WORKING PAPER 404/COMPUTER MANUAL 23

SPATIAL POPULATION ANALYSIS USING MOVEMENT DATA AND ACCOUNTING METHODS: THEORY, MODELS, THE 'MOVE' PROGRAM AND EXAMPLES

PHILIP REES

LARGE

Geography

A.0.029
LEE|W

WORKING PAPER

School of Geography
University of Leeds

Stamp indicates date for RETURN. Fines for late returns will be charged in accordance with the regulations.

Books required by another reader will be recalled. This also applies during vacations.

02 SEP 2009		

Classmark

Geography
A.0.029 LEE/W

30106 006960602

Working Paper 404/Computer Manual 23

SPATIAL POPULATION ANALYSIS USING MOVEMENT DATA AND ACCOUNTING METHODS: THEORY, MODELS, THE 'MOVE' PROGRAM AND EXAMPLES

Philip Rees

School of Geography
University of Leeds

August 1984

Also published in
Working Paper Series
Netherlands Interuniversity Demographic Institute
PO Box 955, Prinses Beatrixlaan 428
2270 AZ Voorburg, The Netherlands

CONTENTS

Preface
Acknowledgments
Abstract
List of Tables
List of Figures

1. GETTING STARTED WITH MOVEMENT ACCOUNTS

1.1 Age and time
1.2 Regions
1.3 What data do we need?
1.4 A model for constructing movement accounts
1.5 Producing movement accounts and projections
1.6 The tables of movement accounts
1.7 Tables of projected populations

2. THE THEORY OF MOVEMENT ACCOUNTS

2.1 Movement accounts
2.2 Meaning of the diagonal term
2.3 Accounting models for historical or base period accounts
2.4 Accounting models for projection

3. THE BUILDING BLOCKS OF THE ACCOUNTS

3.1 Initial populations (subroutine BEGPOP)
3.2 Populations at risk (subroutine PARA)
3.3 Deaths (subroutine DEATHS)
3.4 Emigrations (subroutine EMIG)
3.5 Internal migrations (the MIG subroutine)
3.6 The residual balancing terms (subroutine RES)
3.7 Immigrations (subroutine IMMIG)
3.8 Final populations (subroutine ENDPOP)
3.9 Other marginal totals
3.10 Iteration and convergence of the accounts model (subroutine VERGE)
3.11 Row and column constraints (subroutine CONSTR)
3.12 Births (subroutine BIRTHS)
3.13 Putting the building blocks together

4. A GUIDE FOR USERS OF THE PROGRAM

4.1 The instructions file
4.2 The component data files
4.3 The work file
4.4 The projected populations file
4.6 The survivorship rates file
4.7 Running the program

5. EXAMPLES OF MOVEMENT ACCOUNTS MODELS FOR BASE PERIODS

5.1 The forecast unconstrained model
5.2 The backcast unconstrained model
5.3 The no-emigrations model
5.4 The both populations model

5.5 The constrained model
5.6 Other models for the base period

6. EXAMPLES OF MOVEMENT ACCOUNTS MODELS FOR PROJECTION

6.1 A model with gross external migration rates
6.2 A model with gross external migration flows
6.3 A model with net external migration rates
6.4 A model with net external migration flows
6.5 A model with emigration rates and immigration flows
6.6 A model with emigration flows and immigration rates

7. EXAMPLES OF ALTERNATIVE MODELS

7.1 A model incorporating independent projections of migration
7.2 A model without migration
7.3 A model without external migration
7.4 A single region model with net migration flows
7.5 A single region model with net migration rates
7.6 A closed model for the whole world
7.7 An aggregate model
7.8 Other models

8. CONCLUDING REMARKS

8.1 A general guide to constructing population analyses
8.2 An example of input estimation
8.3 An example of output tabulation
8.4 Future developments

REFERENCES

APPENDIX: A PROGRAMMER'S GUIDE

PREFACE

A word is, I think, needed to explain why I have written such an extensive document describing a set of population models and associated computer program. Firstly, I wanted to put down on paper all the details lest I forget them. Secondly, if the program was to be used in another collaborating institution, it was essential to record all the features of the program so that other researchers could use and, if necessary, modify it. Thirdly, since the program is intended for use with my undergraduate students, I had to write an easy to understand and use manual, which will, I hope, help students to acquire some of the tools of spatial population analysis described herein. Fourthly, if users other than these were to have the opportunity of using the program, the description had to be extremely explicit.

The manual is a lengthy one so that busy users will be grateful for guidance on reading. Everyone should read section 1, GETTING STARTED WITH MOVEMENT ACCOUNTS, if only to find out whether the models and program are of any utility. The casual user can probably use section 1 and one of the example models in sections 5, 6 or 7, the EXAMPLES sections, to prepare a set of accounts and projections, with occasional excursions into the relevant part of section 4, A GUIDE FOR USERS. My students should read and gain a good grasp of section 2, THE THEORY OF MOVEMENT ACCOUNTS, as I will be asking them questions about the concepts involved. The full details of section 3, THE BUILDING BLOCKS OF THE ACCOUNTS, and section 4, A GUIDE FOR USERS OF THE PROGRAM, are for users fully committed to producing the best population accounts and projections for their system of interest.

The EXAMPLE sections are to be consulted by the user looking for an information system and model similar to his own. Three key figures summarize diagrammatically the choices available in the EXAMPLE sections: Figure 2.1 covers the MODELS FOR BASE PERIODS of section 5; Figure 6.1 covers the MODELS FOR PROJECTION of section 6; and Figure 7.1 covers the ALTERNATIVE MODELS of section 7 which are designed to link the user back to prior practice so that the improvements effected by adopting an accounting perspective can be evaluated.

The CONCLUDING REMARKS should definitely be left by the reader to the end. Otherwise you might never start on the accounting adventure!

ACKNOWLEDGMENTS

My thanks are due to Frans Willekens of the Netherlands Interuniversity Demographic Institute who persuaded me that the development of movement accounts and projection models was both useful and non-trivial, and who invited me to make two all-too-short visits to NIDI where his hospitality was without match, to work on the program and a set of accounts for the Netherlands. I am grateful too for the help in using the NIDI computer provided by Trix Hummelinck.

ABSTRACT

This paper describes how population accounts can be constructed and regional population projections carried out using migration data of the movement type. A detailed description is given of the computer program for implementing movement accounts, and the wide variety of alternative models that can be built with this model building kit is illustrated.

LIST OF TABLES

1.1 The period-cohort (type II) classifications used in the movements accounts examples.
1.2 Components of the movement accounts input as data.
1.3 Movement accounts tables for selected cohorts, 1976-81.
1.4 Net components of change 1976-81.
1.5 Projected populations, Greater London, 1981-2006.

2.1 A movement accounts table for cohort x in one time interval.
2.2 Movement accounts for the 20-24 to 25-29 cohort, assuming mobility levels 2.5 times those observed.

4.1 The analysis or n parameters: formats.
4.2 Tabulation level settings using the NTAB parameter.
4.3 The region name records.
4.4 The model or i parameter records: record one.
4.5 The model or i parameter records: record two.
4.6 The model or i parameter records: record three.
4.7 The model or i parameter records: record four.
4.8 File names for the component data files.

6.1 Choices for modelling external flows.

LIST OF FIGURES

1.1 Alternative age-time observation plans.
1.2 Listings of data files input to the movement accounts program.
1.3 Instructions to the program about the system being studied and the accounts model to be employed.
1.4 Inputs to and outputs from the second cohort population of Greater London, 1976-81.

2.1 The different ways of assembling movement accounts.
2.2 Age-time observation plan for the last cohort.

3.1 The principal subroutines of the movement accounts program.
3.2 The building blocks of the movement accounts and the subsection in which they are described.

4.1 The system of files needed for producing movement accounts.
4.2 A framework for preparing the instruction file.
4.3 Examples showing the structure of vector component files.
4.4 An example showing the structure of a matrix component file.
4.5 A portion of a constraints file for Greater London and the

Rest of the UK.
4.6 A births and sex proportions file for the West Netherlands, Rest of the Netherlands system.
4.7 The results file: example 1.
4.8 The results file: example 2.
4.9 The results file: example 3.
4.10 An extract from a projected populations file for the Greater London, Rest of the UK system.
4.11 An extract from a survivorship rates file for the West Netherlands, Rest of the Netherlands system, 1976.
4.12 A listing of the executive file that controls the running of the program at the University of Leeds.

5.1 The forecast unconstrained model: model parameters and component data files.
5.2 The backcast unconstrained model: model parameters and component data files.
5.3 The no-emigration model: model parameters and component data files.
5.4 The both populations model: model parameters and component data files.
5.5 The constrained model (forecast version): model parameters and component data files.

6.1 Examples of movement accounts models for projection.
6.2 A model incorporating gross external migration rates: model parameters and component data files.
6.3 A model incorporating gross external migration flows: model parameters and component data files.
6.4 A model incorporating net external migration rates: model parameters and component data files.
6.5 A model incorporating net external migration flows: model parameters and component data files.
6.6 A model incorporating emigration rates and immigration flows: model parameters and component data files.
6.7 A model incorporating emigration flows and immigration rates: model parameters and component data files.

7.1 Alternative models.
7.2 A model incorporating independent projections of migration: model parameters and component data files.
7.3 A model without migration.
7.4 A model without external migration.
7.5 A single region model with net migration flows.

8.1 The time periods involved in a population projection.
8.2 An age-time space diagram for the Netherlands deaths estimation problem.
8.3 A program to tabulate projected populations for a 20 zone system.

A.1 A listing of the main segment of the MOVE program.
A.2 A flow chart of the main segment of the MOVE program.

1. GETTING STARTED WITH MOVEMENT ACCOUNTS

Let us assume that you are interested in using an accounting approach to population analysis and population forecasting. If you aren't, stop reading right now! This approach simply involves being as careful with population figures as you are with your own household budget. Let us also assume you wish to construct a realistic set of population accounts for a region in your country - this could be something as big as "Scotland" or as small as the "Borough of Kensington and Chelsea". What information would you have to collect and how would you use it?

1.1 Age and time

First, you would have to decide about age and time, that is, how many age groups you were going to distinguish and over what period. You would have to choose a plan for observing these quantities. Only one plan is feasible for most population accounts - the one labelled period cohort or projection cohort or Type II in the diagram (Figure 1.1). Note that most demographic agencies don't assemble population data in that way. They use plans of Type I or Type IV. You are going to have to do the work of getting the data into shape. This program won't do that for you because there are too many variations in the "raw data" starting points to be feasibly handled in a general program. Each set of accounts will have its own set of estimation routines.

So back to the original question: what ages, what time intervals? Let us use as a first example the 5 year time period mid-year 1976 to mid-year 1981, and the 5 year age groups 0-4, 5-9, 10-14 and so on to 80-84, 85+. This will give us a relatively fine age breakdown to reasonably elderly ages. However, it isn't quite as easy as that as Table 1.1 shows. These age groups are what is needed for mid-year 1976, but by mid-year 1981 we have added another, so the ending age groups have to be 85-89 and 90+. The diagram lists some 19 projection cohorts labelled by initial and final ages: for example, cohort 6 runs from ages 20-24 to 25-29 or cohort 10 runs from ages 40-44 to 45-49. Note that the first cohort runs from birth to 0-4 and the last from 85+ to 90+.

All the demographic events which we are going to measure and put into the accounts must be classified according to the projection cohorts.

1.2 Regions

Let us choose to analyse the population of Greater London. However, we cannot look at Greater London in isolation because large numbers of people move between Greater London and the rest of the country. Also, as anyone who lives in or has visited London in the recent past will know, the metropolis contains vast numbers who have immigrated to the city from all parts of the globe. So that, if we are to analyse the Greater London population properly, we must set up a system of three regions consisting of Greater London, the Rest of the United Kingdom and the Rest of the World.

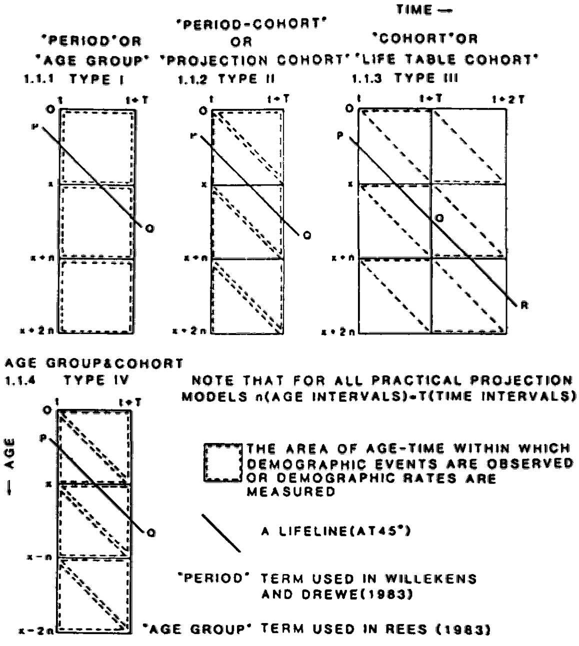


FIGURE 1.1 Alternative age-time observation plans

TABLE 1.1 The period cohort (type II) classifications used in the movement accounts example

Cohort number	Initial populations	Deaths, emigrations, internal migrations, immigrations and births from	to	Final populations	Cohort order in files
1		birth	0- 4	0- 4	19
2	0- 4	0- 4	5- 9	5- 9	1
3	5- 9	5- 9	10-14	10-14	2
4	10-14	10-14	15-19	15-19	3
5	15-19	15-19	20-24	20-24	4
6	20-24	20-24	25-29	25-29	5
7	25-29	25-29	30-34	30-34	6
8	30-34	30-34	35-39	35-39	7
9	35-39	35-39	40-44	40-44	8
10	40-44	40-44	45-49	45-49	9
11	45-49	45-49	50-54	50-54	10
12	50-54	50-54	55-59	55-59	11
13	55-59	55-59	60-64	60-64	12
14	60-64	60-64	65-69	65-69	13
15	65-69	65-69	70-74	70-74	14
16	70-74	70-74	75-79	75-79	15
17	75-79	75-79	80-84	80-84	16
18	80-84	80-84	85-89	85-89	17
19	85+	85+	90+	90+	18

Notes:

1. Births data assembled for cohorts 4 through 10, which refer to mothers' cohort at time of maternity

2. All cohorts are 'projection' cohorts, type II in Figure 1.1

1.3 What data do we need?

1.3.1 Initial populations

We need populations for the start of our base period to begin with. For the time being we use the mid-year 1976 (June 30th/July 1st) population estimates shown in Table 1.2.1. You can tell these are estimates rather than counts because they all end in "00": that is, they have been estimated to the nearest hundred.

1.3.2 Deaths

In any description of population change we must take account of mortality. The estimated numbers of deaths in projection cohorts for the mid-year 1976 to mid-year 1981 period are assembled in Table 1.2.2 for Greater London and the Rest of the UK. Deaths are concentrated at the older ages as is usual in a low mortality country, although the number of deaths in the first cohort is only exceeded when the 10th or 11th is reached.

1.3.3 Emigrations

Population is lost from a region not only through death but also through migration. Migrations can be divided into those internal to a country and those external. Emigrations are migrations from regions within the country being studied to another country.

Table 1.2.3 shows the estimated numbers of such moves from Greater London and the Rest of the UK over 1976-81. Note that the distribution of losses as a result of emigration is very different from that of deaths, being highly concentrated in the young adult ages from 20 to 34.

1.3.4 Internal migrations

Two migration streams are important to account for under the heading internal migration: the migration stream from Greater London to the Rest of the UK and that from the Rest of the UK to Greater London. Estimates of the size of these streams are shown in Table 1.2.4 derived from data collected for the Office of Population Censuses and Surveys in the National Health Service Central Register.

Each internal inter-regional migration flow plays a dual role. It is a debit from the population of the origin region and a credit to the population of the destination region.

1.3.5 Immigrations

Also very important as additions to the population are migrations from abroad. Table 1.2.5 gives estimates of the numbers of immigrations to the two regions.

Interestingly, for this two region system, the number of external migrations (1.931 millions) exceeds that of internal migrations (1.824 millions). Of course, there is a much larger number of migrations over this period whose origins and destinations

TABLE 1.2 Components of the movement accounts input as data

1.2.1 Initial Populations (MY 1976)

Cohort	Age Group	Greater London	Rest of the UK
1	Birth	-	-
2	0- 4	419,900	3,319,100
3	5- 9	491,900	3,954,300
4	10-14	518,700	4,134,600
5	15-19	487,700	3,748,300
6	20-24	497,300	3,373,900
7	25-29	573,200	3,621,700
8	30-34	480,200	3,108,200
9	35-39	413,500	2,795,900
10	40-44	397,800	2,729,500
11	45-49	415,500	2,840,600
12	50-54	449,800	2,976,600
13	55-59	422,800	2,723,100
14	60-64	423,200	2,685,200
15	65-69	379,200	2,459,600
16	70-74	287,400	1,953,100
17	75-79	189,100	1,304,700
18	80-84	108,300	722,400
19	85+	72,800	451,100
	Total	7,028,300	48,901,900

1.2.2 Deaths (MY 1976-MY 1981)

Cohort	Initial Age Group	Final Age Group	Greater London	Rest of the UK
1	Birth	0- 4	5,632	39,101
2	0- 4	5- 9	1,343	9,940
3	5- 9	10-14	569	4,876
4	10-14	15-19	1,038	7,821
5	15-19	20-24	1,802	12,945
6	20-24	25-29	1,689	11,348
7	25-29	30-34	2,060	12,685
8	30-34	35-39	2,478	16,667
9	35-39	40-44	3,225	23,381
10	40-44	45-49	5,452	40,557
11	45-49	50-54	10,036	75,384
12	50-54	55-59	17,775	130,775
13	55-59	60-64	27,155	197,682
14	60-64	65-69	38,988	285,791
15	65-69	70-74	55,445	405,566
16	70-74	75-79	65,078	487,200
17	75-79	90-84	62,680	468,911
18	80-84	85-89	51,984	369,029
19	85+	90+	53,842	354,429
	Total		409,271	2,954,158

TABLE 1.2 continued

1.2.3 Emigrations (MY 1976-MY 1981)

Cohort	Initial Age Group	Final Age Group	Greater London	Rest of the UK
1	Birth	0- 4	7,815	28,072
2	0- 4	5- 9	12,515	48,936
3	5- 9	10-14	10,276	44,109
4	10-14	15-19	12,724	60,131
5	15-19	20-24	28,804	101,776
6	20-24	25-29	52,568	136,443
7	25-29	30-34	51,687	130,579
8	30-34	35-39	29,804	83,999
9	35-39	40-44	16,766	50,065
10	40-44	45-49	10,343	30,694
11	45-49	50-54	6,836	19,001
12	50-54	55-59	6,949	17,643
13	55-59	60-64	4,917	11,475
14	60-64	65-69	2,635	5,706
15	65-69	70-74	1,974	4,723
16	70-74	75-79	1,211	3,490
17	75-79	80-84	792	2,411
18	80-84	85-89	415	1,227
19	85+	90+	260	730
Total			259,291	781,210

1.2.4 Internal Migrations (MY 1976-MY 1981)

From Greater London to: Rest of the UK	From Rest of UK to: Greater London
34,800	23,070
63,705	40,235
52,090	30,425
64,881	59,159
138,394	141,931
188,928	165,090
152,815	112,759
88,834	58,730
49,706	33,591
36,507	24,338
32,056	19,744
32,745	16,585
33,880	12,764
33,823	10,862
25,794	9,222
12,149	5,694
6,647	3,470
3,593	1,823
2,128	1,049
1,053,475	770,541

1.2.5 Immigrations (MY 1976-MY 1981)

Greater London	Rest of the UK	Cohort
8,758	24,578	1
14,598	45,216	2
11,082	40,956	3
21,268	54,695	4
52,746	90,884	5
65,904	101,907	6
47,142	83,677	7
24,564	54,397	8
13,970	32,219	9
9,256	21,475	10
6,003	14,777	11
5,040	14,150	12
2,766	9,111	13
1,159	4,971	14
989	4,020	15
790	2,799	16
573	1,929	17
304	998	18
192	598	19
287,104	603,356	

TABLE 1.2 continued

1.2.6 Births (MY 1976-MY 1981)

Cohort	Initial Age Group	Final Age Group	Greater London	Rest of the UK
1	Birth	0- 4	0	0
2	0- 4	5- 9	0	0
3	5- 9	10-14	0	0
4	10-14	15-19	11,337	91,791
5	15-19	20-24	87,019	679,522
6	20-24	25-29	145,881	1,089,354
7	25-29	30-34	129,616	870,172
8	30-34	35-39	54,367	319,848
9	35-39	40-44	13,059	72,321
10	40-44	45-49	2,267	13,283
11	45-49	50-54	0	0
12	50-54	55-59	0	0
13	55-49	60-64	0	0
14	60-64	65-69	0	0
15	65-69	70-74	0	0
16	70-74	75-79	0	0
17	75-79	90-84	0	0
18	80-84	85-89	0	0
19	85+	90+	0	0
Total			443,546	3,136,291

1.2.7 Final Populations (MY 1976-MY 1981)

Greater Age Group	Greater London	Rest of the UK
0- 4	397,000	3,055,100
5- 9	393,018	3,300,582
10-14	466,182	3,978,113
15-19	534,146	4,143,754
20-24	577,292	3,617,807
25-29	526,352	3,280,740
30-34	520,796	3,604,604
35-39	425,498	3,179,403
40-44	386,607	2,820,893
45-49	377,075	2,731,426
50-54	396,047	2,816,455
55-59	413,079	2,899,919
60-64	372,899	2,607,002
65-69	353,464	2,453,836
70-74	302,936	2,103,564
75-79	215,036	1,491,263
80-84	124,064	841,335
85-89	55,814	365,585
90+	24,086	146,316
Total	6,851,401	49,437,702

are wholly within either Greater London or the Rest of the UK.

1.3.6 Births

In each period a fresh set of people, infants born during the time interval, are added to the population, and we require estimates of these. Table 1.2.6 displays the numbers of births in the two regions, disaggregated by age of the mother at time of maternity.

1.3.7 Final populations

The components of population change already listed above are sufficient to determine the net outcome that is, the population at the end of the period. However, if we have independent knowledge of these final populations it is useful to include them in our information system as a check on the accuracy of our other estimates. Table 1.2.7 sets out mid-year estimate populations for 1981 for Greater London and the Rest of the UK based on the Census of April 1981.

1.3.8 Data files on the computer

All of these data on the components of change must be made accessible to the computer program that is going to "process" them into sets of population accounts. Figure 1.2 shows listings of the contents of some seven disk files into which the data have been entered allowing 10 columns for each item and right justifying the numbers within these 10 column "fields". The ordering and layout of the numbers is exactly the same as that in the tables with two exceptions. Firstly, in the deaths, emigrations, internal migrations, immigrations and final populations files, the figures for the first cohort are placed last for computational reasons. The second exception is in the internal migrations file which is arranged as a series of two by two matrices with zeroes in the diagonal for within region migrations.

1.4 A model for constructing movement accounts

However, just supplying the the accounts program with data on the population stocks and flows is not enough. We must also tell the program what system the data refer to and what the program should do with the data. What does the program need to know?

Four sets of information must be supplied.

(i) Firstly, you need to describe in detail the system being studied, the model being used and the projections to be carried out so that the printout of the results of the exercise can be labelled unambiguously for future reference and use.

(ii) You need to supply the program with information about how many regions you are studying, how many time periods are involved, how many age groups there are and so on. Some 12 items of information are required in all.

(iii) The program requires the names of your regions.

(iv) You need to specify the type of model that you want the program to use in constructing the accounts. There are some fourteen characteristics of the model that need to be specified. A model needs to be designed for each period studied.

FILE: DISGL1 PIDATA A

419900	3319100
491900	3954300
518700	4114600
487700	3748300
497300	3371900
573200	3621700
480200	3108200
413500	2795900
397800	2729500
415500	2840600
449800	2976600
422800	2723100
423200	2685200
379200	2459600
287400	1953100
189100	1304700
108300	722400
72800	451100

FILE: DISGL1 EDATA A

12515	44535
10276	44105
12724	50131
28804	101776
52568	115443
51687	130579
29804	97499
16766	50065
10343	30644
6635	13001
6949	17563
4917	11475
2635	5796
1974	4723
1211	1490
792	2411
415	1227
260	730
7815	29072

FIGURE 1.2

<u>Listings of data files input to the movement accounts program</u>

FILE: DISGL1 DDATA A

1343	9940
569	4576
1038	7821
1802	12945
1689	11348
2060	12685
2478	16667
3225	23381
5452	40557
10036	75384
17775	130775
27155	197682
38988	285791
55445	405566
65078	487200
62680	453911
51984	369099
53842	354429
5632	19101

FILE: DISGL1 MDATA A

0	53705
40235	0
0	52090
30425	0
0	54981
59159	0
0	133394
141931	0
0	108928
165090	0
0	152915
112759	0
0	88834
58730	0
0	49706
33591	0
0	16507
24335	0
0	32056
13744	0
0	32745
16585	0
0	11890
12764	0
0	13823
10862	0
0	25794
9222	0
0	12149
5694	0
0	6647
3470	0
0	1593
1823	0
0	2128
1049	0
0	14900
23070	0

FILE: DISGL1 IDATA A

14598	45216
11082	40956
21260	54695
52746	90684
65904	101907
47142	53677
24564	54397
13970	32219
9254	21475
6003	14777
5040	14150
2766	9111
1159	4971
389	4020
790	2749
573	1928
304	994
192	598
8758	24578

FILE: DISGL1 PFDATA A

383018	3300582
466182	3978118
534146	4143754
577292	3617807
526162	3290740
520776	3604604
425498	3179403
386607	2820893
377075	2711426
396047	2815455
413079	2999319
372899	2607032
353454	2453936
302936	2103554
215036	1431263
124064	841135
55814	345585
24086	146316
397000	1055100

FILE: DISGL1 BDATA A

11337	91791
87019	679522
145881	1099354
129616	870172
54367	319848
13059	72321
2267	13283
1000	19000

FIGURE 1.2

<u>Continued</u>

All this information is assembled in another disk file. Figure 1.3 shows what the one for our Greater London/Rest of the UK system looks like. The first four lines of the file are set aside for the general identification of the analysis. The next three lines contain the information about the characteristics of the system being studied. For example, "NINREG= 2" means that the number of regions internal to the system (that is, within our country) will be 2. Similarly, "NSEX= 1" indicates that we are working with a single sex population of "persons". The eighth line of the file contains the abbreviated names of the regions. The ninth to twelfth lines of the file supply the model characteristics for the first period of analysis. For example, "IPOP= 1" means that the program should expect to find the initial population on a separate disk file. Later on in the instructions you will find that "IPOP= 0": this means that the program should use the final populations of the previous period as the initial populations of the current period. The set of "I-parameter" instructions, thirteen in all over 4 lines, is repeated six times in the file, once for each period of analysis.

