to year t, change their HIV stage or remain in their current HIV stage. This matrix therefore includes mortality rates by age, gender, race and HIV status. We adapt the death rates from the ASSA model to determine the death rates in the SAGE-H database.

- Number of adults in each HIV stage. The ASSA model shows the number of HIV-positive adults in each stage (Stages 1, 2, 3 and 4). From the ASSA model we use the share of the number of people in each stage to distribute the total number of HIV-positive adults across stages in the SAGE-H database.

2. CREATING THE ACTIVITIES MATRIX IN THE BASE YEAR

This section describes the creation of the activities matrix (ACT) for the base year. No single source of data includes all the information required to create the ACT matrix. I use an indirect route to infer the number of adults distinguished by labour-market activity, age, gender, race and HIV status.

2.1. Creating the age dimension for the unemployment and PDL activities

I begin by constructing an activities matrix based on LFS data for the base year. For all *employment* activities I have a matrix describing for each occupation the age, gender and race composition of all employed persons. Call this matrix $WAP\_E^{LFS}_{(o,a,g,r)}$ where $WAP\_E^{LFS}_{(o,a,g,r)}$ is the number of people employed in each occupation o by age a, gender g and race r.[4]

For the unemployment and PDL activities, the LFS provides data by gender and race. Call this matrix $WAP\_UP^{LFS}_{(m,g,r)}$ where $WAP\_UP^{LFS}_{(m,g,r)}$ is the number of people in short-term unemployment, long-term unemployment and PDL activities (m) by gender g and race r.[5]

Our task at hand is to create an age dimension for $WAP\_UP^{LFS}_{(m,g,r)}$. To create this age dimension we construct an activities matrix for the *unemployment* and "*permanently departed from the labour-market*" activities based on Census data. This matrix describes the

[4] $RACE$ is a set containing two elements, African and Other. The Other population group is calculated as $WAP\_E^{LFS}_{(o,a,g,Other)} = \sum_{p} WAP\_E^{LFS}_{(o,a,g,p)}$ where p includes the number of Coloured, Indian and Whites.

[5] At this stage PDL is equivalent to the number of people who are not economically active (NEA). Once I have created the HIV dimension, people with a Stage 4 status are also considered PDL.

number of people in labour-market activity m by age a, gender g and race r. Call this matrix $WAP\_C^{Census}_{(m,a,g,p)}$ where $WAP\_C^{Census}_{(m,a,g,p)}$ is the number of people in short-term and long-term unemployment and PDL activities by age a, gender g and race r.

For each labour-market activity m, the gender and race-specific age shares are calculated via (E.1):

$$AGE\_SHR_{(m,a,g,r)} = \frac{WAP\_C^{Census}_{(m,a,g,r)}}{\sum_{a \in AGE} WAP\_C^{Census}_{(m,a,g,r)}} \quad \text{(E.1)}$$

$m \in UP$;[6] $a \in AGE$; $g \in GEN$ and $r \in RACE$.

where $AGE\_SHR_{(o,a,g,r)}$ is the age-specific share for each labour-market-specific activity m by gender g and race r. Note that $\sum_{a \in AGE} AGE\_SHR_{(m,a,g,r)} = 1$

To create the age dimension, these shares are then multiplied with $WAP\_UP^{LFS}_{(m,g,r)}$:

$$WAP^{LFS}_{(m,a,g,r)} = WAP\_UP^{LFS}_{(m,g,r)} * AGE\_SHR_{(o,a,g,r)} \quad \text{(E.2)}$$

$m \in UP$; $a \in AGE$; $g \in GEN$ and $r \in RACE$.

where $WAP^{LFS}_{(m,a,g,r)}$ is the number of people in all unemployment and PDL activities by age a, gender g and race r. $WAP^{LFS}_{(m,a,g,r)}$, summed across age, is consistent with published unemployment and NEA data (Statistics South Africa, 2009a).
The outcome of this procedure is a matrix describing the number of people in each labour-market activity o, by age a, gender g and race r.

2.2. Creating the HIV status dimension of the activities matrix

In this section we describe the creation of the HIV dimension of the activities matrix. This HIV dimension shows the number of HIV-positive people in each labour-market activity o by age, gender and race. Our aim is to capture the age, gender and race-specific pattern of the epidemic observed in South Africa. To create the HIV dimension the following steps are followed:

- I first calculate the number of people who are HIV positive based on the

[6] UP is a set containing three elements: short-term unemployment S, long-term unemployment L and PDL.

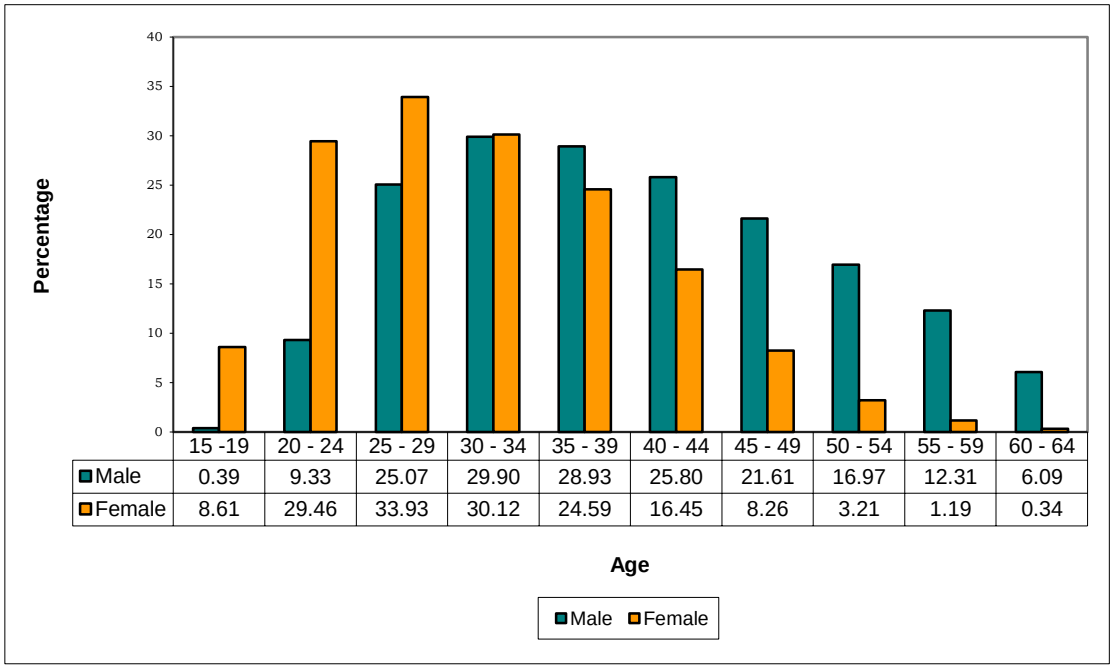
age, gender and race-specific prevalence rates adopted from the ASSA model. I multiply these prevalence rates with the number of people by age, gender and race, in the labour force. The outcome of this step is the total number of HIV-positive and negative adults by age, gender and race.

- The next step is to determine the labour-market activity of HIV-positive and negative adults, i.e. what activities do people perform given their HIV status? Based on recent research I assume that a certain percentage of people in each labour-market activity is HIV-positive. I assume that the prevalence rates for employment activities are lower than for the unemployed or PDL activities. Within occupations I assume that the prevalence rates are higher for unskilled than skilled occupations.
- Finally, I determine for those who are HIV positive, the HIV stage in which they fall. These stages refer to the clinical stages that describe the spectrum of HIV symptoms.

2.2.1. Determining the number of HIV-positive adults by age, gender and race

In this section we describe the determination of the number of HIV-positive people by age, gender and race. The prevalence rate is the result of cumulative exposures to HIV that occurred not only recently but mostly in the past (Shisana *et al.*, 2005: 49). I adopt from the ASSA model the age, gender and race-specific prevalence rates implied for the base year. These rates are illustrated in Figures 2 and 3. The prevalence rate is defined as the proportion of adults who are HIV positive in a given year. Figures 2 and 3 clearly illustrate that HIV prevalence rates follow a distinct pattern by age, gender and race. The following features are noted: (1) the HIV prevalence rates are higher for Africans than for the "Other" population group; (2) HIV prevalence increases dramatically for females and peaks at 34 per cent for the 25–29 age group. Male prevalence peaks at 30 per cent for the age group 30–34. Among young adults aged 15–24, females have almost four times the HIV prevalence of males. Among older age groups, the prevalence rate is higher for males.

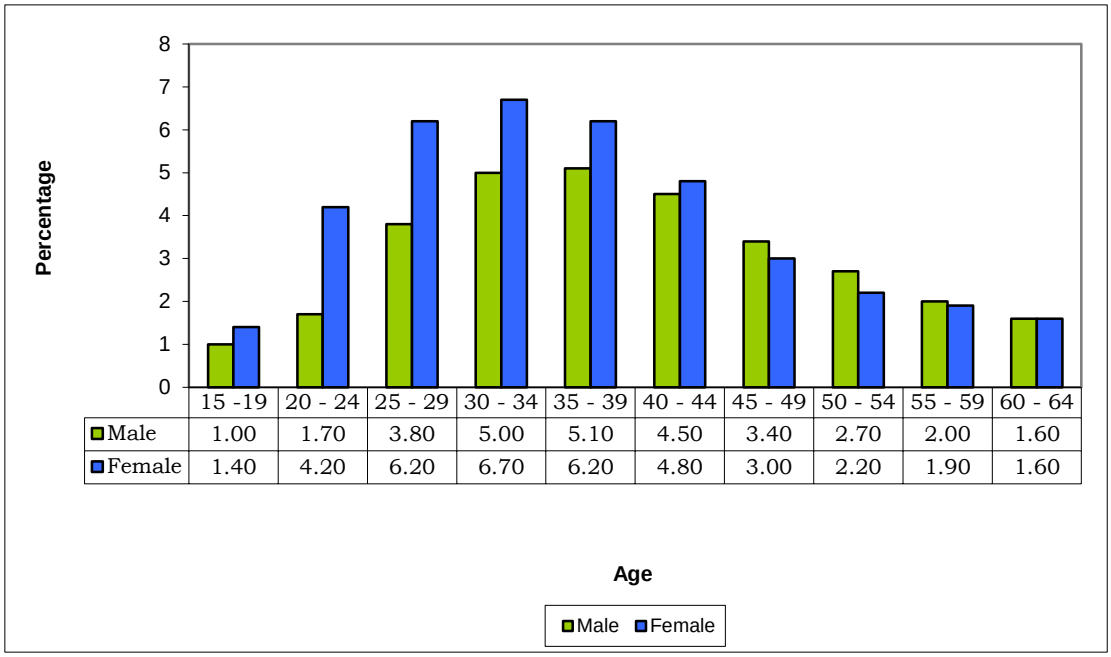
Figure 2. HIV prevalence for Africans by age and gender (2002) (percentage)



	15 -19	20 - 24	25 - 29	30 - 34	35 - 39	40 - 44	45 - 49	50 - 54	55 - 59	60 - 64
Male	0.39	9.33	25.07	29.90	28.93	25.80	21.61	16.97	12.31	6.09
Female	8.61	29.46	33.93	30.12	24.59	16.45	8.26	3.21	1.19	0.34

Source: Own calculation, based on ASSA outputs.

Figure 3. HIV prevalence for "Others" by age and gender (2002) (percentage)



	15 -19	20 - 24	25 - 29	30 - 34	35 - 39	40 - 44	45 - 49	50 - 54	55 - 59	60 - 64
Male	1.00	1.70	3.80	5.00	5.10	4.50	3.40	2.70	2.00	1.60
Female	1.40	4.20	6.20	6.70	6.20	4.80	3.00	2.20	1.90	1.60

Source: Own calculation, based on ASSA outputs.

The number of HIV-positive adults by age, gender and race is calculated via Equation (E.3).

$$WAP_{(a,g,r,HIV+)} = \sum_{o \in ACT} WAP_{(o,a,g,r)} * PREV_{(a,g,r)} \qquad \text{(E.3)}$$

$o \in ACT$, $a \in AGE$; $g \in GEN$ and $r \in RACE$.

where $PREV_{(a,g,r)}$ is the age, gender and race-specific prevalence rates adopted from the ASSA model.

The HIV-negative proportion of the working age population, by age, gender and race is calculated as the difference between the total working age population by age, gender and race and the number of HIV-positive adults.

$$WAP_{(a,g,p,HIV-)} = \sum_{o \in ACT} WAP_{(o,a,g,p)} - WAP_{(a,g,p,HIV+)} \quad \text{(E.4)}$$

$o \in ACT$, $a \in AGE$; $g \in GEN$ and $r \in RACE$.

The following is noted:

- Equation (E.3) does not determine the number of HIV-positive people by labour-market activity. We want to capture the fact that the some labour-market activities are more vulnerable than others and therefore have higher prevalence rates. We also want to reflect the HIV prevalence patterns in all labour-market activities that are also highly differentiated in terms of age, gender and race; and
- Equation (E3) only determines the number of HIV-positive adults and not the clinical stage (Stages 1–4) in which they may fall.

In the SAGE-H database, 4.4 million out of 28 million people, between the ages of 15 and 65, are HIV positive – 15 per cent of the working age population. The results further suggest that HIV is more prevalent among females than males and in young adults, especially African females.

2.2.2. Determining the number of HIV-positive adults in each labour-market activity

HIV/AIDS has a negative impact on the labour-market in terms of morbidity and mortality. However, a number of studies note that in the absence of detailed occupation and industry-specific information, it is difficult to determine the impact of HIV/AIDS on labour supply (Booysen *et al.*, 2003; Vass, 2005).

In this section we explain the calculation of the number of HIV-positive people in each labour-market activity. Our aim is to specify the prevalence rate for each employment occupation, unemployment activity and PDL activity. In determining the HIV prevalence across labour-market activity we consider the following:

- the South African labour market is highly differentiated by skill, gender and race. We also know that HIV prevalence follows a distinct age, gender and race pattern. We therefore expect HIV prevalence by age, gender and race to be reflected in across all labour-market activities;
- the relationship between educational attainment and HIV prevalence; and
- adults' vulnerability because of the nature of the occupation.

2.2.2.1. HIV prevalence and the labour market

The South African labour market is highly differentiated by education level, gender and race. There is high unemployment in general[7], with particularly high unemployment rates for unskilled labour and low unemployment rates among highly skilled labour. Historical imbalances in terms of educational attainment and employment discrimination are captured in the race and gender profile of the different occupations. Unskilled labour tends to be poorly educated and mainly performed by the African population group. People employed in skilled occupations, predominantly performed by the White population group, report higher levels of education (Statistics South Africa, 2009b).

Survey data suggests that HIV prevalence rates follow a specific age, gender and race pattern (Shisana *et al.*, 2002, 2005, 2009). HIV prevalence in the labour market is dependent on the age, gender and race profile of the different occupations (Laubscher *et al.*, 2001; Vass, 2005). I expect occupations employing high-risk groups to have higher HIV prevalence rates. In contrast, HIV prevalence is expected to be the lowest for the *White* population, which generally has a higher level of education and is employed mainly in skilled occupations. (Vass, 2005: 573).

2.2.2.2. HIV prevalence and the level of education

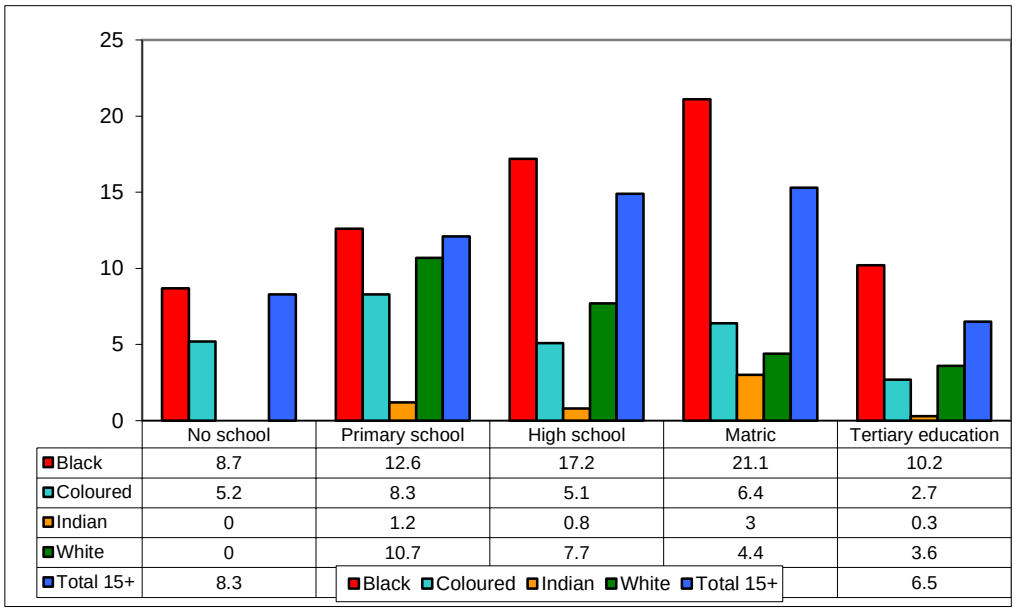
Many studies suggest that there is an inverse relationship between HIV prevalence and educational attainment. Higher level of educational attainment is associated with lower HIV prevalence rates. A possible reason is that higher levels of education are associated with an increase in HIV awareness and knowledge. Also, educated people may have access to more HIV information and improve their ability to act on prevention messages (Laubscher, 2000;

[7] The total unemployment rate for South Africa in 2002 was 41 per cent.

