STOCK AND FLOW MODELS
FOR THE SUDANESE EDUCATIONAL SYSTEM

by

Leila Fuad Aboulela

London School of Economics and Political Science

Thesis submitted to the University of London for the degree of Master of Philosophy.

June 1991

ABSTRACT

The thesis examines the possibility of applying a Markov planning model to the Sudanese educational system. The limitations of the available data published by the Sudanese Ministry of Education is examined and the quality of the data discussed.

Study of the system establishes the presence of a bottleneck between secondary and higher education due to shortages of places in the latter. Two adaptations of the simple Markov model are proposed in which the flow of students into higher education is determined by the number of vacancies. The first model considers the case when a capacity constraint exists in the first grade of a particular higher education institute. In the second model it is the total size of higher education which is assumed to be fixed and expansion or contraction of the capacity constraints is allowed. For both models, it is shown that a steady-state exists and can be evaluated.

A serious limitation of the available data is the lack of flow rates which therefore must be estimated. The estimation methods available assume a system that is constant over time. As the Sudanese educational system is expanding an extension of the original regression method was developed to account for growth. The procedure was used to obtain estimates of the transition rates of students in different parts of the Sudan. The fit of the model was good in the majority of the cases and validating the prediction of the model with newly published data was successful.

Lastly, a simulation program was developed which generated artificial data sets from which transition rates were estimated. Sampling distributions of these estimates were then obtained by repetitive simulations. Studying these distributions showed the estimation technique to be effective in terms of ability to estimate the true transition rates and make reasonable predictions.

ACKNOWLEDGEMENTS

I wish to thank my supervisors Professor David Bartholomew and Dr.Celia Phillips for their guidance and advice during the course of my work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Review of the literature

Modelling social and economic phenomena often requires models which are stochastic rather than deterministic. This is due to the unpredictability of human behaviour and the uncertain nature of the social environment. Models that are based on stochastic processes are therefore often used to describe systems, gain insight into their dynamics and predict their future behaviour under certain assumptions. The success of modelling social and economic processes is somewhat reflected in the diversity of applications; from consumer behaviour Massy and co-workers(1970) and Gupta(1986) to geographical mobility Tziafetas(1980) and states of credit accounts Frydman, Kallberg and Kao(1985). More examples are given in Bhat(1984) who also provides examples of applications in the biological sciences. Bartholomew(1977a) gives an introduction to the analysis of data arising from stochastic processes while a theoretical account of the field and a full bibliography is given in Bartholomew(1982).

One of the largest field of application of stochastic models has been manpower and educational planning. Early examples go back to Seal(1945) and Vajda(1947). Many of the applications have made use of the discrete time Markov chain. Time is assumed to be discrete either because changes in the process occur only at discrete points of time (as in the case of education), or because a process which develops continuously is observed only at discrete time points.

Individuals are grouped together in categories or grades (usually hierarchial), according to characteristics they share in common such as age, seniority etc. It is assumed that they move independently between the grades and into and out of the system with fixed transition probabilities. The central point of interest is the distribution of numbers in the grades and the prediction of future stocks and flows under various conditions. An account of the application of the Markov model to manpower planning is given in Bartholomew(1971) and in the textbook by Bartholomew and Forbes(1979) which contains a full bibliography. Examples of the application of the Markov model are many; it was used by Young and Almond(1961) and later Leeson(1980) for predicting distributions of staff while Sales(1971) applied it for a branch of the Civil Service. Gani(1963) used it for projecting student enrollment in university. Other applications in education include Thonstad(1969), Clough and MacReynolds(1966), Kamat(1968), Armitage,Phillips and Davies(1970) and Armitage,Smith and Alper(1969). More applications in education are given by Stone(1965) and (1972), Balinsky and Reisman(1972) and (1973) and Britney(1975).

The second main class of models of hierarchically graded manpower systems are renewal models. In systems where the grade sizes are fixed, renewal theory provides the mathematical foundation. The application of renewal theory to manpower systems in introduced by Bartholomew(1963),(1976) and a full account is given in Bartholomew(1982) and practical applications in Bartholomew and Forbes(1979).

Attempts have been make to generalize Markov models in order to achieve more realistic applications. Moya-Angeller(1976) considers a system which acts in an intermediate manner between the Markov and Renewal model because there are capacity constraints which place limits on the grade sizes. Once these limits are reached the surplus has to be relocated in other grades. This is an extension of the bottleneck models introduced earlier by Armitage and co-workers(1969). Young and Vassiliou(1974) also consider the problem of capacity constraints by allowing the numbers promoted to depend not only on the stock from which they come but also on the size of the destination grade. A system in which transition probabilities are changing over time cannot adequately be described by a simple Markov chain model and requires a generalization. Kalamatianou(1984) and (1988) discusses a model for responding to promotion blockages where she assumes that 'pressure' is created when there is an increase in the numbers of eligible employees that are passed over for promotion. When managers are faced with a high 'pressure' for promotion they respond by changing the promotion policy.

In addition to the widespread application of Markov models, research has been undertaken on more theoretical aspects; on the limiting properties of the model Vassiliou(1981),(1982) and Woodword(1983a), the variances and covariances of the grade sizes Vassilou and Gerontidis(1985) and on forecasting grade, age and length of service distribution Woodward(1983b). De Stavola(1988) gives a number of tests for departure form time homogeneity in Multistate Markov processes. On the geometric probabilistic relationship in a Markov manpower model see Davies(1983) and Wegner(1985) argues for the use of simulation in corporate manpower planning.

Significant new advances have been made in the area of modelling social processes. Bartholomew(1983) reviews recent developments and gives a brief summary of the general literature on the subject. Developments have taken place in the application of continuous time models Plewis(1981) and Tuma, Hannan and Groenveld(1979). Examples of semi-Markov models were discussed by McClean(1980), Thompson(1981) and Davies(1985). Hassani(1980) discusses the application of semi-Markov models in manpower planning and his thesis contains a full bibliography. A more up-to- data analysis in given in a collection of works edited by Jansen(1986). Another interesting development has been the investigation of non-linear models; Bartholomew(1984) gives a review of the relevant literature. An example of non-linearity is provided by Conlisk(1976),(1978) who first introduced the term 'interactive' Markov chain. One of the basic assumptions of the Markov chain model is that individuals move independently of each other. However in reality the behaviour of individuals is affected by the behaviour of others. Conslink proposed to allow for this interaction between individuals by allowing the transition probabilities of the Markov chain to be a function of the state probability vector.

An area which has attracted a great deal of research is the theory of control of Markov models. Abdellaoui(1985) reviews the development in the subject with regard to stochastic control in manpower planning and gives a full bibliography. A full treatment is given in Bartholomew(1982) which is developed from earlier work in Bartholomew

(1977b), Forbes(1971), Davies(1973) and Vajda(1975). Bartholomew and Forbes(1979) gives an elementary account of the theory in the context of practical manpower planning. Some examples of recent work includes Davies(1982), Haigh(1983) on the maintainability of manpower structures and Abdallaoui(1987) on the probability of maintaining or attaining a structure in one step.

One of the practical problems in manpower planning applications is the estimation of the parameters of the Markov model. The techniques in use vary with the type of data that is available. Collins(1974) discusses the estimation of Markov transition probabilities when micro-unit (flow data) is available. The estimators in this case are the maximum likelihood estimators; the observed proportions of flows or in other words the ratio of the total flow to the total stock. More on the methods of maximum likelihood for this type of application is given by Anderson and Goodman(1957). The general advances in data collection and in particular the availability of detailed flow data has made this method the most widespread in empirical applications of Markov chains.

Sometimes ,however, what is only available is aggregate stock (macro) data giving the proportions or numbers observed for each state at each moment of time. Lee, Judge and Zellner(1970) presented a number of different estimators and compared them by simulations. Their main approach was based on regression analyses and it led to many empirical applications to mention a few; in consumer behaviour Sherif and Thompson(1980), Kelton and Kelton(1982) and voter transition behaviour McCarthy and Ryan(1977). The work of Lee and co-workers(1970) also gave rise to work on the statistical problems involved;see for example MacRae(1977), Kelton(1981), Van der Plas(1983), McLeish(1984) and Kelton and Kelton(1984). Kalbfleish and Lawless(1984) introduced a weighted least-square estimator and compared it with previous estimators. Lawless and McLeish(1984) compared the information content of macro data with that of complete micro data. They were able to show that in some instances, aggregate data can give good estimation of equilibrium distributions and mean occupancy numbers for states in a chain. More evidence on the possibility of obtaining good estimates form macro data is given in asymptotic terms by Thorburn(1982) and with simulation earlier by Lee,Judge and Zellner(1970). Leeflang(1974) suggested the combination of micro and macro data in the estimation of transition probabilities. An example in which both micro and macro data is available is the area of consumer behaviour where micro data is obtained from panels of households and macro data from retail store audits. More work on using such combined data is given in Rosenqvist(1986).

The above review is by no means inclusive but many of the references given will lead to others. Of note Bartholomew(1982) and (1983) contains many references on the stochastic modelling of social processes that have not been mentioned. With regard to the area of parameter estimation of the Markov model, Rosenqvist(1986) provides a full bibliography. For a more comprehensive but non-statistical study on manpower planning in the Sudan see Ali(1986), which leads to many references specific to the Sudanese application.

1.2 Objectives and hypotheses of the research

The following section reviews the objectives and hypothesis of the thesis. My first aim was to carry out an application of a Markov planning model on the educational system in the Sudan with emphasis on the movement of students between secondary and higher education. This would be the first time such an application had been made on Sudanese education and could pave the way for future stochastic modelling of the system. Education at all levels is limited; only 40% of all those aged 7-18 go to school. Thus education is neither universal nor compulsory which makes a demographic approach to planning, based on predicting school-age population, unrealistic. What can be done is study the stocks and flows of the system, describe its dynamics and make predictions based on the present propensities.

A study of the data available (in the form of annual official statistics published by the Sudanese Ministry of Education), revealed serious limitations that posed problems in applying the model. The major limitation was the unavailability of flow data. While detailed stock data was available and presented in terms of the total numbers in each grade by region and by sex, flow figures on the numbers being promoted, those repeating and any wastage rates were unavailable. It was therefore necessary to concentrate on the problem of estimating transition probabilities from a sequence of stock data assuming an underlying Markov process.

Studying the data on secondary and higher education revealed that shortages of places in the latter restricts the entrance of students. The demand for higher education is greater than its capacity limits and this excess of demand over supply results in an 'overspill' of eligible students. Of this 'overspill' a proportion leave the system and a proportion decide to remain and increase the number of applicants in the following year. Such a bottleneck is reflected in the steadily falling proportion of eligible students who are admitted each year into higher education. This breaks down the Markovian assumptions of 'push flows' (where the impetus for change resides in the conditions in the state in which the flow originates) and there is a need to consider adaptations to the simple Markov model.

Taking the above two points the hypotheses being tested can be summarized as follows:

1. Within secondary schools, the movement of students follows a discrete time Markov process. Differences might exist, however, between the transition rates of boys and girls and between provinces which makes it necessary to model each separately.

2. Due to shortages of places in higher education, the movement of students into higher education is restricted and the Markov assumptions do not hold. Such a system with a bottleneck can be modelled in terms of an intermediate model between the Markov model and the Renewal model where flows take place only to fill vacancies.

3. The unavailability of flow data is not an obstacle to applying a flow model as the transition probabilities can

be estimated adequately. The estimation methods available that assume a constant-sized system can be extended in order to account for expansion as is the case for the Sudanese system.

Through testing the above hypotheses, the aim is also to generalize bottleneck models in order to allow expansion or contraction of the capacity limits, and extend the method of estimating transition probabilities from stock data for use on expanding systems.

1.3 Structure of the thesis

The thesis is made up of six chapters. Chapter I as an introduction reviews the relevant literature, states the objectives and hypothesis of the research and outlines the structure of the thesis.

Chapter II has two aims : to give a general description of the Sudanese educational system as well as highlight particular problems that are examined in more detail in the following chapters. These are the presence of a bottleneck between secondary and higher education and the lack of detailed flow data. The chapter also includes a discussion on the quality of the available data.

In Chapter III two bottleneck models are proposed to the model the movements of students from secondary to higher education. The first considers the case when a capacity constraint exists in the first grade of a particular higher educational institute. In the second model, it is the total size of higher education which is assumed to be fixed. The model allows for the expansion or contraction of the capacity constraints and is thus a more generalized bottleneck model. For both models it is shown that a steady-state exists and can be evaluated. Numerical examples using hypothetical data are also given.

Chapter IV gives an overview of the available methods for estimating transition probabilities from aggregate stock data. An extension of the original regression method for expanding systems is presented. The procedure is applied to the data and transition probabilities of students in different parts of the country are estimated. The predictions of the model are then verified with newly published data.

In Chapter V a simulation program is developed in order to assess the effectiveness of the estimation procedure presented in Chapter IV. The simulation set generates artificial data upon which the model is fitted and prediction errors are calculated. The chapter then studies the sampling distributions of the estimates obtained from repetitive simulations under various conditions.

As a final chapter, Chapter VI presents the conclusions of the research and offers some suggestions for future research.

CHAPTER II

THE SUDANESE EDUCATIONAL SYSTEM

The Sudan is the largest country in Africa covering an area of 2.5 million square kilometres, nearly one-tenth that of the continent. It shares borders with Egypt and Libya to the north, Chad, Central African Republic and Zaire to the west, Uganda and Kenya to the south and Ethiopia and the Red Sea to the East.

The Sudan is thinly populated with a population of about 22.2 million and has an annual growth rate of about 2.8%. The age structure is very young with 46% of the population aged under 15 and only 3% aged 65 and over. The country is predominantly rural with 69% of the population living in rural communities and 11% of them nomads.

The Sudan is one of the world's twenty-five least developed countries with a per capita GNP of $320. This is reflected in a poor infrastructure; transport is slow and communications inefficient. The economy is largely based on agriculture with cotton constituting 95% of all exports.

Formal education began in the Sudan in 1956. Since then the expansion has been rapid with an aim of reaching universal and compulsory education by the year 2000. This is however unlikely for as will be shown later only 39.7% of those aged 7-18 in 1986/87 went to school.

2.1 THE GENERAL STRUCTURE OF THE SUDANESE EDUCATION SYSTEM

2.1.1 Levels of education and types of schools.

General education in the Sudan is divided into 3 levels:

-Primary education (6 years) from ages 7-13.

-Intermediate education (3 years) from age 13-16.

-Secondary education (3 years) from age 16-19.

Primary education may be preceded non-formally by pre-schools (present in cities and towns) or religious schools (khalwas) in the rural areas. Secondary education is divided into academic secondary of 3 years duration and secondary technical education of 4 years duration. At the secondary level there is also the option of Teacher Training Institutes of 4 years duration.

Student movement within each level is determined by passing examinations set by the school authorities. If a student fails the examination he is allowed to repeat. Movements from one level to the next, however, are not automatic. As a minimum requirement, students must pass standard national examinations set by the Ministry of Education. If a student passes this examination he obtains a school certificate depending on the level he has reached. For example a student who has completed primary school and passed the primary national examination obtains a Primary School Certificate. As places are limited from one level to the next students are expected to achieve more than the minimum requirements and the examinations are highly competitive. If a student fails the national examination he is not normally allowed to repeat but may transfer to another school.

There are broadly 4 types of schools:

(i) GOVERNMENT SCHOOLS: These comprise the majority of schools and the bulk of all students (about 94% of all students and 92% of all schools). They are financed by the central government as well as regional and local councils. The administration is in the hands of the respective regional governments who follow the educational policies set by the central Ministry of Education. The curricula at all levels is standard and is set by the central Ministry of Education which also publishes all school books.

(ii) AIDED SCHOOLS: These schools are mainly found in rural areas and are partially financed by the government. About 6% of all schools in the Sudan are aided and they are attended by 4% of the total number of students,

(iii) PRIVATE EDUCATION: This only consists of about 2% of all students and schools.

(iv) CATHOLIC AND EGYPTIAN MISSION SCHOOLS: These schools are financed by privation donations and the Egyptian government respectively with a nominal contribution from the Sudanese Ministry of Education. They cover all levels of education including a few secondary technical schools yet they are less than 1% of the total number of schools and students.

2.1.2 <u>Student Enrollment by Level and Sex.</u>

Table 2.1.1 below shows the enrollment ratios of students by level and sex for the year 1986/87 which is the latest available published statistics. Higher education is also included. Although the number of males is always grater than the number of females at all levels, the female enrollment ratio is high compared to other African countries. The fact that the enrollment ratio remains nearly constant between primary, intermediate and secondary education might indicate that the transition rates between these stages are equal for both males and females. With regard to higher education ,however, the ratio drops indicating a difference between the sexes in the transition rate from secondary school to higher education.

The table also shows the distribution of students among the various levels. The number of students drops the higher they climb the educational ladder. Although this is a reflection of the population structure of which more will be shown later, it is also related to the shortages of places and the bottlenecks between the levels.

Table 2.1.1 Student Enrollment by Level and Sex

LEVEL	MALES	FEMALES	TOTAL=100%
PRIMARY	1081295 59.07%	749282 40.93%	1830577
INTERMED.	211638 56.05%	165960 43.95%	377598
SECONDARY	94224 56.82%	71602 43.18%	165826
HIGHER	25114 63.59%	14380 36.41%	39494
TOTAL	1412271 58.52%	1001224 41.48%	2413495

Source: Official educational statistics 1986/87 published by the Ministry of Education.

2.1.3 Enrollment to Population Ratios

Table 2.1.2 compares between student population and school age population from the years 1961/62 and 1986/87. The aim is to show the percentage of the population that goes to school as well as the expansion that has taken place in the 25 years.

Table 2.1.2 **Comparison of enrollment to population ratios between 1961/62 and 1984/85.**

	1961/62	1986/87
PRIMARY ED.(7-13)		
No. of students	335089	1830577
Population 7-13	1373000	3559083
%Rate of Enroll	24.4%	51.4%
INTERMEDIATE(13-16)		
No. of students	56714	377598
Population 13-16	1201000	1325297
%Rate of Enrol.	4.8%	28.5%
SECONDARY (16-19)		
No. of students	18063	195708
Population 16-19	1050000	1172647
%Rate of Enrol.	1.7%	16.7%
TOTAL (7-19)		
No. of students	409866	2403883
Population 7-19	3624000	6057027
%Rate of Enrol.	11.3%	39.69%

Source: Statistics 1986/87 published by the Ministry of Education.

As shown by the table, the percentage rate of enrollment is higher at the primary level and then drops at the intermediate and secondary levels. This characteristic has not changed in the 25 year period although considerable expansion has taken place. The total rate of enrollment has more than trebled in the period, however still less than half of those aged 7-19 receive some kind of education. The rate is small and shows that it will be a long time before universal education can be reached. One of the factors involved is the rise in population. For the total age range considered (7-19), the population has increased by 67.1% and so although the number of students has increased by 486.5% the effect in terms of rate of enrollment is not as dramatic.

Table 2.1.3 below shows the population and the ratios of enrollment to population within the various provinces of the country. The figures are given separately for each level and for each sex.

Table 2.1.3 <u>Ratios of enrollment to population by province, sex and level of education.</u>

PROVINCE	PRIMARY		INTERMEDIATE		SECONDARY	
	BOYS	GIRLS	BOYS	GIRLS	BOYS	GIRLS
Northern	94.5	88.9	67.8	69.1	42.3	37.7
Khartoum	90.4	81.4	65.3	64.0	42.9	44.6
Central	77.1	64.6	42.3	35.0	24.5	16.4
Eastern	52.3	36.9	26.1	20.9	15.2	9.9
Kordofan	58.0	34.9	26.6	20.2	17.1	9.7
Darfur	53.7	28.1	18.4	9.6	10.8	5.0
Upper Nile	34.1	10.5	9.3	3.4	5.6	1.2
B.-elGhazal	13.4	5.5	5.9	2.1	4.9	1.5
Equatoria	50.4	29.6	20.1	8.4	14.0	4.1
All Sudan	59.4	43.1	31.3	25.5	19.4	13.9

Source: Official statistics 1986/87 published by Ministry of Education.

The table shows that education is not distributed equally among the difference geographical areas of the country. The differences are great between the areas; although 88.9% of girls attend primary school in the Northern province, only 5.5% attend in the Bahr-el-Ghazal province. The Northern province and the capital Khartoum have the highest enrollment to population ratios at all educational levels and for both sexes. The most heavily populated (for the relevant age group) Central province also has a high enrollment ratio. This reflects a high concentration of schools in the capital Khartoum and its neighbouring provinces (the Northern and central provinces). The differences among the levels reflects what was shown in table 1.2 that the enrollment ratios drop further along the educational ladder. This characteristic applies to both girls and boys although the enrollment ratios for girls is as expected from table 1.1 always smaller.

2.2 STATISTICS FOR SECONDARY AND HIGHER EDUCATION 1979/80-86/87

2.2.1 Secondary Education (government schools)

The following discussion is related to government secondary schools. As was shown in section 2.1.1, government schools make up the overwhelming bulk of all students and schools and are therefor the most important. Table 2.2.1 below gives the total number of students in each of the three grades that make up secondary education. The number of candidates who at the end of grade 3 have sat for the Secondary School Examination is also given.

Table 2.2.1 <u>Secondary education by grade for 1979/80-1986/87 and the number of candidates for the Secondary School Examination.</u>

YEAR	GRADES 1	2	3	TOTAL	CANDIDATES
79/80	22003	22607	21692	66302	22981
80/81	27484	26479	27598	81561	25839
81/82	32056	28530	28207	88790	25149
82/83	32060	32558	27120	91730	28363
83/84	33318	30817	33316	97451	33052
84/85	37240	31636	31658	100534	30978
85/86	37314	36924	32296	106534	32176
86/87	40841	38105	38056	117002	38605

Source: Official educational statistics 1979/80-1986/87 published by the Ministry of Education.