1.5 Producing movement accounts and projections

Having assembled the files of demographic data and program instructions, what do you do to produce sets of movement accounts and associated projections? Very simply, you run the program.

On the University of Leeds Amdahl mainframe under the CMS operating system this boils down to
(i) connecting your userid to the Geography permanent library disk
(ii) issuing the command

MOVE DISGL1 20 225

MOVE is the name of an EXECutive file that contains the necessary instructions for accessing the movement accounts program and for connecting it to your data files, your instruction file and your output files. Figure 4.12 later gives a listing of this EXEC file. You need not worry what all the lines mean, unless soemthing goes wrong. More details are given in section 4 on this.

DISGL1 is the name of the file containing the instructions to the program about your particular piece of spatial population analysis, which we have just discussed. DISGL1 stands for "Demographic Information System for Greater London run 1". The number "20" that follows the filename indicates to the program the record length in bytes of the component data files. This will depend on the number of regions being studied. For 2 regions the data files will be of length "20". The number "225" that follows is the number of records needed on a direct access file that is used for temporary data storage by the program. This number is again a function of the number of (internal) regions in your system. For 2 regions you need 225 records, for 20 regions you need 720 records, for example.

What happens then? If everything is set up correctly, you should see the lines of the EXEC file listed on your terminal. This listing will be followed by the message "EXECUTION BEGINS ..." and then on the terminal will appear messages informing you about the

```
FILE: DISGL1    DATA      A    LEEDS UNIVERSITY VM/SP 2.05

DEMOGRAPHIC INFORMATION SYSTEM FOR GREATER LONDON   PHIL REES SEPT 1983
MOVEMENT ACCOUNTS AND PROJECTIONS USING NHSCR MIGRATION DATA FOR 1976-81
MODEL - UNCONSTRAINED FORECAST ABM 1976-81   CONSTANT RATES PROJECTION
PROJECTION  5 PERIODS 1981-2006   CLD RATES 1981 POPULATION BASE
 NINREG=    2 NPRIOD=    6 NCOHRT=   19    NSEX=    1
  NITER=   10   NTAB=    0  NYEAR=1976 NTIMIN=    5
 NAGEIN=    5  NPSEX=    1  NALPA=    4   NBETA=   10
    GL       RUK
   IPOP=   1    IPAR=   1    IDTH=   1   IEMIG=   1 INTMIG=   1
   IRES=   0    IMIG=   1   IPPOP=   0   IVERG=   1 IACCNT=   1
   ICON=   0    ITOL=   0 ITERBP=    0   ISURV=   0 IBIRTH=   1
 ISEXP=    1  IAUTO=    1
   IPOP=   1    IPAR=   1    IDTH=   2   IEMIG=   2 INTMIG=   2
   IRES=   0    IMIG=   2   IPPOP=   0   IVERG=   1 IACCNT=   1
   ICON=   0    ITOL=   0 ITERBP=    0   ISURV=   0 IBIRTH=   2
 ISEXP=    0  IAUTO=    1
   IPOP=   0    IPAR=   1    IDTH=   2   IEMIG=   2 INTMIG=   2
   IRES=   0    IMIG=   2   IPPOP=   0   IVERG=   1 IACCNT=   0
   ICON=   0    ITOL=   0 ITERBP=    0   ISURV=   0 IBIRTH=   2
 ISEXP=    0  IAUTO=    0
   IPOP=   0    IPAR=   1    IDTH=   2   IEMIG=   2 INTMIG=   2
   IRES=   0    IMIG=   2   IPPOP=   0   IVERG=   1 IACCNT=   0
   ICON=   0    ITOL=   0 ITERBP=    0   ISURV=   0 IBIRTH=   2
 ISEXP=    0  IAUTO=    0
   IPOP=   0    IPAR=   1    IDTH=   2   IEMIG=   2 INTMIG=   2
   IRES=   0    IMIG=   2   IPPOP=   0   IVERG=   1 IACCNT=   0
   ICON=   0    ITOL=   0 ITERBP=    0   ISURV=   0 IBIRTH=   2
 ISEXP=    0  IAUTO=    0
   IPOP=   0    IPAR=   1    IDTH=   2   IEMIG=   2 INTMIG=   2
   IRES=   0    IMIG=   2   IPPOP=   0   IVERG=   1 IACCNT=   0
   ICON=   0    ITOL=   0 ITERBP=    0   ISURV=   0 IBIRTH=   2
 ISEXP=    0  IAUTO=    0
```

FIGURE 1.3 Instructions to the program about the system being studied and the accounts model to be employed progress of the calculations:

```
ACCOUNTS FOR COHORT =  2  FINISHED
ACCOUNTS FOR COHORT =  3  FINISHED
          .
          .
          .
ACCOUNTS FOR COHORT = 19  FINISHED
ACCOUNTS FOR COHORT =  1  FINISHED
ACCOUNTS FOR PERIOD =  1  COMPLETED
ACCOUNTS FOR COHORT =  2  FINISHED
```

and so on until the message

MOVEMENT ACCOUNTS FOR 6 PERIODS COMPLETED

appears followed by

EXECUTION ENDS

These messages are to keep you from worrying as the computation progresses. The elapse time for the execution of the program can extend over several minutes depending on how fast and how busy your computer is.

1.6 The tables of movement accounts

So where, you ask, are the results of your fine efforts to date? The results have been written to two files. The first contains the accounts tables for each cohort for the periods requested; the second contains tabulations of the projected populations.

Accounts tables for selected cohorts are given in Table 1.3 for our Greater London example. The first table is for the first cohort to appear on the printout: that for persons aged 0-4 at mid-year 1976 who are aged 5-9 five years later, assuming survival. From the initial Greater London population of 419,900 in this cohort are subtracted 1,343 deaths, 12,515 emigrations, 63,705 out-migrations to the Rest of the UK, leaving a residual term of 342,337 to which are added 40,235 in-migrations to Greater London from the Rest of the UK and 14,598 immigrations from abroad. Figure 1.4 shows these as inputs (stippled shading) and outputs (pecked shading) to the second Greater London projection cohort.

Table 1.3.2 presents the accounts for the sixth cohort, the one exhibiting maximum movement, and Table 1.3.3 shows the accounts for the last cohort where mortality dominates. The accounts tables for cohorts that exist at the start of the period are followed by that for those born in the period (Table 1.3.4). Finally, Table 1.3.5 summarizes the movements involved in an aggregate accounts table summed over all ages. Note that from these accounts can be computed the simpler net components of change. Table 1.4 shows those tables for the same cohorts as Table 1.3. The summary table shows that over the period that Greater London lost population, due to a net migration loss to the Rest of the UK only partially compensated by a natural increase and a gain from external migration. Conversely, the Rest of the UK gained in population through net internal

TABLE 1.3 Movement accounts tables for selected cohorts, 1976–81

TABLE 1.3.1

THE MOVEMENT ACCOUNTS MATRIX
PERIOD = 1 COHORT = 2 AGES = 0- 4 TO 5- 9 SEX = 1 ITERATION = 2

	GL 1	RUK 2	EMIGRTNS 3	DEATHS 4	TOTALS 5
GL	342337.	63705.	12515.	1343.	419900.
RUK	40235.	3219989.	48936.	9943.	3319100.
IMMIGTNS	14598.	45216.	0.	0.	59814.
TOTALS	397170.	3328910.	61451.	11283.	3798814.

TABLE 1.3.2

THE MOVEMENT ACCOUNTS MATRIX
PERIOD = 1 COHORT = 6 AGES = 20-24 TO 25-29 SEX = ITERATION = 2

	GL 1	RUK 2	EMIGRTNS 3	DEATHS 4	TOTALS 5
GL	254115.	188928.	52568.	1689.	497300.
RUK	165090.	3061019.	136443.	11349.	3373900.
IMMIGTNS	65904.	101907.	0.	0.	167811.
TOTALS	485109.	3351854.	189011.	13037.	4039011.

TABLE 1.3.3

THE MOVEMENT ACCOUNTS MATRIX
PERIOD = 1 COHORT = 19 AGES = 85-89 TO 90-94 SEX = 1 ITERATION = 2

	GL 1	RUK 2	EMIGRTNS 3	DEATHS 4	TOTALS 5
GL	16570.	2128.	260.	53842.	72800.
RUK	1049.	94892.	730.	354429.	451100.
IMMIGTNS	192.	598.	0.	0.	790.
TOTALS	17811.	97618.	990.	408271.	524690.

TABLE 1.3.4

THE MOVEMENT ACCOUNTS MATRIX
PERIOD = 1 COHORT = 1 AGES = -5--1 TO 0- 4 SEX = 1 ITERATION = 2

	GL 1	RUK 2	EMIGRTNS 3	DEATHS 4	TOTALS 5
GL	395299.	34800.	7815.	5632.	443546.
RUK	23070.	3046048.	28072.	39101.	3136291.
IMMIGTNS	8758.	24578.	0.	0.	33336.
TOTALS	427127.	3105426.	35887.	44733.	3613173.

TABLE 1.3.5 Movement accounts for all ages, 1976–81

	State before move / State after move	Greater London	Rest of UK	Rest of World	Death	Totals
Existing cohorts	Greater London	5,355,510	1,018,675	402,639	402,639	7,028,500
	Rest of UK	747,471	44,486,234	753,138	2,915,057	48,901,900
	Rest of World	278,346	578,778	0	0	857,124
Infant cohort	Greater London	395,299	34,800	7,815	5,632	443,546
	Rest of UK	23,070	3,046,048	28,072	39,101	3,136,291
	Rest of World	8,758	24,578	0	0	33,336
	Totals	6,808,454	49,189,113	1,040,501	3,362,429	60,400,497
		6,851,401	49,437,702			

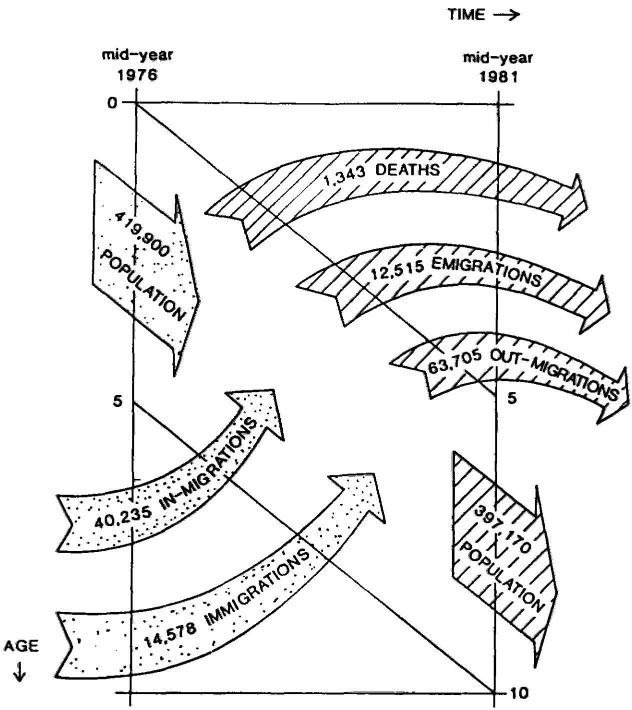


FIGURE 1.4 Inputs to and outputs from the second cohort populations of Greater London, 1976-81

TABLE 1.4 Net components of change, 1976-81

1.4.1 Cohort 2: 0-4 to 5-9

Region	Initial populations	Deaths	Net migration		Final populations
			Internal	External	
Greater London	419,900	-1,343	-23,470	2,083	397,170
Rest of the UK	3,319,100	-9,940	23,470	-3,720	3,328,910
UK	3,739,000	-11,283	0	-1,637	3,726,080

1.4.2 Cohort 6: 20-24 to 25-29

Region	Initial populations	Deaths	Net migration		Final populations
			Internal	External	
Greater London	497,300	-1,689	-23,838	13,336	485,109
Rest of the UK	3,373,900	-11,348	23,838	-34,536	3,351,854
UK	3,871,200	-13,037	0	-21,200	3,836,963

1.4.3 Cohort 1: Birth to 0-4

Region	Initial populations	Deaths	Net migration		Final populations
			Internal	External	
Greater London	443,546	-5,632	-11,730	943	427,127
Rest of the UK	3,136,291	-39,101	11,730	-3,494	3,105,426
UK	3,579,837	-44,733	0	-2,551	3,532,553

1.4.4 All ages

Region	Initial populations	Births	Deaths	Net migration		Final populations
				Internal	External	
Greater London	7,028,300	443,546	408,271	-282,934	27,813	6,808,454
Rest of the UK	48,901,900	3,136,291	2,954,158	282,934	-177,854	49,189,113
UK	55,930,200	3,579,837	3,362,429	0	-150,041	55,997,567

migration and natural increase but lost substantially through net external loss.

1.7 Tables of projected populations

These accounts tables can be produced for future time periods if you wish, but normally you would be more interested in the population stocks. So,these are gathered together in one file as the main accounts program moves through the projection periods (Table 1.5). The user can arrange the projected populations in suitable tables of statistics by writing his or her own tabulation program that use these data as input.

TABLE 1.5 Projected populations, Greater London, 1981-2006

Age group	1981	1986	1991	1996	2001	2006
0-4	397.0	449.4	442.0	417.4	381.3	357.8
5-9	383.0	374.6	421.5	417.0	395.2	361.5
10-14	466.2	367.9	358.7	402.2	399.2	379.2
15-19	534.1	471.3	374.3	362.8	404.3	403.7
20-24	577.3	563.6	507.6	408.0	391.2	430.5
25-29	526.4	551.0	561.4	514.7	417.9	397.1
30-39	520.8	482.3	509.6	529.7	489.6	399.3
35-39	425.5	483.7	446.7	474.2	498.2	462.2
40-44	386.6	406.1	461.3	425.3	452.8	478.5
45-49	377.1	369.8	390.0	442.7	407.8	434.9
50-54	396.0	357.0	350.9	371.1	421.1	387.5
55-59	413.1	365.5	330.2	324.9	344.5	390.7
60-64	372.9	364.9	323.4	292.7	288.4	306.3
65-69	353.5	317.9	311.6	276.7	250.7	247.3
70-74	302.9	286.0	257.7	253.0	225.0	204.2
75-79	215.0	227.2	214.7	193.7	190.5	169.6
80-84	124.1	139.9	147.9	139.8	126.3	124.4
85-89	55.8	62.4	70.4	74.4	70.4	63.7
90+	24.1	19.9	20.6	22.7	24.2	23.7
Totals	6,851.4	6,660.4	6,500.5	6,343.0	6,178.6	6,022.1

2. THE THEORY OF MOVEMENT ACCOUNTS

2.1 Movement accounts

Movement accounts are tables which show how the population moves between states of interest over a single period. The movement accounts table which uses counts of moves between regions over a fixed time interval for a type II, projection or period cohort age-time observation plan is set out in Table 2.1.

Different letters are used to represent the various components of population change. There are some 8 components in all. Initial populations are represented as P(i,*,x) where the * indicates summation over all subsequent states and the region i and age x labels refer to state at the start of the time interval. From these populations are subtracted regional deaths, D(i,x), in the projection cohort x, regional emigrations, M(i,O,x), where i represents the region prior to emigration and O the outside world (outside the system or country of interest), and out-migrations M(i,j,x) to other internal regions j. These subtractions yield an account balancing item, R(i,x), which we place in the principal diagonal of the table that starts at the top left. These balancing terms are defined simply as

$$R(i,x) = P(i,*,x) - D(i,x) - M(i,O,x) \quad \sum_{j \neq i} M(i,j,x) \qquad (2.1)$$

To these balancing terms are added the in-migrations from other regions, M(i,j,x), and the immigrations, M(O,i,x), from the outside world to yield the population, P(*,i,x), in the various regions:

$$P(*,i,x) = R(i,x) + \sum_{j \neq i} M(j,i,x) + M(O,i,x) \qquad (2.2)$$

2.2 Meaning of the diagonal term

In other accounting frameworks, for example, the transition framework, the diagonal term in the corresponding position in the table (the exist-survive portion of the accounts matrix) consists of surviving stayers - that is, persons who survive the time interval and start and finish it in the same region. In migration matrices, in which the concern is with predicting the volume of migration "trips" between origins and destinations, the diagonal terms are migrations that occur within each region.

In the movement accounting framework no such substantive meaning can be attached to the diagonal term. In fact, in extreme cases negative values can occur whereas in the case of transition accounts and migration matrices there is a lower bound of zero (no surviving stayers, no intraregion migrations). This lack of a substantive meaning may be a little difficult to accept so some additional justification is provided.

Suppose, for example, that Britons were 2.5 times more mobile than they currently are observed to be. This would make them just a little more mobile than Americans. The movement accounts table for Greater London and the Rest of the UK for the period cohort 20-24 to 25-29 would then look like Table 2.2. The diagonal contains

TABLE 2.1 A movement accounts table for cohort x in one time interval

State before move		State after move Destinations: Internal regions 1	2	N	Outside world O	Death	Totals
Origins	1	R(1,x)	M(1,2,x)	M(1,N,x)	M(1,O,x)	D(1,x)	P(1,*,x)
Internal regions	2	M(2,1,x)	R(2,x) .	M(2,N,x)	M(2,O,x)	D(2,x)	P(2,*,x)
			.				
	N	M(N,1,x)	M(N,2,x)	R(N,x)	M(N,O,x)	D(N,x)	P(N,*,x)
Outside World	O	M(O,1,x)	M(O,2,x) .	M(O,N,x)	ø	ø	M(O,*,x)
Totals		P(*,1,x)	P(*,2,x) . .	P(*,N,x)	M(*,O,x)	D(*,x)	T(*,*,x)

Notes

1 The infant accounts, for x = -n, follow the same structure except that variables B(1,*) .. B(N,*) replace P(1,*,-n)... P(N,*,-n)

2 The x subscript refers to a "projection cohort" occupying the age-time space x,t, d+n,t; x+n,t+T; x+2n, t+T in figure 1.1 (type II observation plan).

Definitions

POPULATION

P(i,*,x) population in region i aged x to x+n at start of time interval

P(*,j,x) population in region j aged x+n to x+2n at end of time interval

DEATHS

D(i,x) = deaths in region i to persons in cohort x during time interval

D(*,x) total deaths in all internal regions

EMIGRATIONS

M(i,O,x) = emigrations from region i by persons in chort x to the outside world, O, during the time interval

M(*,O,x) = total emigrations from all internal regions

INTERNAL MIGRATIONS

M(i,j,x) = migrations from region i to region j by persons in cohort x during the time interval

RESIDUAL TERMS

R(i,x) . residual accounts balancing term for region i

IMMIGRATIONS

M(O,j,x) = immigrations from outside world, O, to region j by persons in chort x during the time interval

M(O,*,x) = total immigrations to all internal regions

T(*,*,x) = total of all inflows/inputs to the internal system
= total of all outflows/outputs from the internal system

TABLE 2.2 Movement accounts for the 20-24 to 25-29 cohort, assuming mobility levels 2½ times those observed.

State before move \ State after move		Destinations: Greater London	Rest of UK	Rest of World	Death	Totals
Origins	Greater London	-108.1	472.3	131.4	1.7	497.3
	Rest of UK	412.7	2,608.8	341.1	11.3	3,373.9
	Rest of World	164.8	254.8	0	0	419.6
	Totals	469.4	3,335.9	472.5	13.0	4,290.8

Note 1. The numbers are in 1000's

negative entries.

There is, however, a connection between the movement accounts diagonal term and that of transition accounts:

$$R(i,x) = K(i,i,x) - \sum_{j \neq i} M(i,j,x)(\text{surplus}) - M(i,0,x)(\text{surplus}) \quad (2.3)$$

where $K(i,i,x)$ are the numbers of surviving stayers in period cohort x and the "(surplus)" postscript means the migrations in excess of those needed to account for the change in location between the start and finish of a period of persons in a transition account.

2.3 Accounting models for historical or base period accounts

How do we construct sets of movement accounts? We simply put into the accounts table the items we know or can estimate from available official statistics, and use accounting relationships to work out missing elements. If we have surplus information we must decide on what data to believe and what to alter. Thus, there are a variety of accounting models possible. Figure 2.1 sets out some of the main alternatives, which are now described.

2.3.1 The forecast model

The following components are entered in the accounts (Figure 2.1.1): initial populations, $P(i,*,x)$, deaths, $D(i,x)$, emigrations, $M(i,0,x)$, for all regions i, and internal out-migrations, $M(i,j,x)$, for all regions i of origin and j of destination. Then the residual balancing terms, $R(i,x)$, are computed using equation (2.1). The internal in-migrations, $M(j,i,x)$, are added to the balancing items, as are immigrations, $M(0,i,x)$, as in equation (2.2) to yield estimates of the end of period population, $P(*,i,x)$. The model is labelled forecast because it works forward from the initial population, and is the one that will be used in a variety of ways, in modified form, in forecasting the population.

2.3.2 The backcast model

In the backcast model (Figure 2.1.2) we start at the opposite end of the time interval with the final populations. The immigrations and internal in-migrations are subtracted from the final populations to give an estimate of the balancing terms

$$R(i,x) = P(*,i,x) - M(0,i,x) - \sum_{j \neq i} M(j,i,x) \quad (2.4)$$

To these are then added the internal out-migrations, emigrations and deaths to yield an estimate of the initial populations in the accounting time interval:

$$P(i,*,x) = R(i,x) + \sum_{j \neq i} M(i,j,x) + M(i,0,x) + D(i,x) \quad (2.5)$$

This model is likely to be used when the final population derives from a recent, accurate census and when the only information available for the initial population is a rather unreliable mid-decade estimate. In the case of the first cohort, the backcast

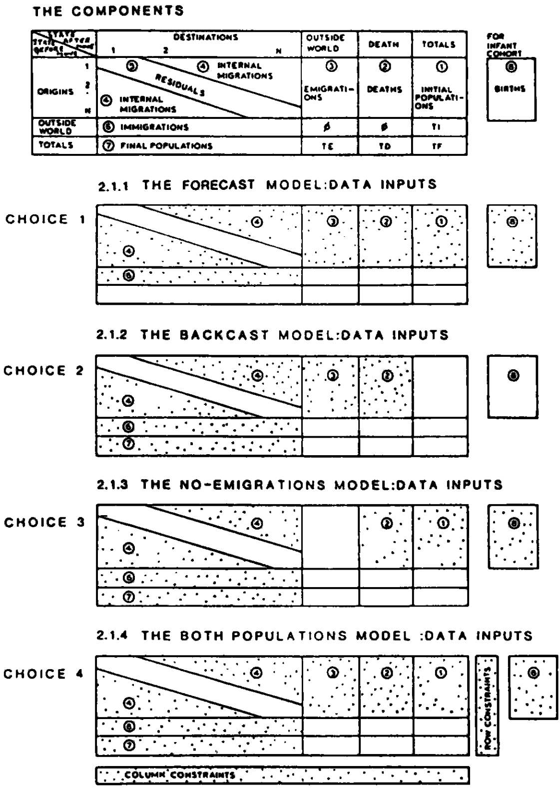


Figure 2.1 The different ways of assembling movement accounts

model would be used to make an estimate of the number of births in the time interval

$$B(i,-n) = R(i,-n) + \sum_{j \neq i} M(i,j,-n) + M(i,0,-n) + D(i,-n) \quad (2.6).$$

However, for developed countries, as good estimates should be available for births as the migration and deaths flows in equation (2.6) so a forecast model might be used instead.

2.3.3 The no-emigration model

In this model all data components for the accounts are known except for emigrations (Figure 2.1.3). Equation (2.4) is used first to derive the residual balancing terms and then emigrations are computed as residuals

$$M(i,0,x) = P(i,*,x) - D(i,x) - \sum_{j \neq i} M(i,j,x) - R(i,x) \quad (2.7).$$

However, this model is likely to be used rather rarely in the movement accounts context as the registration systems or survey methods which provide immigration estimates are equally likely to provide emigration estimates. It is rather more applicable when transition accounts are being constructed.

2.3.4 The both populations model

A final situation is the one (Figure 2.1.4) in which all components, apart from the balancing terms, are known. This might apply when the initial and final populations derive from censuses five years apart.

Two equations are available to estimate the residual terms, either equation (2.3) (the forecast equation) or equation (2.4) (the backcast equation). If either is used, the accounts table will be inconsistent in the sense that either the elements in the interior fail to sum to the column totals or they fail to sum to the row totals.

2.3.5 Constrained accounts

The accounts elements can be made consistent through adjustment. Let us assume the accounts elements have been estimated using one of the residual equations and can be represented as terms A(i,j,x) referring to row i and column j of the accounts matrix. We then compose row constraints, V(i,x), from the known initial populations, P(i,x), and the immigrations total, the sum of M(0,j,x) over j

$$V(i,x) = P(i,x) \quad \text{for all } i=1,2, \ldots ,N \quad (2.8)$$

$$V(N+1,x) = \sum_{j=1}^{N} M(0,j,x) \quad \text{for } i=N+1 \quad (2.9)$$

where N is the number of regions internal to the system. Column constraints, W(j,x), are composed of the final populations,

$P(*,j,x)$, total emigrations, $M(*,O,x)$, and total deaths, $D(*,x)$:

$$W(j,x) \quad = \quad P(*,j,x) \quad \text{for } j=1,2, \ldots ,N \tag{2.10}$$

$$W(N+1,x) = \sum_{i=1}^{N} M(i,O,x) \tag{2.11}$$

$$W(N+2,x) = \sum_{j=1}^{N} D(j,x) \tag{2.12}$$

Constrained re-estimates of the accounts elements are obtained by iteration through the following set of equations

$$A^{q+1}(i,j,x) = A^{q}(i,j,x) \quad (V(i,x)/\sum_j A^{q}(i,j,x)\ ) \tag{2.13}$$

$$A^{q+2}(i,j,x) = A^{q+1}(i,j,x) \quad (W(j,x)/ \sum_j A^{q+1}(i,j,x)) \tag{2.14}$$

with q indicating the step of the procedure and $A^{q+2}(i,j,x)$ values bein reset to $A^{q}(i,j,x)$ at each iteration until all elements differ by less than a small amount, say "half an event" from one iteration, p, to the next, p+1, that is:

$$\text{absolute value of } (\ A^{p+1}(i,j,x) \quad A^{p}(i,j,x)\ ) \ < 0.5 \tag{2.15}$$

for all i and j from 1 to N+1, and for j = N+2.