Laubscher *et al.*, 2001; ILO, 2004; Shisana *et al.*, 2002).

This is consistent with findings from the 2002 HIV household survey which suggests that the prevalence of HIV among those with tertiary degrees was significantly lower than those with no schooling, primary or high school (Shisana *et al.*, 2002: 54). However, this pattern is reversed when only the African population group is considered. Figure 4 shows that for the African population there is a significant increase in the HIV prevalence rate as education levels increase. Figure 4 shows, for those who are 15 years and older, the HIV prevalence rate between different levels of education. For the population as a whole the results for all education levels below tertiary education, suggest that there is a positive relationship between the level of education and HIV prevalence. This result is mainly driven by the HIV prevalence pattern observed among the African population. Africans show the highest HIV prevalence rates at Matric level. HIV prevalence trends among the Coloured, Indian and White population groups seem to decline with higher levels of education. This is in line with the general trend found in other studies (Shishana *et al.*, 2002; Vass, 2005).

Figure 4. HIV prevalence rate of adults 15 years and older by educational level and race in South Africa (2002) (percentage)



	No school	Primary school	High school	Matric	Tertiary education
Black	8.7	12.6	17.2	21.1	10.2
Coloured	5.2	8.3	5.1	6.4	2.7
Indian	0	1.2	0.8	3	0.3
White	0	10.7	7.7	4.4	3.6
Total 15+	8.3				6.5

Source: Shisana *et al.*, 2002: 54.

The Bureau of Economic Research (BER) (2001) projects large disparities in prevalence rates across skills. They project an increase in the prevalence rates for skilled and semi-skilled labour from 12.1 per cent and 14.3 per cent respectively in 2000 to 25.4 per cent and 27.6 per cent in 2015. They also suggest that HIV prevalence for those who are unemployed, are up to 30–50 per cent higher compared to those who are employed. This partly reflects the race, gender and age composition of those who are unemployed (cited in Laubscher *et al.*, 2001: 7). A study by ING BARINGS, cited in Laubscher (2000) also indicates large disparities

in the HIV prevalence within various skill categories. They project that the HIV prevalence rate for semi-skilled and unskilled occupations could be three times higher than the HIV prevalence rate for the skilled group. Their projections suggest that HIV prevalence peaks at 13.1 per cent for the highly skilled group, at 22.8 per cent for the skilled groups and 32.8 per cent for the semi- and unskilled groups (cited in Laubscher, 2000: 9).

Survey data for Botswana is consistent with the BER estimates. The data suggests that students and highly skilled workers (office/professional) have lower prevalence rates than lower skilled workers (domestic workers/labourers/store workers and farmers). The prevalence rates for 2005 vary between 20 per cent for students to over 40 per cent for labourers, store workers and domestic workers (Jeffris *et al.*, 2007: 11).

Evian *et al.* (2004) analysed results of 34 workforce surveys to determine HIV prevalence among formally employed, largely male populations. These surveys include 44,000 employees and were conducted in South Africa, Botswana and Zambia during 2000–2001. The overall prevalence rate for the entire sample is 16.6 per cent. Country-wide prevalence is 14.5 per cent for South Africa, 17.9 per cent for Zambia and 24.6 per cent for Botswana. In terms of industrial sectors, the mining and metal-processing sectors have the highest HIV prevalence rates at 18 per cent and 17.3 per cent respectively. Contract, unskilled and semi-skilled workers are more likely to be infected at 23 per cent, 18.3 per cent and 18.7 per cent respectively. Skilled workers and managers have the lowest HIV prevalence rates at 10.5 per cent and 4.5 per cent respectively.

Some studies found that skilled workers are likely to be more vulnerable to contracting the HIV virus than unskilled workers. Skilled workers have a higher level of education and generally experience improvements in their socioeconomic environment. This can lead to behaviour that increases the risk of their exposure to the HIV virus (ILO, 2004). Fortson (2008) used data from the Demographic and Health Surveys for Burkina Faso, Cameroon, Ghana, Kenya and Tanzania to investigate the relationship between HIV status and socioeconomic status. She finds that education is positively related to risk factors that increase exposure to HIV (Fortson, 2008: 303). The surveys suggest that adults with six years of schooling are as much as three percentage points more likely to be infected with HIV than adults with no schooling. A study of the construction industry in South Africa finds that permanently employed, operators and drivers and skilled employees have the highest HIV prevalence rates (Bowen *et al.*, 2008: 834–835).

2.2.2.3. HIV prevalence and the nature of an occupation

Vulnerability in contracting the HIV virus may increase due to the nature of one's occupation (Laubscher *et al.*, 2001; Vass, 2005). A study based on a nationwide

representative sample of medical professionals and non-professional health workers found that an estimated 15.7 per cent of the sample who were employed in the public and private heath facilities was living with HIV. For health workers aged 18–35, the prevalence rates was higher at 20 per cent. For professional health workers the HIV prevalence was 13.7 per cent and for non-professional health workers, 20.3 per cent (Shisana *et al.,* 2004). Another vulnerable occupation is truck drivers. As mentioned before, a study based on HIV prevalence in the construction industry finds that prevalence rates are particularly high for drivers. Drivers and operators are away more often from home compared to other occupations within the construction industry, which increases their risk of infection. They also earn higher incomes enabling them to pay for sex (Bowen *et al.*, 2008: 834–835). They are very mobile because they transport commodities across regions, which allow them to come in contact with women working as sex workers. This places them at risk of contracting HIV.

2.2.2.4. HIV prevalence by labour-market activities in SAGE-H

Our task at hand is to distribute the total number of HIV-positive adults, determined in Equation (E.3), across all employment, unemployment and PDL activities. To do this we create a matrix $\left(PERC_{(o,HIV+)}\right)$ by labour-market activity o. I assign a percentage to each labour-market activity that determines the number of people who are HIV positive. These percentages are reported in Appendix 6A. In determining the number of HIV-positive people across labour-market activity I assume that:

- The prevalence rates are not uniform across labour-market activities.
- The prevalence rates are lower for those who are employed than those who are unemployed or PDL.
- While there is evidence that in some African countries infection levels increase with socioeconomic status, the balance of evidence in South Africa shows that the prevalence rates are lower for skilled people than for unskilled people. Data also suggests that there is a downward trend in infections levels as people move from semi-skilled occupation categories to middle and senior management amongst the African population (Laubscher *et al.*, 2001: 7).

The number of HIV-positive people is calculated via Equation (E.5):

$$WAP_{(o,HIV+)} = \sum_{a \in AGE} \sum_{g \in GEN} \sum_{r \in RACE} WAP_{(o,a,g,r)} * PERC_{(o,HIV+)} \quad \text{(E.5)}$$

$o \in EUP$;[8] $a \in AGE$; $g \in GEN$ and $r \in RACE$.

The number of HIV-negative people is calculated via Equation (E6.6):

$$WAP_{(o,HIV-)} = \sum_{a \in AGE} \sum_{g \in GEN} \sum_{r \in RACE} WAP_{(o,a,g,r)} - WAP_{(o,HIV+)} \quad \text{(E.6)}$$

$o \in EUP$; $a \in AGE$; $g \in GEN$ and $r \in RACE$.

The SAGE-H data suggests that 8 per cent of skilled employees, 15 per cent of semi-skilled and 17 per cent of unskilled workers are HIV positive. The prevalence rates by skill are slightly lower when compared to the prevalence for skilled and unskilled workers projected by Abt Associates/Aids Research Unit Metropolitan Life. They project that in 2000 HIV prevalence for highly skilled, skilled, semi- and unskilled workers are 7.2, 12.1 and 14.3 per cent respectively (Laubscher *et al.*, 2001: 12).[9] What is consistent between our HIV prevalence by skill specification and the HIV profiles defined by Abt Associates, is that HIV prevalence is the lowest for skilled occupations and the highest for unskilled occupations and for those who are unemployed. In the SAGE-H database, the number of HIV-positive people in the unemployment and PDL activity is 17 per cent. HIV prevalence for the unemployed is very high due to a predominantly young, African and female profile (Laubscher *et al.,* 2001: 12). The distribution of HIV-negative and positive people across all labour-market activities is summarised in Appendix 2.

2.2.3. Distributing the number of HIV-positive adults in each labour-market activity by age, gender and race

In this section we describe how we distribute the activity-specific number of HIV-positive adults across age, gender and race. Each labour-market activity is unique in its age, gender and race profile. Some studies assume that the HIV incidence and prevalence rates are

[8] *EUP* is the set of all employment, short and long-term unemployment and "permanently departed from the labour force" (PDL) activities.

[9] It should be noted that skills defined by occupations are different between the Abt Associates study and SAGE-H. In the Abt Associates study skills categories are defined according to the 1996 Census classification. They define highly skilled occupations as professionals, semi-professional and technical occupations, managerial, executive and administrative occupations and certain transport occupations. Skilled occupations include clerical, service and sales occupations, farmers, farm managers, artisans, apprentices and related occupations, production foremen and production advisors. All other occupations are defined as either semi- or unskilled (Laubscher *et al.*, 2001: 10). For SAGE-H the occupations defined by skill are listed in Appendix 6A.

uniform across skill categories (cited in Vass, 2005: 573). This implies that all workers, irrespective of their age, gender or race, face the same level of exposure to the HIV virus. In our study, we want to capture the age, gender and race-specific features of the labour market, and also reflect the age, gender and race-specific HIV patterns observed in Figures 6.2 and 6.3. To capture these features we calculate a number of shares. Each share captures a unique feature of either the labour-market composition or HIV epidemic in South Africa. The first share indicates the number of HIV-positive adults in each labour-market activity according to their HIV status. This share shows that some occupations, such as the unskilled occupations, have a larger share of HIV-positive adults as compared to the skilled occupations.

$$S1_{(o,h)} = \frac{WAP_{(o,h)}}{\sum_{h \in H2} WAP_{(o,h)}} \qquad \text{(E.7)}$$

$o \in EUP$ and $h \in H2$.[10]

where
- $WAP_{(o,h)}$ is the number of adults of in each labour-market activity o given their HIV status h; and
- $\sum_{h \in H2} WAP_{(o,h)}$ is the total number of adults, summed over HIV status, in each labour-market activity.

The second set of shares indicates the number of HIV-positive adults by gender:

$$S2_{(g,h)} = \frac{WAP\_AP_{(g,h)}}{\sum_{h \in H2} WAP\_AP_{(g,h)}} \qquad \text{(E.8)}$$

$g \in GEN$ and $h \in H2$.

where
- $WAP\_AP_{(g,h)}$ is the number of adults by gender g and HIV status h. The values are calculated as the sum over age and race of $WAP_{(a,g,p,h)}$, calculated in (E6.4) and (E6.5); and
- $\sum_{h \in H2} WAP\_AP_{(g,h)}$ is the sum over HIV status h.

[10] $H2$ is a set including two elements: HIV positive and HIV negative.

The third set of shares indicates the number of adults by gender and race based on their HIV status. This share captures the gender and race-specific distribution of HIV and is calculated as:

$$S3_{(g,r,h)} = \frac{WAP\_A_{(g,r,h)}}{\sum_{r \in RACE} WAP\_A_{(g,r,h)}} \qquad \text{(E.9)}$$

$r \in RACE$ and $h \in H2$.

where
- $WAP\_A_{(g,r,h)}$ is the number of adults by gender g, race r and HIV status h. This matrix is calculated from $WAP_{(a,g,p,h)}$, summed over age a; and
- $\sum_{r \in RACE} WAP\_A_{(g,r,h)}$ is the sum over race r.

The final set of shares indicates the number of adults by age, gender and HIV status. This share captures the age and gender-specific distribution of HIV and is calculated as:

$$S4_{(a,g,h)} = \frac{WAP\_P_{(a,g,h)}}{\sum_{a \in AGE} WAP\_P_{(a,g,h)}} \qquad \text{(E.10)}$$

$a \in AGE$; $g \in GEN$ and $h \in H2$.

where
- $WAP\_P_{(a,g,h)}$ is the number of adults by age a, gender g, and their HIV status h. This matrix is calculated from $WAP_{(a,g,p,h)}$, summed across race r; and
- $\sum_{a \in AGE} WAP\_P_{(a,g,h)}$ is the sum across age a.

These shares show that younger females have a higher prevalence rate than males of the same age. Older men have a higher prevalence than females in the same age group. We multiply the above shares with each other to calculate the proportion of adults in o,a,g,r that is HIV positive:

$$S_{(o,a,g,r)} = S1_{(o,HIV+)} * S2_{(g,HIV+)} * S3_{(g,r,HIV+)} * S4_{(a,g,HIV+)} \qquad \text{(E.11)}$$

$o \in EUP$; $a \in AGE$; $g \in GEN$; $r \in RACE$ and $h \in H2$.

The distribution of the proportion of HIV-positive adults across age, gender and race is determined via Equation (6.12).

$$WAP_{(o,a,g,r,HIV+)} = S_{(o,a,g,r)} * WAP_{(o,HIV+)} \quad \text{(E.12)}$$

$o \in EUP$; $a \in AGE$; $g \in GEN$; $r \in RACE$ and $h \in H2$.

where
- $WAP_{(o,a,g,r,HIV+)}$ is the number of HIV-positive adults in each labour-market activity o of age a , gender g and race r .

The number of HIV-negative adults is calculated as the difference between the total number of adults less the portion of adults who are HIV-positive:

$$WAP_{(o,a,g,r,HIV-)} = WAP_{(o,a,g,r)} - WAP_{(o,a,g,r,HIV+)} \quad \text{(E.13)}$$

$o \in EUP$; $a \in AGE$; $g \in GEN$ and $r \in RACE$.

where
- $WAP_{(o,a,g,rHIV-)}$ is the number of HIV-negative adults in each labour-market activity o by age a , gender g and race r ; and
- $WAP_{(o,a,g,r)}$ is the number of adults in each labour-market activity o by age a , gender g and race r .

This matrix, containing information on the number of people in each labour-market activity by age, gender, race and HIV status, is scaled to official data by using the RAS procedure.

The final step is to distribute the number of HIV-positive people into an HIV stage. These stages are based on the WHO Clinical Staging System, which defines the stages through which an infected person moves from infection to death.

2.2.4. Dividing the HIV-positive adults into different HIV stages

To determine the number of people in each HIV stage, we adopt the percentage of HIV-positive people in each HIV stage from the ASSA model.[11] Approximately 34 per cent of HIV-positive adults are in Stage 1, 23 per cent in Stage 2, 34 per cent in Stage 3 and 9 in Stage 4. We use these shares to distribute the number of HIV-positive adults, determined in Equation (E.12), into the four HIV stages.

[11] The total number of HIV-positive adults in each stage is reported in the worksheet named *Results* in the ASSA2003 model.

We further assume that HIV-positive adults who are employed or unemployed can only fall in Stages 1, 2 or 3. Those in Stage 4 are too ill to be part of the labour force. They form part of the PDL. This was the final step in creating the activities matrix ($ACT_{(o,a,g,r,h)}$) for the base year.

2.3. Summary of the activities matrix

The activities matrix contains information on the number of people in each labour-market activity by age, gender, race and HIV status. This matrix serves as our starting point in deriving additional matrices required in the labour-market specification described in Section 1.1. The following features are noted:

- There are 28.5 million people between the ages of 15 and 65 in the base year. Of this total, 4.4 million are HIV positive and 24.1 million are HIV negative. The prevalence rate is 15.5 per cent. See Table 1 in Appendix 6B.
- 11.2 million adults are employed, 8.1 million are unemployed and 9 million are considered PDL. This implies that approximately 40 per cent of people between the ages of 15 and 65 are employed, 28 per cent unemployed and 32 per cent PDL. The unemployment rate is 41.8 per cent.
- There are 14.6 million (51.3 per cent) females and 13.9 million (48.7 per cent) males between the ages of 15 and 65. In general more men (57.3 per cent) than women (48.7 per cent) are employed, more women (55.7 per cent) than men (44.3 per cent) are unemployed and more women (58 per cent) than men (42 per cent) are PDL. See Table 2 in Appendix 6B.
- Of the 28.5 million people, 21.9 million (77 per cent) are part of the African population group. Approximately 7.7 million (35 per cent) are employed, 6.9 million (32 per cent) are unemployed and 7.2 million (33 per cent) are PDL. 6.5 million (23 per cent) fall in the “Other” population group. 3.6 million (54 per cent) are employed, 1.1 million (17 per cent) are unemployed and 1.8 million (28 per cent) are PDL.
- Labour-market disparities are clearly visible in terms of employment and unemployment by gender and race. Overall, 76 per cent of adults in the “Other” population group are employed compared to 52 per cent of Africans. For both races, more males than females are employed. For the “Other” population group, 72 per cent of females and 80 per cent of males are employed. Employment levels for Africans are lower at 46 and

59 per cent for females and males respectively.