It can be seen that the total size of the system is expanding steadily and consequently the number of candidates is also increasing. The distribution among the grades seems to point at high repetition rates and low wastage rates. This can be seen by studying a cohort of students as they move up the educational ladder. For example the cohort who were in grade 1 in 79/80 numbered 22003 and in 80/81 when they were in grade 2 their number rose to 26479. Apart from errors, such an increase can only be attributed to high repetition, low wastage or a combination of the two.

The table also shows discrepancies between the numbers enroled in grade 3 and those who sat for the examination at the end of the year, If the latter number is smaller the difference can be attributed to dropouts who although have been registered in grade 3 did not sit for the examination. It is difficult, however, to account for the opposite as in 79/80, 82/83 and 86/87. Most likely the discrepancy is due to data collection errors; with the number of candidates being more reliable.

2.2.2 Examination Results

Table 2.2.2 below gives the number of candidates from Secondary schools who sat for the Secondary School Examinations and the percentage of those who passed by type of school. Pass rates are printed below the number of candidates. External candidates are those who are not registered in a school and are mainly students re-taking the examination. The classification Union schools includes from 1982/83 private as well as Union schools.

A pass in the Secondary School examination is a minimum requirement for proceeding into higher education. Once a student has achieved this minimum requirement he is not allowed to re-enrol in a government school and must register as an external student or in union or aided schools in order to re-take the examination again.

The expansion in secondary education is reflected in the rising number of candidates sitting for the examinations every year. The numbers have risen by a total of 77.4% in the period 1980-1987. This had a considerable effect on increasing the demand for higher education.

Table 2.2.2 CANDIDATES WHO SAT FOR THE SECONDARY SCHOOL EXAMINATION AND PASS RATES BY TYPE OF SCHOOL.

	TYPE OF SCHOOL					
Year	Gov.	Aided	Union	External	Private	Total
/80	22981	8120	18208	7278	261	56848
	71%	34.8%	40.6%	65.7%	40.6%	55.3%
/81	25839	6292	21963	7748	162	62004
	71.1%	35.8%	43.6%	70.7%	58.6%	58.0%
/82	25149	7020	24484	9175	165	65993
	76.4%	38.1%	42.4%	68.7%	53.9%	58.6%
/83	28363	4945	24731	10943		68982
	79.3%	40.6%	52.1%	75.0%		66.1%
/84	33052	5038n	21991	13992		74073
	77.1%	37.9%	58.1%	76.7%		68.9%
/85	30978	4863	27150	15409		78377
	77.6%	34.0%	58.4%	76.8%		68.1%
/86	32176	5404	28776	15498		81854
	79.3%	35.5%	55.0%	77.3%		67.5%
/87	38605	6677	36630	18920		100832
	70.6%	30.3%	48.5%	70.1%		59.8%

Source: Official educational statistics 79/80-86/87 published by the Ministry of Education.

It can be seen that the pass rate has been rising steadily from 1980 to 1986 and no doubt contributing to the increased demand for higher education. However 1987 shows a sudden drop in the pass rate in all types of school. It would be necessary to have the figures for 1988 and onwards to determine if such a drop is significant and perhaps a deliberate step by the authorities to curb the rising demand.

The number of external students has more than doubled in the 8 year period (a rise of 169%). This has been the largest rate of increase in the period; government schools have increased their candidates by 68%, Union schools by 101% and the number of 'Aided' school candidates has fallen. External students' relative proportion in terms of the total number of candidates has also risen from 12.8% in 1979/80 to 18.7% in 1986/87. It is also notable that their pass rates like those of the government school candidates are higher.

As external students are made up mostly of students re-taking the examination in the hope of obtaining a place in higher education, their growth is a reflection of the great demand for higher education and the shortages of places available. They are of central importance in any bottleneck analysis as they represent the 'overspill' of frustrated applicants to higher education.

2.2.3 Higher Education

Table 2.2.3 gives the total number of students in each university for the period 1972/73 to 1984/85. The percentage of females in higher education is only available at some years and is given below the total figure. Gezira University and Juba University were only established in 1977 with a total size corresponding to their intake. It can be seen that higher education has expanded rapidly over the period with the total size more than doubling. This has been partly as a result of the new universities (Gezira and Juba) but mainly due to the large expansion of some of the already established institutions. Most notable is Cairo University Khartoum branch (with about 50% of the total student body), which has increased in size by 161% in the period 1972-1985. The Islamic university (established in the late 60's) has also increased by 333% while other higher institutions have increased their numbers by 86.5%.

Table 2.2.3 Total Number of students in higher education by type of institution 1972/73-1984/85

INSTITUTIONS

Year	Khartoum	Gezira	Juba	Cairo	Islamic	Others	Total
/73	5811			7708	504	2185	16328
/74	6359			11656	609	2313	18762
/75	6942			13012	754	1702	22069
/76	7235			10200	1016	2134	21324
/77	7276			10288	1162	2228	10887
/78	7912	10	128	12314	1262	2537	24117
/79	8020	215	325	13591	1506	2656	25883
/80	7920	366	409	13808	1585	2928	7016
/81	8111	592	577	14810	1661	2922	28673
/82	8424	797	650	18271	1765	3402	33309
/83	8059 26.6%	899 22.4%	674 17.8%	20385 39.4%	1855 28.8%	3724 27.2%	35596 33.9%
/84		–	–	–	–	–	––
/85	8313 30.7%	965 28.3%	1216 14.2%	20096 42.8%	2184 29.9%	4077 30.3%	36851 36.6%

Source: Official statistics published by the Ministry of Education.

In sharp contrast to these overall high rates of expansion, University of Khartoum has increased by only 43% with very little expansion taking place in the 80's. It therefore appears that the university is not attempting to adjust in order to meet the increasing demand of school leavers for a higher education.

2.3 Evidence of the Presence of a Bottleneck between Secondary and Higher Education.

Table 2.3.1 gives the intake of students into higher education in the period 1977/78-1982/83. These figures show the flow of students between secondary school and higher education. Separate figures for males and females are not available nor are figures after 83/84.

Table 2.3.1 Numbers admitted into higher education by type of institution for the period 1977/78-1982/83

				INSTITUTIONS			
Year	Khartoum	Gezira	Juba	Cairo	Islamic	Others	Total
/78	1988	10	128	4051	359	917	7453
/79	1748	209	169	3412	335	1004	6877
/80	1695	178	81	2995	359	888	6196
/81	1806	205	127	2942	463	1313	6856
/82	1827	227	132	3050	411	1149	6796
/83	1739	202	171	5008	343	1633	9096

Source: Official statistics published by the Ministry of Education.

Unlike the clear picture of expansion presented by the figures on the total size of higher education, the intake of students is shown to fluctuate from year to year. The number of students admitted into University of Khartoum has fallen from 1988 to 1695 in 1978-1980 and the rise in intake in 80/81 and 81/82 did not continue into 82/83. Cairo University shows similar fluctuation with a falling intake until 81 and a sharp rise taking place in '83. For the relatively new universities Gezira, Juba and the Islamic University, the numbers are in general rising. The same is true for the other institutions made up mostly of Technical Institutes. Thus the picture of expansion conveyed by Table 2.2.3 is misleading in the sense that higher demand is not being met by an increase in intake. Such an expansion in the size of higher education institutions unmatched by an expansion in intake can only be explained by high repetition, lower wastage rates or both in higher education.

The demand for University of Khartoum is the highest mainly because it is the oldest and the fact that it is not expanding sufficiently leads to a large 'overspill' of frustrated applicants. Such students will often proceed to the other institutes of education but part of the 'overspill' will re-take the examination in the following year as an external student. Due to this, the following discussion will distinguish between the flow of students into Khartoum University and the flow into all higher education institutes.

Using the above flow figures it is possible to estimate the proportion of students who succeed in making the move from secondary school into higher education. Table 2.3.2 gives two types of proportions; the proportion of all candidates and the proportion of all those who have passed (i.e. met the minimum requirement for admission) who enter University of Khartoum or any higher institution including Khartoum University. The proportions are presented in terms of three decimal places because of their small magnitude.

Table 2.3.2 Proportions of candidates and those who passed who enter higher education and those who enter University of Khartoum.

	Proportion = flow/candidates		Proportion = flow/passes	
Year	Khartoum U.	Higher Ed.	Khartoum U	Higher Ed.
/77	0.060	0.226	0.126	0.437
/78	0.043	0.170	0.068	0.266
/79	0.036	0.131	0.063	0.229
/80	0.030	0.114	0.054	0.205
/81	0.028	0.103	0.047	0.176
/82	0.023	0.128	0.040	0.218

The fact that only a small proportion of students make the transition to higher education does not by itself indicate the presence of a bottleneck. The important factor determining the presence of a bottleneck is whether there is a significant change in the transition of students over time and if that transition is determined by the scarcity of places. From Table 2.3.2 it can be seen that the proportion of students gaining admission into University of Khartoum or in general into any institution of higher education, is falling steadily. The proportions of students who after passing their exams gain entrance are naturally higher than the proportions of candidates however their pattern is identical.

The fact that the transition proportions between secondary and higher education are not constant will result in a poor fit of the simple Markov model. This is because constant transition probabilities over time is one of the basic assumptions of the Markov model. For a model to adequately describe the Sudanese system, it must take account of the presence of a bottleneck. For this purpose, bottleneck models that could describe the Sudanese system are discussed in Chapter III.

2.4 Limitations of the Data

In the sections above, examples have been given of the type of data that is officially published each year by the Ministry of Education. The data is in the form of stock figures that are given by sex, type of school, geographical areas etc. No flow figures however are available with the exception of the figures on the intake into higher education. Even so such figures are not sufficiently detailed as they give only the numbers accepted into university without specifying the origin of the students. It is unknown what percentage of students who enter university are from government schools and whether they are first-time or second-time repeaters.

The omitted flow figures include repetition rates which must be taken into account as repetition is allowed at all school levels. Other omitted figures include wastage rates by grade, pass rates, applications to university by type of college etc. Also unknown is the destination of school leavers by sex, region, type of school etc.

There are also no available figures relating to age.. The numbers of students by age who leave the educational system is unknown. The numbers who enter higher education by age is also unknown. Future expected numbers based on population projections are also not available.

The type of data available imposes limitations on the kind of model that can be used for educational planning purposes. The unavailability of flow data makes it necessary to use methods for estimating repetition and wastage rates before proceeding with any form of flow modelling. Chapter IV is devoted to the problem of estimating transition rates from stock data as this is a major problem in the present application of modelling the education system in the Sudan.

2.5 Errors in the data

Data collection errors restrict the application of statistical techniques and cast doubts on the results obtained. This section describes the errors in the data that make it difficult to apply a model which relies on the changes in the numbers of students over time. Examples are given of the typical data collection errors in secondary schools' statistics as it is important to assess the quality of the data before proceeding with any analysis.

Table 2.5.1 gives the numbers in each grade in the Northern province by sex and the number of schools for the period /78-/86.

Table 2.5.1 Northern Province

	Boys GRADES				Girls GRADES			
	1	2	3	Sch	1	2	3	Sch
/78	1430	1143	895	7	887	516	390	5
/79	1284	1253	1218	8	974	742	415	7
/80	1315	1306	1414	9	1135	930	752	10
/81	1399	1324	1331	9	1297	1280	1149	11
/82	1401	1392	1270	9	1433	1306	1291	11
/83	1348	1333	1315	9	1404	1403	1446	11
/84	1336	1253	1290	9	1352	1346	1474	11
/85	1336	1253	1290	9	1352	1346	1474	11
/86	1334	1224	1089	9	1479	1352	1354	11

Source: Official statistics published by the Ministry of Education.

The data for the Northern province illustrates many of the errors that are found repeatedly elsewhere throughout the different provinces. The number of schools from which data is collected as can be seen, differs from year to year. If a new school has been established (and this is the case sometimes), the effect would only be felt in grade 1. For example with regard to the girls' data in /80 the number of schools for which data is collected rises from 7 to 10; the difference is only observed in grade 1 and grades 2 and 3 appear unaffected by these added schools. This confirms that these are new schools.

A problem exists however when an already established school is added or perhaps was not added in the previous year. In this case all the grades are affected. This can be seen by looking at the boys data for the years /79-/80 when the number of schools rises from 8 to 9. It can be seen that all the grades are affected particularly grade 3 which is much larger than grade 2 was in the previous year by an amount which is unlikely to be attributed to repetition. This must be because an already established school was added. For the girls data a rise in the number of schools from 10 to 11 in /80-/81 appears to affect all the grades and must be due to an already established school being added.

This type of error reduces the sequence of years that can be used for modelling purposes. It would be necessary to model for two sets of time periods, one period which had ignored this particular school and a second period which had included this school. Obviously if such an error occurs frequently, there would not be a sufficient sequence of years for estimation purposes.

Another type of error occurs when in a particular year, statistics are not collected and the previous year's figures are used to form the overall national aggregate. In the Northern province, this took place in /85 and it can be seen that the figures are identical to those of /84. Thus in modelling the system the figures /85 must be considered as missing values. This reduces again the sequence of time periods that can be used.

Table 2.5.2 below gives the number of students in each grade of secondary school for the capital province of Khartoum.

Table 2.5.2 Khartoum Province

	Boys				Girls			
	GRADES				GRADES			
	1	2	3	Sch.	1	2	3	Sch.
/78	2250	2083	2326	11	1120	965	915	5
/79	2715	2467	2247	11	1122	1151	1009	5
/80	2081	2675	2556	11	1009	1137	1159	5
/81	3076	2992	3499	17	2295	1660	1795	13
/82	3234	3292	3544	18	2948	2832	2272	17
/83	2366	3038	2452	12	2010	2289	1940	12
/84	3311	2528	3010	15	2834	2474	2488	13
/85	4004	3032	2887	19	3179	2389	2059	19
/86	4294	4281	3249	22	4558	4043	3444	26

Source: Official statistics published by the Ministry of Education.

Khartoum is the capital of Sudan and has 30% of all secondary schools in the country. In spite of the fact that it is an urban area and central, the quality of the data collected is very poor. The number of schools given varies greatly from year to year. In the case when the number of schools drops it is difficult to assess whether a drop in the data is due to natural wastage or the exclusion of a particular schools. Although the first 3 years appear to be free of error, between /80-/82 data on additional schools was being collected as can be seen from the very large discrepancy between the size of grade 2 in a particular year; a discrepancy which is too large to be explained by high repetition. Other urban areas exhibit the same errors as those shown above for the Khartoum province. The Gezira province which has 26% of all secondary schools also shows errors due to different numbers of schools being counted. It is in provinces with small numbers of students that the data seems to be relatively free of errors arising from adding or ignoring particular schools.

In very remote areas especially in the south, the data is again very poor. This is due to the difficulties in communications and the long distances between such areas. Statistics are not collected annually and for a particular year, the previous year's figures might be used to form the overall national aggregate. These provinces include Jonglei, Upper Nile, Lakes, Bahr El Ghazal and East and West Equatoria. In certain areas there are no schools for girls while in other areas statistics for girls were not collected for certain years (Bahr El Ghazal, Upper Nile). In West Equatoria, it appears that schools for both girls and boys were started only in /79 and the process can be seen from the beginning. However the data collected is very poor; for some grades statistics were not collected, at some years no figures are given etc.

From the above it can be seen that using total aggregate figures for the whole country in order to model the system would inevitably result in a poor fit due to the many errors that are present in the data. It can only be possible to model certain provinces for certain time periods and hope that this would make it possible to reach conclusions about the suitability of a particular model in describing the whole system. Therefore in Chapter IV, the attempt to obtain estimates of the transition probabilities of secondary school students is restricted to those provinces in which data collection errors are minimum.

CHAPTER III

FLOW MODELS FOR BOTTLENECK SYSTEMS

3.1 Intermediate Models in Manpower Planning

In discrete time Markov models, individuals are classified into grades according to characteristics they share in common such as age, length of service etc. Movements between the grades and from and to the outside world are assumed to be independent and time-homogenous. It is also assumed that all members within a grade share the same transition probabilities of movement. The expected stock number denoted by n in a particular grade (say grade j) is then related by the difference equation:

$n_j(T+1) = \Sigma n_i(T)p_{ij} + R(T)r_j$ (the summation is over $i=1,2,\ldots s$). (3.1)

p_{ij} is the transition probability of movement from grade i to grade j.

R(T) is the total number of new entrants to the system.

r_j is the probability that the new recruit enters grade j. The total number of grades in the system is s.

This equation is used for predicting future stock sizes and for finding the steady state structure of the system.

The other main class of transition models applied to manpower planning are models based on Renewal theory. Like the Markov models, individuals are classified into grades and transition between grades and from and to the outside world are governed by probability laws.

The main difference is that while in the Markovian models the transition probabilities are fixed, in the Renewal models the grade sizes are fixed. Hence while promotion and wastage can "push" flows in the former, promotion and recruitment in the Renewal models can only take place to fill vacancies and are as such "pull" flows.

Although the Markov chain models have demonstrated robustness in practice, at certain stages in many systems the assumptions of the renewal models are more realistic. In many organisations, the grade sizes are restricted for financial or even practical reasons and even if they were allowed to vary that would entail a considerable time lag and the sizes would be known in advance. The fact that in practice many systems behaved in an intermediate way between Markov models and renewal models, led some writers to develop generalizations of the Markov model which could be termed <u>Intermediate Models.</u> These models in different ways introduced the concept of flow constraints.

Young and Vassiliou(1974) developed a model in which the number promoted does not only depend on the numbers available for promotion as in the Markov model but on the stock of the destination grade. Armitage, Smith and Alper(1969) introduced bottleneck systems for educational planning as an enlargement of simpler models because "movements cannot be at all times and all places entirely free and without restrictions". They distinguished between desired transition proportions and actual transition proportions determined by the provision of places and the selection procedures adopted by admission authorities. Building on their ideas Moya-Angeler(1976) considered a special case of a bottleneck system with the following assumptions:

1. From a certain time onwards the promotions to some grades are determined only by the number of vacancies because there are capacity constraints on the grade sizes.

2. Once the limits of the grade sizes are reached, an overspill occurs which has to be allocated to other grades. The overspill is the excess of the demand over the supply.

3. When a grade has a capacity limit, this value must be reached and from then on maintained.

Thus the system behaves as an ordinary Markov model until the capacity limit is reached in some grade. From then on the size of the grade would be fixed at its capacity limit and the number of promotions and will be determined by the vacancies arising in such a grade. From the total number of people able to be promoted there will be a number that cannot be promoted. Of this number of frustrated promotions there will be a proportion that decide to stay in the system and remain in the grade in which they were. The rest will leave the system.

The two models proposed in this chapter are extensions on Moya-Angeller's bottleneck model described above. In the first model the proportion of frustrated promotions that stay in the system is no longer fixed but is a random variable. The second model allows for expansion and contraction of the capacity constraints.

3.2 <u>Sudanese education as an example of a bottleneck system.</u>

Educational systems are typical of systems where shortages exist usually in the movement of students from secondary school to university. The number of places in universities can be and often is restricted due to a number of reasons and the demand for places very often exceeds the supply. As seen in Chapter II, the Sudanese educational system presents such an example. Although still far from achieving universal education, the system has witnessed massive expansion in the past decades. This was reflected in the increased numbers applying to universities and other institutes of higher education. The expansion of higher education has not matched the number of potential entrants with results that admission qualifications have been increasingly more stringent. As repetition of the Secondary School Certificate (the equivalent of O levels but with which a student may enter university directly) is allowed, the number of repeaters has also increased with the hope of obtaining the necessary qualifications.

Developing extensions to Moya-Angeler's model can be approached through a study of the Sudanese system. In both of the models presented in the following sections, secondary education (which in the Sudan covers 3 years) is treated as the first grades in a hierarchical system which would include university education. Thus in talking about entrants to university we would not be talking about recruitment but about "promotion" from the last year of secondary school. Although such a classification is not intuitive it enables the discussion of the problem in terms of a promotion bottleneck. Such a classification, however, carries the inherent assumption that all students who fulfil the entrance qualifications want to enter university.

In Model One a capacity constraint exists in the first grade of the country's main university, the University of Khartoum. Demand for this university is high and it can be assumed to be a first choice for all students. Because of the capacity constraint not all qualified students are accepted and they must therefore seek alternatives. The proportion of these frustrated promotions who decide to remain in the system is no longer fixed as in the previous models but is a function of the number of alternatives available to them. It is assumed that the more alternatives that are available the less likely students will be to remain at the secondary school level by retaking the examination. Of central interest would be the possibility of the alternatives increasing sufficiently in time to remove the bottleneck. Model Two is a much more realistic and flexible version of Moya-Angeler's model and considers the presence of a bottleneck in systems which are expanding, contracting or remaining constant. In this model a capacity constraint exists on the total size of the higher educational system rather than a particular university.

3.3 Bottleneck MODEL ONE

3.3.1 Description of the system

The system is composed of 7 grades; the three bottom grades representing Secondary Education while the 4 upper grades represent one particular University (call it university K) for which demand is high. "Recruitment" occurs only at the bottom grade and "promotion" takes place only to the next higher grade. In general movement between grades is automatic (students pass exams and move on to the following grade) and hence follows a Markov Chain. However, because of the limited availability of places in University K, movements to it from secondary school i.e. between grade 3 and 4 is determined by the availability of places.