Now this set of equations has a convergent solution only when the sum of the row constraints is equal to the sum of the column constraints:

$$\sum_i V(i,x) \quad = \quad \sum_j W(j,x) \tag{2.16}$$

Otherwise the estimates will not converge.

If equation (2.16) does not hold then one or more elements of the constraints set must be adjusted. The problem we have is how to distribute the difference or error, E, involved

$$E \ = \sum_i V(i,x) \qquad \sum_j W(j,x) \tag{2.17}.$$

There are a large number of possible ways of distributing E so that equation (2.16) is satisfied. We could, for example,

(i) distribute E proportionately across the $V(i,x)$ elements, the row constraints;
(ii) distribute E proportionately across the $W(j,x)$ elements, the column constraints;

(iii) distribute half of E to each of the row and column constraints; or one could
(iv) load the error on only one component term of either set of constraints that was felt to be unreliable.

All sorts of intermediate possibilities exist. In the end distribution of the error must depend on assessment of the degree of likely error in the terms that make up the row and column constraints. In constructing many sets of transition accounts for British or French regions I have always laid the blame at the door of the emigration terms which could only be crudely estimated. At times the adjustments needed can be drastic (see Rees and Convey, 1984).

The blame is more difficult to allocate in the case of movement accounts. There is also the danger of circularity in adjustment procedure. Ideally, the population estimates used for initial and final populations should be independent. Frequently, however, they are not because final populations may have already been estimated using a set of accounting equations. If row and column constraints do not agree in this situation either some error has been made by the office charged with estimating area populations or some change has been made between the two dates (from enumerated to usually resident, for example) or some special category of the population has been treated differently at the two points in time.

2.4 Accounting models for projection

One of the main purposes of building accounts for multiregional populations is to project those populations into the future. There are many ways of using the information in the accounts as input to a projection model. Examples of some of the more useful will be reviewed systematically in section 6 of the manual. Here we concentrate on converting the forecast accounts model described above into a projection model by measuring rates of "movement" for all the components.

It turns out that a projection model based on movement (or transition) accounts can be specified in either an algebraic, iterative form or in a matrix, non-iterative form. The algebraic iterative form is very convenient as the basis of a computer algorithm to project movement accounts, while allowing the user the maximum flexibility in choosing his or her own model. The non-iterative matrix formulation is convenient for derivation of long run properties (stability, stationarity, convergence) of the population system.

2.4.1 An iterative, accounts based model for projection

The first step is to define the rates equivalent to each of the components of the accounts. To accomplish this we need a specification for the population at risk (or the person-years lived during the observation interval by individuals at risk of experiencing the event of interest divided by the length of the time interval). If we assume that demographic events are evenly distributed over the time interval then we should use the average population of the time interval as the population at risk, PAR(i) for region i, cohort x:

$$PAR(i,x) = (1/2)\ (P(i,*,x) + P(*,i,x)) \tag{2.18}$$

We can then straightforwardly compute death rates, d(i,x), using lower case letters for rates

$$d(i,x) = D(i,x)/(1/2)(P(i,*,x) + P(*,i,x)) \tag{2.19}$$

emigration rates, m(i,O)

$$m(i,O,x) = M(i,O)/(1/2)(P(i,*,x) + P(*,i,x)) \tag{2.20}$$

and internal out-migration rates, m(i,j)

$$m(i,j,x) = M(i,j,x)/(1/2)(P(i,*,x) + P(*,i,x)) \tag{2.21}$$

Emigration and internal out-migration rates, computed using equations (2.20) and (2.21), involve the populations of the origin regions and are thus "transmission" rates. These are the rates normally used in projection models.

However, there is the difficulty that when we define immigration rates we don't necessarily know or are unable to make a good estimate of the population of the rest of the world. One alternative is to use the population at risk of the destination region

$$m(O,i,x) = M(O,i,x)/(1/2)(P(i,*,x) + P(*,i,x)) \tag{2.22}$$

These immigration rates are "admission" rather than transmission rates.

How should these movement rates be employed in a multiregional population projection model? All that is needed that the rate equations be rearranged into the equivalent flow equations for all regions i=1, ... , N

$$\begin{aligned}
D(i,x) &= d(i,x)\ (1/2)\ (P(i,*,x) + P(*,i,x)) \\
M(i,O,x) &= m(i,O,x)\ (1/2)\ (P(i,*,x) + P(*,i,x)) \\
M(i,j,x) &= m(i,j,x)\ (1/2)\ (P(i,*,x) + P(*,i,x)) \\
&\quad \text{for all } j=1, \dots, N \text{ except where } i=j \\
M(O,i,x) &= m(O,i,x)\ (1/2)\ (P(i,*,x) + P(*,i,x))
\end{aligned} \tag{2.23}$$

These are the items we need to enter the accounts table to estimate the residual balancing terms

$$R(i,x) = P(i,*,x) - D(i,x) - M(i,O,x) - \sum_{j \neq i} M(i,j,x) \tag{2.24}$$

and the final populations

$$P(*,i,x) = R(i,x) + \sum_{j \neq i} M(j,i,x) + M(O,i,x) \tag{2.25}$$

Note that the M(j,i,x) terms will derive from the multiplication of the migration rates by the populations of regions other than region i

$$M(j,i,x) = m(j,i,x)\ (1/2)\ (P(j,*,x) + P(*,j,x)) \qquad (2.26)$$

However, when we use these equations for the first time the $P(*,i,x)$ terms in the populations at risk are not known. The solution is to enter a first estimate of zero or of $P(i,*,x)$ the first time equation set (2.23) is used, to work out a final population estimate using equation (2.25) and then to return to the top of equation set (2.23) with this new estimate of the final population. The sequence of calculations is repeated until the population at risk estimates cease to change significantly.

The infant cohort (persons born during the time interval) should be computed, in this scheme, after all the "existing" cohorts (that is, those alive at the start of the time interval) have been projected for the time interval. Births will then be projected as

$$B(i,x) = b(i,x)\ (1/2)\ (P(i,*,x) + P(*,i,x)) \qquad (2.27)$$

if a single sex model is being used or as

$$B(i,x) = b(i,x)\ (1/2)\ (P(i,*,x,F) + P(*,i,x,F)) \qquad (2.28)$$

if a female (subscript F) dominant model is being used and if the population of both sexes is being projected. The births then need multiplication by the proportions female to derive numbers born to each sex

$$B(i,x,M) = s(M)\ B(i,x) \qquad (2.29)$$

$$B(i,x,F) = s(F)\ B(i,x) \qquad (2.30)$$

where $s(M)$ is the proportion of births that are male and $s(F)$ the proportion that are female. Equations (2.29) and (2.30) apply to all fertile cohorts $x= a \ldots b$.

The range of cohorts to which equations (2.27) and (2.28) apply will be that which covers the principal fertile age range of 15 to 45. If a five year age and time interval is being used then the youngest fertile cohort will be those aged 10-14 at the start of the time interval and 15-19 at the end, and the oldest might be those aged 45-49 at the start and 50-54 at the end. Or it might be convenient to include the few births before age 15 in the 15-19 to 20-24 cohort and the few births after 45 in the 40-44 to 45-49 cohort. For single year cohorts the first will be 15 to 16, the last 49 to 50, with the same inclusion of earlier and later births.

Finally, the projected births are summed over age

$$B(i,*) = \sum_{x} B(i,x) \qquad (2.31)$$

and these are equivalent to the "initial populations" of the accounts for infant cohorts. The flow equations corresponding to equation set (2.23) are

$$D(i,-n) \quad = d(i,-n)\ (1/2)\ (B(i,*) + P(*,i,-n))$$

$$M(i,0,-n) = m(i,0,-n)\ (1/2)\ (B(i,*) + P(*,i,-n))$$

$$M(i,j,-n) = m(i,j,-n)\ (1/2)\ (B(i,*) + P(*,i,-n))$$

$$M(*,i,-n) = m(0,i,-n)\ (1/2)\ (B(i,*) + P(*,i,-n)) \quad \text{for all regions } i=1, \ldots, N \qquad (2.32).$$

Residual terms for the infant cohort are:

$$R(i,-n) = B(i,*) - D(i,-n) - M(i,0,-n) - \sum_{j \neq i} M(i,j,-n) \qquad (2.33)$$

and final populations are given by

$$P(*,i,-n) \quad R(i,-n) + \sum_{j} M(j,i,-n) + M(0,i,-n) \qquad (2.34)$$

One point to be noted about the rates involved for the infant cohort is that they refer to a time interval of exposure that is only half of that in the "existing" cohorts. If they are to be compared with those for existing cohorts or included in the set of rates to which model schedules are to be fitted then they should be multiplied by 2 for fitting purposes, and the model rates for infants subsequently divided by 2 for use in projection. If infant mortality is being studied in detail the assumption that deaths are evenly distributed between birth and the end of the time interval will be very approximate and suitable modifications should be made.

The last cohort in the accounts is a little different from the others in that it refers to a portion of the age-time diagram that extends beyond the normal parallelogram (Figure 2.2), the pecked area defined by age-time coordinates (z,t), (w,t), (w,t+T) and (z+n,t+T). Here we are interested in projecting the population aged z and over. Thos age group might be "85+" or "90+", for example. Normally, the initial population of the next period in projection is obtained from the final population of the previous period

$$\overset{t}{P}(i,*,x) = \overset{t}{P}(*,i,x-1) \qquad (2.35)$$

where the left hand side t refers to the start of the new time interval and the right hand side t to the end of the old time interval.

In order to make reasonable estimates of the terms in the accounts for the last cohort, the input components must be estimated to one age interval beyond "z+". Final populations are required for age group "z+n+", and reasonable estimates of the other terms from period observations require "z" and "z+n+" age group-period data. So, for example, if z were 85, period deaths for age group 85-89 and 90+ would be needed to make an estimate of the last cohort deaths.

2.4.2 A non-iterative model for projection

Iteration through the projection model equations is not strictly necessary. To obtain a non-iterative model we proceed as follows. First, we replace R(i,x) in equation (2.25) by the right hand side of equation (2.24)

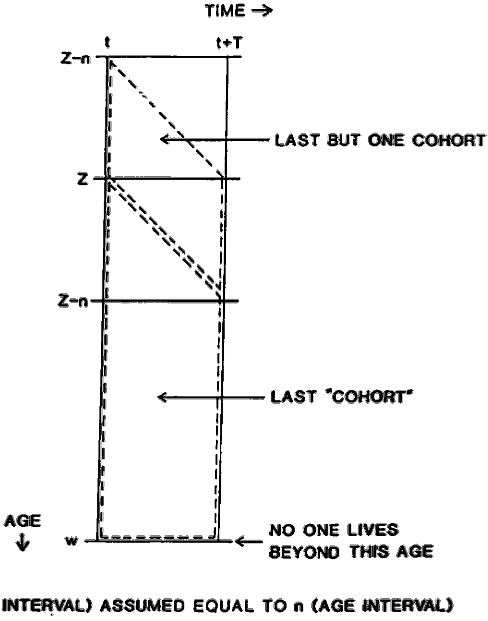


Figure 2.2 <u>Age-time observation plan for the last cohort</u>

$$P(*,i,x) = (P(i,*,x) \quad D(i,x) \quad M(i,O,x) - \sum_{j\neq i} M(i,j,x))$$
$$+ \sum_{j\neq i} M(j,i,x) + M(O,i,x) \qquad (2.36)$$

We then substitute for the flow terms their occurrence-exposure rates multiplied by the population at risk equivalents (that is, the right hand sides of equation set 2.23):

$$
\begin{aligned}
P(*,i,x) = P(i,*,x) - \; & d(i,x) \; (1/2) \; (P(i,*,x) + P(*,i,x)) \\
& m(i,O,x) \; (1/2) \; (P(i,*,x) + P(*,i,x)) \\
- & \sum_{j\neq i} m(i,j,x) \; (1/2) \; (P(i,*,x) + P(*,i,x)) \\
+ & \sum_{j\neq i} m(j,i,x) \; (1/2) \; (P(j,*,x) + P(*,j,x)) \\
+ & \; m(O,i,x) \; (1/2) \; (P(i,*,x) + P(*,i,x))
\end{aligned}
\qquad (2.37)
$$

Collecting together all P(*,i,x) and P(*,j,x) terms on the left hand side of the equation and all terms with P(i,*,x) or P(j,*,x) on the right hand side, the following equation for region i is obtained:

$$
\begin{aligned}
& \{1 + 0.5\, d(i,x) + 0.5\, m(i,O,x) + 0.5 \sum_{j\neq i} m(i,j,x) \\
& -0.5\, m(O,i,x)\}\, P(*,i,x) - 0.5 \sum_{j\neq i} m(j,i,x)\, P(*,j,x) = \\
& \{1 - 0.5\, d(i,x) - 0.5\, m(i,O,x) - 0.5 \sum_{j=i} m(i,j,x) \\
& + 0.5\, m(O,i,x)\}\, P(i,*,x) + 0.5 \sum_{j\neq i} m(j,i,x)\, P(j,*,x)
\end{aligned}
$$

(2.38).

We have N of the equations, one for each of the regions. As equation (2.38) is written there is still dependence of the population of region i on those of regions j.

Equation (2.38) is more conveniently written as a matrix equation, and in order to do this we set up a matrix of occurrence exposure rates of the following form:

$$
\underline{M}(x) =
\begin{bmatrix}
m(1,1,x) & -\, m(2,1,x) & \ldots & -\, m(N,1,x) \\
m(1,2,x) & m(2,2,x) & \ldots & -\, m(N,2,x) \\
\cdot & \cdot & & \cdot \\
-\, m(1,N,x) & -\, m(2,N,x) & \ldots & m(N,N,x)
\end{bmatrix}
\qquad (2.39)
$$

where

$$
m(i,i,x) = d(i,x) + m(i,O,x) + \sum_{j \neq i} m(i,j,x) \; - \; m(O,i,x) \qquad (2.40)
$$

Equation (2.38) can be rewritten in matrix terms as

$$
[\underline{I} + 0.5\, \underline{M}(x)]\, \underline{P}^{t+1}(x) = [\underline{I} - 0.5\, \underline{M}(x)]\, \underline{P}^{t}(x) \qquad (2.41)
$$

where I is an N x N identity matrix, $\underline{P}^{t+1}(x)$ is a column vector of final populations and $\underline{P}^{t}(x)$ is a column vector of initial populations. The t refers to the point in time at the start of a time interval and t+1 to the point in time at the end.

Rearranging equation (2.41) yields the non-iterative projection model equation for all projection cohorts from the second to the next but last

$$
\underline{P}^{t+1}(x) = [\underline{I} + 0.5\, \underline{M}(x)]^{-1}\, [\underline{I} \quad 0.5\, \underline{M}(x)]\, \underline{P}^{t}(x) \qquad (2.42)
$$

To convert the model into a more recognizable multiregional cohort survival form, a matrix S(x) is defined that contains the rates that transform the initial into the final populations:

$$
\underline{S}(x) = [\underline{I} + 0.5\, \underline{M}(x)]^{-1}\, [\underline{I} - 0.5\, \underline{M}(x)] \qquad (2.43)
$$

so that the projection model for existing cohorts up to the last reads

$$
\underline{P}^{t+1}(x) = \underline{S}(x)\, \underline{P}^{t}(x) \qquad (2.44)
$$

For the last age group of interest, the projected population is given by

$$
\underline{P}^{t+1}(z) = \underline{S}(z-n)\, \underline{P}^{t}(z-n) + \underline{S}(z)\, \underline{P}^{t}(z) \qquad (2.45)
$$

For the infant cohort, using a female dominant model of births, births are given by

$$
\underline{B}(x) = \underline{b}(x)\; 0.5\, [\underline{P}^{t}(x,F) + \underline{P}^{t+1}(x,F)]
$$

$$= \underline{b}(x)\ 0.5\ [\underline{I} + \underline{S}(x)]\ \underline{P}^{t}(x,F)$$

$$= \underline{b}(x)\ 0.5\ [\underline{I} + \{\underline{I} + 0.5\underline{M}(x,F)\}[\underline{I} - 0.5\underline{M}(x,F)]]\ \underline{P}^{t}(x,F) \tag{2.46}$$

where $\underline{b}(x)$ is a diagonal matrix of age specific fertility rates, $\underline{B}(x)$ is a column vector of births by region and cohort of mother at time of maternity, and $S(x,F)$ is the matrix of "transformation" rates for females. These births have then to be sexed and then summed over all fertile ages

$$\underline{B}(X) = \underline{s}(X) \sum_{x} \underline{B}(x) \tag{2.47}$$

where $s(X)$ is a diagonal matrix with sex proportions for sex X in the principal diagonal. The $B(x)$ vector then enters the usual equation to yield the end of period population for the first cohort.

$$\underline{P}^{t}(-n,X) = [I + 0.5\underline{M}(-n,X)]^{-1}\ [I - 0.5\underline{M}(-n,X)]\ \underline{B}(X) \tag{2.48}$$

The matrix equations for all cohorts can be collected together in one matrix equation (following Rees and Wilson, 1977, Chapter 12 and Willekens and Drewe, 1983) as

$$\begin{bmatrix} \underline{P}^{t+1}(F) \\ \\ \underline{P}^{t+1}(M) \end{bmatrix} = \begin{bmatrix} \underline{H}(F) & \underline{O} \\ \\ \underline{F} & \underline{H}(M) \end{bmatrix} \begin{bmatrix} \underline{P}^{t}(F) \\ \\ \underline{P}^{t}(M) \end{bmatrix} \tag{2.49}$$

where

$\underline{P}^{t}(X)$ = a vector of initial populations for sex x by region and age

$\underline{H}^{t}(F)$ = a matrix with transformation rate matrices for females, $S(x,F)$, in the N subdiagonals and with matrices incorporating the product of all terms in the infant cohort equations for daughters on the N first rows

$\underline{H}^{t}(M)$ = a matrix with transformation rate matrices for males, $S(x,M)$ on the N subdiagonals

$\underline{F}(M)$ = a matrix with the products of all terms in the infant cohort equations for sons on the first N rows.

Each infant cohort matrix involves the following product of component matrices:

$$\underline{F}(x,X) = \underline{S}(X)\ \underline{b}(x)\ 0.5\ [\underline{I} + [\underline{I} + 0.5\underline{M}(x,F)]^{-1}\ [I - 0.5\underline{M}(x,F)]$$

(2.50).

The reader should, if he or she is unsure of what the contents and layout of these matrices are, take a large piece of paper and list all the elements in suitable arrays.

The models described here are only two of a much wider set that can be constructed. The task of building such alternative models is, however, fairly simple if the model is constructed component by component, making one choice from a few alternatives for each component. These choices are described in the next section together with the associated parameter values expected by the program and which should be entered in the instructions file.

3. THE BUILDING BLOCKS OF THE ACCOUNTS

The program for constructing movement accounts consists of a sequence of subroutines called by the main segment that either read in or extract from store or compute each component in turn. Figure 3.1 shows the sequence, the associated components, variables and control parameters, and Figure 3.2 places them in the framework of a movement accounts table.

3.1 Initial populations (subroutine BEGPOP)

There are four possible ways in which information on initial regional populations can be entered into a movement account.

3.1.1 The previous period's initial population (ipop=0)

The normal way in which initial populations are fed into a series of projected movement accounts is by transfering the final populations of the previous period for the previous cohort (from store), as in equation (2.35) for cohorts x=0 to x=z-n

$$P^{t}(i,*,x) = P^{t}(*,i,x-1) \qquad (3.1).$$

For the final cohort, the transfer is of the sum of the final populations of the last and last but one cohort

$$P^{t}(i,*z) = P^{t}(*,i,z-n) + P^{t}(*,i,z) \qquad (3.2).$$

To effect this option, parameter ipop must be set to 0.

3.1.2 The "initial populations" of the first cohort (jcohrt=1)

As discussed in the previous section, the totals of births computed after the program has constructed accounts for existing cohorts be used as the initial populations for the first cohort (after "sexing" has been carried out)

$$P(i,*.-n,X) = B(i,*,X) \qquad (3.3).$$

The ipop parameter setting is ignored when the first cohort accounts are being computed, that is, when jcohrt = 1.

3.1.3 Initial populations read in from a data file (ipop = 1)

In the first of a sequence of accounts over several time periods, the initial populations will be read in from a data file for the cohorts 0, n, 2n, ... , z-n, z+

$$P(i,*,x) = P(i,*,x) \text{ [data]} \qquad (3.4)$$

The parameter ipop is set to 1 to implement this choice. However, it is sometimes appropriate to interrupt a sequence of projected populations and to input exogeneous regional populations. For example, in the projections described in Rees (1983) the projections for all models start with freshly input mid-year 1981 populations, not with the final populations of the 1976-81 base period. The ipop

Subroutine	Component	Variable	Control parameter
BEGPOP	Initial populations	P(i,*,x)	ipop
↓			
PARA	Populations at risk	PAR(i,x)	ipar
↓			
DEATHS	Deaths	D(i,x)	idth
↓			
EMIG	Emigrations	M(i,0,x)	iemig
↓			
MIG	Internal migrations	M(i,j,x)	intmig
↓			
RES	Residual balancing terms	R(i,x)	ires
↓			
IMMIG	Immigrations	M(0,i,x)	immig
↓			
ENDPOP	Final populations	P(*,i,x)	ifpop
↓			
VERGE	(Tests for convergence)		
⋮			
CONSTR	Row and column constraints	V(i,x), W(j,x)	icon
⋮			
BIRTHS	Births & sex proportions	B(i,x), S(x)	birth, isexp

FIGURE 3.1 The principal subroutines of the movement accounts program

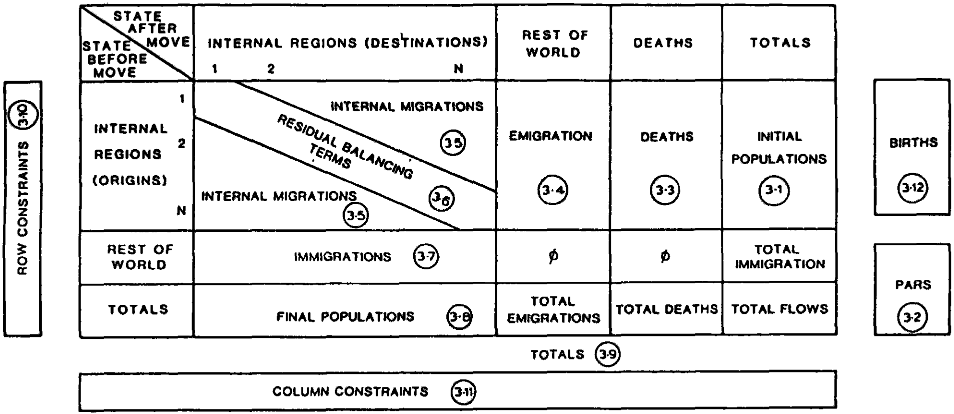


Figure 3.2 The building blocks of the movement accounts and the subsections in which they are described

parameter is again set to 1.

3.1.4 Initial populations computed from the row accounting equation (ipop = 2)

It may be appropriate in some cases to use a backcast model to estimate the base period movement accounts (see section 2.3.2). In this case equation (2.5) is used to compute initial populations

$$P(i,*,x) = R(i,x) + \sum_{j \neq i} M(i,j,x) + M(i,O,x) + D(i,x) \quad (3.5)$$

The ipop parameter is set to 2.

3.2 Populations at risk (subroutine PARA)

Before any other components are introduced into the accounts, populations at risk are computed, so that occurrence-exposure rates can be worked out as each component is reached. The parameter governing the choice of population at risk is ipar.

3.2.1 Populations at risk at the first iteration (jiter = 1)

On the first iteration of the model, final populations are unknown. Rather than compute populations at risk as averages of initial populations, already input, and zeroes (for the unknown final populations), the initial populations are used for the first iteration

$$PAR(i,x) = P(i,*,x) \quad (3.6)$$

3.2.2 Populations at risk on the second and subsequent iterations (ipar = 1)

During the second and subsequent iterations of the accounts based model, final populations are known so that equation (2.18) can be implemented

$$PAR(i,x) = (1/2)\ (P(i,*,x) + P(*,i,x) \quad (3.7).$$

The ipar parameter is set to 1.

3.2.3 Other populations at risk

In a previous manual (Rees, 1981) describing a transition accounts program, several other choices of population at risk were defined: the mid-period population, the multiregional population at risk, the initial population, the final population, and combinations of initial origin and final destination populations (for computing rates of internal inter-regional migration). None of these has been implemented in the current version of this program. Only the mid-period population is a possible candidate. If mid-period populations were to be used, they would be input directly as populations at risk. The problem would then be to compute initial populations using a modified accounting equation, where PAR(i,x) here refers to mid-period population in region i and cohort x

$$P(i,*,x) = PAR(i,x) + (1/2)\ D(i,x) + (1/2)\ M(i,O,x)$$
$$+ (1/2) \sum_{j \neq i} M(i,j,x) - (1/2) \sum_{j \neq i} M(j,i,x) - (1/2)\ M(O,i,x) \qquad (3.8)$$

with final populations being computed using a complementary equation:

$$P(*,i,x) = PAR(i,x) - (1/2)\ D(i,x) - (1/2)\ M(i,O,x)$$
$$- (1/2) \sum_{j \neq i} M(i,j,x) + (1/2) \sum_{j \neq i} M(j,i,x) + (1/2)\ M(O,i,x) \qquad (3.9)$$

The residual terms would be defined as

$$R(i,x) = PAR(i,x) - (1/2)\ D(i,x) \quad (1/2)\ M(i,O,x)$$
$$(1/2) \sum_{j \neq i} M(i,j,x) + (1/2) \sum_{j \neq i} M(j,i,x) \quad (1/2)\ M(O,i,x) \qquad (3.10)$$

with again the possibility of being negative and without substantive meaning. Successive mid-period populations could be connected together using terms from accounts for two successive periods. The simpler alternative of estimating component flows from mid-year to mid-year from data for two successive years was employed in constructing United Kingdom movement accounts (Rees, 1983).

3.3 Deaths (subroutine DEATHS)

Here four choices are available. These form a set which is common to all of the flow components of the accounts, although the choice for each component is independent of the choices for the other.