- Labour-market disparities are also clearly observed in terms of employment by skill. In general more males than females are employed in skilled occupations. Females tend to be employed in service-oriented occupations such as Clerks and the Domestic occupation. Craft workers and Machine operators tend to be male dominated.
- Inequalities in the labour market are most starkly reflected in the employment by race (Appendix 6B, Table 6). Of the total number of people employed, approximately 22 per cent are employed in skilled occupations, 50 per cent in semi-skilled occupations and 28 per cent in unskilled occupations. The majority of skilled occupations are held by the "Other" population group especially the Legislators and Professionals occupations. The majority of the semi-skilled and unskilled workers are Africans with 85 per cent of the unskilled occupations performed by Africans.
- Finally, the youngest age group (15–24) and oldest age group (55–65) report the largest percentage of those who are PDL at 53 and 47 per cent respectively. Unemployment is the highest for the younger age group (15–35). The employment level is the highest for ages 35–54.

3. DERIVING THE INITIAL MATRICES REQUIRED IN THE LABOUR-MARKET SPECIFICATION MODEL

Several additional matrices are derived from the activities matrix in the base year. In deriving these matrices, a program code developed by Dixon and Rimmer is adopted for our purposes. These matrices include:

- The lagged activity matrix (ACT_L) for each labour-market activity o by age a gender g, race r and health status h. This is the activities matrix for year $t-1$.
- The categories matrix at the start of year t (CAT). This matrix is important because it (1) indicates how many people survived from year $t-1$ to year t, (2) indicates how many people are part of a labour force at the start of year t and (3) forms the basis for deriving planned and actual flows matrices.

3.1. Calculating categories at the start of the base year

The categories matrix is defined as $CAT_{(o,a,g,r,h)}$ and shows the number of adults in each labour-market category o by gender g and race r who are allocated to age a and HIV status h at the start of year t. Categories are dependent on (1) the number of people in each labour-market activity by age, gender, race and HIV stage in year $t-1$ and (2) the transition matrix which allows people in each labour-market activity by gender and race to move from age aa to a and from HIV stage hh to h. Categories are determined via Equation (E6.14) and calculate the number of adults who survive from year $t-1$ to the start of year t:

$$CAT_{(o,a,g,p,h)} = \sum_{aa \in AGE} \sum_{hh \in STG} ACT\_L_{(o,aa,g,p,hh)} * T_{(o,aa,g,p,hh,a,h)} \qquad \text{(E.14)}$$

$o \in EUP$; $a \in AGE$; $g \in GEN$; $r \in RACE$ and $h \in STG$.[12]

where

- $ACT\_L_{(o,aa,g,r,hh)}$ is the activity matrix in year $t-1$. This activities matrix shows the number of people in each labour-market activity o by age aa gender g, race r and HIV stage hh in year $t-1$. $ACT\_L_{(o,aa,g,r,hh)} = \left(ACT_{(o,a,g,r,h)} * \frac{1}{1.02} \right)$; and
- $T_{(o,aa,g,p,hh,a,h)}$ is the transition matrix and shows the proportion of people in each labour-market activity o by gender g and race r, surviving from age aa to a and moving from HIV stage hh to h. In other words, it is the proportion of people in labour-market activity o by gender g and race r, who are allocated to category c at the start of year t.

Notice that Equation (E.14) does not allow adults to change their labour-market activity o, gender g or race r from year $t-1$ to the start of year t. The transition matrix allows adults to only change their age from aa to a and HIV stage from hh to h.

[12] STG is a set containing all HIV stages: HIV negative, Stage 1, Stage 2, Stage 3 and Stage 4.

3.2. Calculating the number of new entrants at the start of the base year

During the baseline forecast simulation, new entrants to the labour force are exogenously added at the beginning of each year. However, for the base year, we need to determine the initial value for the number of new entrants to the labour force. We assume that the number of new entrants is the difference between the activities undertaken during year 2002 and the categories at the beginning of year 2002. New entrants are calculated via Equation (E.15):

$$N_{(a,g,r,h)} = \sum_{EUP \in o} ACT_{(o,a,g,r,h)} - \sum_{EUP \in o} CAT_{(o,a,g,r,h)} \qquad \text{(E.15)}$$

$a \in AGE$; $g \in GEN$; $r \in RACE$ and $h \in STG$.

where

- $\sum_{EUP \in o} ACT_{(o,a,g,r,h)}$ is the activities matrix derived in the previous section. This matrix shows the number of adults, summed across all labour-market activities o, by age, gender, race and HIV stage; and
- $\sum_{EUP \in o} CAT_{(o,a,g,r,h)}$ is the categories matrix at the start of the base year. This matrix is calculated via (E.14) and shows the number of adults, summed across all labour-market categories o, by age, gender, race and HIV stage.

4. MATRICES DESCRIBING FLOWS FROM CATEGORIES TO ACTIVITIES

The *OFFERS* matrix shows people's desire, given their age a, gender g, race r and HIV stage h, to move from category c to activity a. I assume that most people would like to be employed.[13] There is, however, no guarantee that all offers from category c (occupations, unemployed and new entrants) to activity a, where a refers to all occupations, will be accepted. The offers that are accepted and those that are rejected are derived in the actual flows matrix (H), explained in Section 4.2. The following similarities between the matrices are noted:

- Both matrices determine for all age a, gender g, race r and HIV stage h, the flow of people from all category c to all activity a.

[13] The pool of people who offer their labour is restricted to those in the employment, short and long-term unemployment and new entrants. Adults in the PDL category do not offer to employment or unemployment activities.

- Their respective dimensions in terms of labour-market categories c, activities o, age a, gender g, race and HIV stage h are similar.

Figure 5. Specifying the flows from all categories to all activities

Categories \ Activities		Employment		Unemployment		PDL
		Occupation 1 (a,g,r,h)	Occupation 2 … 11 (a,g,r,h)	Short-term unemploy-ment (a,g,r,h)	Long-term unemploy-ment (a,g,r,h)	PDL (a,g,r,h)
Employment	Occupation 1 (a,g,r,h)	1	3	2	0	6
	Occupation 2..11 (a,g,r,h)	1		2	0	6
Unemployment	Short-term unemployment (a,g,r,h)	4		0	5	6
	Long-term unemployment (a,g,r,h)	4		0	5	6
	New entrant (a,g,r,h)	4		5	0	6
	PDL (a,g,r,h)	0		0	0	7

Because the matrices are very large and capture different behaviour of adults moving from categories to activities, I construct them in a number of steps. In Figure 5 the rows represent the categories from which the flows originate and the columns represent the activities to which these flows are destined. The numbers in each block are for reference purposes only. We use the numbers in Figure 5 to identify the different flows discussed in the sections below. Zero indicates the flows from categories to activities that are prohibited.

4.1 Creating the planned flows matrix

In linking the categories at the start of year t to the activities undertaken during year t, we specify a flow for all age a, gender g, race r and HIV status h elements, from each

category c to each activity o, $OFFER_{(c,a,g,r,h,o)}$. Notice that $OFFER_{(c,a,g,r,h,o)}$ includes all flows from all categories (employment, unemployment, new entrant and PDL) to all activities (employment, unemployment and PDL). In constructing the *OFFER* matrix, we include assumptions regarding:

- the strength of the offers from unemployment and new entrant categories to all employment activities;
- the number of people who voluntarily move from employment categories c to employment activities a where $c \neq a$;
- the flow of people between all employment categories and unemployment activities; and
- adults in all categories with a Stage 4 HIV stage.

The assumptions are discussed in the relevant sections below.

4.1.1 Planned flows from employment category o to employment activity m, that is, non-diagonal flows (Figure 5; Area 1)

In this section I describe the movement of adults from occupation *o* to occupation *m,* where *o* is different from *m*. The non-diagonal employment flows are determined via (E6.16):

$$OFFER_{(o,a,g,r,h,m)} = \left(1-\Omega_{(o,m)}\right) * DELTA_{(o,a,g,r,h)} * CAT_{(o,a,g,r,h)} * \left[\frac{CAT_{(m,a,g,r,h)}}{BOT_{(o)}}\right] \quad \text{(E6.16)}$$

$o \in OCC$; $a \in AGE$; $g \in GEN$; $r \in RACE$; $h \in H3$[14] and $m \in OCC$.

where
- $OFFER_{(o,a,g,r,h,m)}$ is the offers of people of age a, gender g, race r and HIV stage h, from employment categories o to employment activities m;
- $\left(1-\Omega_{(o,m)}\right)$ activates or deactivates a particular equation. $\Omega_{(o,m)}$ takes a value of 0 or 1. If $\Omega_{(o,m)}$ is 0, then the equation is activated because the term between the brackets is 1. Alternatively, if Ω is set to 1, the term between the brackets is zero, thereby deactivating the equation;
- $DELTA_{(o,a,g,r,h)}$ is the proportion of people of all age a, gender g, race r and HIV stage h, preferring to move from occupation category o to

[14] $H3$ is a set containing four elements: HIV negative, Stage 1, Stage 2 and Stage 3.

employment activities m. DELTA is uniformly set at 7 per cent;

- $CAT_{(o,a,g,r,h)}$ is the number of people in employment category o, gender g and race r, allocated to age a and HIV stage h at the start of the year;
- $BOT_{(o)}$ is the total offers from all employment categories except those in o;[15] and
- $\frac{CAT_{(m,a,g,r,h)}}{BOT_{(o)}}$ is the share of occupation m in the total offers to occupation o.

Equation (E.16) shows that the offers from occupation o to occupation m are dependent on the number of people wanting to move to a different occupation and the share of offers from occupation *m* in *o*.

I further assume that:

$$OFFER_{(o,a,g,r,Stage4,m)} = 0 \qquad \text{(E.17)}$$

$o \in OCC$; $a \in AGE$; $g \in GEN$, $r \in RACE$ and $m \in OCC$.

Equation (E.17) states that offers of people of all age a, gender g, race r with a Stage 4 status, moving from employment category o to employment activity m, are zero. People in Stage 4 are too ill to be part of the labour force and will permanently leave the labour force. They move into the PDL activity where they will remain until they die. This flow is determined via (E.26).

The following stylised features are built into the non-diagonal flows:

- I assume that there is occupation persistence. This means that the majority of people employed in occupation *o* during year *t*-1 choose to be employed in the same occupation during year *t*. A small proportion of people will choose to offer their labour to a different occupation. This movement from occupation *o* to occupation *m* is regulated by *DELTA* in Equation (E6.16).

---

[15] $BOT_{(o)} = \left(\sum_{m \in OCC} CAT_{(m)} \right) - CAT_{(o)}$ where m is all employment categories. For example, assume there are three occupations (occ1, occ2 and occ3). *BOT* for occupation 1 is the sum of all the employment categories except occupation 1.

- Workers employed in a specific occupation will only offer to an occupation with a similar skill level. To determine this flow, we adopt the skill attainment classification by Statistics South Africa (2006). The skills associated with each occupation are presented in Table 2.

Table 2. Skill level and educational attainment

Major group (Occupation)	Skill level	Description
1. Legislators 2. Professionals	4	Education that begins at the age of 18 or 19, lasts about 3, 4 or more years, and leads to a university or post-graduate university degree.
3. Technicians	3	Education that begins at the age of 17 or 18, lasts about 1 to 4 years and leads to an award not equivalent to a first university degree.
4. Clerks 5. Service workers 6. Skilled agricultural workers 7. Crafts workers 8. Plant and machine operators	2 2 2 2 2	Secondary education that begins at the age of 13 or 14 and last about five years. A period of on-the-job training and experience may be necessary.
9. Elementary occupations 10. Domestic workers 11. Occupation unspecified	1 1 1	Primary education that generally begins at the age of 6 or 7 and lasts about 7 years. Including persons with no formal primary education, or with incomplete primary education.

Source: Statistics South Africa, 2006b: 49.

I assume that occupations with a skill level of 1 and 2 only offer to occupations with a similar skill level. Similarly, occupations with skill levels of 3 and 4, only offer to occupations with a similar skill level. For example, people employed as Clerks (row 4) are classified as having a Level 2 skill level. I assume that they will offer to any occupations in rows 4–11, that is, occupations with a similar skill level. They do not offer to occupations in rows 1–3.

4.1.2. Planned flows from employment category o to unemployment activity S (Figure 5; Area 2)

The voluntary flows from occupation o to short-term unemployment S are calculated via (E.18):

$$OFFER_{(o,a,g,r,h,S)} = CAT_{(o,a,g,r,h)} * EMPtoS_{(o,a,g,r,h)} \quad \text{(E.18)}$$

$o \in OCC$; $a \in AGE$; $g \in GEN$; $r \in RACE$ and $h \in H3$.

where
- $OFFER_{(o,a,g,r,h,S)}$ is the number of people of age a, gender g, race r and HIV stage h, offering to move from employment category o to short-term unemployment (S);

- $CAT_{(o,a,g,r,h)}$ is the number of people in employment category o, gender g and race r, allocated to age a and HIV stage h at the start of the year; and
- $EMPtoS_{(o,a,g,r,h)}$ is the proportion of people, given a set of age a, gender g, race r and HIV stage h characteristics, choosing to voluntarily move to short-term unemployment.

I assume that:

$$OFFER_{(o,a,g,r,Stage4,S)} = 0 \qquad \text{(E.19)}$$

$o \in OCC$; $a \in AGE$; $g \in GEN$ and $r \in RACE$.

Equation (E.19) states that offers from employment category o to unemployment activity S, for all age a, gender g, race r with a Stage 4 HIV status, are zero.

The following features are incorporated to regulate $EMPtoS$:

- People in employment categories only offer to the short-term unemployment activity and not the long-term unemployment activity.
- Nobody with a Stage 4 HIV status offers to the short-term unemployment activity. They permanently move to the PDL activity.
- I assume that an HIV-positive Stage 3 person offers more strongly to short-term unemployment than an HIV-negative person or an adult in Stage 1 or Stage 2. The proportion of people voluntarily moving to the short-term unemployment activity is set at 3 per cent for HIV negative, Stage 1 and Stage 2, and 5 per cent for those in Stage 3. People in Stage 4 do not offer to short-term unemployment.

4.1.3. Planned flows from employment category o to employment activity o, that is, diagonal flows (Figure 6.5; Area 3)

The diagonal flows are the number of people in occupation o at the start of the year less those who chose to move away from occupation o. People may leave occupation o and offer to occupation m or move to short-term unemployment S. The diagonal flows are determined via (E6.20):

$$OFFER_{(o,a,g,r,h,o)} = CAT_{(o,a,g,r,h)} - OFFER_{(o,a,g,r,h,m)} - OFFER_{(o,a,g,r,h,S)} \quad \text{(E.20)}$$

$o \in OCC$; $a \in AGE$; $g \in GEN$; $r \in RACE$ and $h \in H3$.

where
- $OFFER_{(o,a,g,r,h,o)}$ is the number of people of age a, gender g, race r and HIV stage h, offering to move from employment category o to employment activity o;
- $CAT_{(o,a,g,r,h)}$ is the number of people in employment category o, gender g and race r, allocated to age a and HIV stage h at the start of the year;
- $OFFER_{(o,a,g,r,h,m)}$ is the number of people of age a, gender g, race r and HIV stage h, offering to move from employment categories o to employment activities m. These are the non-diagonal flows determined via (E6.16); and
- $OFFER_{(o,a,g,r,h,S)}$ is the number of people of age a, gender g, race r and HIV stage h, offering to move from employment category o to short-term unemployment (S). This flow is derived via (E.18).

Again, I assume that:

$$OFFER_{(o,a,g,r,Stage4,o)} = 0 \quad \text{(E.21)}$$

$o \in OCC$; $a \in AGE$; $g \in GEN$ and $r \in RACE$.

Equation (E.21) states that offers from employment category o to employment activity o, for all age a, gender g, race r with a Stage 4 HIV status, are zero.

4.1.4. Planned flows from unemployment and new entrant categories c to employment activities o (Figure 6.5; Area 4)

Offers from the unemployment and new entrant categories are determined via (E6.22):

$$OFFER_{(c,a,g,r,h,o)} = GAMA_{(c,h)} * CAT_{(c,a,g,r,h)} * \left[\frac{CAT_{(c,a,g,r,h)}}{\sum_{mm \in OCC} CAT_{(mm,a,g,r,h)}} \right] \quad \text{(E6.22)}$$

$c \in UN$;[16] $a \in AGE$; $g \in GEN$; $r \in RACE$ and $h \in H3$.