3.3.2. Assumptions of the Model

1. Because promotion within the University system is a direct result of passing exams, "promotion control" to achieve a desired structure, cannot be exercised by the authorities. It is only through determining the number of entrants that the authorities can have control over the size of the university system. It is assumed that the authorities have fixed the size of grade 4 (first year of university) and thus the number of entrants is determined by the vacancies arising in grade 4. The remaining grades including those of secondary school are allowed to vary.

2. The number of students qualified to enter University K and wish to do so is always greater than the available vacancies.

3. Grade 3 is assumed to include those registered in schools as well as external students who are repeating the examination. In other words, grade 3 also includes 'overspill' students.

4. Following the previous assumption , the size of grade 3 is affected by the overspill of students who decide to sit for the examinations again. The proportion of students who decide to remain

in the system is assumed to vary inversely with the number of alternatives available for them in other institutes of higher education. This assumption implies that all students unable to obtain admission to their first choice and given an alternative will proceed to that alternative. Furthermore it is assumed that students would prefer such an alternative over moving to the outside world.

3.3.3 Notation

$\bar{n}_i(T)$ denotes the expected number of elements in Grade i at time T.

p_{ij} denotes the probability that a member of grade i moves to grade j assuming there is no capacity limit at grade j.

$R_i(T)$ denotes the number of new recruits to grade i at time T.

$\beta_3(T)$ denotes the proportion of frustrated promotions (qualified students) that decide to remain in the system at time T.

$\delta_4(T)$ denotes the proportion of students qualified to enter University K but cannot be allocated.

3.3.4 Basic Relations

The expected structure of the system at time T can be obtained from the following set of equations:-

$$\bar{n}_1(T) = p_{11}\bar{n}_1(T-1) + R_1(T) \qquad (3.4.1)$$

$$\bar{n}_2(T) = p_{22}\bar{n}_2(T-1) + p_{12}\bar{n}_1(T-1) \qquad (3.4.2)$$

$$\bar{n}_3(T) = (p_{33} + \beta_3(T)\delta_4(T))\bar{n}_3(T-1) + p_{23}\bar{n}_2(T-1) \qquad (3.4.3)$$

where $\delta_4(T)$= No. of students who are qualified but cannot be allocated/initial size of grade 3.

$$\delta_4(T) = S(T)/\bar{n}_3(T-1) \qquad (3.4.4)$$

$$S(T) = p_{34}\bar{n}_3(T-1)-E_4 \qquad (3.4.5)$$

where E_4 is the expected number of losses in grade 4 due to wastage and to promotion into the next higher grade i.e.

$$E_4=n_4(1-p_{44}) \qquad (3.4.6)$$

where n_4 is the fixed size of grade 4.

$$\beta_3=(S(T)-D(T))/S(T) \qquad (3.4.7)$$

where D(T) is the number of available places at other institutions of higher education and is thus an exogenous variable.

$$\bar{n}_4(T) = p_{44}n_4 + E_4 = n_4 \qquad (3.4.8)$$

$$\bar{n}_5(T) = p_{55}n_5(T-1) + p_{45}n_4 \qquad (3.4.9)$$

$$\bar{n}_6(T) = p_{66}\bar{n}_6(T-1) + p_{56}\bar{n}_5(T-1) \qquad (3.4.10)$$

$$\bar{n}_7(T) = p_{77}\bar{n}_7(T-1) + p_{67}\bar{n}_6(T-1) \qquad (3.4.11)$$

EXAMPLE 3.4

For illustration, a number of simple examples are given. A program has been written that computes using the above equations, the successive structures and the steady state of any system. The program has been applied to an imaginary educational system made up of secondary education(3 grades) and a higher educational institute(University K. made up of 4 grades). Values for the initial structure of the system and the matrix of transition probabilities are chosen so as to be typical of an educational system. Assume the former to be (3000, 2000, 1000, 200, 160, 150, 140). Recruitment takes place only in the bottom grade and is constant at 2500. The number of alternative places at other institutes of higher education is fixed at D(T) = 200. Movement takes place in the system according to the following matrix:

$$\begin{matrix} .10 & .60 & & & & & \\ & .10 & .65 & & & & \\ & & .20 & .60 & & & \\ & & & .15 & .61 & & \\ & & & & .11 & .71 & \\ & & & & & .10 & .70 \\ & & & & & & .05 \end{matrix} = P$$

The results for successive values of T are given below.

GRADES	T = 0	T = 1	T = 2	T = 3	T=9
1	3000	2800	2780	2778	2778
2	2000	2000	1880	1856	1852
3	1000	1730	2314	2703	3785
4	200	200	200	200	200
5	160	140	137	137	137
6	150	129	112	109	108
7	140	112	96	83	80
$\delta_4(T)$		0.43	0.50	0.53	0.55
$\beta_3(T)$		0.54	0.77	0.84	0.90

It can be seen that as the overspill from grade 4 rises, the value of $\beta_3(T)$ rises. This causes an expansion in the size of grade 3 which again gives rise to a large overspill. The situation continues with $\beta_3(T)$ approaching but never reaching its maximum possible value of 1. As grade 4 is fixed at its capacity constraint grades 4,5,6,7 can be regarded as a separate system in which recruitment is fixed.

3.3.5 Steady-State

Because Grades 1 & 2 are not affected by the bottleneck their limit satisfies:

$$n^* = n^*P + Rr \qquad (3.5.1)$$

This can be solved to give:

$$n^*_1 = R/(1-p_{11})$$

$$n^*_2 = p_{12}n_1/(1-p_{22})$$

As shown in the previous section, the expected size of grade 3 is given by:

$$\bar{n}_3(T) = (p_{33} + \beta_3(T)\delta_4(T))\bar{n}_3(T-1) + p_{23}\bar{n}_2(T-1)$$

This has a steady state of:

$$n^*_3 = (p_{33} + \beta_3\delta_4)n^*_3 + p_{23}n^*_2 \qquad (3.5.2)$$

which can be solved to give:

$$n^*_3 = \frac{\beta_3 n^*_4(1-p_{44}) + p_{23}n^*_2}{1-\beta_3 p_{34} - p_{33}} \qquad (3.5.3)$$

When $\beta_3(T) = (S(T)-D(T))/S(T)$

$$n^*_3 = [p_{23} - D - n^*_4(1-p_{44})]/w_3 \qquad (3.5.4)$$ where $\omega_3 = 1 - p_{33} - p_{34}$

As $\beta_3(T)$ is the proportion of frustrated promotions that decide to remain in the system and take the examination again, it cannot take negative values.

Hence, if $D(T) > S(T)$ $\beta_3(T) = 0$. (3.5.5)

As can be seen a steady-state size for grade 3 exists only when D(T) the number of alternative places available at other institutions of higher education settles in the long run to being constant at D. With D(T) settling at D, $\beta_3(T)$ will have a steady-state value at β_3. The equilibrium value of grade 4 is its fixed size n_4. Hence as mentioned before, the grades 4,5,6, and 7 can be regarded as a separate system to which recruitment is fixed and they are not affected by the bottleneck. Their steady state values are then simply:

$$n^*_i = (p_{i-1,i}\, n^*_{i-1})/(1-p_{ii}) \text{ for } i=5,6,7 \qquad (3.5.6)$$

EXAMPLE 3.5

Using the same data as in example 1.1, the steady-state of the system is reached at T=5 and T=50. It is printed below along with the theoretical steady-state.

GRADE	T = 0	T = 5	T = 50	THEORETICAl STEADY-STATE
1	3000	2777.8	2777.7	2777.7
2	2000	1851.9	1851.9	1851.8
3	1000	3233.2	4168.4	4168.4
4	200	200.0	200.0	200.0
5	160	137.0	137.0	137.1
6	150	108.2	108.0	108.1
7	140	79.8	79.7	79.7
$\beta_3(T)$		0.88	0.91	0.91

As expected all the grades with the exception of grade 3 reach their steady-state values fairly rapidly due to the small diagonal elements in P. However, due to the overspill caused by the bottleneck, grade 3 continues to expand rapidly at first and then more slowly until it reaches its steady state at T=50.

3.3.6 Conditions for the bottleneck to exist

$\beta_3(T)$ is the proportion of frustrated promotions (qualified students) that decide to remain in the system and take the examinations again. Thus $\beta_3(T)$ can assume values between 0 and 1 but would not practically reach 1 as there will always be a number of frustrated promotions that decide to leave the system (in this model go to other institutions).

For an overspill to take place, S(T) (the number of students who are qualified but cannot be allocated in University K) must be greater that the number of alternative places available at other institutes of higher education. In other words:-

$$p_{34}\bar{n}_3(T-1) > D(T) + n_4(1-p_{44}) \qquad (3.6.1)$$

The right hand side represents all possible vacancies in higher education. For the bottleneck to be resolved and $\beta_3(T)$ to reach 0:

$$p_{34}\bar{n}_3(T-1) \leq D(T) + n_4\ (1-p_{44}) \qquad (3.6.2)$$

The following examples consider the possibility of resolving the bottleneck with regard to different values of D (T).

EXAMPLE 3.6.1

Using the same data as in previous examples but with D(T) increased to 500, the results show that although $\beta_3(1)=0$ and there is no bottleneck $\beta_3(2)>0$ and goes on increasing with time. Only the sizes of grade 3 are printed below because the remaining grades are unaffected and remain the same as in the previous examples.

GRADE	T = 0	T = 1	T = 2	T = 5	T = 44
3	1000	1500	1830	2253.4	2668.5
$\beta_3(T)$		0	0.32	0.55	0.65

EXAMPLE 3.6.2

In this example D(T) starts at 200 but increases at a rate of 100 every year.

GRADE	T = 0	T = 1	T = 2	T = 3	T = 4	T = 5	T = 6
3	1000	1730	2214	2423	2475	2414	2265
D(T)		200	300	400	500	600	700
β_3(T)		0.54	0.65	0.66	0.61	0.54	0.45

GRADE	T = 7	T = 8	T = 9	T = 10
3	2046	1770	1558	1515
D(T)	800	900	1000	1100
β_3(T)	0.33	0.15	0	0

It is only at T = 9 when D (T) reaches 1000 that β(T) reaches 0 and the bottleneck is resolved. The examples show that once a bottleneck has developed it would take considerable expansion to resolve it and restore the system to a pre-bottleneck state.

3.4 BOTTLENECK MODEL TWO

3.4.1 Description of the System

The system is the same as that of Model One. However, the university system which includes grades 4,5,6,7 can be regarded as the whole higher education system and not one single university.

3.4.2 Assumptions of the Model

1. As before the authorities have control only in the number of entrants to higher education, i.e. those who move from Grade 3 to 4. However the bottleneck in this case occurs not because there is a capacity constraint in grade 4 but because the total size of Higher Education (consisting of grades 4,5,6,7) is fixed. Thus the individual grade sizes of Higher Education are allowed to vary within a fixed global total. The number of entrants to University is then a random variable composed of two parts; those entering to fill any new vacancies arising from growth in the system and those who replace leavers.

2. Assuming that there is no growth in University K's system, the number of those qualified to enter is always greater than those admitted. However, the possibility of the University system expanding sufficiently to remove the bottleneck can be considered by the model.

3. Assumption (4) of MODEL ONE regarding the variability of $\beta_3(T)$ (the proportion of qualified students that decide to remain in the system) is here relaxed and is assumed to be a fixed proportion. This model is more realistic and flexible than the previous one. It makes is possible to consider the behaviour of the total system under expansion or contraction of the university system.

3.4.3 Basic Relations

$$\bar{n}_1(T) = p_{11}\bar{n}_1(T-1) + R(T) \qquad (4.3.1)$$

$\bar{n}_2(T)=p_{22}\bar{n}_2(T-1) + p_{12}\bar{n}_1(T-1)$ (4.3.2)

$\bar{n}_3(T)=(p_{33}+\beta_3\delta_4(T))\bar{n}_3(T-1)$

$+ p_{23}\bar{n}_2(T-1)$ (4.3.3)

where $\delta_4(T) = S(T)/\bar{n}_3(T-1)$ (4.3.4)

where $S(T)=p_{34}\bar{n}_3(T-1) - E(T)$ (4.3.5)

where E(T) is the expected number of entrants to Higher Education and is given by:

$E(T) = M(T) + \Sigma\bar{n}i(T-1)wi$.

The summation is from i=4 to i=7. (4.3.6)

where $M(T)=N(T)-N(T-1)$. (4.3.7)

Here N(T) is the total size of the university system i.e. total sizes of grades 4,5,6, and 7 combined. If there is no growth in the system M(T)=0. If there is a contraction M(T) will be negative.

$\bar{n}_4(T) = p_{44}\bar{n}_4(T-1) + E(T)$ (4.3.8)

$\bar{n}_5(T) = p_{55}\bar{n}_5(T-1) + p_{45}\bar{n}_4(T-1)$ (4.3.9)

$\bar{n}_6(T) = p_{66}\bar{n}_6(T-1) + p_{56}\bar{n}_5(T-1)$ (4.3.10)

$\bar{n}_7(T) = p_{77}\bar{n}_7(T-1) + p_{67}\bar{n}_6(T-1)$ (4.3.11)

3.4.4 STEADY-STATE

In this model the total size of Higher Education is fixed and thus grades 4,5,6 and 7 can be seen as a separate system to which the number of entrants is a random variable. This random variable is composed of two parts; those who replace leavers and those who enter to fill new vacancies created by an expansion of University K's system. The steady-state of the whole system differs whether there is an expansion, contraction or the size of Higher Education is constant.

STEADY-STATE WHEN SIZE OF HIGHER EDUCATION IS CONSTANT

Assuming that there is no expansion or contraction in the total size of Higher Education and M(T)=0, the steady-state structure of the grades satisfies:-

$$n^*_4=p_{44}n^*_4 + \Sigma n^*_i w_i \qquad \text{for } i=4,5,6,7 \qquad (4.4.1)$$

$$n^*_5=p_{55}n^*_5 + p_{45}n^*_4 \qquad (4.4.2)$$

$$n^*_6=p_{66}n^*_6 + p_{56}n^*_5 \qquad (4.4.3)$$

$$n^*_7=p_{77}n^*_7 + p_{67}n^*_6 \qquad (4.4.4)$$

with Σn^*_i for i=4,5,6,7 constant at a given value of N.

Once the above grade sizes have reached their steady-state it would be possible to evaluate the steady-state structure of grade 3 as:

$$\bar{n}^*_3 = \frac{p_{23}n^*_2 - \beta_3 \Sigma n^*_i w_i}{1 - p_{33} - \beta_3 p_{34}} \qquad (4.4.5)$$

Where n_2 is the steady-state structure of grade 2 obtained by:

$$n^*_2 = p_{12}n_1/(1-p_{22}) \text{ and} \qquad (4.4.6)$$

$$n^*_1 = R/(1-p_{11}) \qquad (4.4.7)$$

The two bottom grades are not affected by the bottleneck.

EXAMPLE 4.4.1

The same data as in the previous examples is here used with recruitment constant at R(T)=2500, β_3 = 0.5 and M(T)=0; i.e no growth in the university system. The results for successive values of T are given below. The steady-states of the system is reached at T=5 and T=25. It is printed below along with the theoretical steady-state.

GRADES	T = 0	T = 1	T = 5	T = 25	THEORETICAL STEADY-STATE
1	3000	2800	2777.7	2777.7	2777.7
2	2000	2000	1851.9	1851.9	1851.9
3	1000	1680	2184.6	2196.9	2196.9
4	200	270	247.5	247.7	247.7
5	160	140	162.7	169.8	169.8
6	150	129	135.6	133.9	133.8
7	112	140	104.2	98.7	98.7
E(T)		240	212.1	210.5	211
N(T)	650	650	650	650	650

STEADY-STATE UNDER EXPANSION OR CONTRACTION

In the case of expansion, α can represent the rate of expansion (it will be negative if the system is contracting) and so:

$$M(T) = \alpha\, N(T-1) \qquad (4.4.8)$$

$$N(T) = (1+\alpha)N(T-1) \qquad (4.4.9)$$

Introducing the proportions $q_i(T) = \bar{n}_i(T)/N(T)$ the equations become:-

$$(1+\alpha)q_4(T) = p_{44}q_4(T-1) + \Sigma q_i\,(T-1)w_i + \alpha$$

$$\text{for } i=4,5,6,7 \qquad (4.4.10)$$

$$(1+\alpha)q_5(T) = p_{55}q_5(T-1) + p_{45}q_4(T-1) \qquad (4.4.11)$$

and so on for grades 6 and 7.

This has a stationary structure satisfying :-

$$(1+\alpha)q_4 = p_{44}q_4 + \Sigma q_i w_i + \alpha \qquad (4.4.12)$$

$$(1+\alpha)q_5 = p_{55}q_5 + p_{45}q_4 \qquad (4.4.13)$$

and so on for grades 6 and 7.

For an expanding or contracting system, the steady size of grade 3 is given by:

$$\bar{n}_3(T-1) = \frac{p_{23}\bar{n}_2(T-1) - \beta_3(M(T) + \Sigma\bar{n}_i\ (T-1)w_i)}{1 - p_{33} - \beta_3 p_{34}} \qquad (4.4.14)$$

Its limit varies whether the system is expanding or contracting as shown below.

In the Case of Expansion:

As T approaches infinity, the total size of the higher education expands steadily, the number of entrants increases and $\delta_4(T)$ approaches 0. When $\delta_4(T)=0$ the bottleneck is resolved and the structure of grade 3 approaches its steady-state which can be evaluated as:

$$n^*_3 = p_{23}n^*_2 + p_{33} = n^*_3 \qquad (4.4.15)$$

In the Case of Contraction:

As T approaches infinity, M(T) approaches 0. When M(T) =0, the structure of grade 3 approaches its steady-state which can then be evaluated as in the case of a constant size for the higher educational system as shown above.

<u>EXAMPLE 4.4.2</u>

Using the same data as in the previous example but with α=0.005, i.e. the higher educational system expanding at a rate of 0.5%, the results are given below. It can be seen that despite the small size of α , grade 3 does approach (very slowly) the steady-state given above. With larger values of α , grade 3 reaches its steady-state more quickly (for example with α=0.10 the steady-state reached at T=19). Grades 1 and 2 are not printed as they are unaffected and their values are the same as in example 4.4.1.

GRADES	T = 0	T = 1	T = 4	T = 99	T=200
3	1000	1678	2168.6	2061.74	1835.36
4	200	273	242.1	408.33	675.75
5	160	139	174.7	278.34	460.57
6	150	129	143.9	218.34	361.33
7	140	112	102.4	160.04	264.85
N(T)	650	653	663.1	1065.01	1762.49
E(T)		243	204.1	347.39	574.89
δ_4(T)		0.4	0.5	0.4	0.3

GRADES	T = 291	T = 294	THEORETICAL STEADY-STATE
3	1506.8	1504.7	1504.6
4	1063.9	1079.9	1079.2
5	725.1	736.0	735.2
6	568.9	577.5	577.4
7	417.0	423.3	422.5
N(T)	27774.9	2816.7	
E(T)	905.1	918.8	
δ_4(T)	0.0	0.0	

EXAMPLE 4.4.3

In this example the higher education system is contracting by -10% and so α= -0.10. The results are given below and show that grade 3 approaches its steady state much slowly than in the case of an expansion. It can also be seen that M(T) need not reach 0 for the theoretical steady-state of grade 3 defined above to be a good approximation. Once again the same data is used as in previous examples and so grades 1 and 2 are not printed as they are unaffected.

GRADES	T=0	T = 1	T = 10	T = 20	T=30
3	1000	1712.6	2336.59	2382.9	2398.86
4	200	204.8	75.34	26.2	9.12
5	160	139.6	57.70	20.2	7.04
6	150	128.6	51.04	17.9	6.25
7	140	112.0	42.57	14.8	5.15
N(T)	650	585.0	226.64	79.1	27.55
E(T)		174.8	62.88	21.8	7.60
δ_4(T)		0.43	0.57	0.59	0.60

GRADES	T=39	T=40	T=41	STEADY-STATE
3	2404.1	2404.4	2404.70	2404.56
4	3.5	3.1	2.86	3.29
5	2.7	2.5	2.21	2.24
6	2.4	2.2	1.96	1.82
7	1.9	1.8	1.61	1.29
N(T)	10.7	9.6	8.65	
E(T)	2.9	2.7	2.38	
δ_4(T)	0.6	0.6	0.60	

3.5 Limitations of Model One and Model Two.

Model One assumes that β_3(T) is a linear function of the number of alternatives available to students denoted by D(T). The assumption is simplistic and empirical evidence is needed in order to determine the exact relationship if any between $\beta_3(T)$ and D(T). This needs detailed statistics which, however, are unavailable.

The alternative of assuming that $\beta_3(T)$ is constant, as was done in Model Two, is unlikely to hold in the very long run. It is expected that with time, the difficulties encountered by students in overcoming the bottleneck will lower the proportions that are willing to repeat the examination again. This point is somewhat connected to an inherent limitation in all flow models which is that they do not incorporate a 'feed-back' mechanism. It is possible that in a bottleneck situation a feed-back effect of a reduction in demand might take place. However, it is necessary to have detailed data over a considerable time period for use as empirical evidence.

The models assume that first-time and higher-time repeaters have the same transition possibilities of promotion and wastage as other students. This comes about from the first-time order of the Markov process. It might be possible that such an assumption is invalid, however detailed empirical data which is unavailable in the present application is needed to support or disprove this assumption.

CHAPTER IV

ESTIMATING TRANSITION PROBABILITIES IN SECONDARY SCHOOLS

4.1 Methods of estimating transition probabilities from stock data.

As was shown in Chapter II, the data collected by the Sudanese Ministry of Education is in the form of a sequence of stock data made up of the sizes of each grade in the system over a number of time units. With the unavailability of flow data, the problem is then to estimate the transition probability of movement assuming that the stock data has been generated by an underlying Markov process.