3.3.1 Deaths input from the previous period (idth = 0)

In this choice deaths from the previous period are used in the current

$$D(i,x)\ [t,t+1] = D(i,x)\ [t-1,t] \qquad (3.11)$$

The idth parameter is set at zero. This choice will be rarely used but is included to maintain comparability with treatment of the migration components where the alternative may be used.

3.3.2 Deaths input from a data file (idth = 1)

When concstructing base period accounts regional deaths by projection cohort will be estimated externally from available mortality statistics and input to the program from a data file

$$D(i,x) = d(i,x)\ [data] \qquad (3.12)$$

To effect this choice the idth parameter is set to 1 in the instructions data file.

3.3.3 Death rates from a previous period used (idth = 2)

If projections are being carried out under the assumption that current rates persist unchanged into the future, then death rates from the immediately previous period will be input from store

$$d(i,x)\ [t,t+1] = d(i,x)\ [t-1,t] \qquad (3.13).$$

The idth parameter is assigned a value of 2 when this choice is required. Even if this choice is not adopted in the short or medium term in a projection, it will usually be adopted in the long run.

3.3.4 New death rates for the current period input (idth = 3)

Over the short or medium term in projections it will be customary to input new death rates for each period, the rates changing in some secular fashion:

$$d(i,x) = d(i,x)\ [\text{projected}] \qquad (3.14).$$

The idth parameter is given the value 3 for this option. A common method of trending death rates is to assume they are declining in a negative exponential fashion.

3.4 Emigrations (subroutine EMIG)

Four choices, exactly parallel to those for deaths, are available.

3.4.1 Emigrations input from the previous period (emig = 0)

Emigration flows from the period prior to the current are used

$$M(i,O,x)\ [t,t+1] = M(i,O,x)\ [t-1,t] \qquad (3.15).$$

The iemig parameter is set to 0. This choice will be used more frequently for emigrations than the equivalent one for deaths.

3.4.2 Emigrations input from a data file (iemig = 1)

Externally estimated emigration flows are input to the current period accounts when iemig is set to 1:

$$M(i,O,x) = M(i,O,x)\ [\text{data}] \qquad (3.16).$$

3.4.3 Emigration rates from a previous period used (iemig = 2)

A constant rates model demands input from store of emigration rates from the immediately prior period

$$m(i,O,x)\ [t,t+1] = m(i,O,x)\ [t-1,t] \qquad (3.17)$$

when iemig will be given the value 2.

3.4.4 New emigration rates for the current period input (iemig = 3)

UNIVERSITY
LIBRARY
LEEDS

Projected emigration rates, based on time series extrapolation or ideas about the way opportunities abroad are evolving, can be input to the projected accounts from a data file

$$m(i,O,x) = m(i,O,x) \text{ [projected]} \qquad (3.18).$$

The iemig parameter is set to 3 in this choice.

3.4.5 Emigrations as residuals (iemig = 4)

This model is implemented by setting the iemig parameter to 4, although it is more likely to be used in the equivalent transition rate program (Rees, 1981).

3.5 Internal migrations (the MIG subroutine)

Internal migrations are input to the program in the same four forms as used for deaths and emigrations, but as matrices for each cohort rather than as vectors. In each case, the terms on the principal diagonal are input as zeroes.

3.5.1 Internal migrations input from the previous period (intmig = 0)

The internal migration flows of the period immediately previous to the current may be used in flow form:

$$M(i,j,x)\ [t,t+1] = M(i,j,x)\ [t-1,t] \qquad (3.19)$$

if the intmig parameter is set equal to zero. Such an option is useful if comparison needs to be made with other single region, net migration flow models. Using gross migration flows in a multiregional context is equivalent to using net migration flows in a single region context

$$IN(i,x) = \sum_{j \neq i} M(j,i,x) - \sum_{j \neq i} M(i,j,x) \qquad (3.20)$$

for internal regions i and j, where IN(i,x) is the internal net migration for region i, cohort x. Normally, this internal net migration term is coupled with an equivalent external one

$$EN(i,x) = M(O,i,x) - M(i,O,x) \qquad (3.21)$$

so that when parameter intmig is set equal to zero, so normally are parameters iemig and immig.

3.5.2 Internal migrations read in from a data file (intmig = 1)

The normal choice for historical, base period accounts will be to input internal inter-regional migrations from a data file using a parameter setting of intmig equal to 1:

$$M(i,j,x) = M(i,j,x) \text{ [data]} \qquad (3.22)$$

This option can also be used to introduce new migration flows using a spatial interaction or log-linear model.

3.5.3 Internal migration rates from the previous period used (intmig = 2)

In projections where the internal migration rates are assumed constant, the values of the previous period will be read in from store:

$$m(i,j,x)\,[t,t+1] = m(i,j,x)\,[t-1,t] \tag{3.23}$$

with an intmig setting of 2.

3.5.4 New internal migration rates input from a data file (intmig = 3)

If the internal migration rates are themselves projected, for example, through use of a log-linear decomposition into level, generation and distribution components (Willekens and Baydar, 1982, 1983; Stillwell, 1983), then the resulting projected migration rates should be input from a data file

$$m(i,j,x) = m(i,j,x)\,[\text{projected}] \tag{3.24}$$

with intmig set to 3.

3.6 The residual balancing terms (subroutine RES)

No fresh data are required by this part of the program but there are two ways in which the residual balancing terms can be computed.

3.6.1 Residual terms computed using a row accounting equation (ires = 0)

The residual terms for each region-cohort population are computed using equation (2.1):

$$R(i,x) = P(i,*,x) \quad D(i,x) - M(i,O,x) \quad \sum_{j \neq i} M(i,j,x) \tag{3.25}$$

when a forecast model is being used. The controlling parameter ires should be set to zero.

3.6.2 Residual terms computed using a column accounting equation (ires = 1)

When a backcast model is employed, the column accounting equation is utilized (equation 2.4):

$$R(i,x) = P(*,i,x) \quad M(O,i,x) - \sum_{j \neq i} M(j,i,x) \tag{3.26}.$$

The parameter ires must be set to 1 to select this option.

3.7 Immigrations (subroutine IMMIG)

Four choices are available, as with the other flow components.

3.7.1 Immigrations from a previous period input (immig = 0)

A decision may be taken to project immigration using a constant input of flow data on the argument that this migration stream is responsive to legal restrictions in the destination country rather than to the size of the population at risk in either origin or destination. In this case immigrations can be input from store for a previous period

$$M(O,i,x)\ [t,t+1] = M(i,O,x)\ [t-1,t] \qquad (3.27).$$

The immig parameter is set to 0.

3.7.2 Immigrations input from a data file (immig = 1)

For the base time period or for any time period for which new immigration flows are projected, the numbers of immigrations to a region should be input from a data file, having been suitably estimated:

$$M(O,i,x) = M(O,i,x)\ [\text{data}] \qquad (3.28).$$

The immig parameter is set to 1.

3.7.3 Immigration rates from a previous period used (immig = 2)

If immigration admission rates are used in the accounting model, and a constant rates projection is being carried out, immigration rates are input from store from the previous period

$$m(O,i,x)\ [t,t+1] = m(O,i,x)\ [t-1,t] \qquad (3.29)$$

with the immig parameter assigned a value of 2.

3.7.4 New immigration rates input from a data file (immig = 3)

When immigration rates are projected independently, the projected rates can be introduced into the program from a data file in the current period

$$m(O,i,x) = m(O,i,x)\ [\text{projected}] \qquad (3.30).$$

The immig parameter is set to 3 in this case.

3.8 Final populations (subroutine ENDPOP)

3.8.1 Final populations computed from the column accounting equation (ifpop = 0)

In most accounting models and projection applications the final populations for each cohort in each region will be computed from the column accounting equation, equation (2.2)

$$P(*,i,x) = R(i,x) + \sum_{j \neq i} M(j,i,x) + M(O,i,x) \qquad (3.31).$$

The parameter ifpop is set to zero to effect this choice.

Final populations for the last cohort are computed using equation (3.31) for cohort z+

$$P(*,i,z+) = R(i,z+) + \quad M(j,i,z+) + M(O,i,z+) \qquad (3.32)$$

and the additions of the last but one and last cohort populations left until conversion of the final populations into initial populations for the next period

$$P^{t}(i,*,z+) = P^{t}(*,i,z-n) + P^{t}(*,i,z+) \qquad (3.33).$$

3.8.2 Final populations input from a data file (ifpop = 1)

An alternative to computing the final populations via the column accounting equations listed above is to input to the accounts fresh data on final populations:

$$P(*,i,x) = P(*,i,x)\ [\text{data}] \qquad (3.34)$$

Such an alternative is used in the backcast model (Figure 2.1.2) and in the "both populations" model. Normally, in the latter model the researcher would proceed to adjust the accounts so that the total of row sums and column sums are the same. However, even if this is not done, this alternative of inputting final populations can often be useful if the figures are reliable as then the best estimate of the population at risk can be made.

3.9 Other marginal totals

3.9.1 Total immigrations

These are derived purely by summation over all the immigrations to internal regions:

$$M(O,*,x) = \sum_{j} M(O,j,x) \qquad (3.35).$$

3.9.2 Total emigrations

Similarly, the total of emigrations is derived by summation

$$M(*,O,x) = \sum_{i} M(i,O,x) \qquad (3.36).$$

3.9.3 Total deaths

The total of deaths is found in the program by adding up regional deaths

$$D(*,x) = \sum_{i} D(i,x) \qquad (3.37).$$

Normally, if the internal regions together constitute the national territory, these totals should also be the national figures for these demographic flows, and may well have been used in the estimation of fully age-disaggregated regional figures.

3.9.4 Total flows

This total, formed by summing either the row totals

$$T(*,*,x) = \sum_i P(i,*,x) + M(0,*,x) \quad (3.38)$$

or the column totals

$$T(*,*,x) = P(*,j,x) + M(*,0,x) + D(*,x) \quad (3.39)$$

of the accounts. This figure, for each period cohort, is the total number of inputs to the demographic system of interest over the time interval of concern, and the total number of outputs from the system and time interval.

3.10 Iteration and convergence of the accounts model (subroutine VERGE)

After all the accounts components have been assembled the program returns to the first subroutine and recomputes any of the components that need to be recalculated again. Which these will be will depend on the sequence of choices made using the parameters detailed above.

The model will then cycle through the component subroutines until satisfactory convergence is achieved. The criterion for convergence is very simple: the populations at risk (average populations) are compared at successive iterations and if for all regions

$$| P(i,x) [p] - P(i,x) [p-1] | < 1/2v \quad (3.40)$$

then the model is regarded as having converged. The convergence parameter, v, is set by the modeller (parameter IVERG): normally it should be set to 1 so that the convergence criterion refers to 1/2 a person.

When base period, historical accounts are being constructed, only two iterations are required by the program. If all items are input as flows or stocks, then no iterations are really required but two iterations still occur because of the convenience of the general iterative structure of the model. In projection, normally 3 to 6 iterations of the model occur before convergence.

3.11 Row and column constraints (subroutine CONSTR)

3.11.1 No adjustment to constraints (icon = 0)

In a projection context it will not normally be appropriate to impose constraints on the accounts matrix, unless it is desired to fit the projections to target populations. The parameter, icon, is set to zero if no adjustments are required.

3.11.2 External constraints (icon = 1)

When there is disagreement between the final population generated by the accounts model and those form an external source (for example, a Census), it will be appropriate to prepare external constraints and to input them to the program. Where, in fact, do the constraints come from?

The row constraints consist of the initial populations of the regions and the total of immigrations to the internal system, each adjusted by the estimated error:

$$V(i,x) = P(i,*,x) + e(i,*,x) \quad \text{for } i=1, \dots, N \qquad (3.41)$$

$$V(N+1,x) = \sum_j M(0,j,x) + e(N+1,*,x) \qquad (3.42)$$

The column constraints are made up of the final populations plus the total of emigrations from the total of deaths in the internal set of regions:

$$W(j,x) = P(*,j,x) + e(*,j,x) \quad \text{for } j=1, \dots, N \qquad (3.43)$$

$$W(N+1,x) \quad \sum_i M(i,0,x) + e(*,N+1,x) \qquad (3.44)$$

$$W(N+2,x) = \sum_i D(i,x) + e(*,N+2,x) \qquad (3.45)$$

3.11.3 Constraints constructed internally with all the blame put on the emigration estimates (icon = 2)

In many countries the least reliable of the demographic components that enter accounts are the emigrations. The difference between the sums of initial estimates of the row and column constraints (specified as above) is calculated and assigned purely to emigration:

$$E = (\sum_i P(i,*,x) + \sum_j M(0,j,x)$$

$$- (\sum_j P(*,j,x) + \sum_i M(i,0,x) + \sum_i D(i,x)) \qquad (3.46)$$

$$V(i,x) = P(i,*,x) \qquad (3.47)$$

$$V(N+1,x) = \sum_j M(0,j,x) \qquad (3.48)$$

$$W(j,x) = P(*,j,x) \quad \text{for } j=1, \dots, N \qquad (3.49)$$

$$W(N+1,x) = \sum_i M(i,0,x) + E \qquad (3.50)$$

$$W(N+2,x) = \sum_i D(i,x) \qquad (3.51)$$

This can be accomplished by the accounts program if the icon parameter is set to 2.

3.11.4 The balancing factor routine

The routine used in the program to adjust the accounts matrix to marginal constraints is ordered in a slightly different way from that described earlier (section 2.3.4). The adjustment is achieved through use of balancing factors

$$A(i,j,x) = a(i,x)\ b(j,x)\ A(i,j,x) \tag{3.52}$$

where

$$a(i,x) = V(i,x) / \sum_j b(j,x)\ A(i,j,x) \tag{3.53}$$

$$b(j,x) = W(j,x) / \sum_i a(i,x)\ A(i,j,x) \tag{3.54}$$

recalling that $A(i,j,x)$ was our general notation for the accounts elements for cohort x in row i and column j of the accounts matrix, $V(i,x)$ was the external constraint for row i of the x cohort accounts matrix and $W(j,x)$ the external constraint for column j of the accounts matrix for cohort x. The terms $a(i,x)$ and $b(j,x)$ are the row and column balancing factors respectively.

The calculation proceeds by first setting all the $b(j,x)$ terms to 1 and by using equation (3.53) to solve for the $a(i,x)$'s. These are then used in equation (3.54) to estimate the $b(j,x)$'s. Equations (3.53) and (3.54) are then repeated for as many times as necessary to achieve the following condition for all i's and j's

$$\text{abs}\ \{\ (a^{p+1}(i,x) - a^{p}(i,x)) / a^{p}(i,x)\ \} < 10^{-itol} \tag{3.55}$$

$$\text{abs}\ \{\ (b^{p+1}(j,x) - b^{p}(j,x)) / b^{p}(j,x)\ \} < 10^{-itol} \tag{3.56}$$

where abs refers to the absolute value function, p and p+1 to successive iterations and itol to an integer that determines the degree of relative difference in balancing factors which is to be achieved before the adjustment is considered to have converged. When this occurs equation (3.52) is used to compute adjusted elements, $A(i,j,x)$.

3.12 Births (subroutine BIRTHS)

The choices of equation for birth are in part the same as those for the other flow components, but there are further decisions to be made relating to the arrangements made for sex disaggregation.

3.12.1 Births input from the previous period (ibirth = 0)

The program allows the user to make the assumption that births for the current period are the same as those for the previous period:

$$B(i,x)\ [t,t+1] = B(i,x)\ [t-1,t] \tag{3.57}$$

$$\text{for all fertile cohorts } x = a \ldots b$$

This assumption is not a good one, of course, but may be useful to demonstrate that better models, involving fertility rates, make. The ibirth parameter should be set equal to 0 for this alternative.

3.12.2 Births input from the current period (ibirth = 1)

This will be the normal choice when base period accounts are constructed. Births are input from a data file for each fertile cohort

$$B(i,x) = B(i,x) \text{ [data]} \qquad (3.58)$$

when the parameter ibirth is set to 1. In this program, in common with many other analyses, births by age (period-cohort) of mother are not disaggregated by sex. A sex proportion is applied later to the sum of the births to all the fertile cohorts.

3.12.3 Birth rates of the previous period used (ibirth = 2)

When projecting the population, particularly in the long term, it will be common to make the assumption that birth rates continue unchanged into the future. This model, that

$$b(i,x)\ [t,t+1] = b(i,x)\ [t-1,t] \qquad (3.59)$$

can be implemented in the program by setting the ibirth parameter to 2.

3.12.4 New birth rates used (ibirth 3)

Normally, when the purpose is to project the population in the short term, new birth rates, forecast externally to the projection model, will be introduced from the data file

$$b(i,x)\ [t,t+1] = b(i,x) \text{ [data]} \qquad (3.60)$$

for the fertile cohorts of each region. When such new fertility rates are being introduced the ibirth parameter should be given the value 3.

3.12.5 Populations at risk of giving birth

As with the other components, the population at risk of giving birth is taken to be the average population in the time period for the period cohorts in the fertile age range (say between 15 and 50).

When the populations being studied are distinguished by sex, there exist a number of alternative population at risk and associated fertility models. The first is the single sex model in which males give birth to males and females to females or persons to persons. To use this model, accounts are constructed separately for each sex or for persons. The number of sexes parameter, nsex, is set to 1, and the number of the fertile sex, nfsex, is also set to 1.

The preferred alternative to the single sex model is the female dominant model in which only women make up the population at risk. This model is effected in the program by giving nsex a value of 2 and nfsex the sequence number of the female sex (may be 1 or 2, but remember that the choice will determine the order of the sexes in all other components). An equivalent male dominant model can be specified but is not recommended as the classification of births by father's age is rarely given, if at all, for regional populations and is less reliable than the classification by mother's age because of the non-reporting of father's age for many of the children born outside wedlock.

3.12.6 Sexing the infants

Before they are entered into the initial cohort as the row totals or "initial populations", the infants born during the time interval must be divided into males and females. This is done by applying the proportions born into the first sex in each of the regions

$$B(i,*,X) = s(i,X)\ B(i,*) \qquad (3.61)$$

where X refers to sex X and B(i,*) to the total of all births by age of mother summed over age, and s(i,X) is the proportion of births in region i of sex X. These proportions are entered into the births data file at the very end, and the isexp parameter is set to 1. In periods subsequent to the first the same sex proportions can be used and the isexp parameter is set to 0.

3.13 Putting the building blocks together

When a choice has been made for each of the accounts components listed in Figure 3.1 and located in an accounts table in Figure 3.2, then an accounts based model has been constructed. Care must be taken in making sensible combinations of choices. The combination of choices for a base period's accounts model will differ substantially from that for projection, for example. In order to guide the user in the choice of a sensible model sections 5, 6 and 7 of the paper set out a number of examples of models that the population analyst is likely to find useful. However, before those examples can be sensibly understood, the general guidlines for setting up the data files for accounting are detailed in the next section of the paper.

4. A GUIDE FOR USERS OF THE PROGRAM

To use the program to produce multiregional movement accounts the user must assemble, on a suitable disk, the files required for running the program. These are depicted in Figure 4.1. The requirements are not quite as formidable as they might first appear.

The user must have on his or her work disk the instructions file, the component data files (not all of those listed are normally required), the work file (or enough space for it if it is created by the program) and enough room for the projected populations file (if required). The program file and job executive file will normally reside on a library disk. The results file and the survivorship rates file (if required) can be printed out directly rather than stored on disk.

The instructions file contains the specifications of the regional system being studied and the accounting model to be adopted that the user wishes to relay to the program. The component data files contain the estimated numbers of people or events that characterize the regional system being studied. The work file holds the set of multiregional accounts variables for the current period during the program's operation. The results file contains tables of movement accounts for all cohorts and sexes used, for all periods requested. The projected populations file stores the population stocks forecast by the accounting model for further analysis and tabulation. The survivorship rates file gives an array of rates for further use in non-iterative projection models (e.g. Willekens and Drewe, 1983). Finally, the user will receive on his or her terminal messages from the program about the progress through the accounts over the periods of the analysis.

The system of files and program is organized by the executive file. This contains the file definitions, loads the program into the computer and starts the program running. To activate the program all the Leeds user need do is type in the name of the EXEC file followed by three arguments: (i) the common filename for all the component data files, (ii) the common length of record on all these files, and (iii) the number of records in the direct access work file. Then just sit back and let the computer do the rest.

The remainder of this section of the paper gives the detailed instructions on how to prepare each of the input files, and detailed descriptions of what to expect on the output files.

4.1 The instructions file

The user of the movement accounts program should prepare a file of instructions using the editor (filetype = 'DATA' at the University of Leeds; filename = 'INPUT' at NIDI, both with 80 character records). Figure 4.2 gives a suggested framework for the instructions file, which users can copy and edit by replacing the items in the shaded zones by their own choices. At the University of Leeds link to the GEO6LIB library disk and issue the command COPY MOVE DATA D YOUR DATA A if you are an undergraduate and the command COPY MOVE DATA B YOUR DATA A if you are not. The filename 'YOUR' may be an valid filename you wish to give to your analysis.

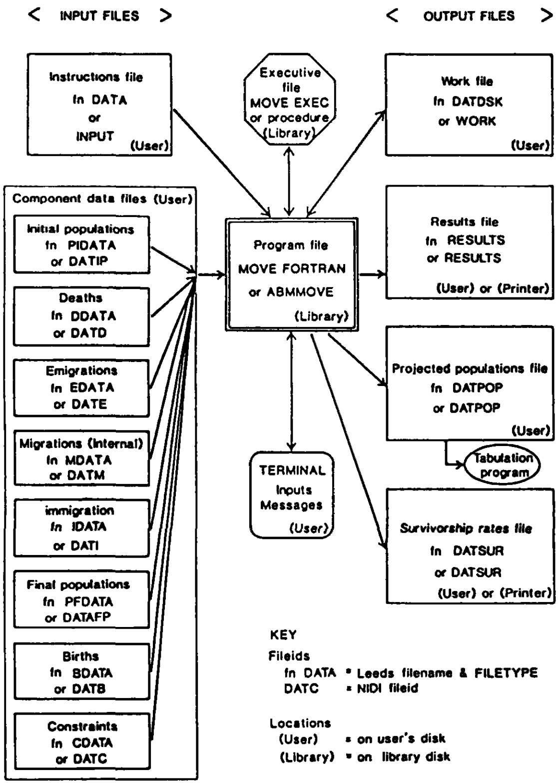


Figure 4.1 <u>The system of files needed for producing movement accounts</u>

The file:

```
SYSTEM OF INTEREST =  THE NETHERLANDS          PERIOD =  1976,77
NAME OF USER       =  PHILIP REES              DATE   =  OCTOBER 1983
BASE PERIOD MODEL  = UNCONSTRAINED FORECAST ABM
PROJECTION MODEL   = EMIGRATION RATES, IMMIGRATION FLOWS
 NINREG=    2 NPRIOD=   2 NCOHRT=   92   NSEX=    2
  NITER=   20   NTAB=   3  NYEAR=1976 NTIMIN=    1
 NAGEIN=    1  NFSEX=   1  NALFA=   17  NBETA=   51
WESTNETHRESTNETH
    IPOP=   1    IPAR=   1     IDTH=   1   IEMIG= '1 INTMIG=   1
    IRES=   0    IMIG=   1    IFPOP=   0   IVERG=  1 IACCNT=   1
    ICON=   0    ITOL=   6 ITERBF=100      ISURV=  1 IBIRTH=   1
   ISEXF=   1   IAUTO=   1
    IPOP=   0    IPAR=   1     IDTH=   2   IEMIG=  2 INTMIG=   2
    IRES=   0    IMIG=   0    IFPOP=   0   IVERG=  1 IACCNT=   1
    ICON=   0    ITOL=   6 ITERBF=100      ISURV=  1 IBIRTH=   2
   ISEXF=   0   IAUTO=   0
```

Comments:

Record	
1–4	title records
5–7	analysis parameter records
8	name record
9–12	model parameter records, period 1
13–16	model parameter records, period 2

Figure 4.2 A framework for preparing the instructions file

4.1.1 The titles

Enter into the instructions file four lines of text - the title records - describing the analysis you are carrying out. You are free to type what you like on these lines, up to column 80 (the format is 10A8 in FORTRAN), but the suggested framework in Figure 4.2 invites you to name the system being studied and the period of study, to give your own name and the date. On the third line you are invited to specify the accounts based model used for the base period (see Figure 2.1), and on the fourth you can enter a description of the model used in projecting the population. You will find this systematic labelling of the analysis very useful when several different analyses are being undertaken.

4.1.2 The analysis or n parameters

The next three lines of the instructions file should contain the values of the parameters that govern the accounting and projection analysis in every period. There are some 12 parameters, 4 to each line, with 12 columns set aside for each parameter. The first 8 chamcaters may be used to label the parameter - use the suggested framework here - and the last four characters contain the number to which the parameter is set. These numbers should all be right-justified, that is, entered in the rightmost columns of the set of four. The parameters are described in turn. A summary is provided in Table 4.1.

4.1.2.1 Parameter NINREG, the number of internal regions

Enter ' NINREG' in columns 1 to 8 of record 5 of the instructions file. Enter the number of regions internal to the system of interest in columns 9 to 12 as an integer. The rest of the world does not count as an internal region.

How many regions will the movement accounts program handle? As currently specified the University of Leeds version will handle up to and including 21 regions. The NIDI version will accommodate up to and including 12 regions. If a larger system of regions is being studied, the program will need modification (see Appendix).

4.1.2.2 Parameter NPRIOD, the number of periods of analysis

Enter ' NPRIOD=' in columns 13 to 20 of record 5. Enter the number of periods which the analysis is intended to cover in columns 21 to 24 (right justified). The minimum number of periods is one; no maximum is specified.

4.1.2.3 Parameter NCOHRT, the number of cohorts in the analysis

Enter ' NCOHRT=' in columns 25 to 32 of record 5. Enter the number of cohorts into which the population is disaggregated in columns 33 to 36 (right justified). By cohort is meant period cohort or projection cohort as shown earlier in Figure 1.1. Care must be taken in counting the cohorts. The first cohort is that involving infants making the move from birth during the period to being in the first age group of the population at the end of the time interval.