[16] UN is a set of all unemployment and new entrant categories.

where

- $OFFER_{(c,a,g,r,h,o)}$ is the number of offers from new entrants, short-term and long-term unemployment category c to employment activity o, for all age a, gender g, race r and HIV stage h;
- $GAMA_{(c,h)}$ captures the strength of these offers from unemployment and new entrants categories c to all employment activities o. The values for $GAMA$ are summarised in Table 6.4;
- $CAT_{(c,a,g,r,h)}$ is the number of people in the unemployment and new entrant category c, gender g and race r, allocated to age a and HIV stage h at the start of the year; and
- $\left[\frac{CAT_{(c,a,g,r,h)}}{\sum_{mm \in OCC} CAT_{(mm,a,g,r,h)}}\right]$ is the share of occupation c in total occupations. If 10 per cent of people are employed in occupation *Legislator,* then 10 per cent of people in the unemployment and new entrant category will offer to *Legislator*.

I further assume that:

$$OFFER_{(c,a,g,r,Stage4,o)} = 0 \qquad \text{(E6.23)}$$

$c \in UN$; $a \in AGE$; $g \in GEN$ and $r \in RACE$.

Equation (E.23) states that offers from new entrant and unemployment category c to employment activity o, for all age a, gender g, race r with a Stage 4 HIV status, are zero.

In assigning values for $GAMA$, which appear in (E.22) and (E.24), I assume that:

- HIV-positive adults make relatively weak offers to employment activities. To capture this assumption, *GAMA* is set at lower values for HIV-positive people than for HIV-negative people.
- Unemployed people offer relatively weakly to employment activities. To capture this assumption, *GAMA* is set at lower values for the unemployment categories than employment categories.
- Those who are long-term unemployed have a lower probability of filling vacancies and make relatively weak offers to employment activities.

Short-term unemployed adults offer relatively stronger to employment activities than long-term unemployed adults, but still weaker than employed adults. *GAMA* for long-term unemployed is set at lower values than for short-term unemployed.

- I assume that all HIV-negative new entrants to the labour force offer strongly to employment activities. *GAMA* is set equal to 1. HIV-positive new entrants offer weakly to employment activities. *GAMA* is set at values less than 1.
- Finally, nobody with a Stage 4 status offers to employment activities. Instead, they offer to the PDL activity. *GAMA* is therefore set at zero.

The values for *GAMA* are summarised in Table 3.

Table 3. Strength of offers from unemployed and new entrant categories to employment activities

	HIV negative	***Stage 1***	***Stage 2***	***Stage 3***	***Stage 4***
New entrant	1.0	0.7	0.7	0.6	0
Short-term unemployed	0.8	0.5	0.5	0.4	0
Long-term unemployed	0.5	0.2	0.2	0.2	0

4.1.5. Planned flows from unemployment and new entrant categories c to unemployment activity u (Figure 5, Area 5)

Equation (E.24) determines the flow of people in unemployment and new entrant categories c, to unemployment activity u.

$$OFFER_{(c,a,g,r,h,u)} = \left(1-GAMA_{(c,h)}\right) * CAT_{(c,a,g,r,h)} * DUMMY_{(c,h,u)} \qquad \text{(E6.24)}$$

$c \in UN$; $a \in AGE$; $g \in GEN$; $r \in RACE$ and $h \in H3$.

where

- $OFFER_{(c,a,g,r,h,u)}$ is the number of offers from both the unemployment and new entrant category c to the unemployment activity u, for all age a, gender g, race r and HIV stage h;
- $\left(1-GAMA_{(c,h)}\right)$ captures the strength of offers from unemployment and new entrant categories c to an unemployment activity u, given an adult's HIV stage h;
- $CAT_{(c,a,g,r,h)}$ is the number of people in the unemployment and new

entrant category c, by gender g and race r, who are allocated to age a and HIV stage h at the start of the year; and

- $DUMMY_{(c,h,u)}$ takes a value of 1 or 0. If $DUMMY$ is set to 1, it allows for the calculation of the flows from unemployment and new entrant categories c to the unemployment activity u. The values for $DUMMY$ are summarised in Table 4.

I further assume that:

$$OFFER_{(c,a,g,r,Stage4,u)} = 0 \qquad \text{(E.25)}$$

$c \in UN$; $a \in AGE$; $g \in GEN$ and $r \in RACE$.

Equation (E.25) states that offers from new entrant and unemployment category c to unemployment activity u, for all age a, gender g, race r with a Stage 4 HIV status, are zero.

Table 4. Values assigned to *DUMMY* to regulate flows from unemployment and new entrant categories to unemployment activities

		HIV stage	Activity	
			S	**L**
Category	**N**	HIV negative	1	0
		Stage 1	1	0
		Stage 2	1	0
		Stage 3	1	0
		Stage 4	0	0
	S	HIV negative	0	1
		Stage 1	0	1
		Stage 2	0	1
		Stage 3	0	1
		Stage 4	0	0
	L	HIV negative	0	1
		Stage 1	0	1
		Stage 2	0	1
		Stage 3	0	1
		Stage 4	0	0

I now focus on the interpretation of the $DUMMY$ parameter. As mentioned before, $DUMMY$ takes a value of 0 or 1. The values assigned to $DUMMY$ are summarised in Table 5. The following assumptions are modelled via the inclusion of the $DUMMY_{(c,h,u)}$ variable:

- Only adults with an HIV stage of HIV negative, Stages 1, 2 or 3 can move from category c to unemployment activity u. For Stage 4,

$DUMMY$ is 0, which means that this flow is not allowed. The movement of people in Stage 4 is determined via (E.26).

- New entrants only move to short-term unemployment activity S and not long-term unemployment activity L.
- People grouped in the short-term unemployment category S only offer to the long-term unemployment activity L.
- People in the long-term unemployment category L only offer to the long-term unemployment activity L.

4.1.6. Planned flows from category c to the PDL activity (Figure 5; Areas 6 and 7)

Equation (E.26) determines the number of people from all categories offering to the PDL activity. There are two flows into the PDL activity. The first flow is determined via the first variable on the RHS of (E.26). This variable shows the number of people in the PDL category at the start of the year. As mentioned in Section 5.1.4 and footnote 7, we assume that people in the PDL category are not part of the labour force and do not move into any employment or unemployment activity. Once a person is PDL they remain PDL, that is, they simply move (offer) from the PDL category to the PDL activity. This is area (7) in Figure 5. The second variable on the RHS of (E.26) is the number of people, in any employment, unemployment or new entrant category o, with a Stage 4 HIV stage. We assume that adults in any category, of age a, gender g and race r and a Stage 4 HIV stage, are too ill to offer to employment or unemployment activities (see Equations (E.17), (E.19), (E.21), (E.23) and (E.25)). Instead, they offer to the PDL activity. This is area (6) in Figure 5.

$$OFFER_{(c,a,g,r,h,PDL)} = CAT_{(PDL,a,g,r,hh)} + CAT_{(m,a,g,r,Stage4)} \qquad \text{(E.26)}$$

$c \in EUNP$;[17] $a \in AGE$; $g \in GEN$; $r \in RACE$; $h \in STG$;[18] $m \in EUN$ [19] and $hh \in H3$.

where

- $OFFER_{(c,a,g,r,h,PDL)}$ is the number of offers from all employment, unemployment, new entrant and PDL categories to the PDL activity, for all age a, gender g, race r and HIV stage h;
- $CAT_{(PDL,a,g,r,h)}$ is the number of adults in the PDL category by gender g and race r, who are allocated to age a and HIV stage h at the start of the year. Notice that HIV stage h includes all HIV stages (HIV negative, Stages 1, 2, 3 and 4); and

[17] $EUNP$ is a set of all employment, unemployment, new entrant and PDL categories.
[18] STG is a set containing five elements: HIV negative, Stage 1, Stage 2, Stage 3 and Stage 4.
[19] EUN is a set of all employment, unemployment and new entrant categories.

- $CAT_{(c,a,g,r,Stage4)}$ is the number of adults in all employment, unemployment and new entrant categories c by gender g, race r and HIV Stage 4, who are allocated to age a at the start of the year.

This concludes the stepwise construction of the *OFFER* matrix.

4.1.7. Summary of the planned flows matrix

This matrix shows the planned labour offers from all categories to all activities. A summary of the *OFFER* matrix is presented in Table 5. The features and assumptions incorporated into this matrix are clearly identifiable. Notice the following:

- I restrict the non-diagonal offers from employment categories c to employment activities a. This reflects the assumption that people can only offer to an occupation which is compatible with their skills. Notice that people with skill levels of 1 and 2 are employed as Legislators, Professionals and Technicians (rows 1–3). If they choose to offer to a different occupation, they can only offer to occupations with the same skill level, i.e. Legislators, Professionals and Technicians (columns 1–3). Similarly, occupations in rows 4–11 can only offer to occupations in columns 4–11. The non-diagonal flows are derived via (E.16).
- Rows 1–11 and columns 1–11 form a sub-matrix that is strongly diagonal. This reflects the assumption that most people prefer be employed in the same occupation as in the previous year. This reflects the diagonal flows that are derived via (E.20).
- Rows 12–14 reflect the flows from the unemployment and new entrant categories to employment activities (columns 1–11). These flows are derived via (E.22).
- Column 12 shows the number of people who voluntarily move from an employment (rows 1–11) and new entrant (row 14) category to short-term unemployment. Notice that there are no flows from employment and new entrant categories to the long-term unemployment activity. This flow is determined via (E.18).
- Column 13 shows the number of people who offered to the long-term unemployed activity. Notice that the only flows to this unemployment activity are from the short and long-term unemployment categories. This flow is determined via (E.24).

- Column 14 shows the number of people from all categories offering to the PDL activity. For rows 1–13 this only refers to those who are in the employment and unemployment category with a Stage 4 HIV status. We assume that they are too ill to remain in the labour force and therefore offer to the PDL activity. Row 15, column 14 shows the number of people who are in the PDL category and remain PDL. This flow is derived via (E.26).
- The row totals (column 15) shows the number of people in each category at the start of the base year.
- The column total (row 16) shows the number of people in each activity. These totals are not consistent with published LFS data because they reflect adults' preference (planned offers) to be employed or unemployed.

Table 5. The *OFFER* matrix summarising the number of people moving from category c to activity a in the base year *

			Activities that people would like to perform during the year													
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
			Leg	Profes	Tech	Clerk	Service	SkildAgr	Craft	PlantMach	Elem	Dom	Unspec	S	L	PDL	Total
Categories at the start of the year	(1)	Leg	677,331	941	1,772	0	0	0	0	0	0	0	0	21,435	0	2,640	704,119
	(2)	Profes	832	467,677	1,128	0	0	0	0	0	0	0	0	14,801	0	1,728	486,166
	(3)	Tech	2,465	1,775	1,124,033	0	0	0	0	0	0	0	0	35,723	0	5,144	1,169,139
	(4)	Clerk	0	0	0	1,036,919	537	246	487	392	815	358	14	33,297	0	6,869	1,079,936
	(5)	Service	0	0	0	547	1,161,525	385	839	684	1,258	386	21	37,330	0	7,779	1,210,752
	(6)	SkildAgr	0	0	0	233	359	653,982	476	398	683	177	12	21,024	0	4,576	681,920
	(7)	Craft	0	0	0	510	863	525	1,355,807	1,173	1,598	211	30	44,090	0	13,135	1,417,942
	(8)	Plant-Mach	0	0	0	394	676	421	1,126	1,071,427	1,313	136	24	34,854	0	10,614	1,120,985
	(9)	Elem	0	0	0	963	1,459	850	1,802	1,542	2,131,441	1,025	44	69,372	0	23,571	2,232,069
	(10)	Dom	0	0	0	346	367	180	195	131	839	786,573	8	25,570	0	8,070	822,278
	(11)	Unspec	0	0	0	12	18	11	25	21	33	7	38,081	1,239	0	431	39,878
	(12)	S	68,990	64,194	190,796	203,058	259,506	155,465	225,535	175,923	504,600	232,574	6,841	0	669,031	22,444	2,778,957
	(13)	L	76,865	71,662	213,232	225,741	289,096	173,170	250,731	195,882	561,255	258,848	7,590	0	2,795,672	52,369	5,172,114
	(14)	N	15,056	14,436	36,184	69,956	95,359	66,303	68,127	42,500	173,029	58,912	2,710	17,657	0	529,962	1,190,193
	(15)	PDL	0	0	0	0	0	0	0	0	0	0	0	0	0	8,405,552	8,405,552
	(16)	Total	841,538	620,684	1,567,145	1,538,681	1,809,764	1,051,537	1,905,151	1,490,072	3,376,866	1,339,207	55,375	356,392	3,464,703	9,094,885	28,512,000

*This matrix is summed across age, gender, race and HIV status and shows the total planned flows from all categories to all activities.

4.2. The actual flows matrix

In determining everyone's activity during the base year, we specify a matrix called $H_{(c,a,g,r,h,o)}$. The H matrix captures the flow from each category c to each activity o, for all age a, gender g, race r and HIV stage h. Notice that $H_{(c,a,g,r,h,o)}$ includes flows from all categories (employment, unemployment, new entrant and PDL) to all activities (employment, unemployment and PDL). This matrix is also constructed in several steps. Allowed flows are represented by a number in Figure 4 and flows indicated by zero are not allowed.

The difference between the *OFFER* and *H* matrix is that the former shows what activities people *would like* to perform during year t, whereas the latter shows who are *successful in securing* employment and what happens to those who do not find employment, that is, actual flows.

4.2.1. Occupation-specific vacancies

To determine who secures employment we need to derive the number of occupation-specific vacancies. Vacancies are defined as the number of jobs (employment activities) less incumbents. The base year values for occupation-specific vacancies are calculated via an iterative process. As a first approximation, vacancies are calculated as those who are fired:

$$VAC\_1_{(o)} = ACT_{(o)} * SACKFRAC \text{ where } o \in OCC \qquad \text{(E.27)}$$

where
- $VAC\_1_{(o)}$ is the first approximation of occupation-specific vacancies;
- $ACT_{(o)}$ is the occupation-specific activities summed across age a, gender g, race r and HIV status h; and
- $SACKFRAC$ is the proportion of people who are fired from each occupation. This parameter is set at 3 per cent.

As a second approximation, vacancies are calculated as the number of people who are employed in an occupation during the year less incumbents:

$$VAC\_2_{(o)} = ACT_{(o)} - Incumbents_{(o)} \text{ where } o \in OCC \qquad \text{(E.28)}$$

with

$$Incumbents_{(o)} = CAT_{(o)} - EMPtoUNEMP_{(o)} - \sum_{m \in OCC} SH\_QI_{(o,m)} * VAC\_1_{(m)} \quad \text{(E.29)}$$

$o \in OCC$

where
- $VAC\_2_{(o)}$ is the second approximation of occupation-specific vacancies;
- $ACT_{(o)}$ is the number of people in each employment activity, summed across age a, gender g, race r and HIV status h;
- $CAT_{(o)}$ is the number of people in each employment category at the beginning of the year, summed across age a, gender g, race r and HIV status h;
- $EMPtoUNEMP_{(o)}$ is the number of people in employment category o moving to short-term unemployment activity S; and
- $\sum_{m \in OCC} SH\_QI_{(o,m)} * VAC\_1_{(m)}$ is summed across all occupations m, the proportion of people moving from occupation o to occupation m. This term refers to the non-diagonal flows.[20]

Occupation-specific vacancies are calculated via an iterative process. Equations (E.28) and (E.29) show that vacancies in occupation o are dependent on the number of employment activities available after the incumbents are taken into account. Incumbent refers to the number of employed adults who return in year t to the same occupation they performed in year $t-1$. Hence, it is the number of people in employment category o less those who voluntarily move to short-term unemployment and less those who find employment in activity m. The number of people who find employment in activity m is dependent on the number of vacancies in activity m. After a number of iterations the occupation-specific values converge. The converged values are then the final value for occupation-specific vacancies.

[20] The proportion of people moving from occupation m to o is defined as $SH\_QI_{(m,a,g,r,h,o)} = \frac{OFFER_{(m,a,g,r,h,o)}}{BOT2_{(o)}}$ with $OFFER_{(m,a,g,r,h,o)}$ referring to all offers from occupation m to occupation o for age a, gender g, race r and HIV stage h. $BOT2_{(o)}$ is all the non-incumbent offers to occupation o. $SH\_QI_{(o,m)}$ is therefore the proportion of people moving from occupation m to occupation o, summed across age a, gender g, race r and HIV stage h. The larger the share of occupation m moving to occupation o, the larger the flow from these m to o.