An up-to-date review on the methods of estimating transition probabilities from stock data is given in Rosenqvist (1986), while a full account of the theory is given in Lee, Judge and Zellner (1970). The methods include regression analysis, maximum likelihood and Bayesian analysis. In this application the regression analysis technique has been used and a short description is given below:

Let $q_j(t)$ be the unconditional probability of being in state j at time t.

Then $p(x_{t-1}=i, \text{ and } x_t=j) = p(x_{t-1} = i)\, p_{ij} = q_i(t-1)p_{ij}$

where x_t is a discrete stochastic process. (4.1.1)

From the addition law of probability:

$$q_j(t) = p(x_t=j) = \Sigma p(x_{t-1}=i)p_{ij} \qquad (4.1.2)$$

$$q_j(t) = \Sigma q_i(t-1)p_{ij} \qquad (4.1.3)$$

the summation is over i and j=1,2,.....,R t = 1, 2,......,T.

If the unconditional probabilities $q_j(t)$ and $q_i(t-1)$ are replaced by the observed proportions (from the stock data) $y_j(t)$ and $y_i(t-1)$ then there will be in general no set of transition probabilities that will satisfy this relation with probability one. Thus if errors are admitted in the equation to account for the difference between the actual and estimated occurrence of y_j then the sample observations may be assumed to be generated by the following linear model:

$$y_j(t) = \Sigma y_i(t-1)p_{ij} + u_j(t) \qquad (4.1.4)$$

or in more conventional matrix notation:

$$\mathbf{y}_j = \mathbf{X}_j \mathbf{p}_j + \mathbf{u}_j \qquad (4.1.5)$$

where $\mathbf{y}_j$ is a TX1 vector of observations reflecting the proportion in state j at time t, $\mathbf{X}_j$ is a TXr matrix of the proportion in state i at time t-1, $\mathbf{p}_j$ is a rX1 vector of unknown transition parameters to be estimated an $\mathbf{u}_j$ is a vector of random disturbances. Applying OLS with **X** as the independent variable and **y** as the dependent variable violates the non-negativity conditions of the transition probabilities. Lee, Judge and Zellner(1970) suggested a quadratic programming approach to deal with this problem. The sum of squares of the function would be minimized subject to the constraints that the rows of the **P** matrix must sum to one and the individual pij's must lie in between 0 and 1. This is the procedure which is referred to as the QP technique in the following sections.

4.2 Estimating the transition probabilities of students in secondary schools

4.2.1 Assumptions and Methods.

The following is an attempt to obtain from stock data estimates of the transition probabilities which govern the movement of students within secondary schools. The aim is to determine whether realistic, acceptable estimates can be obtained from stock data assuming an underlying Markov model. This is related to the effectiveness of the estimation method used and to whether the data is generated from an underlying Markov process.

The analysis is restricted for the present purpose to Government Secondary schools which as can be seen from Chapter II comprise the bulk of all secondary students.

The following assumptions have been made:

1. It is assumed that the three grades of secondary school make up a fixed sized system which (as the data shows) is expanding over time.

2. According to official sources repetition is very low in the first grade and thus it is possible to assume that no repetition exists in grade one. Although perhaps unrealistic, this assumption greatly simplifies the estimation procedure as the size of grade one is now simply determined by the number of entrants into the system.

3.With regard to grade 3 where many frustrated applicants might want to repeat because of the oncoming bottleneck into higher education, repetition is "officially" not permitted and students are expected to transfer to other types of schools or register as external students. The stock data of grade 3 Government schools students is therefore assumed to be free of "overspill" students and to be made up of only those promoted from grade 2 and those who failed the Secondary School Examination and are repeating.

Therefore the movement of students in Government Secondary Schools is assumed to follow a first order Markov process and the sizes of the grades is given by:

$$n_1(T) = \alpha N(T-1) + \Sigma n_i\ (T-1)w_i \qquad (2.1.1)$$

$$n_2(T) = p_{12}n_1(T-1) + p_{22}n_2(T-1) \qquad (2.1.2)$$

$$n_3(T) = p_{23}n_2(T-1) + p_{33}n_3(T-1) \qquad (2.1.3)$$

where $n_i(T)$ is the size of the grade i at time T.

p_{ij} is the transition probability of movement from i to j.

N(T) is the total size of the system at time T,

$$N(T)=n_1(T) + n_2(T) + n_3(T) \qquad (2.1.4)$$

α is the rate of expansion of the system.

As can be seen the size of grade 1 is made up of those who are coming in to fill new vacancies resulting from expansion and those coming in to fill vacancies resulting from wastage.

The quadratic programming technique discussed at length in section 4.1 assumes in its use of the proportion in each grade, a system which is constant over time. When the real system is growing (as in the case of the secondary school system) the quadratic programming technique inevitably produces a transition matrix which under-estimates the grade sizes. One way of overcoming this is through the use of Unrestricted Ordinary Least Squares on the actual grade sizes rather than the proportions. (For grades 2 and 3 $n_i(T)$ would be regressed on $n_i(T-1)$ and $n_{i-1}(T-1)$ and an estimated regression equation passing through the origin would be obtained.

The draw-backs of such a procedure are obvious; there is no guarantee that the resulting estimated matrix fulfils the Markov conditions of non-negativity and less than one estimates as well as the sum of the rows adding to less than or equal to one. However, as will be shown below, the procedure produces acceptable estimates in many cases. In addition to ensuring that the transition rates are not under-estimated, unrestricted least squares avoids the problem of heteroscedasticity and comes about from the use of proportionate data. More important, it makes it possible to test by analyzing the residuals of the models whether the assumptions of a linear relationship between the variables is appropriate. This sheds light on whether the assumption that the data is generated by an underlying Markov process is valid or not. For this the Unrestricted Least Squares procedure must be regarded as a preliminary procedure before using the Adjusted Quadratic Programming technique that will be presented below.

4.2.2 Quadratic programming adjusted by an estimate of expansion.

This technique is an attempt to solve the problem of under-estimation described above. Basically the technique takes account of expansion by estimating the rate of growth of the system assuming that a true underlying growth rate exists. The original data is then rescaled before proceeding with the QP procedure.

The general form of the equations for an expanding stochastic system of given size is presented in Bartholomew and Forbes (1979) as:

$$(1+\alpha)q_{i}(T) = \alpha + \Sigma q_{i}(T-1)w_{i} \qquad (2.2.1)$$

$$(1+\alpha)q_2(T) = p_{12}q_1(T-1) + p_{22}q_2(T-1) \quad (2.2.2)$$

$$(1+\alpha)q_3(T) = p_{23}q_2(T-1) + p_{33}q_3(T-1) \quad (2.2.3)$$

The rate of expansion of the system is α and $q_i(T)$ is the proportion of students in grade i ($q_i(T) = n_i(T)/N(T)$. The expression for the total size of the system N(T) (grades 1+2+3) is:

$$N(T) = (1+\alpha)N(T-1) \quad (2.2.4)$$

Equations (2.2.1) to (2.2.3) can be obtained by dividing both sides of the equations in section 4.2.1 by N(T-1).

If N(T) and N(T-1) are replaced by the actual sizes of the system, there will not be a rate of expansion that would satisfy this relationship with probability one. Thus if errors are admitted in equation (2.2.4) to account for the difference between the actual and estimated value of N(T), the total size of the system may be generated by the following stochastic relation:

$$N(T) = (1+\alpha)N(T-1) + \mu \quad (2.2.5)$$

Rewriting equation (2.2.5) in the form of $Y=\beta X + \mu$ ordinary least squares can be used as a bases for obtaining an estimate of α.

As equation (2.2.1) is determined by the other equations, it is not necessary to estimate its coefficients directly and it can be disregarded. The estimate for α obtained by least squares can then be placed in equations (2.2.2) and (2.2.3). Estimates of the p_{ij} are then obtained by carrying out the QP technique. The original data is thus rescaled by an estimate of α.

The question arises ,however, as to whether the school system is really growing at a constant rate of α. Over a long period of years it is very unlikely that the system would expand at a constant rate. In the present application, education expansion in the Sudan has been higher in the 50's and 60's (formal education was first introduced in 1956) and has been slowing down throughout the 70's and 80's. Consequently the growth curve over a long period of years is unlikely to be linear. However, the present purpose in trying to estimate α is not to forecast the future total size of the system but rather to obtain better estimates of the p_{ij}. Furthermore because the time period used for the estimate is short (7 or 8 years) it is not unrealistic to assume that expansion in constant over this period.

The following sections are examples of the application of the above techniques on a number of secondary school data sets. Each province is considered separately and for each sex, transition rates are estimated. This is done in case differences in pattern of movement exist between the sexes or between the provinces. These particular data sets were chosen so that the number of time periods used for modelling are six or more.

4.3 Estimates for the Nile Province Girls

4.3.1 Unrestricted least squares

The two equations from which transition rates are estimated once again are:

$$n_2(T) = n_1(T-1)p_{12} + n_2(T-1)p_{22}$$

$$n_3(T) = n_2(T-1)p_{23} + n_3(T-1)p_{33}.$$

The estimates for the second equation were unacceptable and so the following analysis is restricted to the first equation.

As a first step it is useful to study the correlation between the variables in the model. The correlation between $n_1(T-1)$ and $n_2(T)$ is high (0.9884) and between $n_2(T-1)$ and $n_2(T)$ is 0.79. This shows that a strong relationship exists between the dependent and independent variables in the equations and it is encouraging to go on with the regression procedure. The effect of repetition is smaller than the effect of promotion however the correlations are not small enough to justify removing p_{22} from the equations.

In addition to checking the correlations between the dependent and independent variables it is also important to study the correlation among the independent variables. Interpretation of a multiple regression equation depends implicitly on the assumption that the explanatory variables are not strongly interrelated. When a linear relationship exists between the independent variables (multi-collinearity), the regression results are ambiguous. In the present application, it is likely that $n_{i+1}(T-1)$and $n_i(T-1)$ are related and it would not be unreasonable to suspect the presence of multi-collinearity. The correlation between $n_1(T-1)$ and $n_2(T-1)$ is high (0.7607). Before, however, the presence of multicollinearity is determined the model specifications must be satisfactory. This will be done by analyzing the residuals of the model.

The estimates for the first equation are $p_{12}=0.86$ and $p_{22}=0.15$. The standardized residual of the model is given by:

$$e_{is} = e_i / s \qquad (3.1.1)$$

where $e_i = y_i - Y_i$ and s is the standard deviation of residuals, the square root of the following:-

$$s_2 = \Sigma(y_i - b_1x_{1i} - b_2x_{2i})^2 / (n-2) \quad (3.1.2)$$

The standardized residuals which have zero mean and unit standard deviation can be plotted against a number of variables. Such graphs often expose gross model violations when they are present. In this and subsequent examples e_{is} is plotted against the fitted value , the independent variables, and the time order in which the observations occurred. In general when the model is correct, the standardized residuals tend to fall between 2 and -2 and are randomly distributed about zero. Any distinct pattern of variation is an indication that the underlying model is inadequate; there is a need for extra terms in the model, or the error variance is not constant as assumed (hetero- scedasticity) and there is a need to carry out a transformation of the y's. In addition, plotting the residuals against time exposes autocorrelation if it is present in the data.

Figure 1.1 shows the standardized residuals plotted against the predicted values whichare given in standardized form. It can be seen that there are no outliers; all the residuals lie between -2 and +2. There appears to be no distinct pattern and the residuals are randomly distributed about zero. This particular plot is a check for violations of the equality of variance assumptions. If the spread of residuals increases or decreases with values of the predicted variable then one would question the assumption of constant variance of Y for all values of X. The plots against the independent variables also do not show an distinct pattern (not shown).

Plotting the residuals against time is shown in figure 1.2. The points neither cluster together nor show signs of being negatively correlated (i.e. when a positive value tends to be followed by a negative one and vice versa). The Durbin-Watson statistic for testing serial correlation is 2.233 a figure which is close to 2 and hence is firmer evidence that there is no autocorrelation present in the errors.

After the analysis of residuals has shown the assumptions of the model to be correct, it is now possible to turn to the problem of multicollinearity. Removing $n_2(T-1)$ from the equation results in an estimate of p_{12} = 0.995. The percentage of variation explained by the model is hardly affected, it drops from R=99.95% to R=99.93% It is obvious that the omitted variable does not improve the model. However multicollinearity if present is not severe because although $n_1(T-1)$ can serve as a proxy for $n_2(T-1)$ the opposite can never take place (it is not possible to assume that no promotion exists). The above symptoms of multicollinearity may only mean that the effect of students moving from grade one to grade two is more than the effect of repetition on the size of grade two.

Figure 1.3 shows the residuals of the no repetition model plotted against time. The points are clustered closely together around origin and the Durbin-Watson statistic is lower at d=1.58. These symptoms of autocorrelation are clearly the result of the variable $n_2(T-1)$ having been omitted from the equation. Because successive values of $n_2(T-1)$ are correlated (the correlation between $n_2(T-1)$ and $n_2(T)$ is 0.07906 as mentioned before), the errors from the estimated model then appear to be correlated.

4.3.2 QP adjusted by the rate of expansion

This procedure involves 2 steps. The first step is finding an estimate for α the expansion rate of the system assuming that $N(T) = (1+\alpha)N(T-1)$. This is a simple regression problem of a straight line passing through the origin ($Y=\beta X$) with the estimated value $\beta=(1+\alpha)$.

The resulting equation $N(T) = 1.0755N(T-1)$ i.e. $\alpha = 0.0755$ is not unacceptable. Plotting the residuals against the predicted variable Y shows no distinct pattern (fig. 1.4). The points lie within the range -2,+2 and are randomly distributed in a horizontal band along zero. Plotting the residuals against the independent variable also shows no distinct pattern and there is no evidence of serial correlation in the plot of residuals against time (fig 1.5).

The above residual analysis is encouraging and gives no reason to reject the proposed model. It is thus possible to use the estimated value of α and carry out a QP procedure for estimating the transition probabilities. The estimated transition matrix is given as

$$P_{adj} = \begin{matrix} 0 & 0.81 & 0 \\ & 0.19 & 0.81 \\ & & 0.28 \end{matrix}$$

The estimates for p_{12} and p_{22} are close to those obtained through unrestricted least squares; although in both cases the latter are larger. The transition matrix obtained using QP without adjustment for expansion is:

$$P = \begin{matrix} 0 & 0.72 & 0 \\ & 0.21 & 0.79 \\ & & 0.18 \end{matrix}$$

It can be seen that adjusting the data for expansion increases the values of the estimates and consequently resolves the problem of under-estimation. This is clearly shown in the column chart of figure 1.6 which compares the expected values obtained by unadjusted QP and adjusted QP. The expected values after adjustment are closer to the true values. In addition to this visual comparison the chi-square statistic can be used as a measure of agreement between observed and expected values of the model. There is little justification for the use of the chi-square test as a test of significance and the only aim is to compare the two techniques. The chi-square statistic dropped from 448.5 to 93.43 when the data was adjusted for expansion.

The observed and expected values obtained from quadratic programming adjusted for expansion are given below. The estimates for grades 2 and 3 are obtained by using the estimated values of the p_{ij}'s in the equation $n_i(T)=n_i(T-1)p_{ii} + n_{i-1}(T-1)p_{i-1,i}$. The actual size of grade one was used in the estimation of $n_2(T)$.

TIME	GRADE 1	GRADE 2	GRADE 3
T = 0 77/78			
OBSERVED	775	513	459
EXPECTED	–	–	–
T = 1 78/79			
OBSERVED	788	758	585
EXPECTED		(727)	(521)
T = 2 79/80			
OBSERVED	729	791	811
EXPECTED		(779)	(708)
T = 3 80/81			
OBSERVED	720	808	675
EXPECTED		(742)	(794)
T = 4 81/82			
OBSERVED	800	717	934
EXPECTED		(727)	(784)
T = 5 82/83			
OBSERVED	1047	800	712
EXPECTED		(789)	(769)
T = 6 83/84			
OBSERVED	1084	1044	868
EXPECTED		(1001)	(816)
T = 7 84/85			
OBSERVED	1280	1071	1075
EXPECTED		(1072)	(998)
T = 8 85/86			
OBSERVED	1080	1230	1071
EXPECTED		(1245)	(1098)

The expected values are close to the observed values and the pattern followed by the two are similar. Sometimes, however, the observed values are substantially higher as in the case of grade 3 for T=4 81/82. The observed value of 934 is much higher than the expected value of 784 indicating a higher than estimated repetition rate in grade 3 for this particular year.

Validating the model by comparing projections for 86/87 with newly published data.

	Grade 1	Grade 2	Grade 3	Total
PROJECTED	1283	1113	1242	3638
ACTUAL	1248	996	1293	3537

The above table compares the projected grade sizes for 86/87 and the actual grade sizes for 86/87 which were published after the estimation procedure had taken place and consequently were not included in it. The estimates for the total size of the system is obtained by using the estimated value of α in equation (2.2.4). The estimates for grades 2 and 3 are obtained by using the estimated values of the p_{ij}'s. The estimate for grade 1 is then obtained by subtraction following the assumption that the system is of a given size.

The discrepancy between the figures is not great and there is no variation in the overall pattern. It is however, noticeable that the projected sizes of grades 1 and 2 as well as the total size are slightly larger than the actual sizes. This is because the expansion rate between 85/86 and 86/87 was 0.046 which is considerably smaller than our estimate of α=0.0755. However, this considerable difference in the estimated and true expansion rate is not reflected in the same magnitude with regard to the grade sizes. This is a credit to the adjusted QP technique and shows that a large variability in the expansion rate does not affect greatly the estimates. More about this is discussed in Chapter V. For the present purpose it is possible to conclude that the technique is successful in projecting the grade sizes for one time unit ahead.

4.4 <u>Estimates for the Kassala province girls.</u>

4.4.1 Unrestricted Least Squares

As for the case of the Nile province, estimates for grade 3 p_{23} and p_{33} (i.e. those produced by the second equation) were unacceptable. They violated the assumptions of non-negativity and less than one conditions. This is in spite of the fact that correlation is high between $n_2(T)$ and $n_3(T)$. Thus such unacceptable estimates do not imply that the linear relationship is inadequate but rather that perhaps repetition in grade 3 is high.

With regard to the first equation, omission of the independent variable $n_2(T-1)$ (hence exploring the possibility of no repetition), has little effect on the performance of the model. The correlation matrix gave evidence of this for while the correlation between $n_2(T)$ was $n_2(T-1)$ high at 0.81, the correlation between $n_2(T)$ and $n_2(T-1)$ was -0.28. The fitted equation was:

$$n_2(T) = 0.89n_1(T-1) + 0.077n_2(T-1)$$

For the variable $n_2(T-1)$, the t-value is low 0.539 which confirms that is can be dropped from the equation. Dropping $n_2(T-1)$ from the equation results in the estimate of p_{12} rising to 0.96. The percentage variation explained by the model drops only slightly when the variable is removed; from 99.8% to 99.8%. The pattern of residuals for the two equations is very similar, there is no evidence of hetero- scedasticity or of serial correlation among the residuals. The Durbin-Watson statistic drops slightly from 3.28 to 3.09 when $n_2(T-1)$ is omitted. For the equation with 2 independent variables see figure 2.1 for the plot of the residuals against the predicted values, and figure 2.3 for the equation with 1 independent variable. The plots against time for the two models (figures 2.2 and 2.4) respectively show the residuals increasing with time. This implies that the variance of the errors is not constant but increases with time. The results thus of the estimates are not very reliable.

4.4.2 QP adjusted by the rate of expansion

For this data set the correlation between N(T) and N(T-1) is not very high 0.65. The consequent poor fit of the model can be seen in the plot of the residuals against the fitted values Y (figure 2.5). Positive residuals correspond to low Y's and negative residuals to high y's. It appears that a term is omitted from the model. As for the present purpose we only need an estimate of α obtained from the equation from $N(T) = (1+\alpha)N(T-1)$, there is no need to continue further with the analysis.

4.5 Estimates for the North Darfur province girls.

4.5.1 Unrestricted least squares

The correlation matrix of the variables (not shown) has high correlations between the dependent and independent variables of the model. However, correlations are also slightly high between pairs of independent variables and it would be important to check for multi-collinearity by removing one variable from the equation. The estimated equations are as follows:

$$n_2(t) = 0.87\ n_1(T-1) + 0.13\ n_2(T-1)$$

$$n_3(T) = 0.87\ n_2(T-1) + 0.17\ n_1(T-1)$$

Plotting the residuals against the predicted values, the independent variables and time do not show a distinct pattern. The residuals are in general well behaved and tend to scatter about zero. (Figure 3.1, 3.2).

Removing repetition from the equation gave an unacceptable estimate for p_{23} but for p_{12} the estimated coefficient rose to 0.999. With regard to the first equation, the value for R_2 hardly changed from 99.95% to 99.92%. Studying the residuals of the model with no repetition shows, however, that omitting the variable does not have a positive effect on the model. While there were previously no outlier, one now exists with a value of -2.21. Two other residuals have negative values of -0.045 and -0.022, the remaining five are all positive. This gives the scatterplot an unusual pattern.

The plot against time is definitely inferior to that produced by the model in which repetition is included. This is reflected in the Durbin-Watson test which drops from 1.93 to 1.77. The scatterplots are shown in figures 3.3, 3.4, and 3.5.