TABLE 4.1 The analysis of n parameter records: formats

Record No	Characters or columns of record	Format	Item	Description or instruction
5	1- 8	a8	'-NINREG='	label for:-
	9-12	i4	ninreg	number of internal regions
	13-20	a8	'-NPRIOD='	label for:-
	21-24	i4	npriod	number of periods of analysis
	25-32	a8	'-NCOHRT='	label for:-
	33-36	i4	ncohrt	number of cohorts
	37-44	a8	'---NSEX='	label for:-
	45-48	i4	nsex	number of sexes
6	1- 8	a8	'-NITER='	label for:-
	9-12	i4	niter	number (maximum allowed) of abm iterations
	13-20	a8	'--NTAB='	label for:-
	21-24	i4	ntab	tabulation level requested
	25-32	a8	'-NYEAR='	label for:-
	33-36	i4	nyear	starting year of 1st period
	37-44	a8	'-NTIMIN='	label for:-
	45-48	i4	ntimin	time interval length (years)
7	1- 8	a8	'-NAGEIN='	label for:-
	9-12	i4	nagein	age interval (years)
	13-20	a8	'--NFSEX='	label for:-
	21-24	i4	nfsex	sequence number of fertile sex
	25-32	a8	'--NALFA='	label for:-
	33-36	i4	nalfa	sequence number of 1st fertile cohort
	37-44	a8	'--NBETA='	label for:-
	45-48	i4	nbeta	sequence number of last fertile cohort

Notes

1. abm accounts based model
2. - blank
3. ' ' indicates that alphanumeric field is to be entered: no quotes needed on record.
4. Record number refers to the sequence number of records in the instructions file.
5. a8 field of 8 alphanumeric characters
6. i4 field of 4 integer number

A few examples will clarify what is involved. Assume we were interested in projecting the population in five year age groups up to age 85+. Cohort 1 would involve the transition "birth to 0-4". Cohort 2 would involve the transition "0-4 to 5-9" and cohort 18 would involve the transition "80-84 to 85-89" and cohort 19 would involve the transition "85+ to 90+". Or, if we were projecting single year populations up to age 90+, cohort 1 would involve the transition "birth to age 0", cohort 2 "age 0 to age 1", ... , cohort 91 "age 89 to 90", and cohort 92 "age 90+ to 91+".

The general rule for computing the number of cohorts involved in the analysis can be stated as follows:

$$NCOHRT = (z/\text{age interval}) + 2 \qquad (4.1)$$

where z is the age that starts the last, semi-closed age group, z+. Formula (4.1) applies also when aggregate accounts are constructed: z+ is 0+ and the number of cohorts is 2.

In preparing the component data files NCOHRT items will be needed for each component except the initial populations where one less, (NCOHRT-1), is required.

4.1.2.4 Parameter NSEX, the number of sexes in the analysis

Enter ' NSEX=' in columns 37 to 44 of record 5 of the instructions file. Enter the number of sexes into which the population is divided in columns 45 to 48 (right justified). This number will be either 1 when just persons are being used (no sex disaggregation) or 2 when males and females are distinguished. The order of the sexes is determined by the NFSEX parameter (see later).

4.1.2.5 Parameter NITER, the number of iterations in the accounts based model

Enter ' NITER=' in columns 1 to 8 of record 6 of the instructions file. Enter the maximum number of iterations to be allowed for the accounts based model in columns 9 to 12 of record 6. In an accounting model in which all the components are input from a data file as fixed quantities only two iterations of the model are required. In a projection context, however, many components will change as the population at risk changes at each iteration, and more iterations will be needed. Setting the maximum to 20 should ensure that all models iterate to convergence (see section 3.10 for a description).

4.1.2.6 Parameter NTAB, the tabulations level requested

Enter ' NTAB=' in columns 13 to 20 of record 6 of the instructions file. Enter the tabulation level, ntab, in columns 21 to 24 of record 6. The ntab level governs the types of tables printed out to the results file (see section 4.4) or types of arrays printed out to the projected populations file (see section 4.5) or survivorship rates file (see section 4.6). The details of what tables are printed out at what ntab setting are given in Table 4.2. Ntab values should range from -2 to 5, but levels 4 and 5 should only be used for single period analyses for program debugging at time of installation since they produce tables at each iteration of

TABLE 4.2 Tabulation level settings using the NTAB parameter

Location of Output	TYPE OF TABLE	NTAB SETTING								Additional period Parameter Settings Needed
		-2	-1	0	1	2	3	4	5	
Results File	Model description: titles, n-parameters, region names and i-parameters	1	1	1	1	1	1	1	1	
Datapop File	Projected final populations	0	1	1	1	1	1	1	1	
Results File	Movement accounts tables	0	0	1	1	1	1	1	1	IACCNT = 1
	Births & fertility rates tables	0	0	0	1	1	1	1	1	
	Movement rates tables	0	0	0	0	1	1	1	1	
	Survivorship rates tables (Survivorship arrays on DATSUR file)	0	0	0	0	0	1	1	1	ISURV = 1
	Movement accounts components tables at each iteration of accounting model	0	0	0	0	0	0	1	1	
	Details of the balancing factors at each iteration of the constraints adjustment model	0	0	0	0	0	0	0	1	ICON = 1

1 = Output produced

0 = No output produced

a biproportional adjustment routine (level 5).

In some instances, the program user may wish to print out many tables for one period of analysis but not for others. This can be accomplished for the movement accounts tables by turning the IACCNT parameter in the model parameter set on (=1) or off (=0). Similarly, if survivorship rates are required (for another use or projection model) the ISURV parameter should be set to 1. And thirdly, the program will only enter the constraints procedure (to which NTAB level 5 refers) if the ICON parameter has been set to 1.

4.1.2.7 Parameter NYEAR, the starting year

Enter ' NYEAR=' in columns 25 to 32 of record 6 of the instructions file. In columns 33 to 36 enter the number of the calendar year that begins the first period of analysis.

4.1.2.8 Parameter NTIMIN, the time interval

Enter ' NTIMIN=' in columns 37 to 44 of record 6. Enter the length of the time interval in years (whole numbers only) in columns 45 to 48 (right justified).

4.1.2.9 Parameter NAGEIN, the age interval

Enter ' NAGEIN=' in columns 1-8 of record 7 of the instructions file. The age interval is the difference in years between the start and the end of the age groups used in the analysis. Enter the length of the age interval in integer years in columns 9 to 12 of record 7 (justified to the right). Note that the age interval should equal the time interval (NTIMIN). If this is not the case then estimates or adjustments to the component data should be made before input to the program.

4.1.2.10 Parameter NFSEX, the sequence number of the fertile sex

Enter ' NFSEX=' in columns 13 to 20 of record 7. Enter the sequence number of the fertile sex in columns 21 to 24 (right justified). If NSEX has been set to 1 just persons or males or females separately are being analysed - NFSEX should also be 1. If NSEX has been set to 2, and a female dominant fertility model is being used, NFSEX should indicate whether the data files have been prepared in order 'females first, males second' when NFSEX should be set to 1 or in the order 'males first, females second' when NFSEX should be set to 2. Undoubtedly "women's libbers" will choose the first strategy, "male chauvinist pigs" the second.

4.1.2.11 The NALFA parameter, the sequaence number of the first fertile cohort

Enter ' NALFA=' in columns 25 to 32 of record 7. Enter the sequence number of the first cohort for which births or fertility data are supplied. The following formula applies

NALFA = (x/age interval) + 2 (4.2)

where x is the age that starts the cohort description "x-x+n to x+n-x+2n", where n is the age interval. For example, if we are using a five year age and time interval (as in the Greater London example) and the first fertile cohort is "10-14 to 15-19", that is, women aged 10-14 at the start of the time interval and aged 15-19 at the end, this cohort has the sequence number

NALFA = (10/5) + 2 = 4 (4.3)

If our population system consisted of single year cohorts and the first fertile age was 15 in the 15 to 16 cohort, then the sequence number of the first fertile cohort would be

NALFA = (15/1) + 2 = 17 (4.4)

4.1.2.12 The NBETA parameter, the sequence number of the last fertile cohort

Enter ' NBETA=' in columns 37 to 44 of record 7. Enter the value of NBETA in columns 45 to 48. This will be the sequence number of the last fertile cohort computed from

NBETA = (x/age interval) + 2 (4.5)

where x is the age that starts the cohort description of the last fertile cohort "x-x+n to x+n x+2n".

4.1.3 The region names

The names of the regions should be entered in the next set of records in the instructions file in fields of 8 characters, 10 to a record (see Table 4.3). The number of records needed will depend on the number of regions being studied. For up to and including 10 regions only one record is needed; for from 11 to 20 regions two records are required; and 21 to 30 regions three records are used.

4.1.4 The model or i parameters

The model used to construct the accounts for each period is governed by a set of model parameters called the i-parameter set because all their program names begin with the letter i. The function of a majority of these parameters has already been discussed in section 3 of the manual. Here that information is gathered together in summary form and details are given of the parameters not covered previously. The user needs to prepare a set of records containing the appropriate values of the i-parameters for each period of analysis until such time as the model becomes constant when the previous period's values can be used by setting the iauto parameter to zero.

4.1.4.1 Model or i parameters: record one

On the first of the i parameter records the user speifies the values he or she wishes to assign to the parameters governing the first 5 components of the accounts based model for the period in question, namely, the initial populations parameter, ipop, the population at risk parameter, ipar, the deaths parameter, idth, the emigrations parameter, iemig and the internal migrations parameter, intmig. Table 4.4 sets out the format in which the record should be prepared. The table also gives a cross-reference to the manual

TABLE 4.3 The region name records

Record number	Characters or columns of record	Format	Item	Description
8	1- 8	a8	nam (1)	Name of region 1
	9-16	"	nam (2)	Name of region 2
	17-24	"	nam (3)	Name of region 3
	25-32	"	nam (4)	Name of region 4
	33-40	"	nam (5)	Name of region 5
	41-48	"	nam (6)	Name of region 6
	49-56	"	nam (7)	Name of region 7
	57-64	"	nam (8)	Name of region 8
	65-72	"	nam (9)	Name of region 9
	73-80	"	nam (10)	Name of region 10
9 (if required)	1- 8	a8	nam (11)	Name of region 11
	9-16	"	nam (12)	Name of region 12
	17-24	"	nam (13)	Name of region 13
	25-32	"	nam (14)	Name of region 14
	33-40	"	nam (15)	Name of region 15
	41-48	"	nam (16)	Name of region 16
	49-56	"	nam (17)	Name of region 17
	57-64	"	nam (18)	Name of region 18
	65-72	"	nam (19)	Name of region 19
	73-80	"	nam (20)	Name of region 20
10 (if required)	1-8	a8	nam (21)	Name of region 21
	⋮	⋮	⋮	⋮ ⋮

Notes

1. a8 field of 8 alphanumeric characters

TABLE 4.4 The model or i parameter records: record one

Record number*	Characters or columns of record	Format	Item	Description of item and choices	Manual section consult
9	1- 8	a8	'^^^IPOP='	label for initial populations parameter	3.1
	9-11	i3	ipop	0= previous period's final population 1= read in from data file 2= computed from row accounting equation for jcohort 1 = births	
	12-19	a8	'^^^IPAR='	label for population at risk parameter	3.2
	20-22	i3	ipar	1= average population in period no other options current	
	23-30	a8	'^^^IDTH='	label for deaths parameter	3.3
	31-33	i3	idth	0= deaths input from previous period 1= deaths input from a data file 2= death rates input from a previous period 3= death rates input from a data file	
	34-41	a8	'^^IEMIG='	label for emigrations parameter	3.4
	42-44	i3	iemig	0= emigrations input from previous period 1= emigrations input from a data file 2= emigration rates input from a previous period 3= emigration rates input from a data file	
	45-52	a8	'^INTMIG='	label for internal migrations parameter	3.5
	53-55	i3	intmig	0= internal migrations input from previous period 1= internal migrations input from a data file 2= internal migration rates input from previous period 3= internal migration rates input from a data file	

Note

* The record number is for when there is only 1 record for region names. Add 1 for each additional region name record.

** ^ blank

section where the alternative models implied by the different parameter settings are discussed.

4.1.5.2 Model or i parameters: record two

The second record of the i parameter set should contain the settings of a further 4 parameters that govern the accounts based model employed, namely ires, the residual balancing items parameter, imig, the immigrations parameter, ifpop, the final populations parameter and iverg, the convergence criterion parameter. The fifth parameter on this record, iaccnt, controls the printing (or not) of the movement accounts matrices for the current period.

Full details of the formats in which these parameters are to be entered are given in Table 4.5.

4.1.4.3 Model or i parameters: record three

The third i parameter record should contain settings of the parameters icon, itol, iterbf, isurv and ibirth in the format specified in Table 4.6.

The constraints control parameter, icon (see also section 3.11) is set to zero if no constraints are to be used in the estimation of the movement accounts matrix, and the associated accounitng model is known as unconstrained. Most projected accounts will be unconstrained and some historical accounts will be unconstrained.

If all accounts components are known and it proves possible to apportion sensibly the difference between the total of row constraints and the total of column constraints, then the constraints adjustment procedure can be invoked by setting the icon parameter to 1. In this situation there must be an additional data file containing the row and column constraints for all cohorts.

If it is felt that the initial difference between row and column constraints can all be apportioned to the emigrations total, then the program provides a means of doing this without having to input fresh data: the icon parameter should be set to 2.

The itol parameter (see also section 3.11) governs whether adjustment to the constraints has been achieved in the form of 10 to the power - itol. If itol is set to 4, for example, the balancing factors (see section 3.11) will be considered to have converged if they all differ, at successive iterations, by less than 1 in 10,000.

Note that, although the process may have converged, the movement accounts table which is written out in whole number form may not be exactly consistent with the marginal constraints. The row and column sums of the adjusted matrix will differ by one or two from the corresponding constraint totals.

The maximum number of iterations allowed in the balancing factor routine is input via the iterbf parameter to ensure that the adjustment process does stop eventually. If the sum of the row constraints does not equal the sum of the column constraints, convergence will never occur. What a reasonable maximum number of iterations within which the adjustments for all cohorts will have

TABLE 4.5 The model or i parameter records: record two

Record number*	Characters or columns of record	Format	Item	Description of item and choices	Manual section consult
10	1- 8	a8	'^^^IRES='	label for residual balancing items parameter	3.6
	9-11	i3	ires	0= computed from row accounting equation 1= computed from column accounting equation	
	12-19	a8	'^^^IMIG='	label for immigrations parameter	3.7
	20-22	i3	imig	0= immigrations input from previous period 1= immigrations input from a data file 2= immigration rates input from previous period 3= immigration rates input from a data file	
	23-30	a8	'^^IFPOP='	label for final populations parameter	3.8
	31-33	i3	ifpop	0= computed from column accounting equation 1= input from a data file	
	34-41	a8	'^^IVERG='	label for convergence criterion parameter	3.10
	42-44	i3	iverg	number governing convergence: see equation (84). Normally set to 1	
	45-52	a8	'^IACCNT='	label for accounts table parameter	see also
	53-55	i3	iaccnt	0= accounts table not printed 1= accounts table printed	4.1 3.6

Notes

* The record number refers to that when there is only 1 region names record. Add 1 for each additional record containing region names.

** ^ = blank

TABLE 4.6 The model or i parameter set: record three

Record number*	Characters or columns on record	Format	Item	Description of item and choices	Manual section to consult
11	1- 8	a8	'···ICON='	label for constraints control parameter	2.3.4
	9-11	i3	icon	0= no constraints applied 1= external constraints applied 2= internally constructed constraints applied	3.11
	12-19	a8	'···ITOL='	label for constraints tolerance parameter	2.3.4
	20-22	i3	itol	integer constant: convergence criteria $= 1/10^{-itol}$ Normally set to 5 or 6.	3.11
	23-30	a8	'ITERBF='	label for	3.11
	31-33	i3	iterbf	maximum number of iterations in balancing factor routine. Set to at least 100	
	34-41	a8	'··ISURV='	label for survival rates parameter	2.4.2
	42-44	i3	isurv	0= no survival rates computed 1= survival rates computed	
	45-52	a8	'··IBIRTH='	label for births parameter	3.12
	53-55	i3	ibirth	0= births input from a previous period 1= births input from a data file 2= birth rates input from a previous period 3= birth rates input from a data file	

Notes

* The record number should be increased if there are more than one region name records.

** · = blank

converged successfully will depend on the setting of the tolerance criterion, the number of regions and the degree of adjustment required. Start with iterbf set to 100 but be prepared to alter it upwards.

The survivorship rates parameter, isurv, should be used if survival rates employed in the non-iterative form of the movement accounts model are employed. The survivorship rates are computed in the way described in section 2.4.2 or in other ways, depending on how the diagonal terms in the M(x) matrix are handled.

$$m(i,i,x) = d(i,x) + \sum_{j \neq i} m(i,j,x) \qquad (4.6)$$

In the NIDI version the current equation used is:

$$m(i,i,x) = d(i,x) + \sum_{j \neq i} m(i,j,x) + m(i,O,x) \qquad (4.7)$$

whereas previously the equation (2.40) was

$$m(i,i,x) = d(i,x) + \sum_{j \neq i} m(i,j,x) + m(i,O,x) \quad m(O,i,x) \qquad (4.8)$$

Each of these equations corresponds to a different way of handling external migration in the projection model.

The births parameter, ibirth, can be set in 4 ways: 1 for births input for the current period from a data file, 2 for birth rates input from the previous period, and 3 for birth rates input for the current period from a data file. Details are explained in section 3.12.

4.1.4.4 Model or i parameters: record four

The details of the last two model parameters are given on Table 4.7.

The sex proportions parameter, isexp, is set to 1 if the sex proportions for each region are to be read in from a data file (the births data file); if sex proportions of a previous period are used isexp is set to zero. Normally just one set of sex proportions are read in for the first period and these are reused.

The automatic parameter, iauto, allows the user to do away with the need to enter any further model parameters in the next period if it is set to zero. When iauto is zero the i model parameters of the current period will be used in the next. When new i parameters are to be read users should ensure that iauto in the previous period has been set to 1. In a typical projection, the base period model parameters will differ from the projection period parameters. There may also be variation in the model parameters in the first few periods. After then a constant model may be adopted. The iauto parameter would be set to zero in the first set of constant model parameters.

TABLE 4.7 The model or i parameter records: record four

Record number*	Characters or columns of record	Format	Item	Description of item and choices	Manual section to consult
12	1- 8	a8	'^^ISEXP='	label for sex proportion parameter	3.12.6
	9-11	i3	isexp	0= sex proportions of previous period used 1= sex proportions read in from data file	
	13-20	a8	'^^IAUTO='	label for automatic parameter	4.1.5
	21-22	i3	iauto	0= i parameters of previous period used 1= i parameters input from instructions file	

Notes

* The record number should be increased if there are more than one record name records.

** ^ blank

4.2 The component data files

Data are supplied to the movement accounts program from a separate data file for each component. This makes the job of preparing the data estimates easier than if all components had to be compressed into one large file.

4.2.1 Record format and length

Currently, the data records are designed as fixed format, fixed record length records. Each data item is assigned 10 characters. The Leeds version allows 7 items per record (a format of 7F10.0); the NIDI version allows 6 items per record (a format of 6F10.0). If disk space is in short supply, the read statements in the program can be modified by the implementing programmer to shorter fixed fields or to free format and data sensitive records, and the file definitions in the EXECuti ve file altered accordingly.

4.2.2 File names

The component data files should be assigned names according to the rules set out iin Table 4.8.

4.2.3 Vector files: initial populations, deaths, emigrations, immigratiions, final populations

Each of these files is organized in the following way.

(1) A set of records is needed that contain the values of the component for each region in 7 fields of 10 columns (Leeds version) or 6 fields of 10 columns (NIDI version). How many records are needed? That will depend on the number of regions being studied. For the Leeds version of the program 1-7 regions require 1 record, 8-14 regions require 2 records, 15-21 regions require 3 records. In the NIDI version 1-6 regions require 1 record and 7-12 regi ons require 2 records.

(2) This set of records is repeated if there is a second sex.

(3) The set of records required to hold one set of variables for the number of sexes used is repeated for all the cohorts.

(4) The order of the cohorts will be for initial populations cohort 2 ... last cohort, and for the other components cohort 2 ... last cohort, cohort 1. Figure 4.3 takes component files that have been listed earlier (Figure 1.2) and annotates them to make their structure clear.

4.2.4 Matrix file: internal migrations file

The matrix file is organized in a similar fashion to the vector files in that each row of the matrix occupies the same number of records as one of the vector components.

(1) A set of records is required to contain one row of the matrix. This set will consist of 1 record if the number of regions is 1-7 (Leeds) or 1-6 (NIDI), 2 records if the number of regions is 8-14 (Leeds) or 7-12 (NIDI), and 3 records if the number of regions

TABLE 4.8 File names for the component data files

Data component	Channel Number	Leeds File Identification	NIDI File Identification
Initial populations	8	fn PIDATA	DATIP
Deaths	9	fn DDATA	DATD
Emigrations	10	fn EDATA	DATE
Migrations (internal)	11	fn MDATA	DATM
Immigrations	12	fn IDATA	DATI
Final Populations	13	fn PFDATA	DATFP
Births	14	fn BDATA	DATB
Constraints	15	fn CDATA	DATC

Notes

1. All Leeds component data files should have a common name which will vary from run to run. The filetypes are fixed.
2. The constraints file is required only if the accounts are to be adjusted to external row and column constraints.
3. The Leeds file names are specified in the MOVE EXEC which may be copied from the library and altered if more than one demographic information system (DIS) is being used.
4. The NIDI filenames are specified in OPEN statements within the program so that if several systems are being studied the filenames listed should be used for the current working copy of the files.
5. The Leeds files should all be sequential access files.
6. The NIDI files should all be direct access files.

An initial population file

rec.	region 1 Greater London	region 2 Rest of UK	cohort	age group at start of period
1	419900	3319100	2	0- 4
2	491900	3954300	3	5- 9
3	518700	4134600	4	10-14
4	487700	3748300	5	15-19
5	497300	3373900	6	20-24
6	573200	3621700	7	25-29
7	480200	3108200	8	30-34
8	413500	2795900	9	35-39
9	397800	2729500	10	40-44
10	415500	2840600	11	45-49
11	449800	2976600	12	50-54
12	422800	2723100	13	55-59
13	423200	2685200	14	60-64
14	379200	2459600	15	65-69
15	287400	1953100	16	70-74
16	189100	1304700	17	75-79
17	108300	722400	18	80-84
18	72800	451100	19	85+
	field 1 cols 1-10	field 2 cols 11-20		

A deaths file

	region 1 Greater	region 2 Rest of UK	cohort	age group at start	end
1	1343	9940	2	0- 4	5- 9
2	569	4876	3	5- 9	10-14
3	1038	7821	4	10-14	15-19
4	1802	12945	5	15-19	20-24
5	1689	11348	6	20-24	25-29
6	2060	12685	7	25-29	30-34
7	2478	16667	8	30-34	35-39
8	3225	23381	9	35-39	40-44
9	5452	40557	10	40-44	45-49
10	10036	75384	11	45-49	50-54
11	17775	130775	12	50-54	55-59
12	27155	197682	13	55-59	60-64
13	38988	285791	14	60-64	65-69
14	55445	405566	15	65-69	70-74
15	65078	487200	16	70-74	75-79
16	62680	468911	17	75-79	80-84
17	51984	369099	18	80-84	85-89
18	53842	354429	19	85+	90+
19	5632	39101	1	Birth	0-4
	field 1 cols 1-10	field 2 cols 11-20			

Figure 4.3 The structure of vector component files

is 15-21 (Leeds).

(2) This set of records is repeated for each region.

(3) The matrix is repeated for the second sex.

(4) The set of records required to hold of set of variables for a cohort and the number of sexes specified is repeated for all cohorts, in the order, cohort 2 to the last, followed by the first.

Figure 4.4 shows the structure of a file containing migration matrices for a two region system for 92 cohorts and 2 sexes. Each matrix has four elements. The entries in the diagonal are migrations internal to the regions concerned. These don't enter the accounts model calculations directly. They may be entered as zeroes if within region migrations are not available.

4.2.5 The constraints file

If a model involving external constraints has been selected, then for each sex and cohort sets of row and column constraints must be input from a separate constraints file. The row constraints consist of the initial populations and the immigrations total; the column constraints consist of the final populations, emigrations total and deaths total. The row constraints occupy (n+1) fields and the number of records occupied is the integer result of {(n+1)/7+6/7} in the Leeds case and of {(n+1)/6+5/6} in the NIDI case. The column constraints occupy (n+2) fields and therefore the number of records will be the integer result of {(n+2)/7+6/7} in the Leeds version and the integer result of {(n+2)/6+5/6} in the NIDI version.

An example of a constraints file is given in Figure 4.5. In this instance the initial populations have been adjusted to bring the sum of the row totals into agreement with the sum of column totals. The reason for choosing to adjust initial populations was because the 1981 Census introduced a new concept for the usually resident population. Whereas previously the national figure had excluded persons usually resident in the UK but absent on Census night abroad and had included overseas visitors present on Census night, in 1981 the definition was changed to include usual residents absent abroad, where the information could be obtained in the case of not wholly absent households, and overseas visitors were excluded. Overall there was a net gain in the UK population of about 230,000, though this was made up of gains and losses in the constituent age groups. Therefore the 1976 estimate populations were adjusted to match in concept the 1981 estimates, which had been derived from the 1981 Census figures.