4.2.2. Creating the actual flows matrix

4.2.2.1. Actual flows from unemployment and new entrant categories m to employment activities o (Figure 5; Area 4)

The actual flows from the unemployment and new entrant categories m to occupation o depend on the number of occupation-specific vacancies and the share of people in category m offering to occupation o. This flow is determined via (E.30):

$$H_{(m,a,g,r,h,o)} = SH\_QI_{(m,a,g,r,h,o)} * VAC_{(o)} \quad \text{(E.30)}$$

$m \in UN$; $a \in AGE$; $g \in GEN$; $r \in RACE$; $h \in H3$ and $o \in OCC$.

where
- $H_{(m,a,g,r,h,o)}$ is the number of people of age a, gender g, race r and HIV stage h, moving from unemployment and new entrant category m to occupation o;
- $SH\_QI_{(m,a,g,r,h,o)}$ determines the proportion of people moving from category m to occupation o. For further details see footnote 28; and
- $VAC_{(o)}$ is the number of occupation-specific vacancies calculated via (E.27) to (E.29).

I assume that:

$$H_{(m,a,g,r,Stage4,o)} = 0 \quad \text{(E.31)}$$

$m \in UN$; $a \in AGE$; $g \in GEN$ and $r \in RACE$.

Equation (E.31) states that the flow of people of age a, gender g, race r with a Stage 4 health status, moving from the unemployment and new entrant category m to occupation o, is zero. Instead they move into the PDL activity. This flow is determined via (E.38).

*4.2.2.2. Actual flows from employment category **o** to employment activity m, where $o = m$ and where $o \neq m$ (Figure 5; Areas 1 and 3)*

Equation (E.32) determines the diagonal and non-diagonal actual flows from employment category o to activity m. The first term in brackets on the RHS of (E.32) activates the determination of diagonal flows. The second term in brackets on the RHS of (E.32)

determines the diagonal flows from occupation o to m, where $o = m$. Diagonal flows are determined as the number of people in each employment category o less those who leave that category. People may leave occupation o and (1) move to a different occupation n or (2) move to short-term unemployment activity S. The final expression on the RHS of (E.32) determines the non-diagonal flows from occupation o to occupation m. The non-diagonal flows are determined by the share of occupation o in occupation m and the number of vacancies in occupation m. This equation calculating diagonal and non-diagonal flows is written as:

$$H_{(o,a,g,r,h,m)} = \left(1 - DUM6_{(o,m)}\right) * \left[CAT_{(o,a,g,r,h)} - \sum_{n \in OCC} SH\_QI_{(o,a,g,r,h,n)} * VAC_{(n)} - EMPtoS_{(o,a,g,r,h)} \right] + DUM6_{(o,m)} * SH\_QI_{(o,a,g,r,h,m)} * VAC_{(m)} \quad \text{(E.32)}$$

$o \in OCC$; $a \in AGE$; $g \in GEN$; $r \in RACE$; $h \in H3$; $m \in OCC$ and $n \in OCC$.

where

- $H_{(o,a,g,p,h,m)}$ is the flow of people of age a, gender g, race r and HIV stage h, moving from employment category o to employment activity m;
- $CAT_{(o,a,g,p,h)}$ is the number of people in each employment category o, gender g, race r and HIV stage h, allocated to age a at the start of the base year;
- $\sum_{n \in OCC} SH\_QI_{(o,a,g,p,h,n)} * VAC_{(n)}$ is the number of people of age a, gender g, race r and HIV stage h, moving from occupation o to occupation n;
- $EMPtoS_{(o,a,g,p,h)}$ is the number of people of age a, gender g, race r and HIV stage h, moving from occupation o to unemployment activity S;
- $DUM6$ is a parameter that takes a value of 1 or 0. With $DUM6$ set to 0 or 1, it activates and deactivates the relevant flows in (E6.32). If $DUM6$ is 0, the relevant flow determined in (E6.32) is the diagonal flow from occupation o to occupation m, where $o = m$; and
- with $DUM6$ set to 1, (E.32) determines the non-diagonal flow of people of age a, gender g, race r and HIV stage h, moving from occupation o to occupation m, i.e. $DUM6_{(o,m)} * SH\_QI_{(o,a,g,p,h,m)} * VAC_{(m)}$.

I assume that:

$$H_{(o,a,g,r,Stage4,m)} = 0 \qquad \text{(E.33)}$$

$o \in OCC$, $a \in AGE$, $g \in GEN$, $r \in RACE$ and $m \in OCC$.

Equation (E.33) states that the flow of people of age a, gender g, race r with a Stage 4 health status, moving from employment category o to employment activity m, is zero.

4.2.2.3. Actual flows from employment category $\boldsymbol{o}$ *to unemployment activity* S *(Figure 5; Area 2)*

Equation (E.34) determines the actual flows from employment category o to unemployment activity S:

$$H_{(o,a,g,p,h,S)} = OFFER_{(o,a,g,p,h,S)} + \mu * DUMSACK_{(m)} * CAT_{(o,a,g,p,h)} \qquad \text{(E.34)}$$

$o \in OCC$; $a \in AGE$; $g \in GEN$; $r \in RACE$ and $h \in H3$.

where

- $H_{(o,a,g,p,h,S)}$ is the flow of people of age a, gender g, race r and HIV stage h, moving from employment category o to short-term unemployment activity S;
- $OFFER_{(o,a,g,p,h,S)}$ is the voluntary offers from people of age a, gender g, race r and HIV stage h, moving from occupation o to short-term unemployment activity S. This flow is determined via (E.18);
- $\mu * DUMSACK_{(m)} * CAT_{(o,a,g,p,h)}$ determines the involuntary flows from employment category o, age a, gender g, race r and HIV stage h to short-term unemployment activity S. μ is the proportion of people fired from occupation o. $DUMSACK_{(m)}$ is a parameter and takes the value of 0 or 1. This term ensures that people move to the correct unemployment activity given their category. People in an employment or new entrant category move to short-term unemployment. People in a short-term or long-term unemployment move to the long-term unemployment activity.

I assume that:

$$H_{(o,a,g,p,Stage4,S)} = 0 \qquad \text{(E.35)}$$

$o \in OCC$; $a \in AGE$; $g \in GEN$ and $r \in RACE$.

4.2.2.4. Actual flows from unemployment category and new entrant category o *to unemployment activity* u*, where* $u = S$ *and* $u = L$
(Figure 6.5; Area 5)

Equation (E.30) determined the actual flows from unemployment and new entrant categories m to occupation o. Those who do not secure employment move to unemployment activity u. The actual flows from unemployment and new entrant categories m to unemployment activity u, are calculated via (E.36):

$$H_{(o,a,g,p,h,m)} = DUMSACK_{(m)} * \left[CAT_{(m,a,g,p,h)} - \sum_{oo \in OCC} SH\_QI_{(m,a,g,p,h,oo)} * VAC_{(oo)} \right] \qquad \text{(E.36)}$$

$o \in UN$;[21] $a \in AGE$; $g \in GEN$; $r \in RACE$; $h \in H3$ and $u \in UNEMP$.[22]

where
- $H_{(m,a,g,p,h,u)}$ is the flow of people of age a, gender g, race r and HIV stage h, moving from unemployment and new entrant category m to unemployment activity u;
- $DUMSACK_{(m)}$ regulates the flows from unemployment and new entrant category m to unemployment activity u. Those who were unemployed and failed to secure employment will move to the long-term unemployment activity. New entrants failing to secure employment move to short-term unemployment;
- $CAT_{(m,a,g,p,h)}$ is the number of people in each unemployment and new entrant category o by age a, gender g, race r and HIV stage h at the start of the base year; and
- $\sum_{oo \in OCC} SH\_QI_{(m,a,g,p,h,oo)} * VAC_{(oo)}$ determines the flow of people who were in an unemployment and new entrant category m, and secured employment in occupation oo. This flow is determined via (E6.30)

Again I assume that:

$$H_{(o,a,g,p,Stage4,u)} = 0 \qquad \text{(E.37)}$$

$o \in UN$; $a \in AGE$; $g \in GEN$; $r \in RACE$ and $u \in UNEMP$.

[21] UN is a set for all unemployment and new entrant categories.
[22] $UNEMP$ is a set containing two elements: short-term and long-term unemployment.

4.2.2.5. Actual flows from all categories o moving to the "permanently departed from the labour force" (PDL) activity
(Figure 5; Areas 5 and 6)

The number of people in the PDL activity is determined by two flows. The first flow is from people who are in the PDL category at the start of the year. As mentioned before, we assume that people in the PDL category are not part of the labour force and do not move into any employment or unemployment activity. Instead they move to the PDL activity as illustrated in Figure 5, area 7. This flow is determined by the first term on the RHS of Equation (E.38).

The second flow to the PDL activity is from people in any employment, unemployment or new entrant category m, with an HIV stage of Stage 4 (Figure 5, area 6). Equations (E.31), (E.33), (E.35) and (E.37) prohibit the flow from any EUN category to any employment and unemployment (EU) activity if the person has an HIV status of Stage 4. People in Stage 4 are too ill to participate in any productive activity, and thus move into the PDL activity. This flow is calculated via the second term on the RHS of Equation (E.38).

$$H_{(c,a,g,p,h,PDL)} = CAT_{(PDL,a,g,p,hh)} + CAT_{(m,a,g,p,Stage4)} \qquad \text{(E.38)}$$

$c \in EUNP$;[23] $a \in AGE$; $g \in GEN$; $r \in RACE$; $h \in STG$; $m \in EUN$ and $hh \in H3$.

where
- $CAT_{(PDL,a,g,p,hh)}$ is the number of people in the PDL category by gender g and race r, allocated to age a and HIV stage h at the start of the base year; and
- $CAT_{(m,a,g,p,Stage4)}$ is the number of people in category o, gender g, race r and HIV status of Stage 4, allocated to age a at the start of the base year.

This concludes the stepwise construction of the actual flows matrix (H).

4.2.3. Summary of the actual flows matrix

A summary of the actual flows matrix is presented in Table 6. The reader will notice that the structure of the H matrix is similar to the *OFFER* matrix. For example, the H matrix, like the *OFFER* matrix, is strongly diagonal. This is because the underlying assumptions between the matrices are the same. These assumptions for example include the regulation of the flows between categories and activities. Notice the following:

[23] $EUNP$ is a set of all employment, unemployment , new entrant and PDL categories.

- I restrict the non-diagonal offers from employment categories *c* to employment activities *a*. This reflects the assumption that people can only offer to an occupation, i.e. compatible with their skills. Notice that people with a skill level of 1 and 2 are employed as Legislators, Professionals and Technicians (rows 1–3). If they choose to offer to a different occupation, they can only offer to occupations with the same skill level, i.e. Legislators, Professionals and Technicians (columns 1–3). Similarly, occupations in rows 4–11 can only offer to occupations in columns 4–11. Actual non-diagonal flow of people to occupation *o* depends on the number of occupation-specific vacancies in *o*. Actual diagonal flows refers to incumbents.
- Rows 1–11 and columns 1–11 form a sub-matrix that is strongly diagonal. This reflects the assumption that most people prefer to continue to perform the same employment activity during year *t* as performed during year *t*-1. This reflects the diagonal flows that are derived via (E.32).
- Rows 12–14 show the actual flows from all unemployment and new entrant categories *m* to employment activities *o* (columns 1–11). These flows are dependent on the number of occupation-specific vacancies as well as the share of category *m* in activity *o*. This flow is derived via (E.30).
- Column 12 shows the number of people moving from an employment or new entrant category to short-term unemployment S. These flows include both voluntary and involuntary flow to S. This flow is determined via (E.34).
- Column 13 shows the actual flow of people from all unemployment categories to the long-term unemployed activity. Notice that the only flows to the long-term unemployment activity are from those who were allocated to the short and long-term unemployment categories and did not secure employment in occupation *o*.
- Column 14 shows the number of people from all categories moving to the PDL activity. For all employment, unemployment and new entrant categories (rows 1–14) this flow reflects those who are in Stage 4. This flow is determined via the second term on the RHS of (E6.38). Row 15, column 14 shows the number of people who are allocated to the PDL category and remain PDL. This flow is derived via the first term on the RHS of (E.38).

- The row totals (column 15) indicate the number of people in each category at the start of the base year.
- Column totals (row 16) indicate the number of people in each activity undertaken during the base year. These totals are consistent with the published LFS data (Statistics South Africa, 2009a).

Table 6. The actual flows matrix summarising the number of people moving from category c to activity a in the base year*

			Activities that people would like to perform during the year														
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
			Leg	Profes	Tech	Clerk	Service	SkildAgr	Craft	PlantMach	Elem	Dom	Unspec	S	L	PDL	Total
Categories at the start of the year	(1)	Leg	658,244	299	457	0	0	0	0	0	0	0	0	42,479	0	2,640	704,119
	(2)	Profes	374	454,439	291	0	0	0	0	0	0	0	0	29,334	0	1,728	486,166
	(3)	Tech	1,107	563	1,091,682	0	0	0	0	0	0	0	0	70,643	0	5,144	1,169,139
	(4)	Clerk	0	0	0	1,007,003	96	44	125	107	147	52	3	65,489	0	6,869	1,079,936
	(5)	Service	0	0	0	111	1,128,685	68	216	187	226	56	5	73,419	0	7,779	1,210,752
	(6)	SkildAgr	0	0	0	47	64	635,505	123	109	123	26	3	41,344	0	4,576	681,920
	(7)	Craft	0	0	0	104	155	93	1,317,575	321	287	31	7	86,234	0	13,135	1,417,942
	(8)	Plant-Mach	0	0	0	80	121	75	290	1,041,379	236	20	5	68,165	0	10,614	1,120,985
	(9)	Elem	0	0	0	196	262	151	464	422	2,071,218	149	10	135,627	0	23,571	2,232,069
	(10)	Dom	0	0	0	70	66	32	50	36	151	763,805	2	49,996	0	8,070	822,278
	(11)	Unspec	0	0	0	3	3	2	7	6	6	1	36,998	2,423	0	431	39,878
	(12)	S	30,987	20,373	49,223	41,277	46,565	27,567	58,062	48,167	90,668	33,751	1,583	0	2,308,290	22,444	2,778,957
	(13)	L	34,524	22,744	55,011	45,888	51,874	30,706	64,549	53,631	100,848	37,564	1,756	0	4,620,648	52,369	5,172,114
	(14)	N	6,763	4,582	9,335	14,220	17,111	11,757	17,539	11,636	31,090	8,549	627	527,021	0	529,962	1,190,193
	(15)	PDL	0	0	0	0	0	0	0	0	0	0	0	0	0	8,405,552	8,405,552
	(16)	Total	731,999	503,000	1,205,999	1,109,000	1,245,003	705,999	1,459,000	1,156,002	2,294,998	844,004	40,999	1,192,175	6,928,938	9,094,885	28,512,000

*This matrix is summed across age, gender, race and HIV status. It shows the total actual flows from all categories to all activities.

5. TRANSITION MATRIX

5.1. Introduction

In this section, I describe the development of the transition matrix, $T_{(o,aa,g,r,hh,a,h)}$. The transition matrix appears in the equation that calculates the number of people in each category at the start of the year. The transition matrix is defined as the product of two matrices via (E.39).

$$T_{(o,aa,g,r,hh,a,h)} = a\_transit_{(aa,g,r,h,a)} * h\_transit_{(o,aa,g,r,hh,h)} \quad \text{(E.39)}$$

$o \in EUNP$; $a \in AGE$; $g \in GEN$; $r \in RACE$ and $h \in STG$.

where
- $T_{(o,aa,g,r,hh,a,h)}$ is the proportion of people in category o, gender g and race r moving from age aa to age a and changing their HIV stage from hh to h;
- $a\_transit_{(aa,g,r,h,a)}$ is the probability of a person of gender g, race r and HIV stage h, moving from age aa to age a in year t; and
- $h\_transit_{(o,aa,g,r,hh,h)}$ is the probability of a person in activity o of age aa, gender g and race r, changing their HIV stage from hh to h in year t.

The transition matrix $\left(T_{(o,aa,g,r,hh,a,h)}\right)$ allows people, given a set of o,g,r characteristics, to do one of the following:

- remain in age aa and remain in HIV status hh;
- move from age aa to age a and change HIV status from hh to h;
- remain in age aa and change HIV status from hh to h;
- move from age aa to age a and remain in HIV status hh; or
- leave the labour force due to death or retirement.

For example, suppose that during year $t-1$ we have a person with the following o,a,g,r,h characteristics: an African female, age 15–24, working as a Clerk with an HIV-negative status, that is, $ACT\_L_{(Legislator,15-24,F,A,HIVn)}$. At the end of year $t-1$ she is allocated to a category via (E.14). Given the options described above, she can move to one of the following categories:

- African, Female, Legislator, aged 15–24, HIV negative, $CAT_{(Legislator,15-24,F,A,HIVn)}$;
- African, Female, Legislator, aged 25–34, HIV positive (Stage1), $CAT_{(Legislator,25-34,F,A,Stage1)}$;
- African, Female, Legislator, aged 15–24, HIV positive (Stage1), $CAT_{(Legislator,15-24,F,A,Stage1)}$;
- African, Female, Legislator, aged 25_34, HIV negative, $CAT_{(Legislator,25-34,F,A,HIVn)}$; or
- leave the labour force due to death or retirement.