4.5.2 QP adjusted by the rate of expansion

For this data set the correlation between N(T) and N(T-1) is high at 0.984. Consequently the fitted equation is good as can be seen in figure 3.6. The plots of residuals show no distinct pattern, there are no outliers, and no evidence of serial autocorrelation (figure not shown). The estimate for α is 0.087. Adjusting the proportions in each grade by the estimated value of α, the following P matrix is obtained using QP:

$$P_{adj} = \begin{array}{ccc} 0 & 0.989 & 0 \\ & 0.009 & 0.80 \\ & & 0.234 \end{array}$$

The estimates are much higher than those produced previously by QP without adjusting for expansion. These were:

$$P = \begin{array}{ccc} 0 & 0.91 & \\ & 0.00 & 0.74 \\ & & 0.00 \end{array}$$

The charts in figures 3.7 and 3.8 compare the values in the grades with the estimates produced by the two methods. It can be clearly seen that QP adjusted for expansion brings the estimates closer to the true values. The chi-square statistic measuring the discrepancy between observed and expected values drops from 135.8 to 11.24 for the adjusted estimates.

The observed and expected values obtained by QP adjusted for expansion are given below. The fit is particularly good especially with regard to grade 2 where many of the expected values are the same as those of the observed figures. The observed data is free of many of the errors that are common in other data sets (see chapter one) and this might account for the exceptionally good fit of the model.

TIME	GRADE 1	GRADE 2	GRADE 3
T=0 77/78			
OBSERVED	227	209	227
EXPECTED	–	–	–
T = 1 78/79			
OBSERVED	305	226	209
EXPECTED		(226)	(221)
T = 2 79/80			
OBSERVED	249	308	226
EXPECTED		(304)	(233)
T = 3 80/81			
OBSERVED	342	249	308
EXPECTED		(249)	(298)
T = 4 81/82			
OBSERVED	404	310	294
EXPECTED		(341)	(269)
T = 5 82/83			
OBSERVED	322	404	310
EXPECTED		(403)	(336)
T = 6 83/84			
OBSERVED	378	339	423
EXPECTED		(322)	(401)
T = 7 84/85			
OBSERVED	447	388	366
EXPECTED		(377)	(352)
T = 8 85/86			
OBSERVED	503	446	380
EXPECTED		(446)	(384)

Validating the model by comparing predictions for 86/87 with newly published data.

	GRADE 1	GRADE 2	GRADE 3	TOTAL
Projected	537	498	411	1446
Actual	491	466	415	1372

The projected values are close to the actual figures; the estimate for Grade 3 is particularly good. The projected values for grades 1 and 2 and the total size are slightly greater than the actual values. This is because the system expanded by $\alpha=0.032$ between 85/86 and 86/87 which is less than the estimated value for $\alpha=0.087$. The difference is large, however fortunately this is not reflected in the same magnitude with regard to the grade sizes.

4.6 Estimates for the South Darfur Province Boys.

For both equations, including a repetition term resulted in unacceptable estimates and Therefore equations with one independent variable were estimated. The unacceptable estimates could be expected from the correlation matrix which showed a negative correlation between $n_2(T)$and $n_2(T-1)$ of -0.303 and a low correlation of 0.012 between $n_3(T)$ and $n_3(T-1)$. Even correlations between $n_2(T)$ and $n_1(T-1)$ and between $n_3(T)$ and $n_2(T-1)$ were not high namely 0.75 and 0.64 respectively. The resulting estimated matrix was as follows:

$$P=\begin{matrix} 0 & 0.94 & 0 \\ & 0 & 0.92 \\ & & 0 \end{matrix}$$

Although the plots of residuals against the fitted Y and against the independent variable do not show any distinct patterns, the plots against time show the residuals dispersing with time (fig. 4.1). It appears that the variance is not constant but increases with time. A transformation is therefore necessary or an introduction of another independent variable that would take into account time.

Fitting the equation $N(T)=(1+\alpha)N(T-1)$ resulted in residuals that tended also to increase with time (see fig 4.2). The correlation between N(T) and N(T-1) was in general low 0.11. There is thus evidence that the proposed model $N(T)=(1+\alpha)N(T-1)$ is inadequate for this data set and it is not possible to proceed with the estimation technique.

4.7 Estimates for the South Darfur Province Girls

4.7.1 Unrestricted Least Squares

The correlation between $n_2(T)$ and $n_1(T-1)$ is very high 0.995 and this might explain why including $n_2(T-1)$ in the model would result in unacceptable estimates The resulting equation with repetition removed is:

$$n_2(T) = 0.98\ n_1(T-1)$$

Plotting the residuals against the fitted value and the independent variable shows no distinct pattern. There are no outliers and the residuals show no distinct pattern. However the plot of residuals against time shows the residuals with the exception of one point increasing with time (figure 5.1). This implies that a linear effect in time should have been included in the model. Alternatively, it could be due to omitting the repetition term from the model.

The results from the second equation is:

$$n_3(T) = 0.91\ n_2(T-1) + 0.14\ n_3(T-1)$$

Plotting the residuals against the fitted values, the independent variables and time shows no distinct pattern and no evidence to suggest that assumptions of the model are violated (see figure 5.2). The t-statistic for $n_3(T-1)$ is very low which indicates that removing it would not affect the model. However removing it from the equation would result in an unacceptable estimate for p_{23}.

4.7.2 QP adjusted for expansion

Fitting the line $N(T)=(1+\alpha)N(T-1)$ gave an estimate of $\alpha=0.08$ (see figure 5.3). The residuals from the fitted line were well behaved and it was possible to use 'α' in order to adjust the proportions in each grade. The resulting P matrix produced by QP was:

$$P_{adj} = \begin{array}{ccc} 0 & 0.82 & 0 \\ & 0.09 & 0.91 \\ & & 0.15 \end{array}$$

These estimates are higher than those produced by QP without adjusting for expansion which are:

$$P = \begin{array}{ccc} 0 & 0.90 & 0 \\ & 0.09 & 0.87 \\ & & 0.11 \end{array}$$

The column charts (figures 5.4 and 5.5) show the improvements in the estimates in Grade 2 and 3 respectively. The chi-square value showed a great improvement. It dropped from 42.96 to 4.81 when the data was adjusted by the estimate of expansion.

The observed and expected values obtained from quadratic programming adjusted for expansion are given below.

TIME	GRADE 1	GRADE 2	GRADE 3
T = 0 78/79			
OBSERVED	165	153	167
EXPECTED	–	–	–
T = 1 79/80			
OBSERVED	157	169	160
EXPECTED		(162)	(164)
T = 2 80/81			
OBSERVED	159	143	162
EXPECTED		(156)	(172)
T = 3 81/82			
OBSERVED	228	152	174
EXPECTED		(157)	(173)
T = 4 82/83			
OBSERVED	220	220	170
EXPECTED		(219)	(231)
T = 5 83/84			
OBSERVED	222	216	233
EXPECTED		(218)	(225)
T = 6 84/85			
OBSERVED	299	217	217
EXPECTED		(219)	(231)
T = 7 85/86			
OBSERVED	279	298	217
EXPECTED		(289)	(234)
T = 8 85/86			
OBSERVED	279	298	217
EXPECTED		(289)	(234)

The expected values are close to the observed ones with many cases having only minor differences between then. This shows that the fit of the model is good.

Validating the model by comparing predictions for 86/87 with newly published data.

	GRADE 1	GRADE 2	GRADE 3	TOTAL
PROJECTED	275	278	304	857
ACTUAL	320	292	324	936

The projected values are close to the actual new figures although they are slightly lower. The expansion between 85/86 and 86/87 is 0.18 more than double the estimated value for α=0.079 which was used in the adjusted QP procedure. This is unlike the data for Nile Province and the North Darfur province (sections 4.3 and 4.5) where the actual expansion has been lower than the estimated expansion. It is therefore, not possible to generalise that there is an overall trend for a drop in the expansion rate. However, what is common is that the estimation procedure is capable of allowing for a large variation in the expansion rate with little effect on the final estimates of **P**.

4.8 Conclusions

The analysis of residuals carried out on the data sets showed that the assumptions of the linear model were **not** grossly violated. There were no outliers as was expected and the standardized residuals tended to fall within the range -2,+2. Because the data was in the form of observations taken in successive time sequence, serial correlation among the residuals was expected. However the majority of residuals showed no signs of correlation. Although the sample size was too small to carry out tests of significance for the Durbin-Watson statistic, the values of the statistic tended to be close to 2 which implied no correlation among the residuals.

Multicollinearity was another problem affecting the analysis as the independent variables were inherently correlated due to their nature. However the correlations were never very high and although the percentage variation explained by the two models were similar, removing repetition affected the residuals of the model. In some instances autocorrelation would appear among the residuals as a result of the omitted variable. Including repetition is thus desirable and true multicollinearity where one variable could act as a substitute for the other is unlikely to exist.

Fitting the linear equation $N(T) = (1+\alpha)N(T-1)$ was successful and the residuals in the majority of the examples behaved well. There was sometimes evidence, however, of the variance of the residuals increasing with time and the analysis for such data sets was discontinued.

Definitely the adjusted QP procedure is an improvement to the QP technique and the grade sizes are no longer under-estimated as when there is no adjustment for expansion. This was shown in the column graphs and in the improvement in the chi-square value (used in this context as a measure of agreement). Validating the model by comparing projected grade sizes for 86/87 with actual newly published figures proved successful. There was only a slight variation between the projected and actual figures although estimates of the expansion rate were very different. It is thus possible to conclude that the results of the estimation procedure can be used for making reasonably accurate projections one time unit ahead.

CHAPTER V

SAMPLING DISTRIBUTIONS OF THE ESTIMATED TRANSITION PROBABILITIES

5.1 The simulation program: Descriptions and Aims.

It has been proposed that the Sudanese secondary school system can be modelled in terms of an expanding stochastic system with given size. Because of the unavailability of flow data, it was necessary to estimate the transition probabilities from stock data. The original QP technique used by Lee, Judge and Zellner (1970) based on minimizing the sum of squares between successive stock proportions resulted in estimates of transition probabilities that always under-estimated the grade sizes. The suggestion put forward in Chapter IV, was to take account of the expansion of the system by estimating α, the rate of growth assuming that a true underlying growth rate existed. The original data would then be rescaled before proceeding with the QP technique. This procedure was carried out on 5 provinces for which data collection errors were minimum and a reasonable sequence of time periods were available.

In order to assess the effectiveness of the above procedure in terms of ability to estimate the true underlying transition matrix and make predictions, a simulation program was developed. Such a simulation set generates artificial data upon which the model is fitted and predictions are made.

Program SIMUL simulates the behaviour of a 3-grade system with a total size expanding at a random rate. The program makes use of the random number generator RN55 which is a double precision function that returns a random number between 0 and 1 from the uniform distribution. For each individual known to be in grade 1 at time T-1, a random number X is generated; if $X < p_{12}$ the individual moves to grade 2 at time T, otherwise he leaves the system. For each individual in grade 2 at the time T-1, other number numbers are generated; if $X < p_{22}$ the individual remains in grade 2 at time T-1, if $p_{22} \leq X \leq p_{22} + p_{33}$ the individual moves to grade 3 at time T-1, a random number X is generated, if $X < p_{33}$ the individual remains in grade 3 at time T otherwise he leaves the system. This procedure is repeated for each individual and for every time period.

In order to obtain an expansion rate which lies between a pre-determined upper and lower bound, the random number was rescaled as follows: If H is the random number between 0 and 1, then the rescaled expansion rate A is given by:

$$A = (H*FAC1) + FAC2 \qquad (5.1)$$

where FAC1 is the difference between the upper and lower bounds of the expansion rate and FAC2 is the lower bound. The total sizes of grades 2 and 3 were then obtained by adding the number of individuals in a particular grade. As no repetition was assumed in grade 1 and as the total size of the system at time T was known following our knowledge of the expansion rate; the number of individuals in grade one were calculated as the difference between the total size and grades 2 and 3.

$$n_1(T) = (1+A)N(T-1) - n_2(T) - n_3(T) \qquad (5.2)$$

where N(T-1) is the total size of the system at time T-1 and

(1+A)N(T-1) is the expression for the total size of the system at time T i.e. N(T).

Using program SIMUL, 3 sets of 100 simulations were carried out. The results obtained are sampling distributions of the estimates under different assumptions.

<u>I.SET ONE: 100 Simulations carried out with an expansion rate varying randomly between 1% and 16% and a 23-time period used to obtain the estimates.</u>

For every simulation the transition matrix used was:

$$\mathbf{P}= \begin{matrix} 0 & 0.90 & \\ & 0.15 & 0.80 \\ & & 0.15 \end{matrix}$$

These probabilities were chosen to be as close as possible to those obtained from the actual data. Because the estimates from the actual data varied between provinces and between boys and girls, the above probability matrix only comes close to the general pattern of low wastage in the first two grades and considerable repetition in grades 2 and 3. For every simulation, the initial vector was set at **n**(0)= [100, 90,90]. Although these numbers are smaller in magnitude that those of the original data, the general pattern of a larger grade one and equal grades 2 and 3 is maintained. The expansion rate α was allowed to vary randomly between 1% and 16%. The data on Secondary Schools showed a wide variation in the expansion rate from year to year. It is typical of a system to contract at one year and expand by as much as 17% on the following year. Thus allowing α to vary randomly between the above 2 bounds would be close to the pattern of the actual data. For every simulation attempt a different seed was used to generate random numbers consequently resulting in different data sets which produced different estimates of **P** and of α.

<u>II.SET TWO: 100 Simulations carried out with fixed expansion and a 23 year period used to obtain the estimates.</u>

These simulations were carried out on the same transition matrix **P** and the same initial starting vector as the first set. However, in this set α was no longer estimated but assumed to be known and fixed at 8% (roughly the average between 1% and 16%). The aim of this was to gain a measure of the variability introduced by estimating (equation) from data that increased annually at a variable rate.

<u>III.SET THREE: 100 Simulations carried out with an expansion rate varying randomly between 1% and 16% and an 11-year period used to obtain the estimates.</u>

Due to data collection errors and the difficulty of obtaining consecutive data on the Sudanese educational system, a period of 7 or 8 years (differing between provinces) was used for estimating the transition probabilities. In this third set of simulations, estimates were obtained from only an 11 year period. The aim of this was to see the effect on the estimates of using a small time period. As in both sets I and II, the same transition matrix **P** and initial starting vector were used. The features of this 3rd simulation set is closer to those of the actual data than the two first sets.

5.2 The Sampling Distribution of the Estimated Transition Probabilities

In this section the sampling distributions of the estimated transition probabilities are presented and comparisons are made between the three sets of simulations. In the first table shown below, table 5.2.1 the measures of central location are compared with each other and with the true values.

TABLE 5.2.1 COMPARISON OF MEANS, MEDIANS AND MODES OF THE TRANSITION ESTIMATES, RESULTS FROM 100 SIMULATIONS.

	TRANSITION PROBABILITIES			
	p_{12}	p_{22}	p_{23}	p_{33}
TRUE VALUES	0.900	0.150	0.800	0.150
VARIABLE GROWTH				
MEAN	0.900	0.149	0.782	0.171
MEDIAN	0.899	0.149	0.794	0.156
MODE	1.000	0.175	0.832	0.108
FIXED GROWTH				
MEAN	0.912	0.136	0.798	0.151
MEDIAN	0.915	0.126	0.812	0.139
MODE	1.000	0.082	0.863	0.000
SMALL SAMPLE				
MEAN	0.909	0.136	0.778	0.174
MEDIAN	0.914	0.132	0.795	0.163
MODE	1.000	0.167	0.845	0.184

As shown by Table 5.2.1 the means of the estimates from the 100 simulations are very close to the true values. This is true for the three set of simulations. For p_{12} and p_{22}, the estimates obtained by simulations with a variable expansion rate are closer to the true values. For p_{23} and p_{33} the estimates obtained with fixed expansion rates are closer to the true value. The results from the small sample data are not very different than those obtained by the other procedures especially if they are rounded to two decimal places.

Thus it is possible to conclude that neither the variability of the expansion rate nor a small data set have any considerable effect on the means of the estimates.

The values of the median and the mode give an idea of the shape of the distributions although more will be known later when comparing the coefficients of skewness and kurtosis. It is interesting that for all three sets and for nearly all the variables, the values for the median are closer to the mean than those of the mode. This is related to a tendency for the values of the mode to be unrealistic in some cases; for p_{12} the mode is 1.00 in the three sets and for p_{33} in the second set the mode is 0. This might be due to the nature of the QP technique which tends at times to over-produce estimates of 1.00 and 0. Consequently such distributions are skewed, negatively skewed when the mode in 1.00 and positively skewed when the mode is 0.00. When estimating from actual data, estimates might be obtained close to the modal value and this would result in an unrealistic estimate of **P**.

TABLE 5.2.2 COMPARISON OF THE VARIANCES AND STANDARD DEVIATIONS. RESULTS FROM 100 SIMULATIONS

	TRANSITION PROBABILITIES			
	p_{12}	p_{22}	p_{23}	p_{33}
VARIABLE GROWTH				
VARIANCE	0.004	0.004	0.004	0.006
STD. DEV.	0.067	0.067	0.067	0.077
FIXED GROWTH				
VARIANCE	0.004	0.005	0.007	0.009
STD. DEV.	0.067	0.069	0.082	0.094
SMALL SAMPLE				
VARIANCE	0.006	0.006	0.008	0.011
STD. DEV.	0.077	0.079	0.089	0.103

In general variability is not high for the majority of the estimates and for the 3 sets of simulations. Fixing growth does not appear to improve the variability among the estimates i.e. figures for the standard deviation are no smaller. For the simulations carried out on a small sample, the values for the standard deviation for all the estimates are higher than those for the first two sets.

Table 5.2.3 shown below displays the correlation matrix of the variables.

TABLE 5.2.3 COMPARISONS OF THE CORRELATIONS BETWEEN THE TRANSITION ESTIMATES. RESULTS FROM 100 SIMULATIONS.

	CORRELATIONS			
	p_{12}	p_{22}	p_{23}	p_{33}
VARIABLE GROWTH				
p_{12}	1.00	-0.988	0.097	-0.079
p_{22}		1.000	-0.13	0.122
p_{23}			1.000	-0.989
p_{33}				1.000
FIXED GROWTH				
p_{12}	1.00	-0.999	0.555	-0.554
p_{22}		1.000	-0.581	0.577
p_{23}			1.000	-0.996
p_{33}				1.000
SMALL SAMPLE				
p_{12}	1.00	-0.983	0.347	-0.318
p_{22}		1.000	0.372	0.336
p_{23}			1.000	-0.986
p_{33}				1.000

As expected from the nature of the variables, p_{12} and p_{22} are highly negatively correlated; if the probability of moving from grade 1 to grade 2 is high then the probability of repeating in grade 2 is low for given size of grade 1 at T-1 and grade 2 at T-1 and at time T. Similarly, p_{23} and p_{33} are also highly negatively correlated. The fact that this is not so is encouraging in the sense that a 'bad' estimate of p_{22} will not have a large effect on the estimate of p_{23} and consequently p_{33}

TABLE 5.2.4 COMPARISON OF THE COEFFICIENTS OF SKEWNESS AND KURTOSIS. RESULTS FROM 100 SIMULATIONS.

	TRANSITION PROBABILITIES			
	p12	p22	p23	p33
VARIABLE GROWTH				
SKEWNESS	-0.412	0.372	-1.239	1.340
KURTOSIS	0.080	-0.014	2.065	2.449
FIXED GROWTH				
SKEWNESS	-0.495	0.580	-0.697	0.635
KURTOSIS	-0.557	-0.328	0.235	0.155
SMALL SAMPLE				
SKEWNESS	-0.617	0.554	-0.767	0.783
KURTOSIS	-0.157	-0.006	0.421	0.311

Table 5.2.4 compares the coefficients of skewness between the 3 sets of simulations. The closer the values for skewness are to 0, the more normal is the observed distribution. Certain features are common among the 3 sets; the distributions of p_{12} and p_{23} are always negatively skewed while those of p_{22} and p_{33} are always positively skewed. Thus many of the estimates of p_{12} and p_{23} tended to be larger than the mean and due to the large negative correlation between p_{12} and p_{22} and between p_{23} and p_{33}, many estimates for p_{22} and p_{33} tended to be smaller than the mean. Another feature common to the three sets is that the distributions of p_{23} and p_{33} are more highly skewed than for p_{12} and p_{22}. This is especially with regard to the simulations carried out under variable growth. The distributions of p_{23} and p_{33} are less skewed under fixed growth implying that the loss of growth variability improves the shape of the distribution. The same is true for simulation set 3 with coefficients of skewness for p_{23} and p_{33} smaller than for the first

set. It is likely that the small time period used has worked to lessen the effect of the large variability in growth and produced distributions closer to the normal shape.

Table 5.2.4 also compares the coefficient of Kurtosis between the 3 sets of simulations. The coefficient of kurtosis measures the extent to which observations cluster around a central point; a value of 0 indicates that the distribution is exactly normal while positive values indicate a distribution that is more peaked than normal. Once again the 3 sets of simulations share a common feature, values of kurtosis for p_{12} and p_{22} tend to be negative while values for p_{23} and p_{33} tend to be positive. Thus the distributions for p_{12} and p_{22} are platykurtic; they cluster less than in the normal distribution and the shape of the distribution is generally flatter. The distributions for p_{23} and p_{33} are more peaked, cases within the distributions cluster than those in the normal distribution and tend to have more observations straggling into the extreme tails. The distributions of p_{23} and p_{33} for the first simulation set (where growth is variable) are highly peaked in comparison to the other simulation sets. For the case when growth is fixed, the value for the kurtosis coefficient drops and is closer to zero. The same to a lesser degree is true for the third set.