4.2.6 The births and sex proportions file

This file has the simplest layout of any of the component data files. For each cohort in the fertile age range a set of records is needed giving the births in each region. The number of records needed in the set will be 1 record for 1-7 regions (Leeds) or 1-6 regions (NIDI), 2 records for 8-14 regions (Leeds) or 7-12 regions (NIDI), and 3 records for 15-21 regions (Leeds).

record	region 1 West Neths WN	region 2 Rest Neths RN	origins	sex	cohort	age at start of period	age at end of period
1	1947.	667.	WN	female	2	0	1
2	431.	2376.	RN	female	2	0	1
3	2131.	727.	WN	male	2	0	1
4	439.	2033.	RN	male	2	0	1
5	1990.	674.	WN	female	3	1	2
6	436.	2563.	RN	female	3	1	2
7	2087.	711.	WN	male	3	1	2
8	422.	2695.	RN	male	3	1	2
.	.	.	.	.	.	.	.
361	137.	36.	WN	female	92	90+	91+
362	27.	175.	RN	female	92	90+	91+
363	72.	20.	WN	male	92	90+	91+
364	15.	111.	RN	male	92	90+	91+
365	1071.	351.	WN	female	1	Birth	0
366	236.	1304.	RN	female	1	Birth	0
367	1116.	300.	WN	male	1	Birth	0
368	223.	1027.	RN	male	1	Birth	0
	field 1 cols 1-10	field 2 cols 11-20					

Figure 4.4 An example showing the strucure of a matrix component file

record						Cohort
1	415127	3201391	59814		Row const.	2
2	383018	3306582	61451	1128	Column "	2
3	492552	3959540	52038		Row "	3
4	466105	3970110	54305	5445	Column	3
5	522461	4161568	75961		Row	4
6	534146	4147254	72055	8454	Column	4
7	483106	3713610	143630		Row	5
8	577292	3617807	135580	14747	Column "	5
9	493464	3347874	167011		Row "	6
10	526262	3306740	109051	12037	Column "	6
.	.	.				.
15	80459	490474	790		Row	19
16	26406	146316	990	10211	Column	19
17	423570	3465800	7326		Row	1
18	397[illegible]	3955100	15007	447	Column	1
	UK popns initial final	RUK popns initial final	total immig total emig	deaths	layout of each cohort's records	

Figure 4.5 A portion of a constraints file for Greater London and the Rest of the UK

265.	394.
530.	787.
530.	787.
530.	787.
530.	787.
2527.	3413.
4524.	6040.
4524.	6040.
4524.	6040.
4524.	6040.
5785.	7516.
7045.	8992.
7045.	8992.
7045.	8992.
7045.	8992.
4784.	6074.
2522.	3155.
2522.	3155.
2522.	3155.
2522.	3155.
1569.	2000.
616.	844.
616.	844.
616.	844.
616.	844.
373.	523.
130.	202.
130.	202.
130.	202.
130.	202.
69.	109.
7.	17.
7.	17.
7.	17.
7.	17.
4.	8.
485.	485.

Figure 4.6 A births and sex proportions file for the West Netherlands, Rest of the Netherlands system

Appended to the end of the birth records are an additional set of records containing the sex proportions per 1000 for the first sex in each region. It is assumed that the proportion of the second sex will be unity minus the proportion of the first sex. Figure 4.6 shows the layout of a births and sex proportions file for the West Netherlands and the Rest of Netherlands example.

4.3 The work file

To save memory space the program uses a work file to house the accounts variables for all the cohorts. If any of the accounts components are required during the next period they can be accessed from the work file.

The user needs to specify a record length for this file, and the program will request, on screen, that this number be typed in. In the NIDI version the FORTRAN 77 system will automatically create a file of required length. In the Leeds FORTRAN 77 system the work file must exist prior to being opened in the program. The user should therefore create a file of required number of records and record length using the editor or a small program. The number of records needed should be computed given values of:

nc = number of cohorts (ncohrt)
ns = number of sexes (nsex)
ni = number of (internal) regions (ninreg)
nb = number of final fertile cohort (nbeta)
na = number of initial fertile cohort (nalfa)

and

m = number of records needed for one vector of values
  = integer part of {ni/7+6/7} (Leeds)
  = integer-part of {ni/6+5/6} (NIDI)

from the following formula

$$n = nc \times ns \times 3 + ((nc \times ns \times (8+ni)) + ((nb-na+1) \times 2) + 3) \times m.$$

So, for example, inn our Greater London Rest of the UK example

nc = 19
ns = 1
ni = 2
nb = 10
na = 4

and therefore

m = 1

so that

$$n = 19 \times 1 \times 3 + (19 \times 1 \times (8+2)) + ((10-4+1) \times 2) + 3) \times 1$$
$$= 264$$

In the Netherlands example

nc = 92
ns = 2
ni = 2
nb = 51
na = 16
m = 1

so that

$$n = 92 \times 2 \times 3 + (92 \times 2 \times (8+2) + (51-16+1) \times 2) + 3) \times 1$$
$$= 2467$$

4.4 The results file

The results produced by the program are normally written to a file so that they can be checked before printing. If there is insufficient room on the user's disk for the results file, results ccan be sent directly to the printer by using the MOVEP exec command rather than the MOVE exec command.

The first unit or block of the results file always contains (Figure 4.7)

(1) the titles of the analysis
(2) the parameters governing the analysis (n parameters)
(3) a list of region names
(4) a list of "magic" parameters used in accessing the work file by the program, including the total number of records in this file.

Then follow

(5) the model parameters for period 1
(6) the tables requested in the order cohort 2 ... last cohort, cohort 1.

Figure 4.8 shows the model parameters for our Greater London example and movement accounts tables for cohorts 8 through 10. These two items are repeated for as many periods as the analysis requires, though the number of tables may be varied from period to period. Figure 4.9 gives the final page from the results file for the UK two region analysis. For the fifth and sixth periods of analysis no accounts tables have been requested (iaccnt = 0) so that only the model parameter settings are printed out.

4.5 The projected populations file

One of the main uses of the movement accounts program is in the projection of multiregional populations. These appear in the bottom row of each accounts table. In order that attention may be focussed on these projected populations, the final populations of each period are written to a separate file, which is always produced.

An extract from the beginning of a projected populatiions file is given in Figure 4.10. Each line (record) of the file contains the final regional populations for a cohort. The cohort populations are arranged in the order in which they are produced in the program: the first cohort is produced last. Note that the final populations for one age group beyond those specified in the initial population file of the base period are output: that is, for age groups 85-89

```
ABM MODEL PARAMETERS FOR PERIOD =     1

PARAMETER FOR:-
INITIAL POPULATON                         IPOP   =  1
POPULATION AT RISK                        IPAR   =  1
DEATHS                                    IDTH   =  1
EMIGRATIONS                               IEMIG  =  1
INTERNAL MIGRATIONS                       INTMIG =  1
RESIDUAL TERMS                            IRES   =  0
IMMIGRATIONS                              IMIG   =  1
FINAL POPULATIONS                         IFPOP  =  0
TOLERANCE LEVEL FOR ABM CONVERGENCE       IVERG  =  1
TABULATION LEVEL FOR ACCOUNTS             IACCNT =  1
CONSTRAINTS CONTROL                       ICON   =  0
TOLERANCE LEVEL FOR BF ROUTINE            ITOL   =  6
MAXIMUM NUMBER OF BF ITERATIONS           ITERBF = 20
SURVIVORSHIP RATES CONTROL                ISURV  =  1
BIRTHS MODEL CONTROL                      IBIRTH =  1
SEX PROPORTIONS CONTROL                   ISEXP  =  1
```

. . .

THE MOVEMENT ACCOUNTS MATRIX
PERIOD = 1 COHORT = 8 AGES = 30-34 TO 35-39 SEX = 1 ITERATION = 2

	GL 1	RUK 2	EMIGRTNS 3	DEATHS 4	TOTALS 5
GL	359084.	88834.	29804.	2478.	480200.
RUK	58730.	2948804.	83999.	16667.	3108200.
IMMIGTNS	24564.	54397.	0.	0.	78961.
TOTALS	442378.	3092035.	113803.	19145.	3667361.

THE MOVEMENT ACCOUNTS MATRIX
PERIOD = 1 COHORT = 9 AGES = 35-39 TO 40-44 SEX = 1 ITERATION = 2

	GL 1	RUK 2	EMIGRTNS 3	DEATHS 4	TOTALS 5
GL	343803.	49706.	16766.	3225.	413500.
RUK	33591.	2688863.	50065.	23381.	2795900.
IMMIGTNS	13970.	32219.	0.	0.	46189.
TOTALS	391364.	2770788.	66831.	26606.	3255589.

THE MOVEMENT ACCOUNTS MATRIX
PERIOD = 1 COHORT = 10 AGES = 40-44 TO 45-49 SEX = 1 ITERATION = 2

	GL 1	RUK 2	EMIGRTNS 3	DEATHS 4	TOTALS 5
GL	345498.	36507.	10343.	5452.	397800.
RUK	24338.	2633911.	30694.	40557.	2729500.
IMMIGTNS	9256.	21475.	0.	0.	30731.
TOTALS	379092.	2691893.	41037.	46009.	3158031.

Figure 4.7 The results file: example 1

```
------------------------------------------------------------------------
DEMOGRAPHIC INFORMATION SYSTEM FOR GREATER LONDON  PHIL REES SEPT 1983
MOVEMENT ACCOUNTS AND PROJECTIONS USING NHSCR MIGRATION DATA FOR 1976-81
MODEL - UNCONSTRAINED FORECAST ABM 1976-81  CONSTANT RATES PROJECTION
PROJECTION  5 PERIODS 1981-2006  OLD RATES 1981 POPULATION BASE
------------------------------------------------------------------------

------------------------------------------------------------------------
PARAMETERS GOVERNING THE ANALYSIS
------------------------------------------------------------------------
NUMBER OF REGIONS                             =   2
NUMBER OF PERIODS                             =   6
NUMBER OF COHORTS                             =  19
NUMBER OF SEXES                                   1
MAX. NUMBER OF ITERATIONS OF ABM              =  10
TABULATION LEVEL REQUESTED                    =   0
BASE YEAR                                     =1976
TIME INTERVAL (IN YEARS)                      =   5
AGE INTERVAL (IN YEARS)                       =   5
SEQUENCE NUMBER OF FERTILE SEX                =   1
SEQUENCE NUMBER OF FIRST FERTILE COHORT       =   4
SEQUENCE NUMBER OF LAST FERTILE COHORT        =  10
------------------------------------------------------------------------

------------------------------------------------------------------------
REGION NAMES
------------------------------------------------------------------------
REGION  1  =      GL
REGION  2  =      RUK
------------------------------------------------------------------------

------------------------------------------------------------------------
MAGIC PARAMETERS FOR DIRECT ACCESS DATA FILE
------------------------------------------------------------------------
NUMBER OF RECORDS NEEDED TO STORE ONE VECTOR    =    1
NUMBER OF RECORDS NEEDED TO STORE ONE MATRIX    =    2
NUMBER OF RECORDS NEEDED TO STORE TOTALS        =    1
NO. OF RECORDS FOR ALL ACCOUNTS VARIABLES
FOR ONE AGE COHORT FOR ONE SEX                  =   11
NO. OF RECORDS FOR ALL ACCOUNTS VARIABLES
FOR ALL COHORTS, FOR ALL SEXES                  =  209
NO. OF RECORDS FOR ALL BIRTHS VARIABLES
FOR ALL FERTILE COHORTS                         =   14
NUMBER OF RECORDS NEEDED IN DIRECT ACCESS FILE  =  225
POPULATION ACCOUNTS STACK POINTERS:
MRECPS=  1  MRECO =  2  MRECEM=  3
MRECM =  4  MRECIM=  6  MRECPF=  7
MRECTT=  8  MRECPR=  9  MRECPK= 10
MRECPP= 11
------------------------------------------------------------------------
```

Figure 4.8 <u>The results file: example 2</u>

```
-----------------------------------------------------------------------
ABM MODEL PARAMETERS FOR PERIOD =      5
-----------------------------------------------------------------------
PARAMETER FOR:-
INITIAL POPULATON                     IPOP    =  0
POPULATION AT RISK                    IPAR    =  1
DEATHS                                IDTH    =  2
EMIGRATIONS                           IEMIG   =  2
INTERNAL MIGRATIONS                   INTMIG  =  2
RESIDUAL TERMS                        IRES    =  0
IMMIGRATIONS                          IMIG    =  2
FINAL POPULATIONS                     IFPOP   =  0
TOLERANCE LEVEL FOR ABM CONVERGENCE   IVERG   =  1
TABULATION LEVEL FOR ACCOUNTS         IACCNT  =  0
CONSTRAINTS CONTROL                   ICON    =  0
TOLERANCE LEVEL FOR BF ROUTINE        ITOL    =  6
MAXIMUM NUMBER OF BF ITERATIONS       ITERBF  = 20
SURVIVORSHIP RATES CONTROL            ISURV   =  1
BIRTHS MODEL CONTROL                  IBIRTH  =  2
SEX PROPORTIONS CONTROL               ISEXP   =  0
-----------------------------------------------------------------------


-----------------------------------------------------------------------
ABM MODEL PARAMETERS FOR PERIOD =      6
-----------------------------------------------------------------------
PARAMETER FOR:-
INITIAL POPULATON                     IPOP    =  0
POPULATION AT RISK                    IPAR    =  1
DEATHS                                IDTH    =  2
EMIGRATIONS                           IEMIG   =  2
INTERNAL MIGRATIONS                   INTMIG  =  2
RESIDUAL TERMS                        IRES    =  0
IMMIGRATIONS                          IMIG    =  2
FINAL POPULATIONS                     IFPOP   =  0
TOLERANCE LEVEL FOR ABM CONVERGENCE   IVERG   =  1
TABULATION LEVEL FOR ACCOUNTS         IACCNT  =  0
CONSTRAINTS CONTROL                   ICON    =  0
TOLERANCE LEVEL FOR BF ROUTINE        ITOL    =  6
MAXIMUM NUMBER OF BF ITERATIONS       ITERBF  = 20
SURVIVORSHIP RATES CONTROL            ISURV   =  1
BIRTHS MODEL CONTROL                  IBIRTH  =  2
SEX PROPORTIONS CONTROL               ISEXP   =  0
-----------------------------------------------------------------------

MOVEMENT ACCOUNTS FOR     6.PERIOD(S) COMPLETED
```

Figure 4.9 The results file: example 3

GREATER LONDON	REST OF UNITED KINGDOM	Cohort	Age group	
397170.	3328910.	2	5- 9	
470472.	3967936.	3	10-14	
520484.	4127065.	4	15-19	
513377.	3720926.	5	20-24	
485109.	3351854.	6	25-29	
526539.	3602169.	7	30-34	1981 final populations
442378.	3092035.	8	35-39	(produced by
391364.	2770788.	9	40-44	unconstrained model
379092.	2691893.	10	45-49	from 1976 base)
392319.	2773304.	11	50-54	
413956.	2858492.	12	55-59	
372378.	2544170.	13	60-64	
359775.	2421635.	14	65-69	
306198.	2069903.	15	70-74	
215446.	1471664.	16	75-79	
123024.	838483.	17	80-84	
54435.	354842.	18	85-89	
17811.	97618.	19	90+	
427127.	3105426.	1	0- 4	
374584.	3065638.	2	5- 9	
367940.	3309146.	3	10-14	
471314.	3967009.	4	15-19	
563533.	4112275.	5	20-24	
551045.	3608134.	6	25-29	
482313.	3264740.	7	30-34	1986 populations
483720.	3579664.	8	35-39	(projected from
406089.	3145747.	9	40-44	1981 base)
369837.	2779873.	10	45-49	
357045.	2665053.	11	50-54	
365508.	2702642.	12	55-59	
364850.	2706550.	13	60-64	
317897.	2348280.	14	65-69	
285953.	2063523.	15	70-74	
227205.	1584797.	16	75-79	
139914.	958349.	17	80-84	
62386.	413221.	18	85-89	
19943.	115759.	19	90+	
449427.	3232977.	1	0- 4	
421535.	3248340.	2	5- 9	
358661.	3075519.	3	10-14	
374270.	3297576.	4	15-19	
507636.	3927145.	5	20-24	
561432.	4071900.	6	25-29	
509561.	3583924.	7	30-34	1991 populations
446736.	3243975.	8	35-39	(projected from
461268.	3542302.	9	40-44	1981 base)
390007.	3097803.	10	45-49	
350855.	2711292.	11	50-54	
330172.	2556124.	12	55-59	
323423.	2520950.	13	60-64	
311608.	2436286.	14	65-69	
257650.	1973503.	15	70-74	
214664.	1554219.	16	75-79	
147852.	1018420.	17	80-84	
70374.	470658.	18	85-89	
20550.	119618.	19	90+	
442049.	3391310.	1	0- 4	

Figure 4.10 An extract from a projected populations file for the Greater London, Rest of the UK system

and 90+ rather than just 85+ in the example.

4.6 The survivorship rates file

A third output file produced by the program is that containing the surviivorship rates suitable for use in a non-iterative, multiregional, cohort survival model for population projection (se section 2.4.2). Survivorship rates are computed and output to a file if the isurv model parameter is set to 1 for the relevant period or periods.

Figure 4.11 shows an extract from the start of such a file for the Netherlands system for 1976. For each cohort a matrix of survivorship rates is printed in order cohort 2 to the last followed by cohort 1. Note that the matrix has been transposed from its layout in the movement accounts tables.

4.7 Running the program

4.7.1 On the University of Leeds Amdahl

To run the program you should prepare the instructions and component data files following the detailed specifications given in sections 4.1 and 4.2, and the naming rules suggested earlier, which are recapped here.

4.7.1.1 Naming rules for input files

All input files should have the same filename, which should describe succinctly the system being studied. For our Greater London - Rest of the UK system we used "DISGL1" as the filename: this meant "Demographic Information System for Greater London, model 1". The instructions file should be of filetype DATA; the component data files should be of filetypes PIDATA, DDATA, EDATA, MDATA, IDATA, PFDATA, BDATA and CDATA where the one or two letter prefix to DATA refers to the component concerned (PI = populations, initial; D = deaths; E = emigrations; M = migrations, internal; I = immigrations; PF = populations, final; B = births; and C = constraints). The PIDATA, DDATA, MDATA, IDATA and BDATA files must have been prepared; accounting model choice will determine which of the EDATA, PFDATA and CDATA files are necessary.

4.7.1.2 The workfile

The workfile should also have the general filename and be of type DATDSK (current period DATa store on DiSK), and should have been created prior to running the program.

4.7.1.3 Linking to the library

Users without automatic access to the library should obtain access by typing

```
LINK GEO6LIB 191 195 RR
ACCESS 195 B/A
```

4.7.1.4 Issuing the EXECutive command

To:	From: West Neths	Rest of Neths	Cohort	Initial age	Final age
WN	.971627	.008610	1	0	1
RN	.017314	.983784	1	0	1
WN	.969852	.008305	2	1	2
RN	.018206	.983654	2	1	2
WN	.971879	.008222	3	2	3
RN	.016892	.984270	3	2	3
WN	.971401	.007544	4	3	4
RN	.017051	.984522	4	3	4
WN	.973504	.007082	5	4	5
RN	.015584	.986170	5	4	5
WN	.973481	.006434	6	5	6
RN	.015400	.986333	6	5	6
WN	.975763	.006398	7	6	7
RN	.014378	.986896	7	6	7
.	.	.	.	.	.
.	.	.	.		.
.	.	.	.	.	

Figure 4.11 An extract from a survivorship rates file for the West Netherlands, Rest of the Netherlands system, 1976

To run the prrogram issue the following command

MOVE argument1 argument2 argument3

where argument1 is the general filename for the instructions and component data files, argument2 is the length in characters of the records in the fixed length component data files (20 for 2 regions ... 80 for 8 or more regions), and argument 3 is the number of records in the work file to be used. The contents of the EXEC file are listed in Figure 4.12.

Alternatively, if the user does not have enough disk space to hold his or her results on file the command

MOVEP argument1 argument2 arrgument3

should be used. This will send the results directly to the user's virtual printer. The second file definition statement in the EXECutive file becomes

FI 2 PRINTER

Users still experiencing difficulties with disk space should consult with the author for further advice.

4.7.1.5 Messages from the program

To keep you informed of progress with the accounts building and population projection the program will tell you when each cohort's account has been completed and when each period's set of cohort accounts have been finished.

4.7.1.6 After_the program has finished

The user's disk will now hold the output files produced by the program: the RESULTS and DATPOP files, the DATSUR file if requested, and the DATDSK file with the last period's accounts. These should be listed on the screen and/or printed out in the normal way.

4.7.2 On NIDI's Data General Eclipse

To run the program you should prepare the instructions and component data files following the detailed specifications in sections 4.1 and 4.2.

4.7.2.1 Naming rules for input files

The instructions file should be named INPUT and the component data files DATIP, DATD, DATE, DATM, DATI, DATFP, DATB and DATC (see Table 4.8): the suffixes to DAT identify the component involved. The accounting model chosen determines which of these component data files should be present. The files are all named within the program. The user should take care when carrying out several runs of the program to save the data files under unique names.

4.7.2.2 The work file

This has the name WORK. It is named and created within the

```
EXEC SETUP FORTRAN
LOAD MOVE
FI 1  DISK &1          DATA     (RECFM F LRECL 80
FI 2  DISK &1          RESULTS  (RECFM F LRECL 132
FI 3  DISK &1          DATDSK   (XTENT &3 RECFM F LRECL &2
FI 4  DISK &1          DATSUR   (RECFM F LRECL &2
FI 5  TERM
FI 6  TERM
FI 7  DISK &1          DATPOP   (RECFM F LRECL &2
FI 8  DISK &1          PIDATA   (RECFM F LRECL &2
FI 9  DISK &1          DDATA    (RECFM F LRECL &2
FI 10 DISK &1          EDATA    (RECFM F LRECL &2
FI 11 DISK &1          MDATA    (RECFM F LRECL &2
FI 12 DISK &1          IDATA    (RECFM F LRECL &2
FI 13 DISK &1          PFDDATA  (RECFM F LRECL &2
FI 14 DISK &1          BDATA    (RECFM F LRECL &2
FI 15 DISK &1          CDATA    (RECFM F LRECL &2
START (CLEAR
```

Figure 4.12 A listing of the executive file that controls the running of the program at the University of Leeds program.

4.7.2.3 Running the program at NIDI

Consult the system manager or an experienced programmer on how to set up the program and data files for execution in a general MACRO.

4.7.3 Further processing of the results

To convert the projected populations being studied into suitable tables for further analysis, a program should be prepared. Several such programs have been prepared but currently they are specific to the number of regions being analyzed, and await generalization. One such program produces tables, for each analysis period, of the populations by age group and region, of the age group percentages for each region, and of the regional percentages for each age group. Alternatively, the user might wish to prepare tables of projected populations or age percentages or regional shares for each region separately.

5. EXAMPLES OF MOVEMENT ACCOUNTS MODELS FOR BASE PERIODS

The user who has read this far in the manual should be able to put together the instructions and component data files for any of a wide variety of different accounting and projection models, each tailored to the information system with which he or she works. However, most users will probably want just to understand a few basic principles and then search for examples of how to use the program that will match their needs. The next three sections are designed to supply these examples.

5.1 The forecast unconstrained model

The simplest way to construct accounts for the base period is to use a model incorporating initial populations, the natural movements of the population, namely births and deaths, the geographic movements of the population, namely, migrations internal to a country, emigrations out of it and immigrations into it (see Figure 6.1). Final populations are generated by the accounting model, and if all movements are correctly measured, the final populations should be the same as a count of populations taken at the end of the period.

Figure 5.1 shows the model parameter records needed for the base period instructions file. The parameters ipop, idth, iemig, intmig, imig, ibirth and isexp are all set to 1 to indicate that data should be supplied for the associated components. The component data files that need to be assembled are listed in Figure 5.1. The parameter ifpop is set to 0 to indicate that no data are expected for final populations, which are computed from the column accounting equations. The icon, itol and iterbf parameters are all set to 0 to indicate that the model is unconstrained.

The other parameters in the list do not involve data inputs but rather model or output choices. They have the following meanings: ipar = 1 means that the average populations are used as populations at risk; ires = 0 means that residual terms are computed from the row accounting equations; iverg = 1 means that the convergence criterion for the accounts based model is less than 0.5 of a person; iaccnt = 1 instructs the program to print out accounts tables for the current period; isurv = 0 instructs the program not to compute survivorship rates; and iauto = 1 tells the program to expect a new set of model parameters in the next time interval after the base period.

5.2 The backcast unconstrained model

Figure 5.2 displays the model parameter settings and data files needed for the backcast unconstrained model (see section 2.3.2 and Figure 6.2).

The settings of three parameters in the previous example have been changed: ipop = 2 means that initial populations are now computed from the row accounting equations; ires = 1 means that the residual terms are computed from the column accounting equations; and ifpop = 1 means that final populations are read in from a data file. The other parameter settings are as explained in the previous example (section 5.1).

5.1.1 Model parameters for the instructions file, base period

```
   IPOP=  1   IPAR=  1   IDTH=  1  IEMIG=  1 INTMIG=  1
   IRES=  0   IMIG=  1  IFPOP=  0  IVERG=  1 IACCNT=  1
   ICON=  0   ITOL=  0 ITERBF=  0  ISURV=  0 IBIRTH=  1
  ISEXP=  1  IAUTO=  1
```

5.1.2 Component data files to be assembled

Component	Leeds	NIDI
Initial populations	fn PIDATA	DATIP
Death	fn DDATA	DATD
Emigrations	fn EDATA	DATE
Internal migrations	fn MDATA	DATM
Immigrations	fn IDATA	DATI
Births	fn BDATA	DATB

FIGURE 5.1 The forecast unconstrained model: model parameters and component data files

5.2.1 Model parameters for the instructions file, base period

```
   IPOP=  2   IPAR=  1   IDTH=  1  IEMIG=  1 INTMIG=  1
   IRES=  1   IMIG=  1  IFPOP=  1  IVERG=  1 IACCNT=  1
   ICON=  0   ITOL=  0 ITERBF=  0  ISURV=  0 IBIRTH=  1
  ISEXP=  1  IAUTO=  1
```

5.2.2 Component data files to be assembled

Component	Leeds	NIDI
Death	fn DDATA	DATD
Emigrations	fn EDATA	DATE
Internal migrations	fn MDATA	DATM
Immigrations	fn IDATA	DATI
Final populations	fn PFDATA	DATPF
Births	fn BDATA	DATB

FIGURE 5.2 The backcast unconstrained model: model parameters and component data files

5.3 The no emigrations model

When no data are available on emigration the model parameters must be assigned the values shown in Figure 5.3 (see section 2.3.3 and Figure 6.3).

The setting of the ipop parameter is 1 because initial populations must be read in from a data file; the iemig parameter must be set to 4 so that emigrations are computed as a residual from the row accounting equations. Otherwise the parameter settings are the same as those of the previous example.

5.4 The both populations model

When reliable information is to hand for all components of the movement accounts, the "both populations" (initial and final) model can be used (section 2.3.4 and Figure 6.4). The model parameter settings and list of component data files to be assembled are given in Figure 5.4. Estimates for all seven data components are required and ipop, idth, intmig, imig, ifpop and ibirth are all set to 1. The ires parameter can be set to either 0 (row equations used) or 1 (column equations used).

This model is usually "inconsistent" in that either terms in columns of the accounts fail to add to the final populations (when ires = 0) or terms in rows of the accounts fail to add to initial populations. However, the model does have the advantage that movement or survivorship rates are computed using average observed populations as in most conventional projection models, which also ignore the consistency problem.

5.5 The constrained model

To solve the inconsistency problem the user should adopt a constrained model (see section 2.3.5). Accounts are initially estimated using one of the unconstrained models. In Figure 5.5 the relevant parameters have been given the settings of the forecast unconstrained model.