Notice that the transition matrix does not allow the labour-market activity o, gender g or race r to change. The remainder of this paper describes the development of the transition matrix by describing the setting of the $h\_transit$ and $a\_transit$ matrices. Once these matrices are developed, the final transition matrix is calculated.

5.2. Health transition $\left(h\_transit_{(o,aa,g,r,hh,h)}\right)$

To determine the health transition rates, I create a matrix called $h\_transit_{(o,aa,g,r,hh,h)}$. This matrix shows the probability of a person in activity o of age aa, gender g and race r, changing their HIV stage from hh to h in year t. The following health transitions are possible:

- remain in the same HIV stage (stay);
- move to the next HIV stage (move); or
- death (die).

The sum of these transitions equals 1. In the health framework the following transitions are allowed:

- Move from HIV negative to Stage 1. This transition is important because it determines the number of new HIV infections.
- Move from Stage 1 to Stage 2. This transition implies that a person is already HIV positive. If they change their HIV stage, the only option is to move to Stage 2.
- Move from Stage 2 to Stage 3. As with Stage 2, this person is HIV

positive. If they change their HIV stage, the only option is to move to Stage 3.

- Move from Stage 3 to Stage 4. This is the final HIV stage a person may move to. Once a person is in Stage 4 they remain in this stage until they die.

The alternative of moving to a different HIV stage is to:

- remain HIV negative;
- remain in Stage 1;
- remain in Stage 2;
- remain in Stage 3;
- remain in Stage 4; or
- die.

All other transitions are not allowed. For example, a person cannot move from Stage 3 to Stage 2 or from Stage 2 to HIV negative. I begin the description of the health transition matrices by first describing the age and gender-specific death rates and then the rates of moving from one stage to the next.

5.2.1. Death rate $\left(1-h\_transit_{(o,aa,g,r,hh,h)}\right)$

HIV/AIDS impacts on the demographic composition via adult mortality. In our framework we need to determine death rates specific to each labour-market activity o by age a, gender g, race r and HIV status h.[24] It is difficult to determine AIDS and non-AIDS mortality rates for South Africa because:

- Not all deaths are registered. It is estimated that in 1996 only 67 per cent of all deaths were registered. This improved in 1997, where an estimated 80 per cent of male deaths and 78 per cent of female deaths were registered (Statistics South Africa, 2006d: 2–3).
- Death registration forms may be incomplete. Estimates of the completeness of adult death registration suggest that the completeness of forms increased to approximately 90 per cent of adult deaths by 2000. To account for incomplete death registration forms, Statistics South Africa applies weights by age and gender for each year and adjusts the data. The adjusted estimates provide better mortality

[24] In terms of HIV status, we refer to AIDS and non-AIDS death rates.

estimates compared to the unadjusted data on deaths (Statistics South Africa, 2006d: xv).

- Finally, death statistics do not provide an accurate picture of AIDS-related deaths in South Africa. This is due to misclassification of HIV/AIDS as a cause of mortality. A study by Groenewald *et al.* (2004) found that a large proportion of AIDS deaths are caused by AIDS-related conditions without reference to HIV as the cause of death (Groenewald *et al.*, 2005).[25] This study identifies nine causes of death that clearly display the age pattern associated with HIV prevalence. Respiratory conditions accounted for the majority of the excess deaths related to AIDS. Similar studies found that tuberculosis is a common cause of death in HIV-positive people.

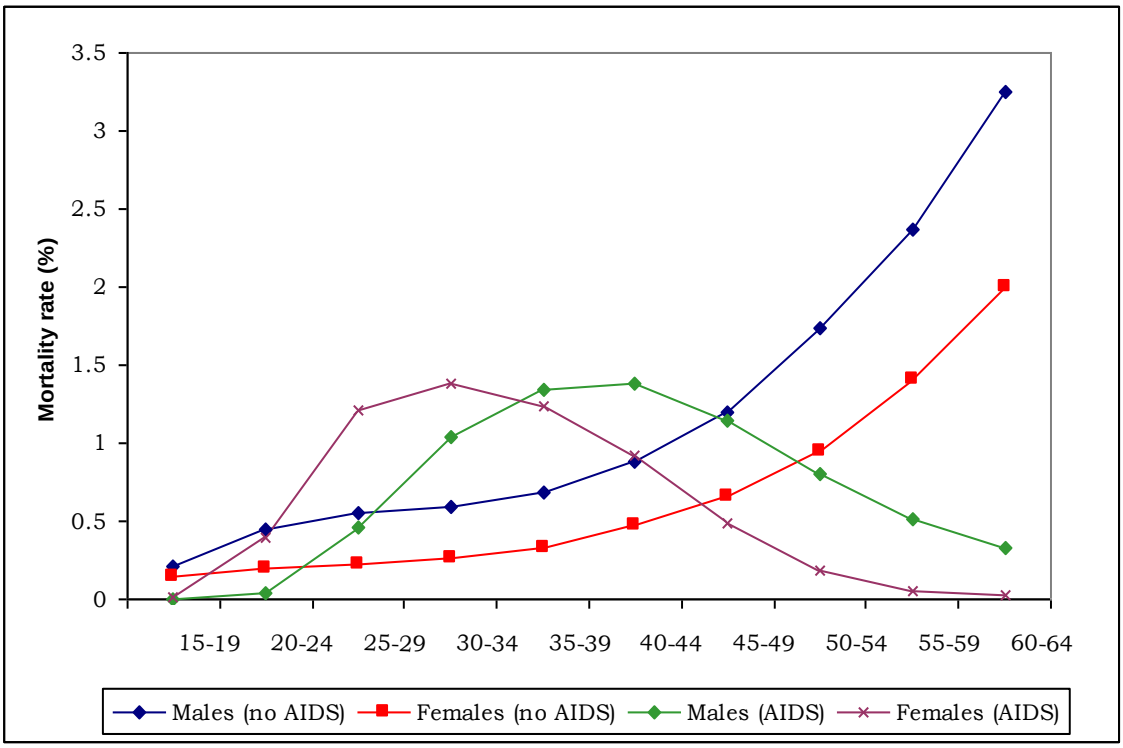
As a first approximation we adopt the mortality data from the ASSA model. The age, gender and HIV status-specific death rates are illustrated in Figure 6. The first set of age and gender-specific death rates are non-AIDS death rates, which shows a smooth increase in the death rate across age, with older age groups showing the highest death rates and the younger age groups the lowest death rates. The second set of age and gender-specific death rates show the AIDS death rates. A distinct age and gender-specific pattern is observed. For both genders, younger age groups have higher death rates compared to the non-AIDS death rates. The high death rates are followed by a rapid decline in the AIDS death rates for older age groups. AIDS deaths peak at 30–34 for females and 35–39 for males. Another distinctive feature is that the female death rate leads by approximately five years.

As mentioned before, we require death rates for each labour-market activity o by age a, gender g, race r and HIV stage h. In this thesis I assume that:

- All AIDS-related death occurs in Stage 4.
- I apply the same death rate for HIV negative, Stages 1, 2 and 3. This assumption can be relaxed in future if we want to show that the relative risk of dying for an HIV-positive person is higher than that of an HIV-negative person, i.e. we introduce unique death rates for Stages 1, 2 and 3.
- The age, gender, race and HIV-specific death rates are uniform across all labour-market activities.

[25] Causes of death are coded based on the International Classification of Diseases (ICD-10) developed by the World Health Organisation (WHO). The ICD-10 contains approximately 8,000 categories of causes of death (Statistics South Africa, 2006e: 13).

Figure 6. AIDS and non-AIDS death rates (2002) (percentage)



Source: Own calculations based on ASSA2003 results.

The following procedure is used to determine the number of deaths. First, I adopt the AIDS and no-AIDS death rates illustrated in Figure 6.6 and multiply them with the number of people in each labour-market activity o by age a, gender g, race r and HIV status h. The number of deaths by labour-market activity, age, gender, race and HIV stage are calculated as:

$$Deaths_{(o,a,g,r,h)} = ACT\_L_{(o,a,g,r,h)} * DRT_{(a,g,hh)} \qquad \text{(E.40)}$$

$o \in EUP$;[26] $a \in AGE$; $g \in GEN$; $r \in RACE$; $h \in STG$ and $hh \in H2$.[27]

- $Deaths_{(o,a,g,r,h)}$ is the number of deaths by labour-market activity o, age a, gender g, race r and HIV stage h;
- $ACT\_L_{(o,a,g,r,h)}$ is the number of people in each labour-market activity o by age a, gender g, race r and HIV stage h in 2001; and
- $DRT_{(a,g,hh)}$ is the AIDS and non-AIDS death rates illustrated in Figure 6.

[26] EUP is a set of all employment, unemployment and PDL activities.
[27] $H2$ is a set containing two elements: HIV negative and HIV positive.

The number of deaths calculated via (E.40) yields too few deaths. I therefore scale the deaths upwards by assuming that 1 per cent of the HIV-negative population and approximately 4.5 per cent of the total HIV-positive population dies. Table 7 summarises the final number of deaths by age, gender and race. The results suggest that the total number of deaths is approximately 440,000, or 1.54 per cent of the working-age population. Non-AIDS deaths amount to 240,000 and contribute 54 per cent to the total number of deaths. AIDS-related deaths are approximately 200,000 deaths and contribute 45 per cent to the total number of deaths. The average death rate for the HIV-negative population is 1 per cent and 4.5 per cent for the HIV-positive population. [28]

Table 7. Number of deaths by age and HIV stage

Age	*HIV negative*	*Stage 1*	*Stage 2*	*Stage 3*	*Stage 4*	*Total*
15–24	32,383	1,024	751	941	58,136	93,235
25–34	34,957	3,917	2,845	3,573	57,434	102,727
35–44	39,324	3,298	2,462	3,091	33,624	81,799
45–54	56,629	2,244	1,706	2,144	15,193	77,916
55–65	76,707	1,129	841	1,057	5,613	85,347
Total	240,000	11,613	8,605	10,807	170,000	441,024

Own calculations

5.2.2. Moving from HIV negative to Stage 1 $\left(h\_transit_{(o,aa,g,r,HIVn,Stage1)}\right)$

In this section, I describe the determination of the transition rates from HIV negative to Stage 1 for all labour-market activities o, age a, gender g and race r. This transition is important because it determines the number of new HIV infections, which is a good indicator of assessing the impact of prevention programs as it better reflects the underlying transmission dynamics of the epidemic in South Africa (Shisana, 2005: 47; 2010: 64). In setting this transition we take the follow into account:

- incidence rates are particularly high for young adults for both genders;
- in terms of race, incidence rates are higher for Africans than for the "Other" population group;
- in terms of labour-market activity, we assume that incidence rates are the highest for those who are "permanently departed from the labour force" and short-term and long-term unemployed;
- in terms of employment activities, we assume that incidence rates are higher for unskilled occupations than skilled occupations.

[28] For non-AIDS death the average death rate is calculated as $\frac{240,000}{24,083,136} * 100 = 1.0$.

For AIDS deaths the average death rate is calculated as $\frac{201,025}{4,428,864} * 100 = 4.5$. Take note that these deaths are summed across Stages 1, 2, 3 and 4.

As a first approximation we adopt data from the ASSA model to determine the age, gender and race-specific incidence rates. These incidence rates are summarised in Table 8.

Table 8. Incidence rates by age, gender and race

Age	*African*		*Other*	
	Female	**Male**	**Female**	**Male**
15–24	5.09	3.09	1.57	0.64
25–34	2.09	3.31	1.54	1.01
35–44	1.22	1.67	1.15	0.71
45–54	0.73	1.49	0.49	0.43
55–65	0.22	1.27	0.19	0.19

Source: Own calculations based on ASSA2003 data.

The reader will notice that in (E.41) we assume that the age, gender and race-specific incidence rates are uniform across all labour-market activities. However, I believe that incidence rates do differ across labour-market activities. I therefore scale the incidence rates in Equations (E.42) and (E.43), so that skilled occupations have the lowest incidence rates and PDL the highest, while ensuring that the initial number of new HIV infections remains the same. The initial number of new HIV infections is calculated as:

$$HIVNEW_{(o,a,g,r)}^{(initial)} = ACT\_L_{(o,a,g,rHIVn)} * INCIDR_{(a,g,r)} \qquad \text{(E.41)}$$

$o \in EUP$;[29] $a \in AGE$; $g \in GEN$ and $r \in RACE$.

- $HIVNEW_{(o,a,g,r)}^{(initial)}$ is the number of new HIV infections by labour-market activity o, age a, gender g and race r;
- $ACT\_L_{(o,a,g,r,HIVn)}$ is the number of people in each labour-market activity o by age a, gender g and race r who are HIV negative; and
- $INCIDR_{(a,g,r)}$ is the initial incidence reported in Table 6.9.

Table 9 presents the total number of new HIV infections by age, gender and race. The total number of new HIV infections is 518,000 people. This implies an overall incidence rate of approximately 2 per cent[30] in the base year.

[29] EUP is a set of all employment, unemployment and PDL activities.

[30] The average incidence rate is determined as the number of new HIV infections given the aggregate number of vulnerable people, that is, $\left(\frac{HIV\_NEW}{HIVnegative} * 100\right)$. The average incidence rate for 2002 is $\frac{517,000}{24,231,116} * 100 = 2.13$ per cent.

Table 9. Number of new HIV cases by age, gender and race

Age	*African*		*Other*		*Total*
	Female	**Male**	**Female**	**Male**	
15–24	146,606	103,559	11,266	4,867	266,297
25–34	52,768	74,750	10,993	7,119	145,630
35–44	22,225	27,523	8,261	5,119	63,127
45–54	10,402	14,718	3,259	2,548	30,927
55–65	2,619	8,223	1,019	899	12,759
Total	234,620	228,773	34,797	20,551	518,741

Table 9 shows that the majority of new HIV infections occur in the African population, with this group accounting for 89 per cent of new HIV infections. The Other population group contributes 11 per cent to the total number of new HIV infections. Females and males contribute 52 and 48 per cent respectively to the total number of new HIV infections. Our matrix shows that HIV incidence rates are the highest for the youngest age groups where people may still be studying or newly employed. The 15–24 age group contributes the most to new HIV infections and 55–65 the least.

Equation (E.41) assumes that the age, gender and race-specific incidence rate is the same for all labour-market activities o. However, I want to show that some occupations, due to their environment, structural factors or the type of labour required, are more vulnerable in contracting the HIV virus. I therefore impose ad hoc higher values for labour-market activities that are more vulnerable. To impose this assumption we create a coefficient called $F_{(o,a,g,r)}$. I assign values for each labour-market activity o, capturing the vulnerability of labour-market activities. These values are uniform across age, gender and race. For skilled occupations we assign a value of 2, semi-skilled a value of 2.5, unskilled occupations a value of 3 and for unemployment and PDL activities a value of 3.5. I then calculate scaling factors to ensure that the number of new HIV infections summed across o is equal to the number of new HIV cases set out in Table 6.10. These scaling factors are calculated via (E6.42):

$$FF_{(a,g,r)} = \frac{NEWHIV\_O_{(a,g,r)}}{\sum_{o \in EUNP} NEWHIV_{(o,a,g,r)}^{(initial)} * F_{(o,a,g,r)}} \qquad \text{(E.42)}$$

$o \in EUP$; $a \in AGE$; $g \in GEN$ and $r \in RACE$.

- $FF_{(a,g,r)}$ is the age, gender and race-specific scaling factor to ensure that the number of new HIV cases summed across the o dimension adds to the numbers presented in Table 6.10;
- $HIVNEW_{(o,a,g,r)}^{(initial)}$ is the number of new HIV cases in each labour-market

activity o by age a, gender g and race r. $NEWHIV\_O_{(a,g,r)}$ is summed across labour-market activity o; and

- $F_{(o,a,g,r)}$ assigns a value for each labour-market dimension o to the capture their vulnerability.

The final scaled number of new HIV infections by labour-market activity o, age a, gender g and race r is calculated via (E.43):

$$HIVNEW^{(final)}_{(o,a,g,r)} = HIVNEW^{(initial)}_{(o,a,g,r)} * F_{(o,a,g,r)} * FF_{(a,g,r)} \quad \text{(E.43)}$$

$o \in EUP$; $a \in AGE$; $g \in GEN$ and $r \in RACE$.

The final incidence rates for each labour-market activity o by age a, gender g and race r is calculated as:

$$h\_transit_{(o,a,g,r,HIVn,Stage1)} = \frac{HIVNEW^{(final)}_{(o,a,g,r)}}{WAP_{(o,a,g,r,HIVn)}} \quad \text{(E.44)}$$

$o \in EUP$; $a \in AGE$; $g \in GEN$ and $r \in RACE$.