It appears that a fixed growth rate improves the shape of the distributions of p_{23} and p_{33} and a small sample size also has a positive effect perhaps indirectly by reducing the effect of the variability in the growth rate. This is however only with regard to the distributions of p_{23} and p_{33}, values of kurtosis for p_{12} and p_{22}

are closer to zero in the first simulation set than in the second set where growth is fixed.

Figures V1-V4 show the histograms of the distributions of p_{12}, p_{22}, p_{33}, and p_{33} respectively for the case of variable growth. Figures F1-F4 show the histograms of the distributions of the estimates for the case of fixed growth. Figures S1-S4 show the histograms in the case where a small sample was used to obtain the estimates. A normal probability curve is superimposed on the histograms. The figures gives a visual representation of the shape of the distributions.

5.3 The Variances and Covariances of the Predicted Grade Sizes

The general formulae for evaluating the expected values, variances and covariances of an expanding system with given size is given in Bartholomew (1982). In the special case when the size of the bottom grade is solely determined by the total number of recruits into the system the formulae becomes:

$$\mu(T+1) = \mu(T)\pi' + \mu'(T+1) \qquad (5.3.1)$$

where $\mu(T)$ is a 1x12 vector with the elements:

$[E(n_1),\ldots,E(n_3),cov(n_1n_1),cov(n_1n_3),\ldots\ldots,cov(n_3n_3)]$

$$\pi' = \begin{array}{cc} \underline{Q} & \underline{X} \\ O & Y \end{array} \qquad (5.3.2)$$

where Q is a 3x3 matrix whose elements q_{ij} are the total probability of a move (of any kind) out of grade i which results in an addition to grade j. Thus:

$$q_{ij} = p_{ij} + w_i r_i \qquad (5.3.3)$$

where r_i is the probability of a recruit allocated in grade i.

For the present purpose as recruitment occurs only in the bottom grade $r_1 = 1$ and $r_2 = r_3 = 0$. Therefore:

$$Q = \begin{matrix} w_1 & p_{22} & \\ w_2 & p_{22} & p_{23} \\ w_3 & & p_{33} \end{matrix} \qquad (5.3.4)$$

O is a 9x3 zero matrix

X is a 3x9 matrix made up of the elements

$(\delta_{j1}\, q_{ij} - q_{ij}\, q_{i1})$

$\delta_{j1} = 1$ when $j = 1$, and is zero otherwise.

Y is a 9x9 matrix which is the direct matrix product of Q.

$\mu'(T+1)$ is a 1x12 vector in which the first 3 elements are the expected number of entrants to each grade at time T+1.

The remaining 9 elements are the covariances of these numbers listed in dictionary order. For the present application $\mu'(T+1)$ contains only one element, namely:

$$\mu'(T+1) = A^*N(T-1)(1,0,0,0,0,........,). \qquad (5.3.5)$$

For each of the 3 sets of simulations, 100 estimated transition matrices were produced using QP. Using the same initial grade size of n(0)=[100,90,90] expected grade sizes for 29 time periods were obtained with each of the 100 estimated transition matrices. The variances and covariances of the predicted grade sizes were also calculated for the 29 time periods using the formulae above. The procedure was carried out on the 3 sets of simulations. The tables below give the average predicted grade sizes and the average variances and covariances for selected time periods.

TABLE 5.3.1 AVERAGE PREDICTED GRADE SIZES. RESULTS FROM 100 SIMULATIONS ROUNDED TO THE NEAREST INTEGER.

	TIME				
	T=0	T=1	T=10	T=19	T=28
VARIABLE GROWTH					
GRADE 1	100	114	227	475	1007
GRADE 2	90	103	217	456	967
GRADE 3	90	86	186	390	827
TOTAL	280	303	630	1321	2801
FIXED GROWTH					
GRADE 1	100	114	217	433	864
GRADE 2	90	103	208	417	835
GRADE 3	90	85	179	358	717
TOTAL	280	302	605	1208	2416
SMALL SAMPLE					
GRADE 1	100	114	223	460	962
GRADE 2	90	103	213	441	921
GRADE 3	90	86	182	377	786
TOTAL	280	303	618	1277	2669

As shown by table 5.3.1, there is hardly any difference between the predicted sizes at T=1 or at T=10. The difference is more marked, however, as time goes by with sets 1 and 3 producing larger grade sizes than set 2. The difference however is not very large and appears to affect the 3 grades equally. Note must be taken however that for simulation set no.2 the growth rate was fixed at 0.08 which is less than the average of the range 0.01 and 0.16 used in the other sets. It is likely that if the average 0.085 rate was used there would be little difference between the three.

TABLE 5.3.2a AVERAGES OF THE VARIANCE-COVARIANCE MATRICES FOR T=1 AND T=10. RESULTS FROM 100 SIMULATIONS

	TIME					
			T=1			T=10
VARIABLE GROWTH						
GRADE 1	25.99	-9.26	-16.73	111.86	-57.43	-54.43
GRADE 2		19.69	-10.44		114.19	-56.76
GRADE 3			27.16			111.19
FIXED GROWTh						
GRADE 1	23.43	-8.29	-15.14	103.22	-53.00	-50.21
GRADE 2		17.78	-9.49		105.25	-52.24
GRADE 3			24.64			102.45
SMALL SAMPLE						
GRADE 1	25.93	-8.42	-17.52	107.57	-55.03	-52.54
GRADE 2		17.72	-9.30		108.98	-53.95
GRADE 3			26.81			106.49

TABLE 5.3.2b AVERAGES OF THE VARIANCE-COVARIANCE MATRICES FOR T=19 AND T=28. RESULTS FROM 100 SIMULATIONS

	TIME					
			T=19			T=28
GRADE 1	234.65	-120.64	-114.01	495.73	-254.57	-241.16
GRADE 2		240.30	-119.66		507.34	-252.78
GRADE 3			233.67			493.94
FIXED GROWTH						
GRADE 1	209.50	-107.98	-101.52	419.37	-216.26	-203.11
GRADE 2		105.25	-106.66		429.97	-213.72
GRADE 3			208.18			416.82
SMALL SAMPLE						
GRADE 1	223.76	-114.78	-108.99	466.80	-239.42	-227.38
GRADE 2		227.57	-112.80		474.92	-235.50
GRADE 3			221.78			462.88

The average variances and covariances of the predicted grade sizes all increase with time. This increase is larger at the beginning of the prediction periods and tends to settle down with time. While the variances and covariances triple between T=1 and T=10, they tend to double in size between T=19 and T=28. Because the variances and covariances are computed from the predicted grade sizes they will always increase even if the system reaches a steady state, however, the magnitude of the increase will diminish as the system approaches equilibrium.

Removing the variability of the growth rate does appear to reduce the prediction errors as can be seen from comparing between simulation sets one and two. It is interesting that a smaller sample size introduces less variability in the estimation process than does a large one.

5.4 Summary of the Results of the Simulations

1. The means calculated from the 100 simulation attempts were very close to the true underlying probability matrix. However the distributions of the estimates are slightly skewed; negatively skewed for estimates of p_{12} and p_{23} and positively skewed for the repetition rate. This must be taken into account when estimating from real data; it is possible to obtain for example an exaggerated high estimate for p_{12} and consequently (due to the large negative correlation) a very low estimate for p_{22}. The same is valid to an even much larger extent when estimating p_{23} and p_{33}.

2. Removing the variability of the expansion rate and reducing the time period had little effect on the sampling distribution of the estimates. The estimates of the means are not different between the

3 sets especially if they are rounded to two decimal places. An improvement however exists in the case of the skewness coefficient for the distributions of p_{23} and p_{33}. They are less skewed under fixed growth and also in the case of set 3 when a smaller period is used for the estimation.

3. The variances and standard deviations of the estimates are not very high. Fixing growth does not appear to improve the variability among the estimates and the variability among the estimates for set 3 (the small time period) is slightly higher than for the first two sets.

4. As expected the correlations between p_{12} and p_{22} and between p_{23} and p_{33} are very highly negative. It is therefore not possible to obtain a 'bad' estimate for p_{12} and a good one for p_{22}. This must be borne in mind when estimating from real data. It is encouraging, however, that the correlation between p_{22} and p_{23} is not very high. Thus a 'bad' estimate (exceptionally high or low) of p_{22} will not have a large effect on the estimate of p_{23} and consequently p_{33}.

5. In terms of predicting grade sizes, the three sets produced on average, values that were not dissimilar. There is hardly any difference between the predicted grade sizes at T=1 and T=10. The difference however is more marked as time goes by with set 1 producing slightly larger grade sizes.

6. The average prediction errors are large and increase greatly with time. There is a slight reduction in the prediction errors when

the variability in growth is removed (set 2). This is expected because a parameter (α) is no longer subject to random variation and hence one source of error has been eliminated. What is unexpected is a similar reduction when a smaller sample is used for the estimation (set 3). In order to justify this there is a need to identify the sources of variation in the model. This requires more simulations using different sample sizes, different growth variation and perhaps comparing with systems in which there is no growth. This point is taken up further in Chapter VI.

CHAPTER VI

CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

6.1 Conclusions

This research has been a first step in the stochastic modelling of the Sudanese educational system. The educational statistics of the late 70's onwards present a picture of an expanding system with a growth rate that is unable to match the growing demand for school places. This high demand is a consequence of a rise in the population and an increase in urbanization. The result is that large numbers of potential students are excluded from the educational process. In particular, statistically speaking, it also leads to bottlenecks between one level of education and the next.

Of interest in the present context is the bottleneck between secondary and higher education. The demand for higher education is greater than its capacity limits and the excess of demand over supply results in an 'overspill' of eligible students. Of this 'overspill' a proportion leave the system and a proportion decide to remain and increase the number of applicants in the following year. The result is a steady fall in the proportion of eligible students who are admitted each year into higher education.

In Chapter III two bottleneck models suitable for modelling the movement of students in secondary and higher education were developed. In both models the flow of students into higher education is not determined by the numbers who are eligible and want to proceed into higher education but by the number of vacancies. The first model considers the case when a capacity constraint exists in the first grade of a particular higher education institute. The second model assumes that it is the total size of higher education which is fixed. It succeeded in generalizing previous bottleneck models by allowing for expansion and contraction of the capacity constraints . As such it is a step further in the theoretical study of bottleneck systems. However, validating the models by data fitting was not possible because of the lack of sufficiently detailed data.

The limitations imposed by unavailable flow rates and the need for their estimation is dealt with in chapter IV. The original procedure used for estimating transition probabilities, (see Lee, Judge and Zellner (1970) and Rosenqvist(1986)) was based on systems that remained constant in size. In this application it was adjusted to allow for expanding systems as was the case in the Sudan. An estimate of the rate of expansion was obtained assuming linear growth. The data was then rescaled by the estimate before the original quadratic programming technique for estimating transition probabilities from stock data was used.

Estimates were obtained of the transition probabilities of students in secondary schools. However, because of many errors in the data, estimates could only be obtained from a limited number of data sets in particular provinces. In the majority of cases the fit of the model was good and validating the predictions of the model with newly published data is highly successful. It is thus possible to conclude that the movements of students in secondary education can be adequately described by a Markov model.

Due to the limited data available, it is not possible to make generalizations about the wastage, repetition and promotion rates of secondary students for the whole country. There is also insufficient evidence of differences in the transition rates between different provinces and/or between sexes.

In Chapter V an attempt was made through simulations to assess the effectiveness of the QP technique adjusted for expanding systems. The results were promising with estimates coming close to the true values. Prediction errors,however, were large showing that predictions over a large period are likely to be poor and should be restricted to only about 5 years ahead. There was slight evidence that the technique was effective in situations where the expansion rate varies widely from year to year. However, more work needs to be done in order to verify the source of variation in the model.

6.2 <u>Suggestions for future research</u>

Future research can take a number of directions. In the theoretical study of bottleneck systems, the models can be generalized for more wider applications in manpower systems which are not necessarily hierarchial. It would be interesting to incorporate a feed-back mechanism which models changes in the behaviour of the 'overspill' promotions as time passes. Also bottleneck systems can be modelled in continuous rather than discrete time; this would be valid in applications to manpower rather than educational systems where changes occur at fixed points in time.

The work carried out in chapters IV and V is only a first step in the estimation of transition rates from stock data of expanding systems. More work can be done in validating the procedure, identifying the sources of variation in the model and considering the effectiveness of the procedure under different assumptions.

For the development of a full model of the Sudanese educational system that can be used for planning purposes, it would be necessary to have more data and preferably flow data. Figure A on page 109 is a flow chart of the entire educational system with arrows denoting the possible movement of students. Students can also at any time move to the outside world (not shown in the figure). Bottlenecks exist at all the levels with the exception of the external students. Secondary government school students who are unable to obtain a higher education place (the 'overspill') and who wish to retake the Secondary School Certificate must move to other secondary schools or become external students.

Of interest would be the total numbers in each of the boxes. These would include repeaters and those who have been promoted to fill either new vacancies or those arising from wastage. The total number of external students would include those from secondary schools who were unable to gain entrance to higher education and who decide to re-take the examination. It will also include those who in the previous year were external students but once again did not gain entrance to higher eduction and decided to re-take the examination.

Another point of interest would be the amount of wastage from the system. Total wastage would be made up of:

-The overspill of students from various kinds of secondary education including external students who were unable to obtain a place in higher education.

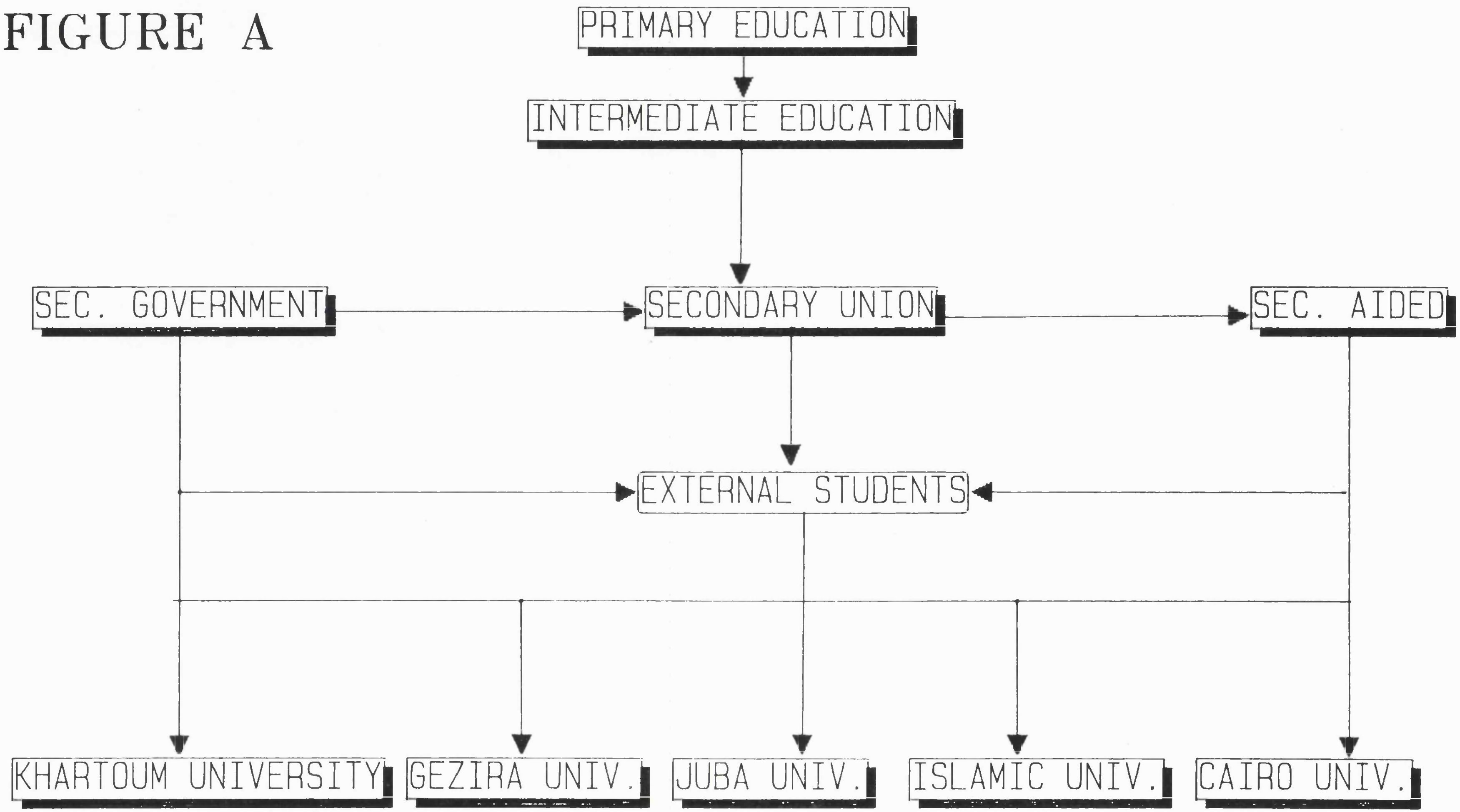
-The overspill of students from intermediate schools who were unable to obtain a place in any kind of secondary school.

-The overspill of students from primary schools who were unable to obtain a place in intermediate schools.

-Natural wastage from all of the boxes in the flow chart. This includes that wastage occurring at the last grade of each box as well as within the boxes.

In order to fit such a model more data needs to be collected on the numbers that apply to higher education, their priorities, as well as the numbers of repeaters among external students. The present research has shown that a simple Markov model can adequately describe the movement of students within the boxes. It has also shown that within the boxes reasonable estimates of transition rates can be obtained from stock data. What is lacking is more data on the flows between the boxes.

FIGURE A


PRIMARY EDUCATION
INTERMEDIATE EDUCATION
SEC. GOVERNMENT
SECONDARY UNION
SEC. AIDED
EXTERNAL STUDENTS
KHARTOUM UNIVERSITY
GEZIRA UNIV.
JUBA UNIV.
ISLAMIC UNIV.
CAIRO UNIV.

APPENDIX A

MAP OF THE SUDAN


THE RED SEA
Northern Region
Eastern Region
DARFUR REGION
Kordofan Region
CENTRAL REGION
Khartoum Province
Bahr EL Ghazal Region
UPPER NILE REGION
EQUATORIA REGION

APPENDIX B

Figures for Chapter IV.

FIGURE 1.1

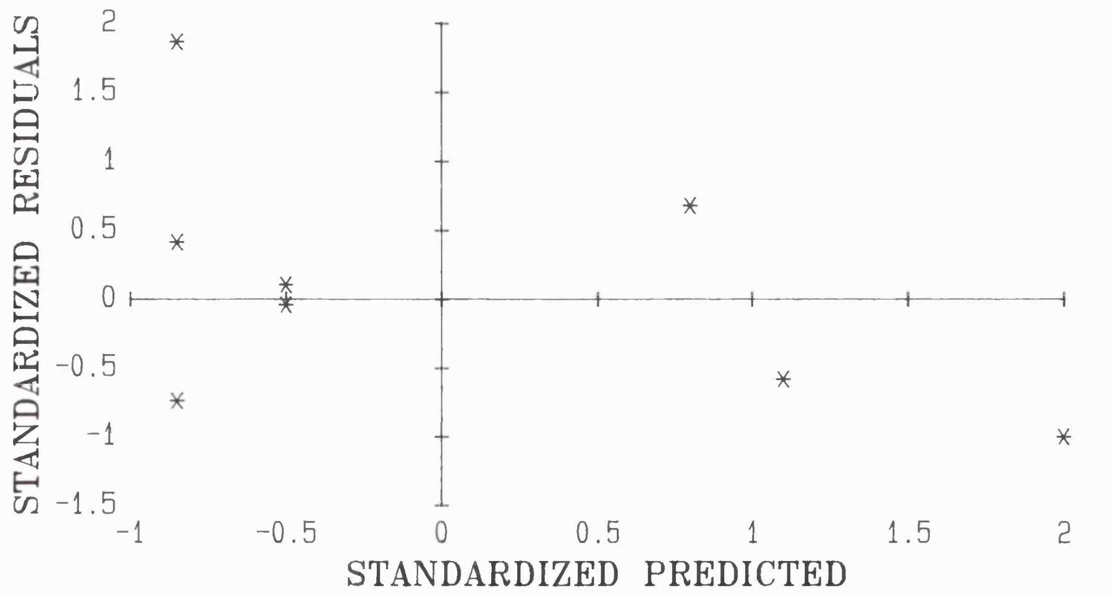
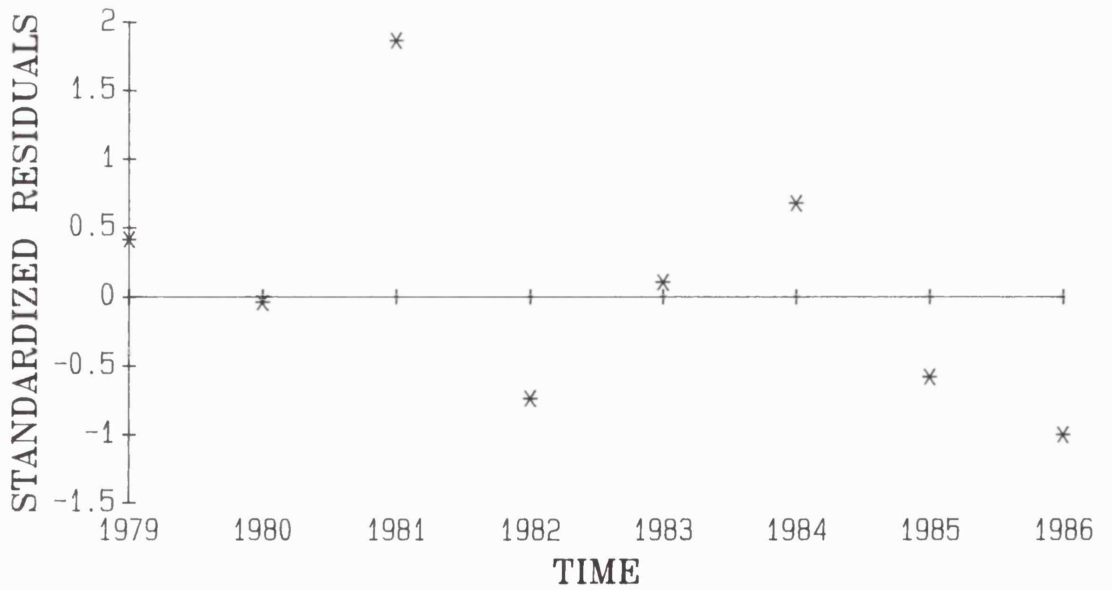