The icon parameter is set to 1 if external constraints are to be applied. These constraints should be input from a CDATA or DATC file. The responsibility for preparing sets of row and column constraints is left to the user.

Two other parameters associated with the constrained model are itol and iterbf. In the example, itol is set to 6 which means the relative error in the test for convergence of the accounts matrix on the row and column constraints is 10 to the power -6. The next parameter, iterbf, is set to 100 which means that the routine which balances the accounts matrix to the constraints is terminated after 100 iterations if convergence has not been achieved.

5.6 Other models for the base period

It is possible to construct other models for the base period. For example, national rates rather than regional flows could be input for the external migration, births and deaths components if regional information was missing. Or if the base period for

5.3.1 Model parameters for the instructions file, base period

```
   IPOP=  1    IPAR=  1    IDTH=  1   IEMIG=  4 INTMIG=  1
   IRES=  1    IMIG=  1  IFPOP=  1   IVERG=  1 IACCNT=  1
   ICON=  0    ITOL=  0 ITERBF=  0   ISURV=  0 IBIRTH=  1
  ISEXP=  1   IAUTO=  1
```

5.3.2

Component data files to be assembled

Component	Leeds	NIDI
Initial populations	fn PIDATA	DATIP
Death	fn DDATA	DATD
Internal migrations	fn MDATA	DATM
Immigrations	fn IDATA	DATI
Final populations	fn PFDATA	DATPF
Births	fn BDATA	DATB

FIGURE 5.3 The no emigratons model:
model parameters and component data files

5.4.1 Model parameters for the instructions file, base period

```
   IPOP=  1    IPAR=  1    IDTH=  1   IEMIG=  1 INTMIG=  1
   IRES=  0    IMIG=  1  IFPOP=  0   IVERG=  1 IACCNT=  1
   ICON=  0    ITOL=  0 ITERBF=  0   ISURV=  0 IBIRTH=  1
  ISEXP=  1   IAUTO=  1
```

5.4.2 Component data files to be assembled

Component	Leeds	NIDI
Initial populations	fn PIDATA	DATIP
Death	fn DDATA	DATD
Emigrations	fn EDATA	DATE
Internal migrations	fn MDATA	DATM
Immigrations	fn IDATA	DATI
Final populations	fn PFDATA	DATPF
Births	fn BDATA	DATB

FIGURE 5.4 The both populations model:
model parameters and component data files

5.5.1 Model parameters for the instructions file, base period

```
  IPOP=  1    IPAR=  1    IDTH=  1   IEMIG=  1 INTMIG=  1
  IRES=  0    IMIG=  1  IFPOP=  0   IVERG=  1 IACCNT=  1
  ICON=  0    ITOL=  6 ITERBF=100   ISURV=  0 IBIRTH=  1
 ISEXP=  1   IAUTO=  1
```

5.5.2 Component data files to be assembled

Component	Leeds	NIDI
Initial populations	fn PIDATA	DATIP
Death	fn DDATA	DATD
Emigrations	fn EDATA	DATE
Internal migrations	fn MDATA	DATM
Immigrations	fn IDATA	DATI
Births	fn BDATA	DATB
Constraints	fn CDATA	DATC

FIGURE 5.5 The constrained model (forecast version): model parameters and component data files

projection was just the last in a sequence of accounting periods the data for the previous period could be used if none were available for the current period for any component. However, these models are rather poor substitutes for either better estimation methodology by the user or better data collection or tabulation efforts by the data collecting agency.

6. EXAMPLES OF MOVEMENT ACCOUNTS MODELS FOR PROJECTION

Once base period accounts have been successfully estimated using one of the models described in section 5, the user will wish to carry out projection of the regional populations for the system of interest. In this section, we assume that the system of interest consists of N regions which together completely partition the country in which they lie, and that internal migration flows between pairs of these regions are modelled by a simple rates model:

$$\text{migration flow} = \text{migration rate} \times \text{population at risk} \qquad (6.1)$$

The choices in model construction are thus restricted to the external flows of emigration and immigration.

Each of these can be modelled as either a flow or as a rate multiplied by a population at risk, in either gross or net fashion. The choices are thus eight in number (Table 6.1), of which the net emigration models are redundant, being exactly equivalent to the net immigration models as long as the relevant numbers are the same in absolute magnitude and opposite in sign.

Figure 6.1 illustrates the nature of each of these six models, focussing on the section of the movement accounts involved with migration (the stippled area in Figure 6.1.0). Each is considered in turn.

6.1 A model with gross external migration rates

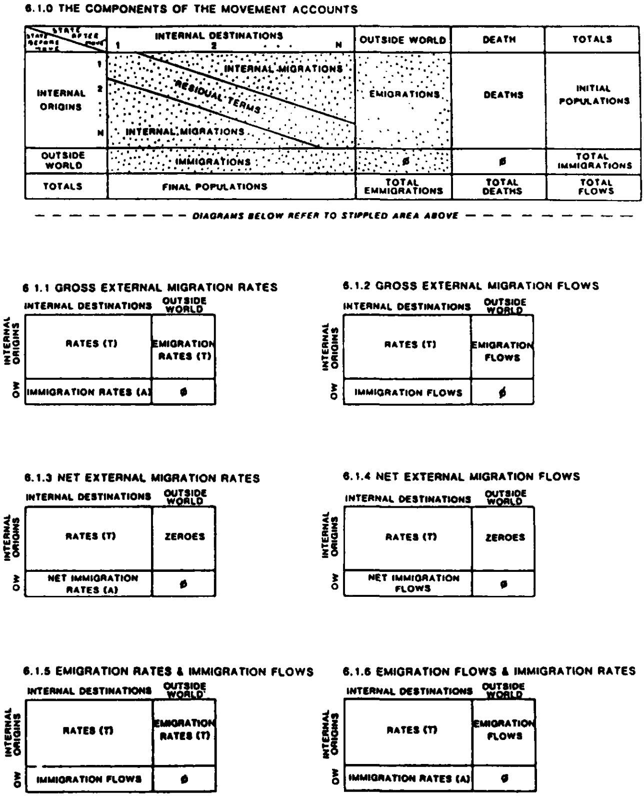
The model parameter settings for this model are given in Figure 6.2. The ipop parameter is set to 0 so that the previous period's final populations are used as the initial populations of the current period. This will be the normal setting for projection. The idth, iemig, intmig, imig and ibirth parameters are set to 2, meaning that the previous period's rates for each of these components is used. Use of the 2 setting for these components for all projection periods will result in a constant rates projection. If the component rates can themselves be forecast, rates can be introduced as fresh data in the relevant component data file, with the associated i parameters set to 3.

The ipar and iverg parameters are set to 1, as currently restricted. The ires parameter should usually be set to 0 for projection because projection models are normally forward looking. The ifpop parameter is set to 0 so that final populations are computed from the accounts. The iaccnt parameter is set to 1 in this first projection period so that the operation of the projection model can be checked at the level of detail provided by the movement accounts tables. It can in subsequent periods be set to 0 so that no further accounts tables are printed. The icon, itol and iterbf parameters are set to 0 because the projection model is unconstrained. The isurv parameter is set to 0 because an iterative accounts based model is being used for projection not a non-iterative cohort survival model. The sex proportions of the base period are used again so that isexp is set to 0. The iauto parameter remains at 1 to request the program to read a fresh set of model parameters for the second projection period.

TABLE 6.1 Choices for modelling external flows

Gross	Emigration Flows	Emigration Rates
Immigration Flows	Figure 6.1.2	Figure 6.1.5
Immigration Rates	Figure 6.1.6	Figure 6.1.1

Net	Emigration Flows	Emigration Rates
Immigration Flows	Figure 6.1.3	-
Immigration Rates	-	Figure 6.1.4



T=TRANSMISSION RATES (=FLOW/ORIGIN POPULATION AT RISK)

A=ADMISSION RATES (=FLOW/DESTINATION POPULATION AT RISK)

Figure 6.1 Examples of movement accounts models for projection

6.2.1 Model parameters for the instructions file, first projection period

IPOP=	0	IPAR=	1	IDTH=	2	IEMIG=	2	INTMIG=	2
IRES=	0	IMIG=	2	IFPOP=	0	IVERG=	1	IACCNT=	1
ICON=	0	ITOL=	0	ITERBF=	0	ISURV=	0	IBIRTH=	2
ISEXP=	0	IAUTO=	1						

6.2.2 Component data files with data entries for this period

None. All necessary data are accessed from the work file containing the previous period's data.

FIGURE 6.2 A model with gross external migration rates: model parameters and component data files

6.3.1 Model parameters for the instructions file, first projection period

IPOP=	0	IPAR=	1	IDTH=	2	IEMIG=	1	INTMIG=	2
IRES=	0	IMIG=	1	IFPOP=	0	IVERG=	1	IACCNT=	1
ICON=	0	ITOL=	0	ITERBF=	0	ISURV=	0	IBIRTH=	2
ISEXP=	0	IAUTO=	1						

6.3.2 Component data files with data entries for this period

Component	Data entries	Leeds	NIDI
Emigrations	Flows	fn EDATA	DATE
Immigrations	Flows	fn IDATA	DATI
Other data accessed from work file			

FIGURE 6.3 A model with gross external migration flows: model parameters and component data files

No further data are needed in the component data files for this projection period for the model specified by the parameter settings of Figure 6.2.1.

6.2 A model with gross external migration flows

The model parameter settings for this model (Figure 6.3) show two changes from the previous one. The iemig and imig parameters are now set to 1 so that forecast gross external migration flows for this period are input from the corresponding component data files to the program. The other parameter settings are as before.

6.3 A model with net external migration rates

Figure 6.4 gives the parameter settings for this model. There are a few changes from the previous model. The imig parameter is now set to 3, and net immigration rates should be entered in the immigrations data file for the current period. The iemig parameter is unaltered at 1 but the emigrations file should contain zeroes for the current period.

6.4 A model with net external migration flows

The model parameter settings, given in Figure 6.5 for this model, are the same as for the model incorporating gross external migration flows. The difference is in the content of the emigrations and immigrations files. The former should contain zeroes and the latter net immigration flows.

6.5 A model with emigration rates and immigration flows

A mixed model in which emigrations are modelled by a simple rates model (equation 6.1) and immigrations are entered as forecast flows is illustrated in Figure 6.6. The iemig parameter is set to 3 and the imig parameter to 1. The emigrations file contains emigrations rates for the current period and the immigrations file holds immigration flows.

6.6 A model with emigration flows and immigration rates

For this model the iemig parameter is set to 1 and the imig parameter to 3, and flows are input from the emigrations file and rates from the immigrations file.

6.4.1 Model parameters for the instructions file, first projection period

IPOP=	0	IPAR=	1	IDTH=	2	IEMIG=	1	INTMIG=	2
IRES=	0	IMIG=	3	IFPOP=	0	IVERG=	1	IACCNT=	1
ICON=	0	ITOL=	0	ITERBF=	0	ISURV=	0	IBIRTH=	2
ISEXP=	0	IAUTO=	1						

6.4.2 Component data files with data entries for this period

Component	Data entries	Leeds	NIDI
Emigrations	Zeroes	fn EDATA	DATE
Immigrations	Net flows	fn IDATA	DATI

FIGURE 6.4 A model with net external migration rates: model parameters and component data files

6.5.1 Model parameters for the instructions file, first projection period

IPOP=	0	IPAR=	1	IDTH=	2	IEMIG=	1	INTMIG=	2
IRES=	0	IMIG=	1	IFPOP=	0	IVERG=	1	IACCNT=	1
ICON=	0	ITOL=	0	ITERBF=	0	ISURV=	0	IBIRTH=	2
ISEXP=	0	IAUTO=	1						

6.5.2 Component data files with data entries for this period

Component	Data entries	Leeds	NIDI
Emigrations	Zeroes	fn EDATA	DATE
Immigrations	Net flows	fn IDATA	DATI

FIGURE 6.5 A model with net external migration flows: model parameters and component data files

6.6.1 Model parameters for the instructions file,
first projection period

```
  IPOP=  0   IPAR=  1   IDTH=  2  IEMIG=  3 INTMIG=  2
  IRES=  0   IMIG=  1  IFPOP=  0  IVERG=  1 IACCNT=  1
  ICON=  0   ITOL=  0 ITERBF=  0  ISURV=  0 IBIRTH=  2
 ISEXP=  0  IAUTO=  1
```

6.6.2 Component data files with data entries for this period

Component	Data entries	Leeds	NIDI
Emigrations	Rates	fn EDATA	DATE
Immigations	Flows	fn IDATA	DATI

FIGURE 6.6 A model with gross external migration rates: model parameters and component data files

6.7.1 Model parameters for the instructions file,
first projection period

```
  IPOP=  0   IPAR=  1   IDTH=  2  IEMIG=  1 INTMIG=  2
  IRES=  0   IMIG=  3  IFPOP=  0  IVERG=  1 IACCNT=  1
  ICON=  0   ITOL=  0 ITERBF=  0  ISURV=  0 IBIRTH=  2
 ISEXP=  0  IAUTO=  1
```

6.7.2 Component data files with data entries for this period

Component	Data entries	Leeds	NIDI
Emigrations	Rates	fn EDATA	DATE
Immigrations	Flows	fn IDATA	DATI

FIGURE 6.7 A model with gross external migration rates: model parameters and component data files

7. EXAMPLES OF ALTERNATIVE MODELS

The models described in sections 5 and 6 have common features: the system of interest consists of two or more regions which partition a country exhaustively and which interact by receiving and sending migrants. The models of section 6 are all driven by the same general process: starting populations are received from the previous period and final populations are estimated from the current period accounts. These features can be altered if comparisons with alternative models are desired. You might, for example, be interested in comparing your multiregional projection with that of your country's official demographic agency. The movement accounts program should be flexible enough to simulate most other demographic projections involving movement data, although no guarantees are offered. For projection models using transition (census) migration data see an earlier program (Rees, 1981).

A range of alternative models are sketched diagramatically in Figure 7.1. In the first, independent migration forecasts are introduced. In the second, migration is ignored entirely. In the third and fourth, the focus is on a single region and migration is modelled in net flow or rate terms. In the fifth, the multiregional system is reduced to a series of biregional systems (as in Rogers, 1976). In the sixth model the outside world is incorporated into the system of interest as the rest of the world and all migration flows are modelled in the same way. Finally, if we run a model with only two cohorts - infants born in a period and those existing at the start - we obtain an aggregate accounts model (the seventh alternative model).

7.1 A model incorporating independent projections of migration

Here we assume that the user has a suitable model that predicts interregional migration flows and that this model requires as input the initial populations of the regions at the start of the projection period.

The movement accounts program would be run one projection period at a time using model parameter settings such as those listed in Figure 7.2. The intmig parameter is set to 1 and forecast internal migration flows are input fom the associated component data file. The user's migration model would extract projected final populations from the DATPOP file and use them in the migration model to prepare new forecasts for the next period. It should be possible for the experienced user to design an executive or macro procedure that made projections for many periods possible in one run of the programs.

7.2 A model without migration

On occasions it might be necessary or useful to assess the differences that migration flows make to projected population outcomes by carrying out a natural increase only projection. This would be accomplished by entering zeroes in the internal migrations, the emigrations and immigrations files for the base period, and then by reusing this information in subsequent projection periods. The model parameter settings for two projection periods are given in Figure 7.3.

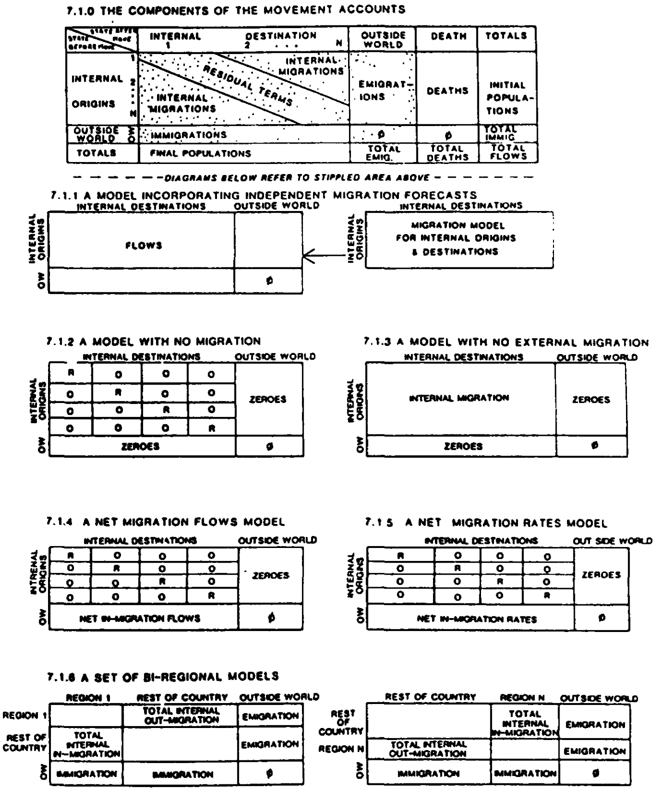
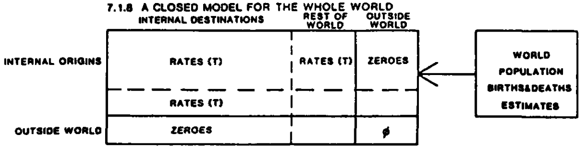


Figure 7.1 Alternative models

7.1.7 3 BI-REGIONAL MODELS ARRANGED IN ONE ACCOUNTS MATRIX

		INTERNAL DESTINATION 1	$\bar{1}$	2	$\bar{2}$	3	$\bar{3}$	OUTSIDE WORLD	DEATHS	TOTALS
INTERNAL ORIGINS	1	R	✓	O	O	O	O			
	$\bar{1}$	✓	R	O	O	O	O			INITIAL POPULATIONS
	2	O	O	R	✓	O	O	EMIGRATIONS	DEATHS	
	$\bar{2}$	O	O	✓	R	O	O			
	3	O	O	O	O	R	✓			
	$\bar{3}$	O	O	O	O	✓	R			
OUTSIDE WORLD				IMMIGRATIONS				ø		
TOTALS		FINAL POPULATIONS						TOTAL EMIGRATIONS	TOTAL DEATHS	TOTAL FLOWS

7.1.8 A CLOSED MODEL FOR THE WHOLE WORLD



7.1.9 AN AGGREGATE MODEL

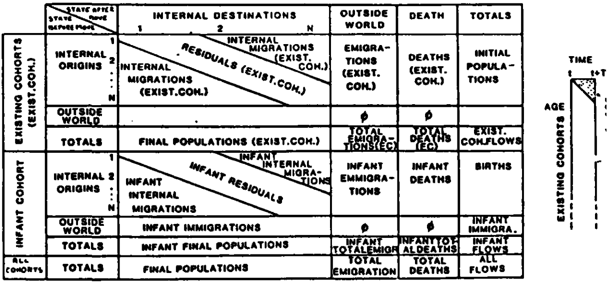


Figure 7.1 Alternative models (continued)

Model parameters for the instructions file, first projection period

IPOP=	0	IPAR=	1	IDTH=	2	IEMIG=	2	INTMIG=	1
IRES=	0	IMIG=	0	IFPOP=	0	IVERG=	1	IACCNT=	1
ICON=	0	ITOL=	0	ITERBF=	0	ISURV=	0	IBIRTH=	2
ISEXP=	0	IAUTO=	1						

Component data files with data entries for this period:

Component	Data entries	Leeds	NIDI
Internal migations	Projected internal migration flows	fn MDATA	DATM

FIGURE 7.2 A model incorporating independent projections of migration:
model parameters and component data files

Model parameters for the instructions file, base period and first projection period

IPOP=	1	IPAR=	1	IDTH=	1	IEMIG=	1	INTMIG=	1
IRES=	0	IMIG=	1	IFPOP=	0	IVERG=	1	IACCNT=	1
ICON=	0	ITOL=	0	ITERBF=	0	ISURV=	0	IBIRTH=	1
ISEXP=	1	IAUTO=	1						
IPOP=	0	IPAR=	1	IDTH=	2	IEMIG=	0	INTMIG=	0
IRES=	0	IMIG=	0	IFPOP=	0	IVERG=	1	IACCNT=	1
ICON=	0	ITOL=	0	ITERBF=	0	ISURV=	0	IBIRTH=	2
ISEXP=	0	IAUTO=	0						

Component data entries for the base period:

Component	Data entries	Leeds	NIDI
Emigrations	Zeroes	fn EDATA	DATE
Internal migrations	Zeroes	fn MDATA	DATM
Immigrations	Zeroes	fn IDATA	DATI

FIGURE 7.3 A model without migration

7.3 A model without external migration

Another comparison that the user might wish to make is between a projection model incorporating external migration flows and one that ignores these flows in order to assess the impact that external migration flows have on the future population. Figure 7.4 shows the parameter settings for the base period and the first projection period for implementing such a model.

7.4 A single region model with net migration flows

Another model commonly used for projection is the cohort survival model for a single region to which forecast net migration flows are added. If the populations of a set of regions are to be projected, the program can be run afresh for each region or the accounts matrix can be set up in the form shown in Figure 7.1.4. The internal migrations matrix is supplied with zeroes, along with the emigrations vector. The immigrations vector is supplied with net in-migration flows into each region. These in-migration flows are the sum of net internal in-migration and net external immigration flows. The model parameter settings for such a model are specified in Figure 7.5.

7.5 A single region model with net migration rates

A simple variant of the net migration model is one in which net migration rates are used rather than flows. The accounts matrix structure for such a model is given in Figure 7.1.5. The model parameter settings are as for the previous model (Figure 7.5), except that the imig parameter for the first projection period is set to 2, indicating that immigration (i.e.total in-migration) rates for the previous period are used.

7.6 A set of bi-regional models

A multiregional population model demands good estimates of the interregional migration flows by age. These are not published or produced for systems with a large number of regions. Two alternative strategies are available.

The first is to estimate from an all-age inter-regional migration matrix and regional totals of internal in- and out-migration by age the requisite region by region by age array.

The second strategy is to divide the system of n regions into n two region systems consisting of each region and the remainder of the country, using the information on total internal in- and out-migration by age that is normally available. This strategy may also be useful if there are memory storage problems in implementing the program at a new installation.

The projected results of the two strategies will, of course, differ in the long run, but according to experiments by Rogers (1976), not by a lot.

There are two ways of using the program to carry out a set of bi-regional model projections. The first would simply be to run the program n times (Figure 7.1.6) with different data sets. The second

Model parameters for the instructions file, base period and first projection period

```
  IPOP=  1    IPAR=  1    IDTH=  1   IEMIG=  1 INTMIG=  1
  IRES=  0    IMIG=  1  IFPOP=  0   IVERG=  1 IACCNT=  1
  ICON=  0    ITOL=  0 ITERBF=  0   ISURV=  0 IBIRTH=  1
 ISEXP=  0   IAUTO=  1
  IPOP=  1    IPAR=  1    IDTH=  2   IEMIG=  0 INTMIG=  2
  IRES=  0    IMIG=  0  IFPOP=  0   IVERG=  1 IACCNT=  1
  ICON=  0    ITOL=  0 ITERBF=  0   ISURV=  0 IBIRTH=  2
 ISEXP=  0   IAUTO=  0
```

FIGURE 7.4 A model without external migration flows

Model parameters for the instructions file,base period and first projection period

```
  IPOP=  1    IPAR=  1    IDTH=  1   IEMIG=  1 INTMIG=  1
  IRES=  0    IMIG=  1  IFPOP=  0   IVERG=  1 IACCNT=  1
  ICON=  0    ITOL=  0 ITERBF=  0   ISURV=  0 IBIRTH=  1
 ISEXP=  1   IAUTO=  1
  IPOP=  0    IPAR=  1    IDTH=  2   IEMIG=  0 INTMIG=  0
  IRES=  0    IMIG=  0  IFPOP=  0   IVERG=  1 IACCNT=  1
  ICON=  0    ITOL=  0 ITERBF=  0   ISURV=  0 IBIRTH=  2
 ISEXP=  0   IAUTO=  0
```

Component data files with special data entries for the base period:

Component	Data entries	Leeds	NIDI
Emigrations	Zeroes	fn EDATA	DATE
Internal migrations	Zeroes	fn MDATA	DATM
Immigrations	Net in-migration flows	fn IDATA	DATI

FIGURE 7.5 A single region model with net migration flows

would be to prepare a rather special internal part of the accounts matrix (Figure 7.1.7) with regions 1, not 1, 2, not 2, and so on to n, not n. The off diagonal blocks of the internal part of the accounts matrix would be set to zero to prevent interaction between the biregional pairs. Of course, in the resulting movement accounts tables the total emigrations, total deaths, total immigrations and total flows terms would not have any substantive meaning. However, the projected populations file (DATPOP) would contain the properly projected populations of each of the regions.

7.7 A closed model for the whole world

Another interesting comparison which the user might wish to make would be with the results of a model in which migrants were as free to move into and out of the country of interest as they were to move within it. This model would simulate the effect of relaxing any immigration or emigration controls.

To implement such a model, the rest of the world would become one of the internal regions, and immigration and emigration would be treated as internal migration. The no external migration model outlined earlier (section 7.3) would be used.

One requirement would be estimates of the world's population, births and deaths. Reference to the latest United Nations "World Population Prospects" should provide these figures. Leeds users can access some crude estimates in a series of library (GEO6LIB) files called WORLDR1 DATA ... WORLDR6 DATA prepared for use with the SIMPLE projection model (see Rees, 1978).

7.8 An aggregate model

In the absence of detailed age breakdowns of the necessary accounts components, the program can be run with age aggregated data. Two cohorts are used: the infant cohort of those existing at the start of the time interval (those aged 1 or more with a 1 year interval, 2 or more with a 2 year interval and so on). The parameters ncohrt, nalfa and nbeta are all set to 2 in the analysis parameter records of the instructions file.

However, as the diagram (Figure 7.1.9) shows, figures for the deaths, emigrations, internal migrations and immigrations data components must include estimates for these two cohorts, the infant and the existing. Fairly, crude deconsolidation factors can be used since the purpose of the model would be to produce all-age projections of the population. It should be noted that the aggregate model only produces reasonable population forecasts if the population system is nearly stable in demographic terms.