- $h\_transit_{(o,a,g,r,HIVn,Stage1)}$ is the labour-market activity o, age a, gender g, race r-specific health transition from HIV negative to Stage 1;
- $HIVNEW^{(final)}_{(o,a,g,r)}$ is the number of new HIV cases by labour-market activity o, age a, gender g and race r; and
- $WAP_{(o,a,g,r,HIVn)}$ is the number of HIV-negative people in labour-market activity o by age a, gender g and race r.

The survival probability of an HIV-negative adult is calculated as:

$$h\_transit_{(o,a,g,r,HIVn,HIVn)} = 1 - h\_transit_{(o,a,g,r,HIVn,Stage1)} - death\ rate_{(a,g,r,HIVn)} \quad \text{(E.45)}$$

$o \in EUP$; $a \in AGE$; $g \in GEN$ and $r \in RACE$.

- $h\_transit_{(o,a,g,r,HIVn,HIVn)}$ is the labour-market activity o, age a, gender g

and race r-specific health transition from HIV negative to HIV negative, i.e. the probability of remaining HIV negative;

- $h\_transit_{(o,a,g,r,HIVn,Stage1)}$ is the labour-market activity o, age a, gender g and race r-specific health transition from HIV negative to Stage 1; and
- $death\ rate_{(a,g,r,HIVn)}$ is the age a, gender g and race r-specific death rates for all HIV-negative persons.

5.2.3. Moving from Stage 1 sequentially through the stages

The ASSA model includes assumptions about the distribution of time adults spend in each HIV stage. Given these assumptions, the median time of infection to death is roughly 11 years for adults when infected under the age of 25, 10.25 years for adults infected between the ages of 25 and 34, and 9.5 years for adults infected when over the age of 34 (Dorrington *et al.*, 2005: 13). This model specifies rates of transition out of each stage according to the duration of the time already spent in each stage. This model allows for people to remain up to 15 years in each stage. These rates are uniform across race and gender (Dorrington *et al.*, 2005: 32–33).

Within the SAGE-H framework, the setting of the transition rates differs in two aspects from the transition rates specified in the ASSA model. Firstly, I require transition rates out of each stage, for each labour-market activity, by age, gender and race. The ASSA model does not model transition rates for any labour-market activities. Secondly, in SAGE-H I do not include or model the time spent in each stage. If I did incorporate a survival time of 15 years in each stage, I would have 60 sub-stages[31] through which people could possibly transit through from the time of infection to death. This has a significant impact on the data requirements and contributes to computational time.

5.2.3.1. Moving from Stage 1 to Stage 2 $\left(h\_transit_{(o,aa,g,r,Stage1,Stage2)}\right)$

The procedure followed to determine the rate of transition between Stage 1 and Stage 2 is as follows:

- I assign labour-market activity o and age a-specific transition rates from Stage 1 to Stage 2.
- The assigned transition rates are not uniform across labour-market activities or age groups, but uniformly set for all genders and race

[31] SAGE-H includes four main stages through which an HIV-positive person can transit through. They are Stages 1, 2, 3 and 4.

groups.

- I calculate the total numbers of people moving out of Stage 1 and compare them with the ASSA estimations. The calculation yields satisfactory results and therefore no adjustment is required.

The following health transition rates are assigned to facilitate the transition from Stage 1 to Stage 2, that is $h\_transit_{(o,aa,g,r,Stage1,Stage2)}$.

Table 10. Assigned transition rates from Stage 1 to Stage 2 (percentage)

Activities	15–24	25–34	35–44	45–54	55–65
Legislator	0.19	0.19	0.24	0.31	0.31
Profes	0.19	0.19	0.24	0.31	0.31
Technical	0.19	0.19	0.24	0.31	0.31
Clerks	0.19	0.19	0.24	0.31	0.31
Service	0.19	0.19	0.24	0.31	0.31
SkilledAgr	0.19	0.19	0.24	0.31	0.31
Craft	0.19	0.19	0.24	0.31	0.31
PlantMach	0.19	0.19	0.24	0.31	0.31
Elementary	0.24	0.24	0.31	0.36	0.36
Domestic	0.24	0.24	0.31	0.36	0.36
Unspecified	0.24	0.24	0.31	0.36	0.36
Short-term unemployed	0.29	0.29	0.34	0.36	0.36
Long-term unemployed	0.34	0.34	0.39	0.39	0.39
Permanently departed from the labour force	0.34	0.34	0.39	0.39	0.39

We can interpret the values in Table 10 as follows: a person in age group 15–24 of gender g, race r, in Stage 1 and employed as a Legislator, has a 19 per cent change to move to Stage 2 (shaded area).

The following features are noted:

- The transitions rates are uniform for all genders and races.
- The transition rates for the first two age groups (15–24 and 25–34) are the same and the transition rates for the last two age groups (45–54 and 55–65) are the same.
- The transition rates are the lowest for the younger age groups and the highest for the older age groups.
- I assume that the transition rates are the lowest for skilled and semi-skilled occupations and the highest for long-term unemployed and PDL activities. We further assume that HIV-positive people who are employed in a skilled occupation are more likely to remain in a specific HIV stage than a person who is employed in an unskilled occupation or who is unemployed. This seems plausible because they are more likely to have access to information and knowledge about treatment and care that allows them to improve their immediate environment. They are also more likely to have access to the latest medicine and private health

care services. For the remaining labour-market activities, the time spent in each stage is shorter.

The transition rate of remaining in Stage 1 is calculated via Equation (E.46):

$$h\_transit_{(o,a,g,r,Stage1,Stage1)} = 1 - h\_transit_{(o,a,g,r,Stage1,Stage2)} - death\ rate_{(a,g,r,Stage1)} \quad \text{(E.46)}$$

$o \in EUP$; $a \in AGE$; $g \in GEN$ and $r \in RACE$.

where
- $h\_transit_{(o,a,g,r,Stage1,Stage1)}$ is the labour-market activity o, age a, gender g and race r-specific transition rates for staying in Stage 1;
- $h\_transit_{(o,a,g,r,Stage1,Stage2)}$ is the labour-market activity o, age a, gender g and race r-specific transition rates from Stage 1 to Stage 2 as summarised in Table 6.9; and
- $death\ rate_{(a,g,r,Stage1)}$ is the age a, gender g and race r-specific death rates.

Based on the assigned transition rates we calculate the number of people moving from Stage 1 to Stage 2 and the number of those remaining in Stage 1. I calculate that the total number of people moving from Stage 1 to Stage 2 is approximately 418,000 people and 970,000 remain in Stage 1. This implies an overall transition rate of 30 per cent.

Note that although the health transition rates for moving from Stage 1 to Stage 2 are the same for all g genders and r races, the transition rates calculated in (E6.47) will not be the same for all genders and races. This is because the death rates are not uniform across age and gender.

5.2.3.2. Moving from Stage 2 to Stage 3 $\left(h\_transit_{(o,aa,g,r,Stage2,Stage3)}\right)$

The procedure followed to determine the rate of transition between Stage 2 and Stage 3 is as follows:

- I assign labour-market activity o and age a-specific transition rates between Stage 2 and Stage 3.
- The assigned transition rates are not uniform across labour-market activities or age groups, but uniformly set for all genders and race groups.
- I then calculate the total number of people moving from Stage 2 to Stage 3 and compare it with the ASSA estimations. I find that the

number of people moving from Stage 2 to Stage 3 is too low. See Appendix 6H for the calculation of the number of people moving to Stage 3 and staying in Stage 2.

- the gender and race-specific number of people moving out of Stage 2 to Stage 3 is scaled to the ASSA numbers. I set a target value for African males and females moving from Stage 2 to Stage 3 and calibrate the transition rates to these set values. I then determine the new transition rates for *Africans* by gender and adopt them for the Other population group. The data used in this procedure is reported in Appendix 6I.

Tables 11 and 12 summarise the scaled transition rates for moving from Stage 2 to Stage 3 by gender. We can interpret the values in Table 6.12 as follows: a Female aged 15–24, race r, who is in Stage 2 and employed as a Legislator, has a 28.18 per cent change to move to Stage 3 (shaded area).

Table 11. Transition rates from Stage 2 to Stage 3 for Females (percentage)

Activities	**15–24**	**25–34**	**35–44**	**45–54**	**55–65**
Legislator	0.2818	0.2818	0.3405	0.3875	0.3875
Professional	0.2818	0.2818	0.3405	0.3875	0.3875
Technical	0.2818	0.2818	0.3405	0.3875	0.3875
Clerks	0.2818	0.2818	0.3405	0.3875	0.3875
Service	0.2818	0.2818	0.3405	0.3875	0.3875
Skilled Agriculture	0.2818	0.2818	0.3405	0.3875	0.3875
Craft	0.2818	0.2818	0.3405	0.3875	0.3875
Plant and Machinery	0.2818	0.2818	0.3405	0.3875	0.3875
Elementary	0.3405	0.3405	0.4110	0.4344	0.4344
Domestic	0.3405	0.3405	0.4110	0.4344	0.4344
Unspecified	0.3405	0.3405	0.4110	0.4344	0.4344
Short-term unemployed	0.3757	0.3757	0.3992	0.4344	0.4344
Long-term unemployed	0.4579	0.4579	0.4579	0.4579	0.4579
Permanently departed from the labour force	0.4579	0.4579	0.4579	0.4579	0.4579

Table 12. Transition rates from Stage 2 to Stage 3 for Males (percentage)

Activities	**15–24**	**25–34**	**35–44**	**45–54**	**55–65**
Legislator	0.3095	0.3095	0.3739	0.4255	0.4255
Professional	0.3095	0.3095	0.3739	0.4255	0.4255
Technical	0.3095	0.3095	0.3739	0.4255	0.4255
Clerks	0.3095	0.3095	0.3739	0.4255	0.4255
Service	0.3095	0.3095	0.3739	0.4255	0.4255
Skilled Agriculture	0.3095	0.3095	0.3739	0.4255	0.4255
Craft	0.3095	0.3095	0.3739	0.4255	0.4255
Plant and Machinery	0.3095	0.3095	0.3739	0.4255	0.4255
Elementary	0.3739	0.3739	0.4513	0.4771	0.4771
Domestic	0.3739	0.3739	0.4513	0.4771	0.4771
Unspecified	0.3739	0.3739	0.4513	0.4771	0.4771
Short-term unemployed	0.4126	0.4126	0.4384	0.4771	0.4771
Long-term unemployed	0.5029	0.5029	0.5029	0.5029	0.5029
Permanently departed from the labour force	0.5029	0.5029	0.5029	0.5029	0.5029

The following features are noted:

- The transition rates are higher than those set for moving from Stage 1 to Stage 2.
- The transitions rates are uniform across race but differ by labour-market activity, age and gender.
- The transition rates for the first two age groups (15–24 and 25–34) are the same and the transition rates for the last two age groups (45–54 and 55–65) are the same.
- The transition rates are the lowest for the younger age groups and the highest for the older age groups.
- The transition rates are the lowest for skilled and semi-skilled occupations and the highest for the long-term unemployed and *PDL* activities.

The transition rate of remaining in Stage 2 is calculated via Equation (E.47):

$$h\_transit_{(o,a,g,r,Stage2,Stage2)} = 1 - h\_transit_{(o,a,g,r,Stage2,Stage3)} - death\ rate_{(a,g,r,Stage2)} \quad \text{(E.47)}$$

$o \in EUP$; $a \in AGE$; $g \in GEN$ and $r \in RACE$.

where
- $h\_transit_{(o,a,g,r,Stage2,Stage2)}$ is the labour-market activity o, age a, gender g and race r-specific transition rates for staying in Stage 2;
- $h\_transit_{(o,a,g,r,Stage2,Stage3)}$ is the labour-market activity o, age a, gender g and race r-specific transition rates allowing people to move from Stage 2 to Stage 3. These rates are summarised in Tables 12 and 13; and
- $death\ rate_{(a,g,r,Stage2)}$ is the age a, gender g and race r-specific death rates.

Based on the above transition rates we calculate the number of people moving from Stage 2 to Stage 3 and the number of those remaining in Stage 2. I calculate that the total number of people moving from Stage 2 to Stage 3 is approximately 422,000 people. This means that the overall transition rate is approximately 42 per cent.

Although the transitions for moving from Stage 2 to Stage 3 are uniform for all races, the transitions calculated in (E.47) will not be the same for all genders and races. This is due to the non-uniform age, gender and race-specific death rates and non-uniform gender-specific transition rates from Stage 2 to Stage 3.

5.2.3.3. Moving from Stage 3 to Stage 4 $\left(h\_transit_{(o,aa,g,r,Stage3,Stage4)}\right)$

The following procedure is followed to determine the rate of transition between the Stage 3 and Stage 4:

- I assign labour-market activity o and age a-specific transition rates between the Stage 3 and Stage 4.
- The assigned transition rates are not uniform across labour-market activities or age groups, but uniformly set for all genders and race groups.
- I calculate the total number of people moving from Stage 3 to Stage 4 and compare it with the ASSA model estimations. We find that the number of people moving from Stage 3 to Stage 4 is too high.
- The gender and race-specific number of people moving out of Stage 3 to Stage 4 is scaled to the ASSA model numbers. We set a target value for African Males and Females moving from Stage 3 to Stage 4 and calibrate the transition rates to these set values. I then determine the new transition rates for Africans by gender and adopt them for the Other population group.

Table 13. Transition rates from Stage 3 to Stage 4 for Females (percentage)

Activities	**15–24**	**25–34**	**35–44**	**45–54**	**55–65**
Legislator	0.1206	0.1206	0.1457	0.1658	0.1658
Professional	0.1206	0.1206	0.1457	0.1658	0.1658
Technical	0.1206	0.1206	0.1457	0.1658	0.1658
Clerks	0.1206	0.1206	0.1457	0.1658	0.1658
Service	0.1206	0.1206	0.1457	0.1658	0.1658
Skilled Agriculture	0.1206	0.1206	0.1457	0.1658	0.1658
Craft	0.1206	0.1206	0.1457	0.1658	0.1658
Plant and Machinery	0.1206	0.1206	0.1457	0.1658	0.1658
Elementary	0.1457	0.1457	0.1759	0.1859	0.1859
Domestic	0.1457	0.1457	0.1759	0.1859	0.1859
Unspecified	0.1457	0.1457	0.1759	0.1859	0.1859
Short-term unemployed	0.1960	0.1960	0.1960	0.1960	0.1960
Long-term unemployed	0.1960	0.1960	0.1960	0.1960	0.1960
Permanently departed from the labour force	0.1960	0.1960	0.1960	0.1960	0.1960

Table 13 and Table 14 summarise the gender-specific transition rates allowing people to move from Stage 3 to Stage 4. We can interpret the values in Table 13 as follows: a 15–24-year-old Female of race r who is employed as a Legislator and in Stage 3, has a 12.06 per cent change of moving to Stage 4 (shaded area).

Table 14. Transition rates from Stage 3 to Stage 4 for Males (percentage)

Activities	15–24	25–34	35–44	45–54	55–65
Legislator	0.1349	0.1349	0.1630	0.1854	0.1854
Professional	0.1349	0.1349	0.1630	0.1854	0.1854
Technical	0.1349	0.1349	0.1630	0.1854	0.1854
Clerks	0.1349	0.1349	0.1630	0.1854	0.1854
Service	0.1349	0.1349	0.1630	0.1854	0.1854
Skilled Agriculture	0.1349	0.1349	0.1630	0.1854	0.1854
Craft	0.1349	0.1349	0.1630	0.1854	0.1854
Plant and Machinery	0.1349	0.1349	0.1630	0.1854	0.1854
Elementary	0.1630	0.1630	0.1967	0.2079	0.2079
Domestic	0.1630	0.1630	0.1967	0.2079	0.2079
Unspecified	0.1630	0.1630	0.1967	0.2079	0.2079
Short-term unemployed	0.2191	0.2191	0.2191	0.2191	0.2191
Long-term unemployed	0.2191	0.2191	0.2191	0.2191	0.2191
Permanently departed from the labour force	0.2191	0.2191	0.2191	0.2191	0.2191

The following features are noted:

- The transitions rates are uniform across race but differ by labour-market activity, gender and age.
- The transition rates for the first two age groups (15–24 and 25–34) are the same and the transition rates for the last two age groups (45–54 and 55–65) are the same.
- The transition rates are the lowest for the younger age groups and the highest for the older age groups.
- The transition rates are the lowest for skilled and semi-skilled occupations and the highest for the long-term unemployed and PDL activities.

The transition rate of remaining in Stage 3 is calculated via Equation (E.48):

$$h\_transit_{(o,a,g,r,Stage3,Stage3)} = 1 - h\_transit_{(o,a,g,r,Stage3,Stage4)} - death\ rate_{(a,g,r,Stage3)} \quad \text{(E.48)}$$

$o \in EUP$; $a \in AGE$; $g \in GEN$ and $r \in RACE$.

where
- $h\_transit_{(o,a,g,r,Stage3,Stage4)}$ is the labour-market activity o, age a, gender g and race r-specific transition rate allowing people to move from Stage 3 to Stage 4; and
- $death\ rate_{(a,g,r,Stage3)}$ is the age a, gender g, race r and HIV stage h-specific death rates.