FIGURE 1.2



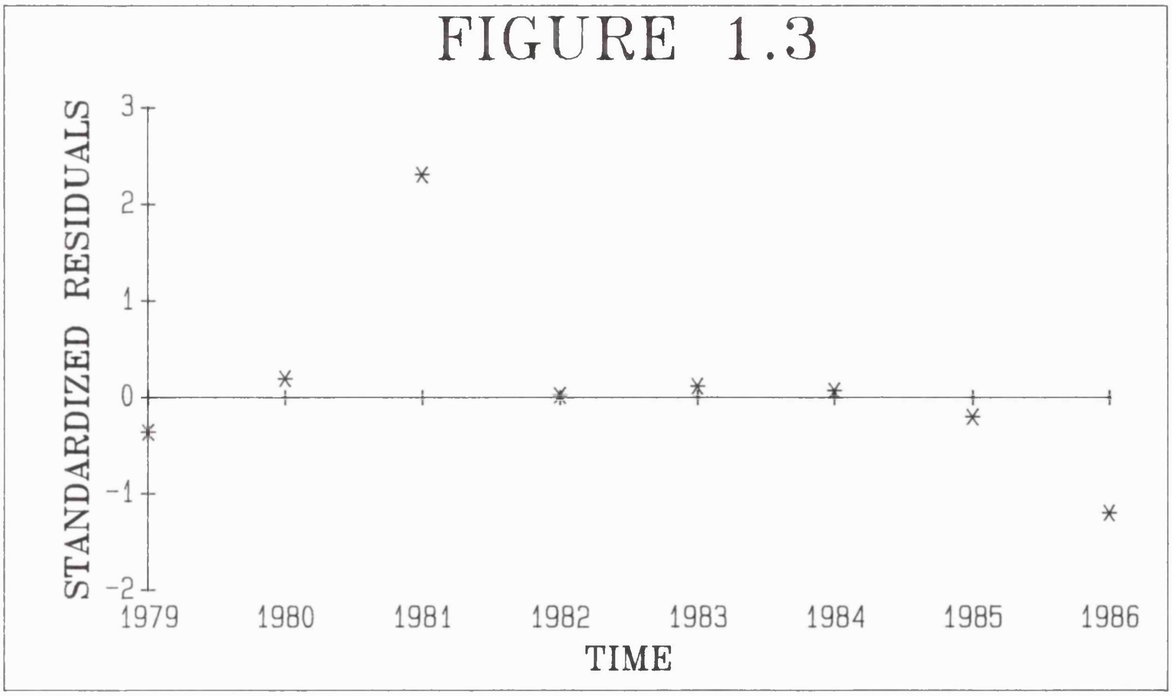


FIGURE 1.3
STANDARDIZED RESIDUALS
3
2
1
0
-1
-2
1979
1980
1981
1982
1983
1984
1985
1986
TIME

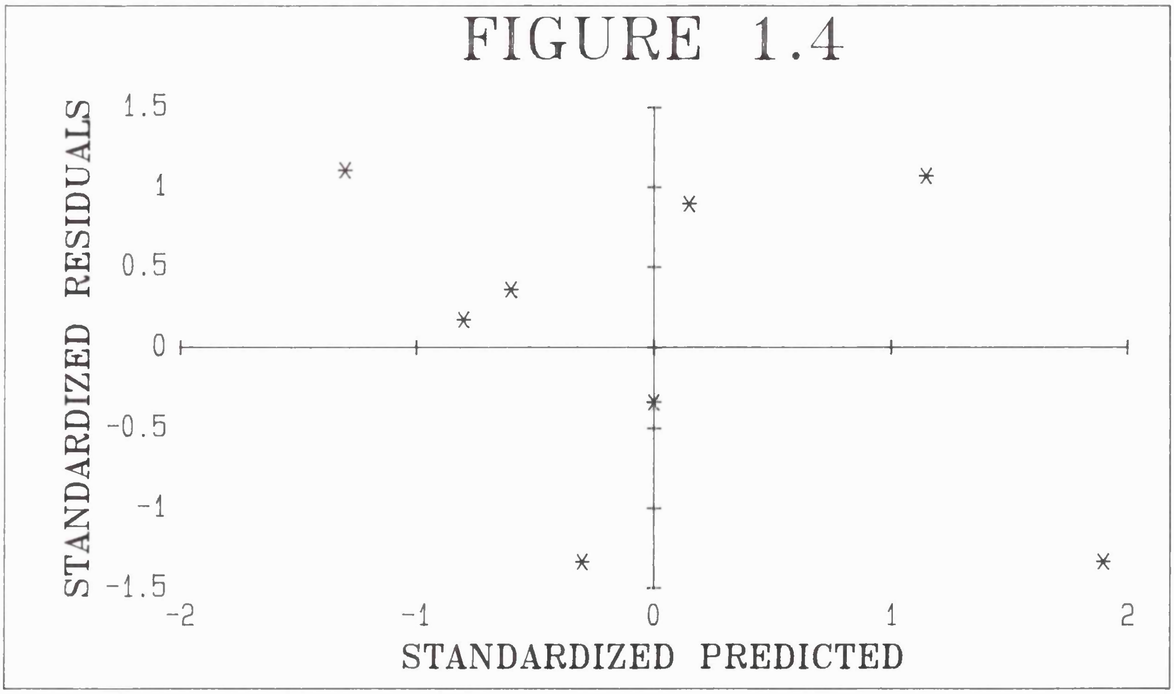


FIGURE 1.4
STANDARDIZED RESIDUALS
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-1
0
1
2
STANDARDIZED PREDICTED

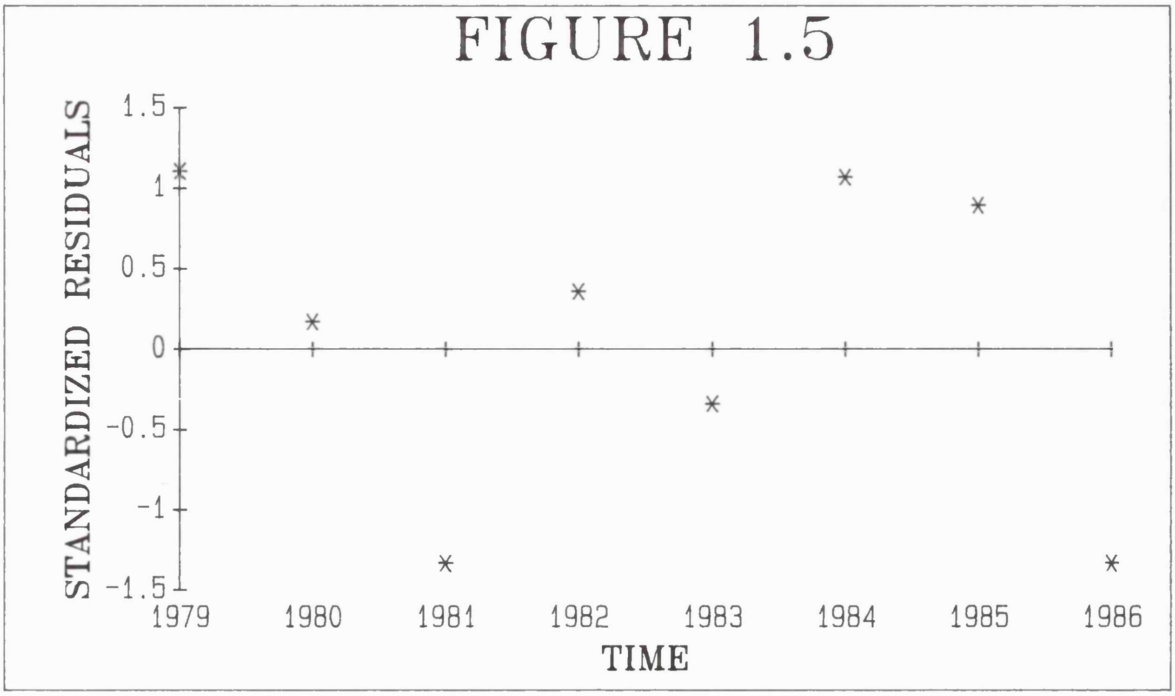


FIGURE 1.5
STANDARDIZED RESIDUALS
1.5
1
0.5
0
-0.5
-1
-1.5
1979
1980
1981
1982
1983
1984
1985
1986
TIME

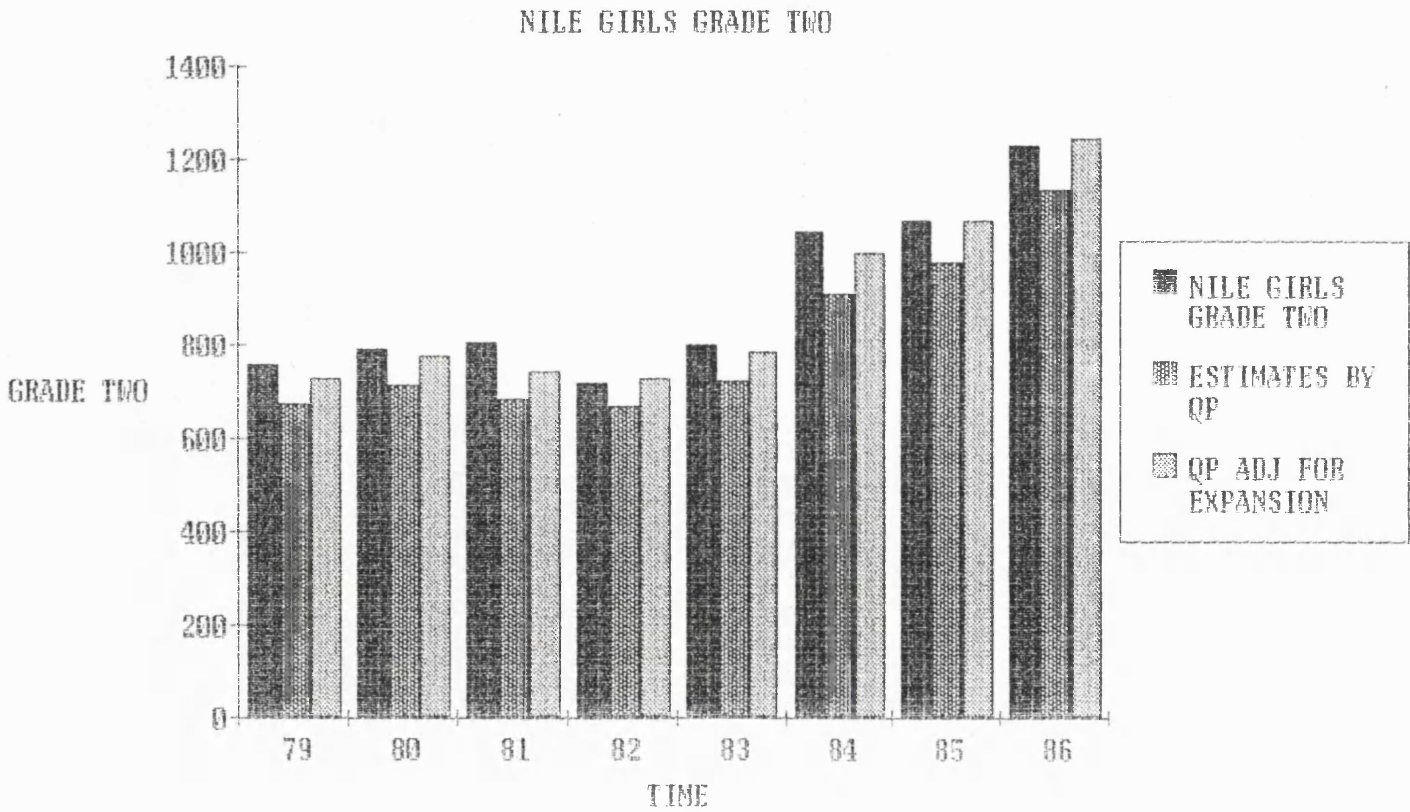
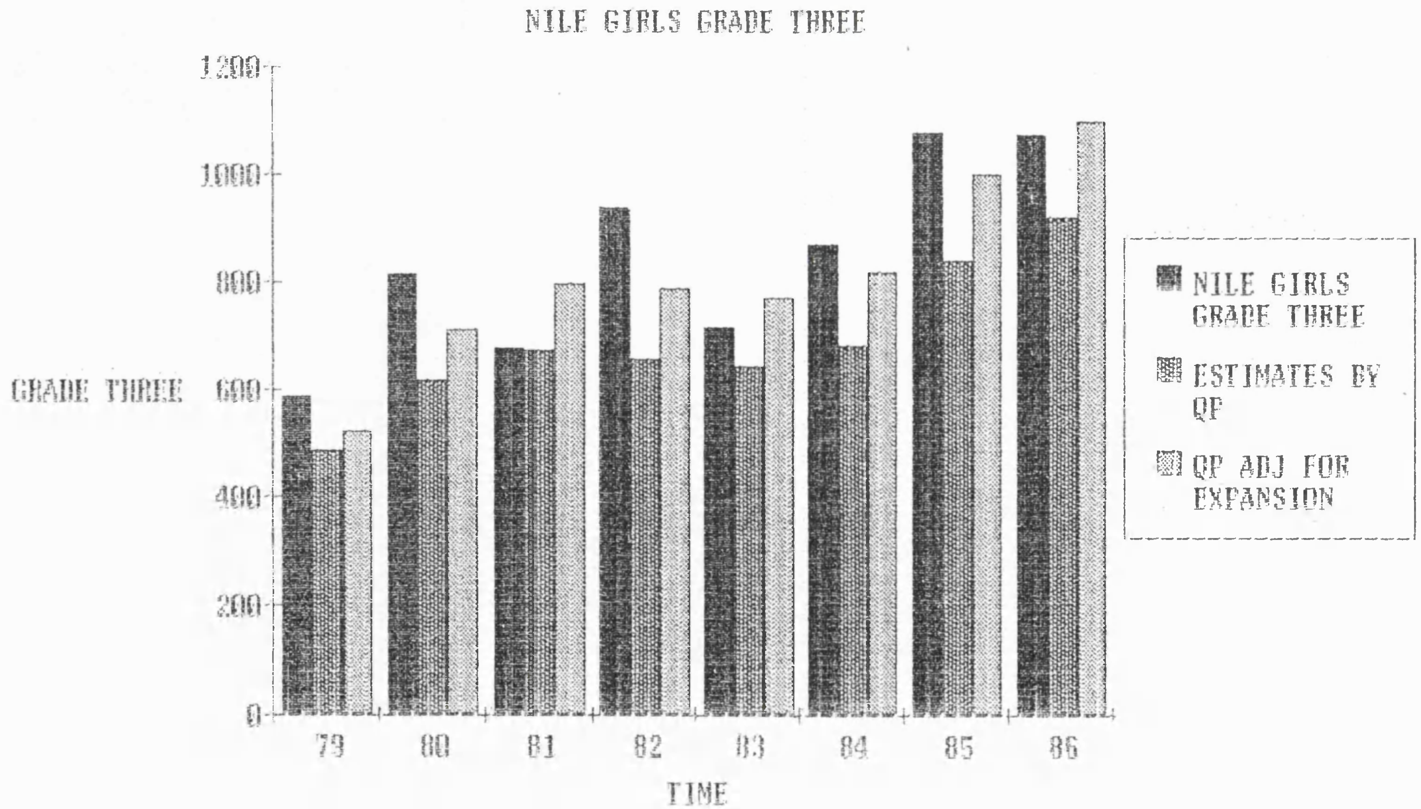

NILE GIRLS GRADE TWO
1400
1200
1000
800
600
400
200
0
GRADE TWO
79
80
81
82
83
84
85
86
TIME
NILE GIRLS GRADE TWO
ESTIMATES BY QP
QP ADJ FOR EXPANSION

Figure 1.6a


NILE GIRLS GRADE THREE
1200
1000
800
600
400
200
0
GRADE THREE
79
80
81
82
83
84
85
86
TIME
NILE GIRLS GRADE THREE
ESTIMATES BY QP
QP ADJ FOR EXPANSION