8. CONCLUDING REMARKS

8.1 A general guide to constructing population accounts and projections

If you have managed to reach this section of the manual, congratulations are in order. It should be clear to you, however, that despite the work that you have put in so far in understanding the models and the program, there is a great deal of work to be done in preparing the input data for the program, and in analyzing its outputs. This is your responsibility, because only you know what types of statistical, graphical or cartographic presentation of the results are required. However, some general guidelines are probably useful, and these are provided below. And in the next two subsections examples of what is involved in estimation of inputs and tabulatiion of outputs are provided.

Figure 8.1 provides a sketch of the time periods involved in an ideal population projection. Accounts for a set of historical periods leading up to the base period should be estimated in order to provide a good time series for forecasting any of the rates involved. Usually the base period will be the latest period for which a reasonable level of information is available. The one or two periods prior to the base period will have associated the best level of demographic information. The earliest periods in the series may have poorer levels of demographic information. It is thus likely that the methods of estimation of input variables for the accounts will vary from period to period, and from component to component.

The first projection periods will show similar variation with respect to exact accounting model used and data inputs supplied. For some m periods fresh data inputs (e.g. forecast rates) will be supplied, though m may vary from component to component. From the m+lst period data inputs and model will be constant.

This ideal historical and projection scenario might, in practice, be reduced to something simpler. The simplest projection might involve just one projection period in which data are input and then a constant scenario adopted thereafter.

8.2 An example of input estimation

8.2.1 The problem

The single year of age accounts for the Netherlands for 1976, which have been used as examples in the manual, require the estimation of regional deaths by single year of age period cohort and by sex. This can be done using the equivalent national data for the Netherlands, and regional deaths data by five year age groups and by sex. The age-time diagram for this estimation is shown in Figure 8.2. Deaths are available by sex for all variables.

The target variable to be estimated can be defined as

D(i,al,*) = deaths in region i (to persons of sex s, not explicitly recognized here) by single year of age al at time t over period t to t+1.

Time Scale	Period	Nature of model and data inputs
	H (1)	Historical period starting series
Past	⋮	⋮
	H (b-1)	Historical period of maximum information
	H (b)	Historical base period
Present →		
	P (1)	First projection period: projection model set up, data inputs
Future	P (2)	Second projection period: model changes, data inputs
	⋮	⋮
	P (m)	Mth projection period: last one with new data inputs
	P (m+1)	(M+1)th projection period: same model and data inputs as Mth
	⋮	⋮
	P (n)	Last projection period

FIGURE 8.1 The time periods involved in a population projection

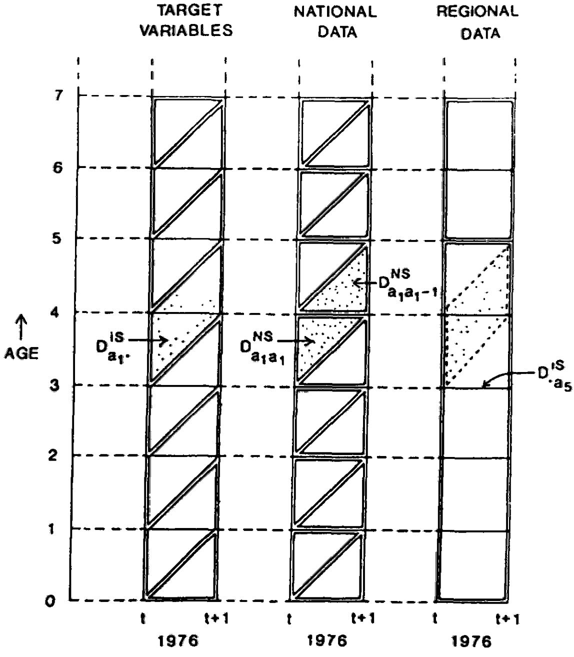


Figure 8.2 <u>An age-time space diagram for the Netherlands deaths estimation problem</u>

This target variable is the sum of two more disaggregated variables

$$D(i,a1,*) = D(i,a1,a1) + D(i,a1,a1+1) \qquad (8.1)$$

where the first age subscript refers to age at time t and the second to age at death during period t to t+1. We call a1,a1 and a1,a1+1 age transitions.

The problem is to estimate variables D(i,a1,a1) and D(i,a1,a1+1) given

D(N,a1,a1) = national deaths for persons who were in age group a1 at time t and a1 at time of death

D(N,a1,a1+1) = national deaths for persons who were in age group a1 at time t and a1+1 at time of death

D(i,*,a5) = deaths in region i of persons by five year age group at time of death.

National and regional deaths are linked in the following way

$$\sum_{a1}\sum_{a5} \{D(N,a1,a1) + D(N,a1-1,a1)\} = \sum_{i} D(i,*,a5) + D(R,*,a5) \qquad (8.2)$$

where R signifies a residual "region" containing the deaths of persons of no fixed abode and "administrative corrections", and where a1 a5 means that single years of age a1 belong to the set of five single years that make up five year age group a5. Equation (8.2) is used to establish what how many residual deaths there are

$$D(R,*,a5) = \sum_{a1}\sum_{a5} \{D(N,a1,a1) + D(N,a1-1,a1)\} - \sum_{i} D(i,*,a5) \qquad (8.3)$$

and then the region R deaths are treated in the same way as the other regional deaths.

To solve this problem we need to compute various mortality rates and for this we need initial and final populations for the Netherlands and its regions by single year of age and sex:

P(N,a1) [t] = population of the Netherlands in single year of age a1 at time t
P(i,a1) [t] = population of region i in single year of age a1 at time t
P(N,a1) [t+1] = population of the Netherlands in single year of age a1 at time t+1
P(i,a1) [t+1] = population of region i in single year of age a1 at time t+1.

As with deaths the residual region populations are worked out by subtracting the sum of the regional populations from the corresponding national populations.

8.2.2 The solution

Step 1: populations at risk

The national populations at risk, PAR, for each age transition are computed using

$$PAR(N,al,al) = 0.75\ P(N,al)\ [t] + 0.25\ P(N,al+1)\ [t+1] \quad (8.4)$$

$$PAR(N,al,al+1) = 0.25\ P(N,al)\ [t] + 0.75\ P(N,al+1)\ [t+1] \quad (8.5).$$

The regional populations at risk are computed similarly.

Step 2: national death rates

The national death rates for each age transition are worked out:

$$d(N,al,al) = D(N,al,al) \ / \ PAR(N,al,al) \quad (8.6)$$

$$d(N,al,al+1) = D(N,al,al+1) \ / \ PAR(N,al,al+1) \quad (8.7).$$

Step 3: initial estimates of regional deaths

The national death rates are applied to the regional populations at risk to obtain first estimates of regional deaths in each age transition

$$D(i,al,al)\ [1] = d(n,al,al)\ PAR(i,al,al) \quad (8.8)$$

$$D(i,al,al+1)\ [2] = d(N,al,al+1)\ PAR(i,al,al+1) \quad (8.9).$$

Step 4: the estimates are constrained to regional deaths by five year age group and to national deaths by single year transition

To adjust the first estimates to the marginal constraints of known regional totals by five year age group we apply

$$D(i,al-1,al)\ [2] = D(i,al-1,al)\ [1] \times \left[\frac{D(i,*,a5)\ [1]}{\sum_{al} \sum_{a5} \{D(i,al-1,al)\ [1] + D(i,al,al)\ [1]\}} \right] \quad (8.11)$$

$$D(i,al,al)\ [2] = D(i,al,al)\ [1] \times \left[\frac{D(i,*,a5)}{\sum_{al} \sum_{a5} \{D(i,al-1,al)\ [1] + D(i,al,al)\ [1]\}} \right] \quad (8.11)$$

These estimates must then be adjusted to national totals

$$D(i,al-1,al)\ [3] = D(i,al-1,al)\ [2] \times D(N,al-1,al) \ / \ \sum_{i} D(i,al-1,al)\ [2] \quad (8.12)$$

$$D(i,al,al)\ [3] = D(i,al,al)\ [2] \times$$

$$D(N,al,al) \;/\; \sum_{i} D(i,al,al)\,[2] \qquad (8.13).$$

The estimates are tested for convergence:

$$\text{if abs } \{\, D(i,al-1,al)\,[3] - D(i,al-1,al)\,[1] \,\} < k \qquad (8.14)$$

and

$$\text{if abs } \{\, D(i,al,al)\,[3] - D(i,al,al)\,[1] \,\} < k \qquad (8.15)$$

for all regions i and age groups at death al, the process is terminated, where k is a small constant (say 0.5). Otherwise, the stage [3] estimates are reintrroduced at stage [1]

$$D(i,al-1,al)\,[1] = D(i,al-1,al)\,[3] \qquad (8.16)$$

$$D(i,al,al)\,[1] = D(i,al,al)\,[3] \qquad (8.17)$$

and the adjustment procedure is repeated.

Have we completed the estimate now? By no means the equations above apply to the second through to the last but one age transition. Adjustments must be made for the first and last transitions in the population at risk:

$$PAR(N,b,0) = 0.25\; B(N)\,[t,t+1] + 0.75\; P(N,*,0)\,[t+1] \qquad (8.18)$$

where b refers to birth in the period and B are births;

$$PAR(N,90+,91+) = 0.5\; P(N,90+)\,[t] + 0.5\; P(n,91+)\,[t+1] \qquad (8.19)$$

if the last age transition is from ages 90+ to 91+. And then all the equations must be converted into a computer program, at least half of which is devoted to checking and cross-checking the constraints data.

8.3 An example of output tabulation

The program listed in Figure 8.3 enables the user to construct easy to read tables of projected populations, age group percentages and zonal percentages for a 20 zone system. It is written in the WATERLOO dialect of BASIC, with plenty of remark statements and printed messages so that any competent programmer should be able to translate it into a more familiar dialect or language and use it with his or her projections.

8.4 Future developments

There are many ways in which this program for analyzing the population dynamics of many, interacting regions might be developed.

These might include:

-- better tabulation of projected populations

-- graphical presentation of the outputs

```
00010 rem *** disppop wbasic ***
00020 ' this program reads in the final projected populations output by
00030 ' disaba fortran and arranges them in useful, easy to read tables
00040 dim p(21,20),pcta(21,20),pctz(21,20),nam$(21),z(21,20)
00050 ' variable definitions  p= popn at end of period by zone and age
00060 ' pcta = pct in age groups   pctz = pct in zones   z = printing var
00070 data ' ni', cc , sr , tw , nr , sv , wv , vhr , em , ea
00080 data  ose   cma    gl    sw    wmc , wmr    gm    m   nwr ,  w
00090 data   ub
00100 for i=1 to 21 % read nam$(i) % next i
00110 ' 20 ages = (order in datpop file) 5-9,.. ,90+,0-4 plus total
00120 rem *** the population data are read in from the file
00130 print  enter the name of the file containing the projected popns
00140 input name$ % filename$=name$+  DATPOP
00150 open£5,filename$,input
00160 filename$=name$+  TABLES (RECFM F LRECL 13
00170 open£6,filename$,output
00180 form1$= ££££££££.  % form2$= ££££££££££
00190 print  please enter the number of projected population vectors
00200 input nt
00210 for t=1 to nt
00220 c=1
00230 c=c+1
00240 for i=1 to 20 % input£5,using form1$,p(i,c); % ne t i
00250 input£5,using form2$,blank$
00260 if c=1 % goto 290
00270 elseif c=19 % c=1 % goto 240
00280 else % goto 230 % endif
00290 rem *** zonal population totals are computed ***
00300 for i=1 to 20 % p(i,20)=0
00310 for a=1 to 19 % p(i,20)=p(i,20)+p(i,a) % ne t a
00320 next i
00330 rem *** age group population totals are computed ***
00340 for a = 1 to 20 % p(21,a)=0
00350 for i=1 to 20 % p(21,a)=p(21,a)+p(i,a) % ne t i
00360 ne t a
00370 rem *** the age group percentages are computed ***
00380 for i=1 to 21 % for a=1 to 20
00390 pcta(i,a)=100*p(i,a)/p(i,20) % next a % next i
00400 rem *** the zonal percentages are computed ***
00410 for a=1 to 20 % for i=1 to 21
00420 pctz(i,a)=100*p(i,a)/p(21,a) % next i % next a
00430 rem *** the tables are printed ***
00440 title$= Table 1. Projected populations for U  zones. Run =
00450 year=1981+(t-1)*5
00460 print£6, 1';title$;name$;   year =  ;year
00470 for i=1 to 21 % for a=1 to 20 % z(i,a)=p(i,a) % next a % next i
00480 form$='>£££££££££' % gosub 620
00490 title$='Table 2.  Age group percentages for U  zones.  Run =
00500 print£6, 1';title$;name$,   year =  ;year
00510 for i=1 to 21 % for a=1 to 20 % z(i,a)=pcta(i,a) % next a % next i
00520 form$=     £££.££  % gosub 620
00530 title$= Table 3.  Zonal percentages for U  zones.  Run =
00540 print£6,'1 ;title$;name$;   year =  ;year
00550 for i=1 to 21 % for a=1 to 20 % z(i,a)=pctz(i,a) % next a % next i
00560 form$=     £££.££  % gosub 620
00570 rem *** end of time loop ***
00580 next t
00590 close£5 % close£6
00600 print 'projected populations have been tabulated
00610 stop
00620 rem *** the table printing routine ***
00630 f1$= -  % f2$=     ££      % f3$=    £££
00640 f4$=    +  ££ +'  +  ££ +  to   + '££ + - +  ££
00650 for x=1 to 2
00660 gosub 800
00670 if x=1 % i1=1 % i2=11 % elseif x=2 % i1=12 % i2=21 % else % endif
00680 print£6,  ':'  age group . ': % for i=i1 to i2 % print£6,&
00690 using f2$,i; % next i % print£6,'
00700 print£6,  ':                :
00710 for i=i1 to i2 % print£6,using f3$,nam$(i);% next i % print£6,
00720 gosub 880
00730 print£6,   : birth to  0- 4 :
00740 for i=i1 to i2 % print£6,using form$,z(i,1); % ne t i % print£6,
00750 for a=2 to 18
00760 a1=(a-2)*5 % a2=a1+4 % a3=a2+1 % a4=a3+4
00770 print£6,using f4$,a1,a2,a3,a4;
00780 for i=i1 to i2 % print£6,using form$,z(i,a);%ne t i % print£6,
00790 ne t a
00800 print£6,   , 85+   to 90+    :
00810 for i=i1 to i2 % print£6,using form$,z(i,19); % next i % print£6,
00820 gosub 880
00830 print£6,   : total              :
00840 for i=i1 to i2 % print£6,using form$,z(i,20); % next i % print£6,
00850 gosub 880
00860 ne t x
00870 return
00880 print£6,   : % for i=1 to 120 % print£6,f1$; % next i % print£6,
00890 return
00900 end
```

Figure 8.3 <u>A program to tabulate projected populations for a 20 zone system</u>

-- use of the survivorship rates in life table and stable population analysis

-- addition of a household projection model to this demographic one.

-- documentation of data estimation methods and programs

-- adjustment of the accounts to more varied constraints

and

-- specification of general methods of constraint construction.

REFERENCES

Rees, P.H. (1978) SIMPLE: a computer program for a single region, closed system, single sex, cohort survival projection model. Computer Manual 6, School of Geography, University of Leeds.

Rees, P.H. (1981) Accounts based models for multiregional population analysis: methods, program and user's manual. Working Paper 295/ Computer Manual 13, School of Geography, University of Leeds.

Rees, P.H. (1983) Choices in the construction of regional population projections. Paper presented at the Anglo-Dutch symposium on Migraton, Soesterberg, Netherlands, 14-16 September, 1983.

Rees, P.H. and Convey, A.L. (1984) Spatial population accounting. In Clarke, J.. (ed) Geography and Population: Approaches and Applications. Pergamon Press.

Rogers, A. (1976) Shrinking large scale population projection models. Environment and Planning A, 8, 515-541..

Stillwell, J.C.H. (1980) IMP2: a computer program for inter-area migration analysis and projection user's manual. Computer Manual 12, School of Geography, University of Leeds.

Stillwell, J..C.H. (1983) Analysis of internal migration in the UK based on transition and movement data. Paper presented at the Anglo-Dutch Symposium, Soesterberg, Netherlands, 14-16 September, 1983.

Willekens, F. and Baydar, N. (1983) Forecasting place-to-place migration with generalized linear models (GLM). Paper presented at the Anglo-Dutch Symposium, Soesterberg, Netherlands, 14-16 September, 1983.

Willekens, F. and Drewe, P. (1984) A multiregional model for regional demographic projection. Chapter 13 in Willekens, F. and Ter Heide, H. (ed) Demographic research and spatial policy. Academic Press, forthcoming.

APPENDIX: A PROGRAMMER'S GUIDE TO THE MOVEMENT ACCOUNTS PROGRAM

A.1 A listing of the program source

A printout of the source listing of the program can be supplied by the author. The program is written in FORTRAN77 for use with the FORTVS (IBM) compiler. Copies of the computer program can be obtained from the author on magnetic tape at a modest charge together with listings of the program text and examples. Many features of the program are specific to the Leeds environment (Amdahl 470 or 580 with the CMS operating system) but the progam has been successfully adapted without too much difficulty to the NIDI computinq environment (Data General Eclipse).

A.2 Structure of the program

A listing of the main segment of the program is given in Figure A.1. The corresponding flow chart is displayed in Figure A.2. The program is organized in four loops: the outermost loop moves the computations from one period to the next; the next loop moves the computations through the cohorts from the second to the last and then to the first (hence its "home made" structure); the third loop effects the computation of accounts for the requested number of sexes; and the innermost loop is the iterative loop of the accounts based model (see section 2).

Prior to the start of the computations the information general to the analysis is input in the PARAM1 subroutine and a set of parameter calculations are made in the MAGIC subroutine. After the end of each period loop the PFSTCK subroutine is used to store various final populations and populations at risk. At the start of each period model parameters are input via the PARAM2 subroutine.

The model subroutines are called within the innermost loop of the program: BEGPOP handles the initial populations and communicates with the PIDATA/DATIP file; PARA computes the populations at risk; DEATHS handles the deaths component of the accounts and communicates with the DDATA/DATD file; EMIG handles the emigration component of the accounts and communicates with the EDATA/DATE file; MIG handles internal migrations and receives inputs from the MDATA/DATM file; RES computes the residual terms of the accounts; IMMIG deals with immigrations and receives data from the IDATA/DATI file; ENDPOP looks after final populations and connects with the PFDATA/DATPF file if necessary; VERGE tests whether the accounts have converged or not.

Once convergence has been achieved a further set of model subroutines are called: ACCASS assembles all the component variables in one array; CONSTR adjusts that array to marginal constraints, if requested, and receives those constraints from the CDATA/DATC file; the ACCNTS subroutine arranges the printing of the movement accounts tables (for each cohort) to the RESULTS file and writes out the projected populations to the DATPOP file; the RATES subroutine calculates movement rates and writes those out, if requested, to the RESULTS file, and also computes, if asked, the survivorship rates and writes them out to the DATSUR file. A final model subroutine, BIRTHS, is called only for the first cohort, and only after all the othe cohorts have been processed. It handles the

```
      IMPLICIT REAL*8 (A-H,O-Z)
      COMMON/A/NINREG,NPRIOD,NCOHRT,NSEX,NITER,NTAB,NAGEIN
      COMMON/B/NYEAR,NTIMIN,NFSEX,NALFA,NBETA
      COMMON/C/NIN,NOUT,NDIR,NSUR,NF,NS,NF
      COMMON/I/IPOP,IPAR,IDTH,IEMIG,INTMIG,IRES,IMIG,IFPOP,
     *IVERG,IACCNT,ICON,ITOL,ITERBF,ISURV,IBIRTH,ISEXP,IAUTO
C--------DESCRIPTION OF THE PROGRAM ----------------------------------
C'MOVE     = ACCOUNTS BASED MODEL FOR MOVEMENT DATA
C'AUTHOR   = PHILIP REES,SCHOOL OF GEOGRAPHY,UNIV.OF LEEDS,UK
C'DATE     = MAY-JULY, 1982  REVISED AUGUST 1983, DEC 1983, JAN 1984
C'USE      = TO CONSTRUCT ACCOUNTS FOR UP TO 21 REGIONS AND
C'           GENERATE SURVIVORSHIP RATES FOR A POPULATION PROJECTION
C'           MODEL. THE PROGRAM CAN ALSO BE USED FOR PROJECTION
C'           WITH MOVEMENT DATA. THE NUMBER OF REGIONS HANDLED
C'           CAN BE INCREASED BY CHANGING THE STORAGE STATEMENTS.
C'THE MAIN SEGMENT ORGANIZES THE FLOW OF WORK IN THE PROGRAM.
C'THE METHOD OF FILE HANDLING MAY NEED MODIFICATION WHEN THE PROGRAM
C'IS IMPLEMENTED AT OTHER INSTALLATIONS.  AT LEEDS THE WORK FILE
C'MUST BE CREATED PRIOR TO RUNNING THE PROGRAM, FOR EXAMPLE
C ---------------------------------------------------------------------
      NT=5
      CALL PARAM1
      CALL MAGIC
      DO 1 JPRIOD=1,NPRIOD
      CALL PARAM2(JPRIOD)
      JCOHRT=1
    9 JCOHRT=JCOHRT+1
      IF(JCOHRT.GT.NCOHRT) GO TO 8
    5 DO 3 JSEX=1,NSEX
      DO 4 JITER=1,NITER
      CALL BEGPOP(JPRIOD,JCOHRT,JSEX,JITER,IPOP)
      CALL PARA   (JPRIOD,JCOHRT,JSEX,JITER,IPAR)
      CALL DEATHS(JPRIOD,JCOHRT,JSEX,JITER,IDTH)
      CALL EMIG   (JPRIOD,JCOHRT,JSEX,JITER,IEMIG)
      CALL MIG    (JPRIOD,JCOHRT,JSEX,JITER,INTMIG)
      CALL RES    (JPRIOD,JCOHRT,JSEX,JITER,IRES)
      CALL IMMIG  (JPRIOD,JCOHRT,JSEX,JITER,IMIG)
      CALL ENDPOP(JPRIOD,JCOHRT,JSEX,JITER,IFPOP)
      CALL VERGE  (JPRIOD,JCOHRT,JSEX,JITER,IVERG,IFLAG)
      IF(IFLAG.EQ.0)GO TO 6
    4 CONTINUE
    6 CALL ACCASS(JPRIOD,JCOHRT,JSEX,JITER)
      CALL CONSTR(JPRIOD,JCOHRT,JSEX,JITER,ICON,ITOL,ITERBF)
      CALL ACCNTS(JPRIOD,JCOHRT,JSEX,JITER,IACCNT)
      CALL RATES  (JPRIOD,JCOHRT,JSEX,JITER,ISURV)
    3 CONTINUE
      WRITE(NS,201) JCOHRT
  201 FORMAT(1H ,'ACCOUNTS FOR COHORT = ',I2,' FINISHED')
      IF(JCOHRT.EQ.1)GO TO 7
    2 GO TO 9
    8 CALL BIRTHS(JPRIOD,IBIRTH,ISEXP)
      JCOHRT=1
      GO TO 5
    7 CALL PFSTCK(JPRIOD)
      WRITE(NS,202) JPRIOD
  202 FORMAT(1H ,'ACCOUNTS FOR PERIOD = ',I2,' COMPLETED')
    1 CONTINUE
      WRITE(NS,200) NPRIOD
  200 FORMAT(//1H ,'MOVEMENT ACCOUNTS FOR ',I4,' PERIOD(S) COMPLETED')
      STOP
      END
```

Figure A.1 <u>A listing of the main segment of the MOVE program</u>

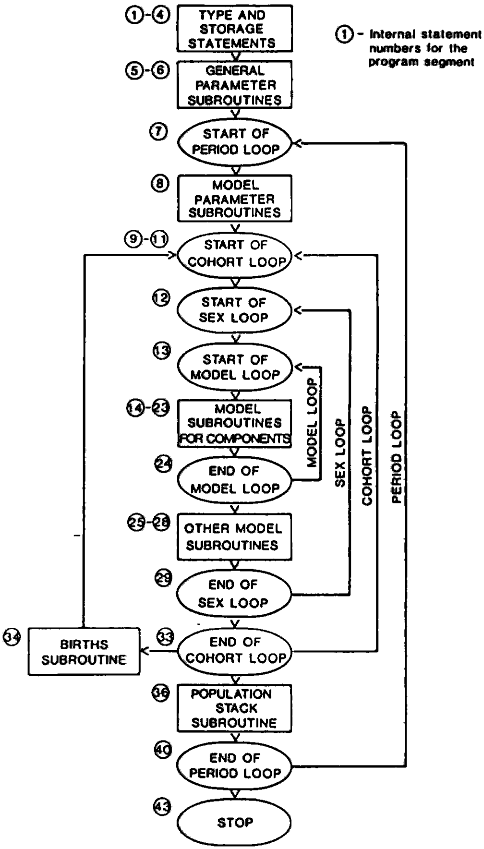


Figure A.2 A flow chart of the main segment of the MOVE program births component and communicates with the BDATA/DATB file.

There are some six other subroutines in the program which handle routine or repeated tasks. The DAFMAT, DAFTOT and DAFVEC subroutines read from and write to the direct access file (DATDSK/WORK), handling accounts matrices, accounts totals and accounts vectors respectively. These subroutines are called by the model subroutines. The INVERT subroutine inverts a matrix and is called in the RATES subroutine if survivorship rates are computed. The MATOUT subroutine writes out a matrix of flows or rates, and is called for that purpose by the ACCNTS and RATES subrroutines. The REWRIT subroutine is called by the CONSTR subroutine to replace the unconstrained accounts variables by their constrained equivalents.

A.3. Items requiring modification for different implementations

(1) All REAL variables are declared as IBM REAL*8. This gives approximately 16 significant digits. Double precision is not required.

(2) The way in which files are handled may need modification. In the NIDI implementation the accounts components files, which in the Leeds implementation are all sequential access files, open throughout the program run, were converted to direct access files opened when necessary, so as to keep within the restricted limits of the number of files that could be simultaneously be open.

(3) The work file in the Leeds implementation contains 7 items per record and the file is of record length 70 bytes. In the NIDI implementation 6 items were carried per record and the record length was 60. If a different record length gives more efficient storage (it depends on the number of internal regions in the system), then this feature can be altered.

(4) In all files data items occupy fields of 10 bytes so as to cater for the largest cohort populations likely to be met. This could be reduced to 8 bytes if the programmer is confident no accounts numbers will exceed 9999999. However, in that case it would not be possible to implement a world model!

12

Produced by

School of Geography

University of Leeds

Leeds LS2 9JT

From whom copies may be obtained