Based on the above transition rates we calculate the number of people moving from Stage 3 to Stage 4 and the number of those remaining in Stage 3. I calculate that the total number

of people moving from Stage 3 to Stage 4 is approximately 232,000 people. This implies that the overall transition rate is 18 per cent.

Although the transitions for moving from Stage 3 to Stage 4 are uniform across all races, the transition rates calculated in (E6.48) will not be uniform for all ages and genders. This is because the death rates are non-uniform across age and gender, and the health transition rates specified in Tables 13 and 14 are non-uniform across gender.

5.2.3.4. Remaining in Stage 4 $\left(h\_transit_{(o,aa,g,r,Stage4,Stage4)}\right)$

As explained before, an HIV-positive person moves sequentially through the different stages until they reach Stage 4. Once they are in Stage 4 they remain in this stage until death. Those who remain in Stage 4 are calculated as:

$$stay_{(o,a,g,r,Stage4)} = ACT\_L^{*}_{(o,a,g,r,Stage4)} \left(1 - death\ rate_{(a,g,r,Stage4)}\right) \qquad \text{(E.49)}$$

$o \in EUP$; $a \in AGE$; $g \in GEN$ and $r \in RACE$.

where
- $ACT\_L_{(o,a,g,r,Stage4)}$ is the number of people in each labour-market activity o by age a, gender g and race r, who were in Stage 4 during year $t-1$; and
- $\left(1 - death\ rate_{(a,g,r,Stage4)}\right)$ is the age a, gender g and race r-specific survival probability of a person in Stage 4.

The following features are noted:

- These survival probabilities are applicable only to the PDL activity. No other labour-market activity includes people in the final stage of the disease.
- The transitions rates are non-uniform across age, gender and race.

The health probability of those remaining in Stage 4 is summarised in Table 15.

Table 15. The probability of staying in Stage 4

	(1)	(2)	(3)	(4)	(5)
	15–24	**25–34**	**35–44**	**45–54**	**55–65**
African					
Female	48.57	36.11	39.98	46.74	53.1
Male	64.56	46.1	34.28	40.26	61.25
Other					
Female	49.95	44.03	40.19	44.62	58.66
Other	64.49	45.82	37.14	44.24	58.33

This concludes the derivation of the health transition matrix. This matrix allows people, in each labour-market activity o by age a, gender g and race r, to move from health status hh to h. The next step is to allow people to move from age aa to a.

5.3. Age transition $\left(a\_transit_{(aa,g,r,hh,a)}\right)$

In this section we introduce a matrix allowing people to move from age aa to age a. To do this I create a matrix called $a\_transit$. I assume that 9 per cent of every age cohort progress to the next age cohort (Table 16). I further assume that this setting is uniform across gender, race and HIV stage.

Table 16. Proportion of people moving from one age group to the next

	15–24	**25–34**	**35–44**	**45–54**	**55–65**	**Total**
15–24	0.91	0.09	0	0	0	1
25–34	0	0.91	0.09	0	0	1
35–44	0	0	0.91	0.09	0	1
45–54	0	0	0	0.91	0.09	1
55–65	0	0	0	0	0.91	0.91

5.4. Final transition matrix $\left(T_{(o,aa,g,r,hh,a,h)}\right)$

The final transition matrix shows the proportion of people in labour-market activity o, by gender g and race r, moving from age aa to age a and health status hh to h. This matrix is calculated as:

$$T_{(o,aa,g,r,hh,a,h)} = h\_transit_{(o,aa,g,r,hh,h)}^{*} \; a\_transit_{(aa,g,r,hh,a)} \qquad \text{(E.50)}$$

$o \in EUP$; $aa \in AGE$; $g \in GEN$, $r \in RACE$, $hh \in H5$, $a \in AGE$ and $h \in H5$.

The following prominent features of the transition matrix are noted:

- The matrix is strongly diagonal implying that the majority of people remain in the same age group and HIV stage.
- The row sums for all labour-market activity are less than 1 because we allow people, of all ages, genders and races, to leave the working age

population through death. The row sums are equal to $1-death\ rate_{(a,g,r,h)}$.

- People move sequentially through the age groups and between HIV stages. All other transition rates are zero.
- People do not change their labour-market activity o, race r or gender g when they are allocated from an activity in year $t-1$ to a category at the start of year t. People can only change their age from aa to a and health status from hh to h.

6. CONCLUDING REMARKS

This paper describes the structure of the labour-market database. Various large matrices are derived. The activities matrix describes the number of people in each labour-market activity o by age a, gender g, race r and HIV stage h. Once the activities matrix is derived, we proceed by deriving the categories matrix. The categories matrix shows the number of people in each labour-market activity o by gender g and race r who are allocated to age a and HIV stage h at the start of the base year.

Two matrices (*OFFERS* and H matrices) describe the flows of people of age a, gender g and race r moving from all categories c to all activities m in the base year. The *OFFERS* matrix shows the activities people *would like* to perform during the year while the actual flows matrix shows who *secures* employment and what happens to those whose offers were rejected. In deriving these matrices several labour-market features are accommodated. For example, I regulate the flows from employment categories to unemployment activities, the flow from unemployment categories to unemployment activities and the flow of people who move to a Stage 4 status. I also allow adults to survive to the next year and change their health status at the end of the year. These transitions are captured in the transition matrix.

REFERNCES

Abt Associates. (2000). *The impending catastrophe: A resource book on the emerging HIV/AIDS epidemic in South Africa*. Kaiser Family Foundation. [www.kff.org/southafrica/20000515a-index.cfm. Accessed in June 2010]

Actuarial Society of South Africa (ASSA). (2005). *ASSA 2003 Full model*. [http://aids.actuarialsociety.org.za/ASSA2003-Model-3165.htm. Accessed in March 2009]

Booysen, F., Geldenhuys, J. & Marinkov, M. (2003). The impact of HIV/AIDS on the South African Economy: A Review of current evidence. Paper prepared for *TIPS/DPRU conference on "The challenges of growth and poverty: The South African economy since democracy".* 8–10 September 2003, Indaba Hotel, Johannesburg.

Booysen, F. le R., Bachmann, M., Van Rensburg, H.C.J., Engelbrech, M., Steyn, F. & Meyer, K. (2002). *The socio-economic impact of HIV/AIDS on households in South Africa: Pilot study in Welkom and Qwaqwa, Free State province.* Final report for USAID on Economic Impact of HIV/AIDS in South Africa. Centre of Health Systems Research and Development, Bloemfontein.

Bowen, P., Dorrington, R., Distiller, G., Lake, H. & Besesar, S. (2008). HIV/AIDS in the South African construction industry: an empirical study. *Construction Management and Economics* (August), Volume 26: 825–837.

Bureau for Economic Research. (2006). *The macroeconomic impact of HIV/AIDS under alternative intervention scenarios (with specific reference to ART on the South African economy).* University of Stellenbosch.

Dorrington, R., Johnson, L. & Budlender, D. (2005). *ASSA2003 AIDS and Demographic models. User guide*. Centre for Actuarial Research, University of Cape Town. [http://aids.actuarialsociety.org.za/ASSA2003-Model-3165.htm. Accessed in March 2009]

Evian, C., Fox, M., MacLeod, W., Slotow, S.J. & Rosen, S. (2004). Prevalence of HIV in workforces in Southern Africa, 2000–2001. *South African Medical Journal*, Volume 94(2): 125–130.

Groenewald, P., Nannan, N. Bourne, D., Laubscher, R. & Bradshaw, D. (2005) Identifying deaths from AIDS in South Africa. *AIDS*, Volume 19: 193–201.

International Labour Organization (ILO). (2004). *HIV/AIDS and work: Global estimates, impact and response.* International Labour Office, Geneva, Switzerland. [http://www.ilo.org/public/libdoc/ilo/2004/104B09_201_engl.pdf. Accessed in December 2011]

Jefferis, K., Nannyonjo, J., Byamugisha, J. & Baine, S. (2007). *Assessing the macroeconomic impact of HIV/AIDS in Uganda*. Phase 1 Report. Literature review: the Macroeconomic impact of HIV/AIDS. Submitted to the Ministry of Finance Planning and Economic Development, United Nations Development Programme.

Laubscher, P. (2000). *HIV/AIDS and the South African economy*. Bureau for Economic Research (BER), Research Note No. 8. University of Stellenbosch, South Africa.

Laubscher, P., Smit, B.W. & Visagie, L. (2001). *The macroeconomic impact of HIV/AIDS in South Africa.* Bureau for Economic Research (BER), Research Note No. 10, University of Stellenbosch, South Africa.

Shisana, O., Hall, E.J., Maluleke. R., Chauveau, J. & Schwabe, C. (2004). HIV/AIDS prevalence among South African health workers. *South African Medical Journal*, Volume 94(10).
[http://www.samj.org.za/index.php/samj/article/viewFile/2779/1989.
Accessed in August 2010]

Shisana, O., Rehle, T., Simbayi, L.C., Parker, W., Zuma, K., Bhana, A., Connoly, C., Jooste, S., Pillay, V. *et al.* (2005). *South African National HIV Prevalence, Incidence, Behaviour and Communication Survey, 2005.* HSRC Press, Cape Town, South Africa.

Shisana, O., Rehle, T., Simbayi, L., Zuma, K., Jooste, S., Pillay-Van Wyk, V., Mbelle, N., Van Zyl, J., Parker, W., Zungu, N.P., Pezi, S. & the SABSSM III Implementation Team. (2009). *South African National Prevalence, Incidence, Behaviour and Communication Survey, 2008. A turning tide among teenagers?* HSRC Press, Cape Town, South Africa.

Shisana, O., Simbayi, L. (Eds.) (2002). *Nelson Mandela/HSRC Study of HIV/AIDS: Full Report South African National HIV prevalence, behaviour risks and mass media. Household survey.* HSRC Press, Cape Town, South Africa.

Shisana, O., Simbayi, L.C., Rehle, T., Zungu, N.P., Ngogo, N., Jooste, S., Pillay-Van Wyk, V., Parker, W., Pezi, S., Davids, A., Nwanyanwu, O., Dinh, T.H. & SABSSM III Implementation Team (2010). *South African National HIV Prevalence, Incidence, Behaviour and Communication Survey, 2008. The health of our children.* HSRC Press, Cape Town, South Africa.

Statistics South Africa. (2003a). *Labour Force Survey September 2002*. Report no. P0210. Statistics South Africa, Pretoria, South Africa.

Statistics South Africa. (2003b). *Census 2001 data*. Pretoria, South Africa.

Statistics South Africa. (2006d). *Adult Mortality (Age 15–64). Based on death notification data in South Africa: 1997–2004.* Report no. 03-09-05. Pretoria, South Africa.

Statistics South Africa. (2006e). *Mortality and causes of death in South Africa, 2003 and 2004. Findings from death notification*. Statistical release P0309.3. Pretoria, South Africa.

Vass, J. (2005). *A review of labour markets in South Africa: The impact of HIV/AIDS on the labour market*. Human Sciences Research Council, Pretoria, South Africa.

World Health Organization. (2007). *WHO case definitions of HIV for surveillance and revised clinical staging and immunological classification of HIV-related disease in adults and children*.
[http://www.who.int/hiv/pub/guidelines/HIVstaging150307.pdf. Accessed in June 2010]

Appendix 1. Sets and elements modelled in the labour-market specification

	All categories at the beginning of year *t* Set EUNP	
	Description	**Short name**
1.	Legislators	Legislator
2.	Professionals	Profes
3.	Technicians	Technical
4.	Clerks	Clerks
5.	Service workers	Service
6.	Skilled agricultural workers	SkilledAgr
7.	Craft workers	Craft
8.	Plant and machine operators	PlantMach
9.	Elementary occupations	Elementary
10.	Domestic workers	Domestic
11.	Occupation unspecified	Unspecified
12.	Short-term unemployed	S
13.	Long-term unemployed	L
14.	New entrants	N
15.	Permanently departed from the labour force	PDL

	All activities undertaken during year *t* Set EUP	
	Description	**Short name**
1.	Legislators	Legislator
2.	Professionals	Profes
3.	Technicians	Technical
4.	Clerks	Clerks
5.	Service workers	Service
6.	Skilled agricultural workers	SkilledAgr
7.	Craft workers	Craft
8.	Plant and machine operators	PlantMach
9.	Elementary occupations	Elementary
10.	Domestic workers	Domestic
11.	Occupation unspecified	Unspecified
12.	Short-term unemployed	S
13.	Long-term unemployed	L
14.	Permanently departed from the labour force	PDL

	Set EU	
	Description	**Short name**
1.	Legislators	Legislator
2.	Professionals	Profes
3.	Technicians	Technical
4.	Clerks	Clerks
5.	Service workers	Service
6.	Skilled agricultural workers	SkilledAgr
7.	Craft workers	Craft
8.	Plant and machine operators	PlantMach
9.	Elementary occupations	Elementary
10.	Domestic workers	Domestic
11.	Occupation unspecified	Unspecified
12.	Short-term unemployed	S
13.	Long-term unemployed	L

Appendix 1 (continue).

	Set EUN	
	Description	**Short name**
1.	Legislators	Legislator
2.	Professionals	Profes
3.	Technicians	Technical
4.	Clerks	Clerks
5.	Service workers	Service
6.	Skilled agricultural workers	SkilledAgr
7.	Craft workers	Craft
8.	Plant and machine operators	PlantMach
9.	Elementary occupations	Elementary
10.	Domestic workers	Domestic
11.	Occupation unspecified	Unspecified
12.	Short-term unemployed	S
13.	Long-term unemployed	L
14.	New entrants	N

	Set UN	
	Description	**Short name**
1.	Short-term unemployed	S
2.	Long-term unemployed	L
3.	New entrants	N

	Set UP	
	Description	**Short name**
1.	Short-term unemployed	S
2.	Long-term unemployed	L
3.	Permanently departed from the labour force	PDL

	Age groups Set AGE	
	Age groups	**Short name**
1.	A15–19	A15–24
2.	A20–24	
3.	A25–29	A25–34
4.	A30–34	
5.	A35–39	A35–44
6.	A40–44	
7.	A45–49	A45–54
8.	A50–54	
9.	A55–59	A55–65
10.	A60–65	

	Race Set RACE	
	Description	**Short name**
1.	Africa	A
2.	Coloured	
3.	Indian	
4.	White	Other

Appendix 1 (continue).

	Gender Set GEN	
	Description	**Short name**
1.	Female	F
2.	Male	M

	HIV Stage Set STG	
	Description	**Short name**
1.	HIV negative	HIVn
	HIV positive – Stages[32]	
2.	Stage 1	Stage1
3.	Stage 2	Stage2
4.	Stage 3	Stage3
5.	Stage 4	Stage4

[32] An HIV-positive person will fall into one of these stages. These stages are based on the WHO classification of HIV progression (WHO, 2007).

Appendix 2. HIV prevalence across all labour-market activities

Table 1 summarises, for each labour-market activity, the number of people by HIV status. Column 1 lists the labour-market activities included in SAGE-H and column 2 lists the skill level associated with each occupation. We assign the percentages in column (3) that determine the percentage HIV-positive adults in each activity. Based on these percentages, the number of HIV-positive people in each labour-market activity is calculated in column 4. The number of HIV-negative people (column 5) is the difference between the total number of people (column 6) minus the number of HIV-positive people (column 4).

Table 1. HIV prevalence across all labour-market activities in the base year

	(1)	(2)	(3)	(4)	(5)	(6)
	Description	**Skill level***	**% HIV positive****	**HIV positive**	**HIV negative**	**Total**
Skilled	Legislators	4	7	51,931	680,068	731,999
	Professionals	4	7	36,330	466,670	503,000
	Technicians	3	9	106,421	1,099,579	1,205,999
Semi-skilled	Clerks	2	13	146,052	962,949	1,109,000
	Service workers	2	13	164,492	1,080,510	1,245,003
	Skilled agricultural	2	13	92,641	613,358	705,999
	Craft workers	2	17	245,928	1,213,072	1,459,000
	Plant and machine operators	2	17	194,613	961,389	1,156,002
Unskilled	Elementary	1	17	388,089	1,906,910	2,294,998
	Domestic	1	17	143,691	700,313	844,004
	Unspecified	1	17	6,866	34,134	40,999
Unemploy-ment	Short-term Unemployed	-	14	159,786	1,032,389	1,192,175
	Long-term Unemployed	-	15	1,071,134	5,857,804	6,928,938
	Permanently departed from the labour force	-	18	1,624,615	7,470,270	9,094,885
	Total	-	15.5	4,432,586	24,079,414	28,512,000

Source: * Statistics South Africa, Sept 2006, Appendix 7, p. 49.
** Own values.