igure 1.6b

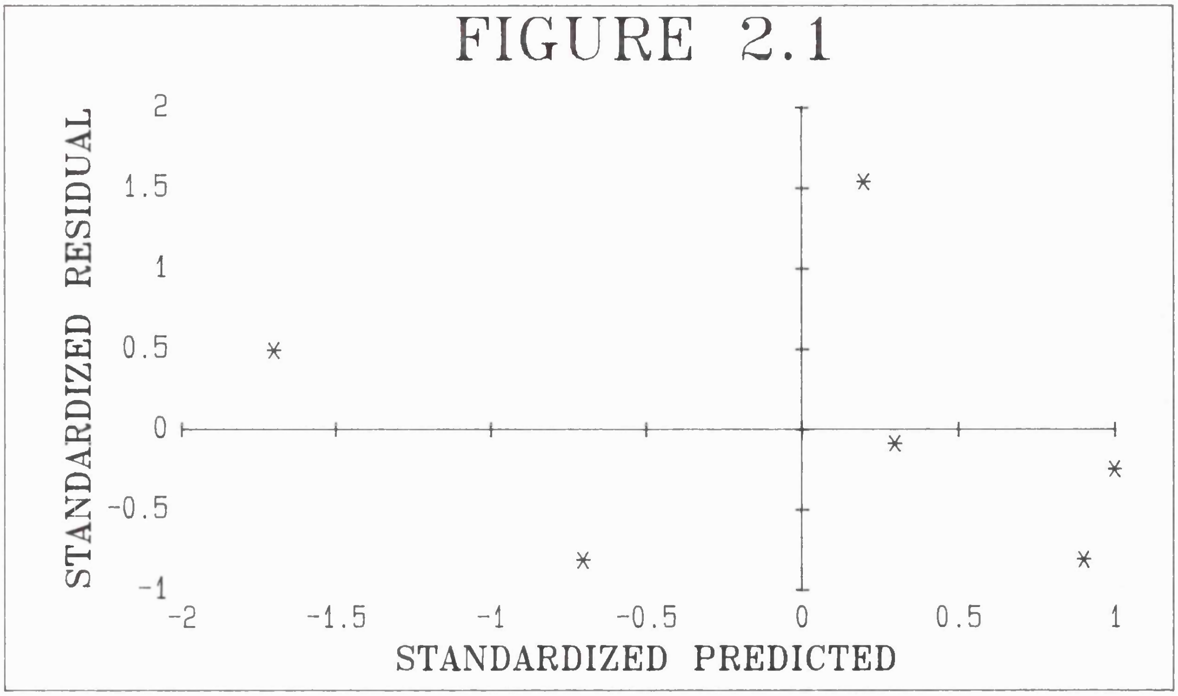


FIGURE 2.1
STANDARDIZED RESIDUAL
STANDARDIZED PREDICTED
2
1.5
1
0.5
0
-0.5
-1
-2
-1.5
-1
-0.5
0
0.5
1

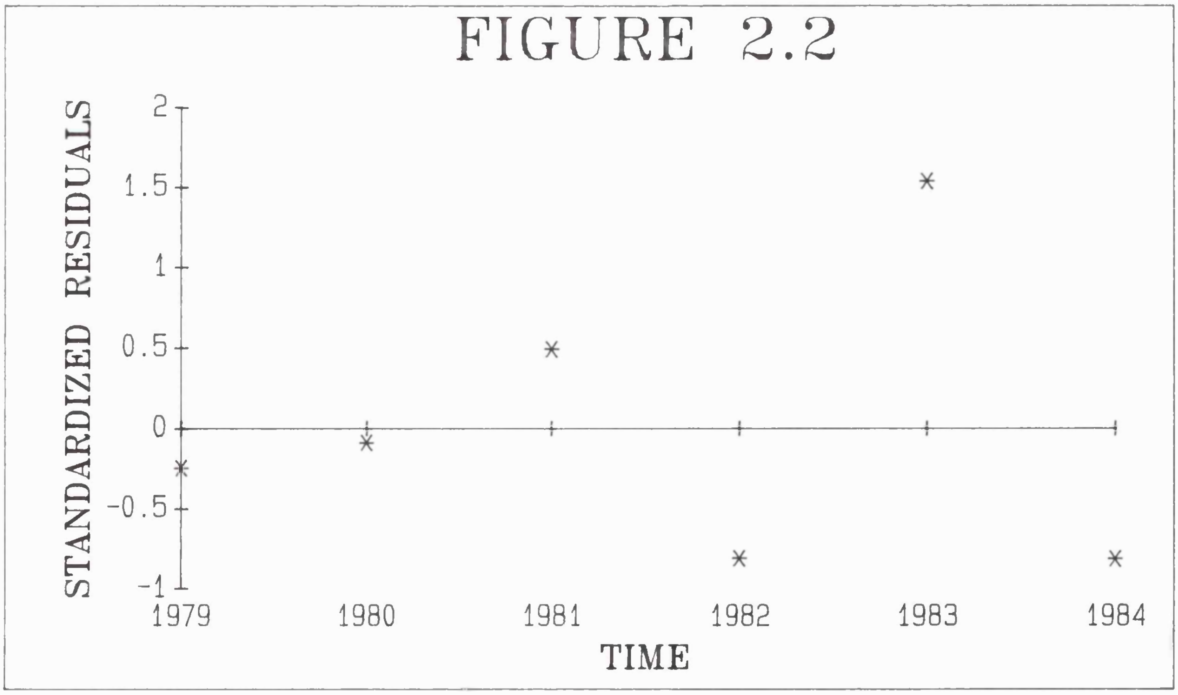


FIGURE 2.2
STANDARDIZED RESIDUALS
TIME
2
1.5
1
0.5
0
-0.5
-1
1979
1980
1981
1982
1983
1984

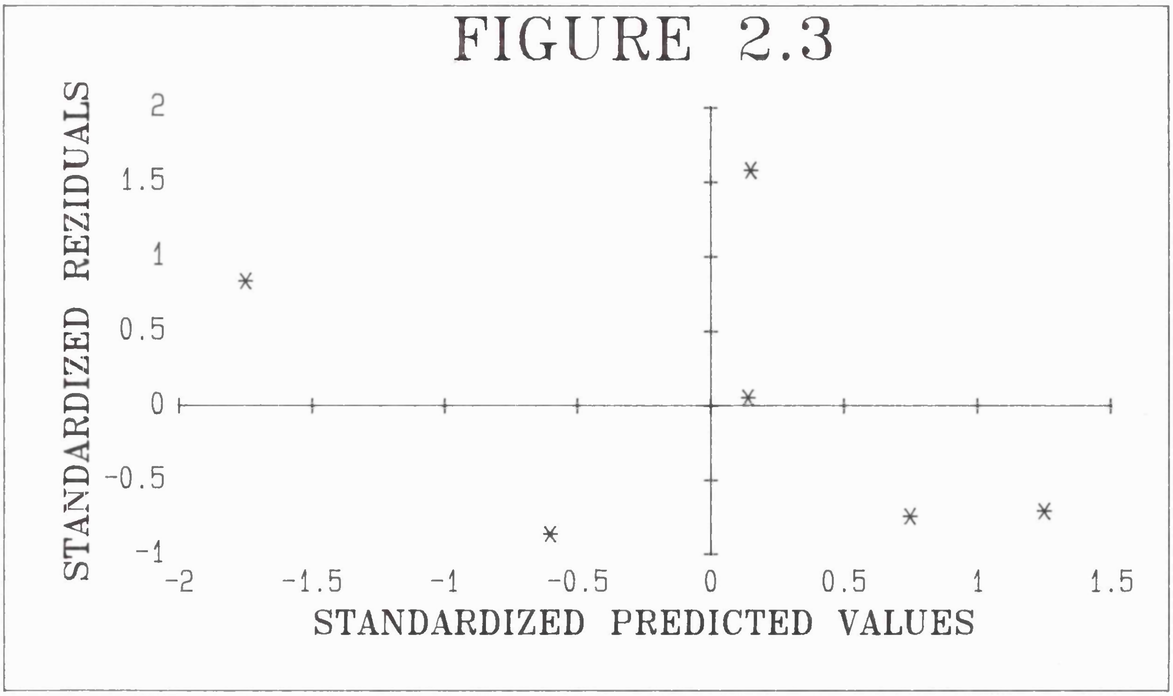


FIGURE 2.3
STANDARDIZED REZIDUALS
2
1.5
1
0.5
0
-0.5
-1
-2
-1.5
-1
-0.5
0
0.5
1
1.5
STANDARDIZED PREDICTED VALUES

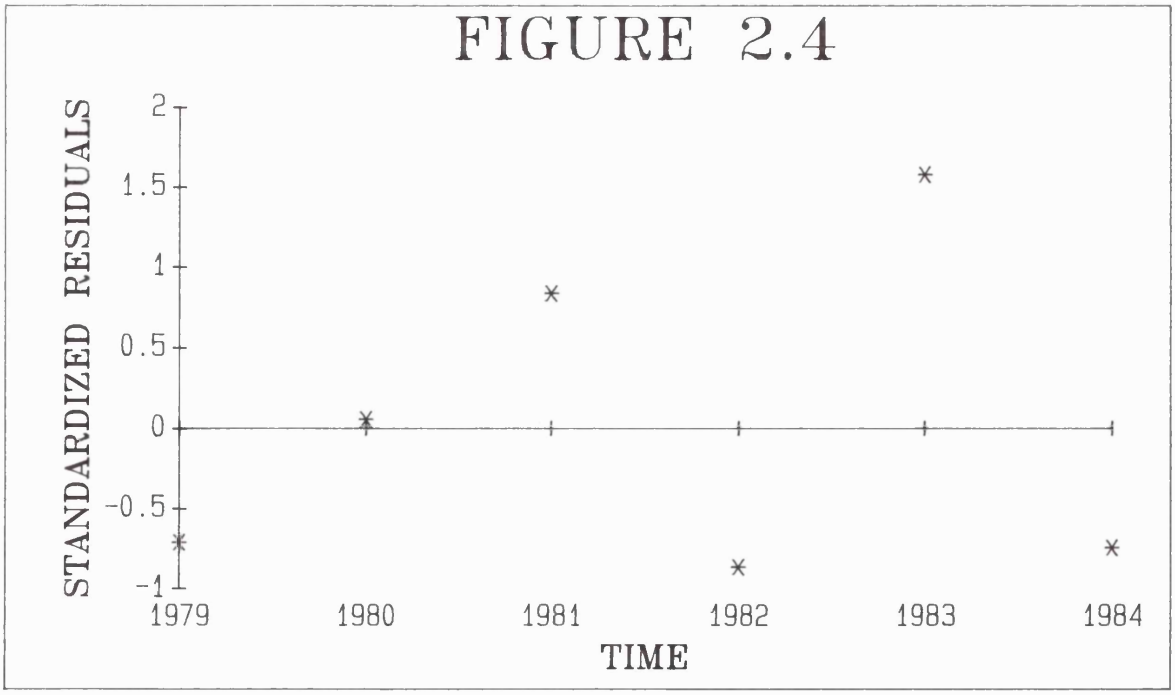


FIGURE 2.4
STANDARDIZED RESIDUALS
2
1.5
1
0.5
0
-0.5
-1
1979
1980
1981
1982
1983
1984
TIME

FIGURE 2.5

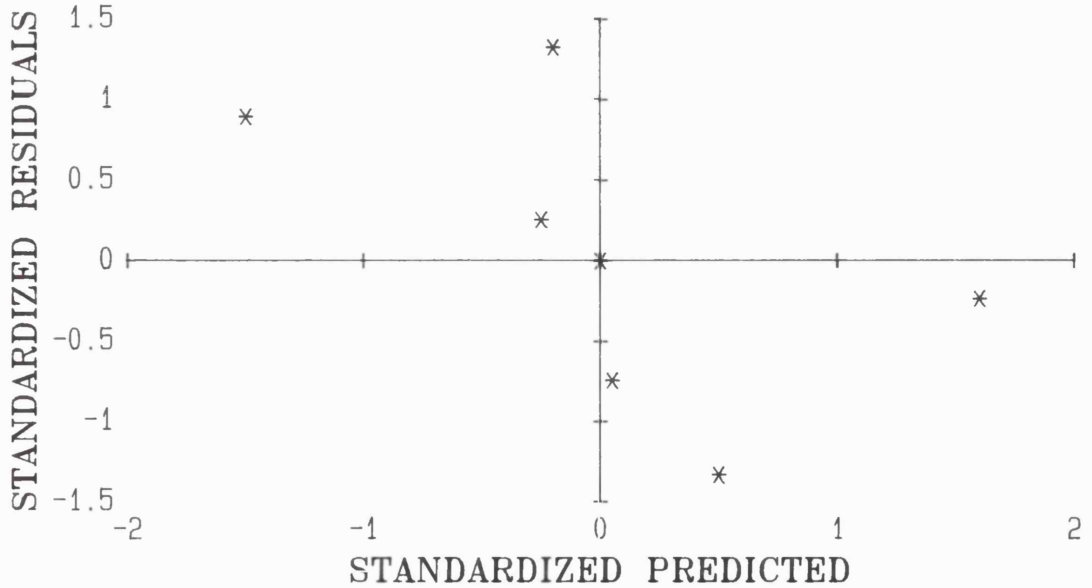

STANDARDIZED RESIDUALS
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-1
0
1
2
STANDARDIZED PREDICTED

FIGURE 3.1

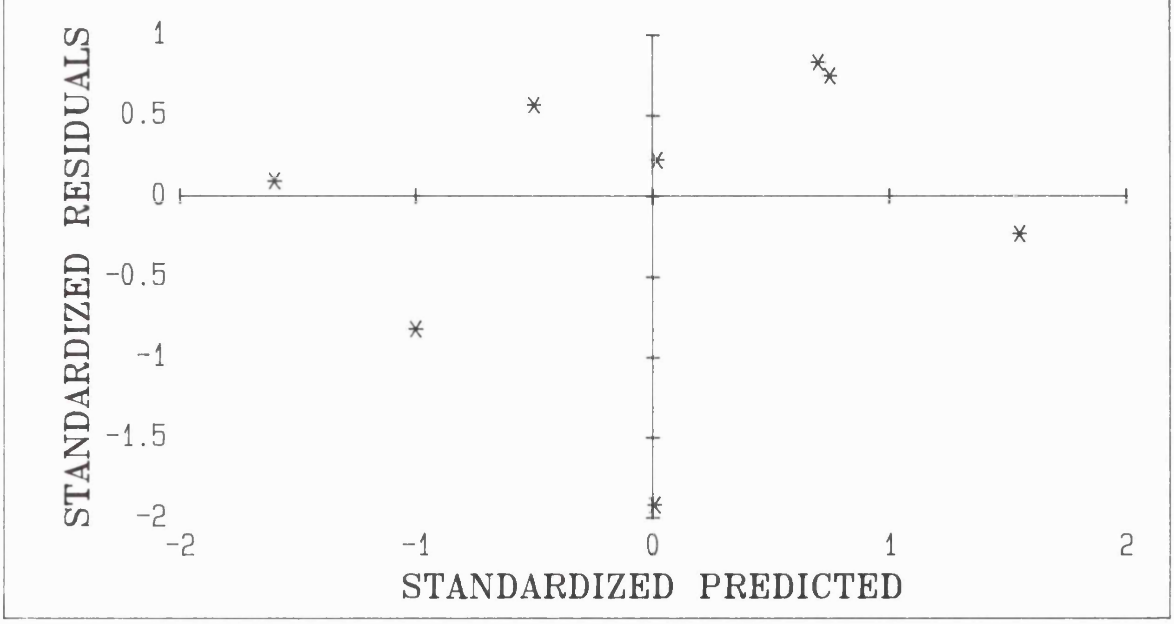
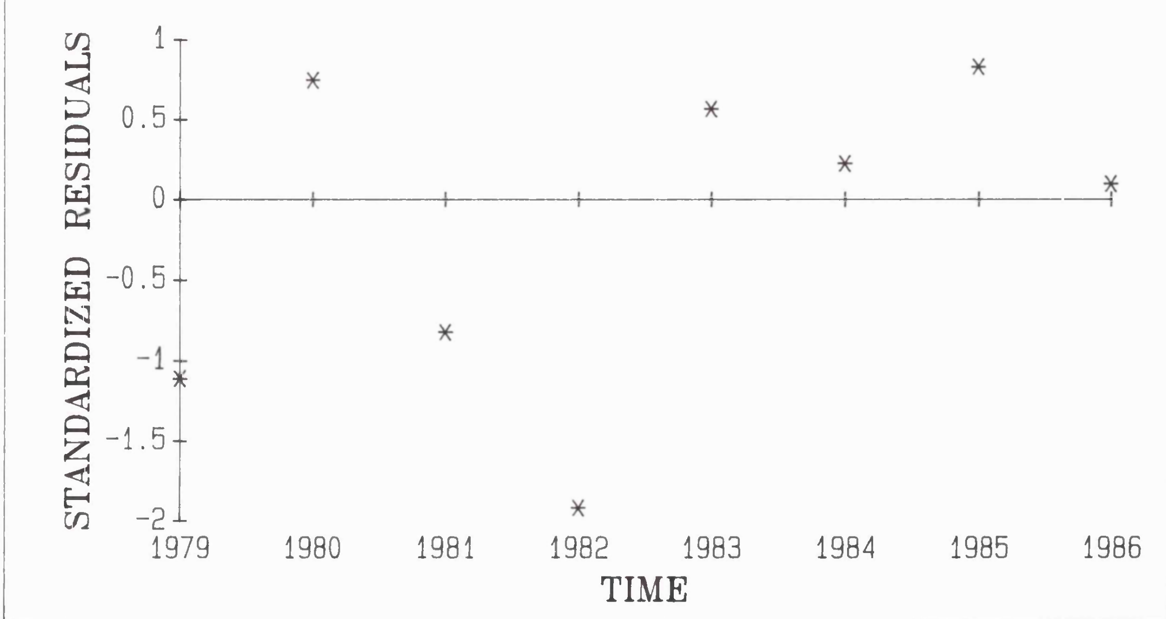


FIGURE 3.2



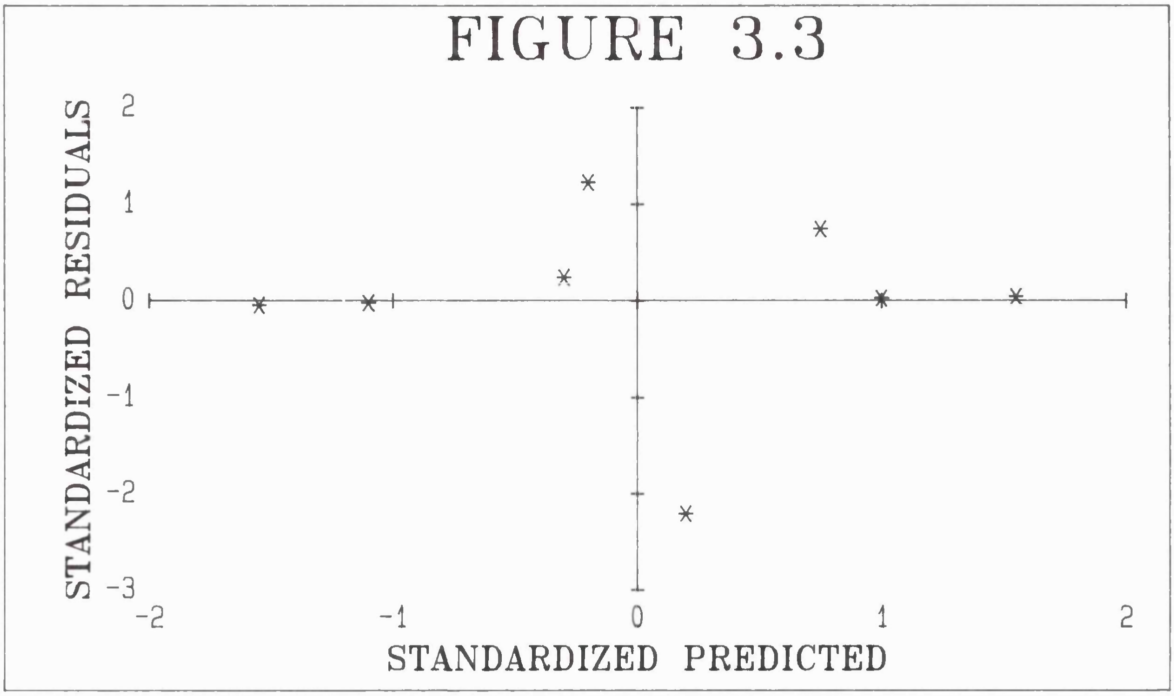


FIGURE 3.3
STANDARDIZED RESIDUALS
2
1
0
-1
-2
-3
-2
-1
0
1
2
STANDARDIZED PREDICTED

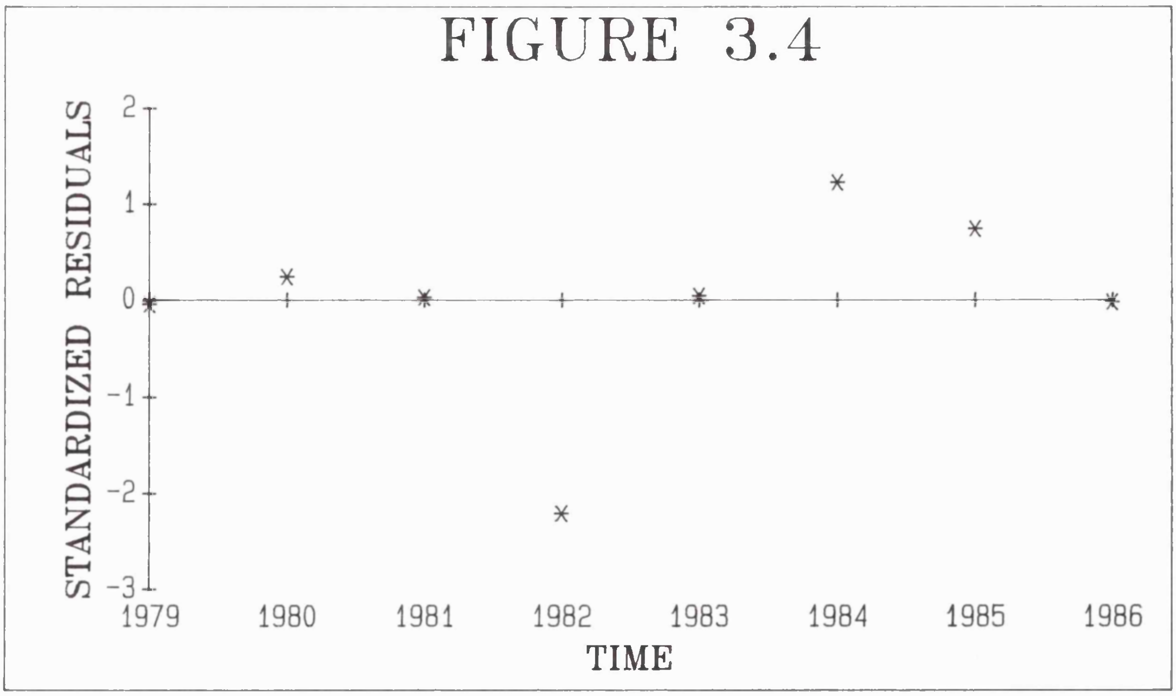


FIGURE 3.4
STANDARDIZED RESIDUALS
2
1
0
-1
-2
-3
1979
1980
1981
1982
1983
1984
1985
1986
TIME

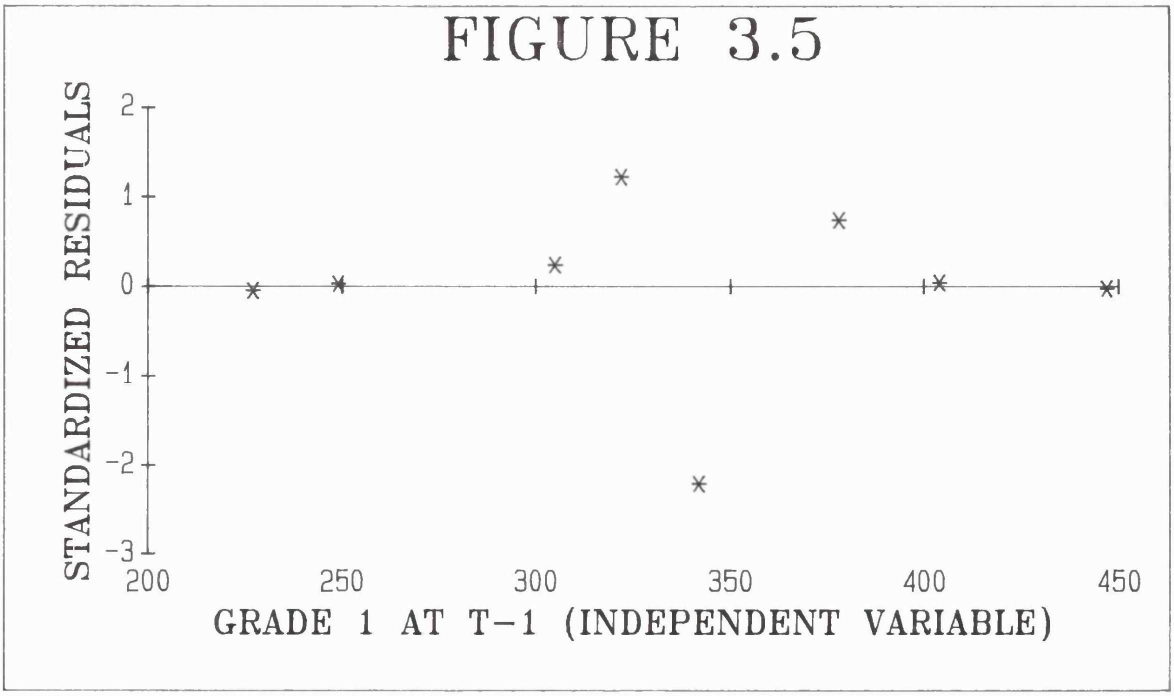


FIGURE 3.5
STANDARDIZED RESIDUALS
2
1
0
-1
-2
-3
200
250
300
350
400
450
GRADE 1 AT T-1 (INDEPENDENT VARIABLE)

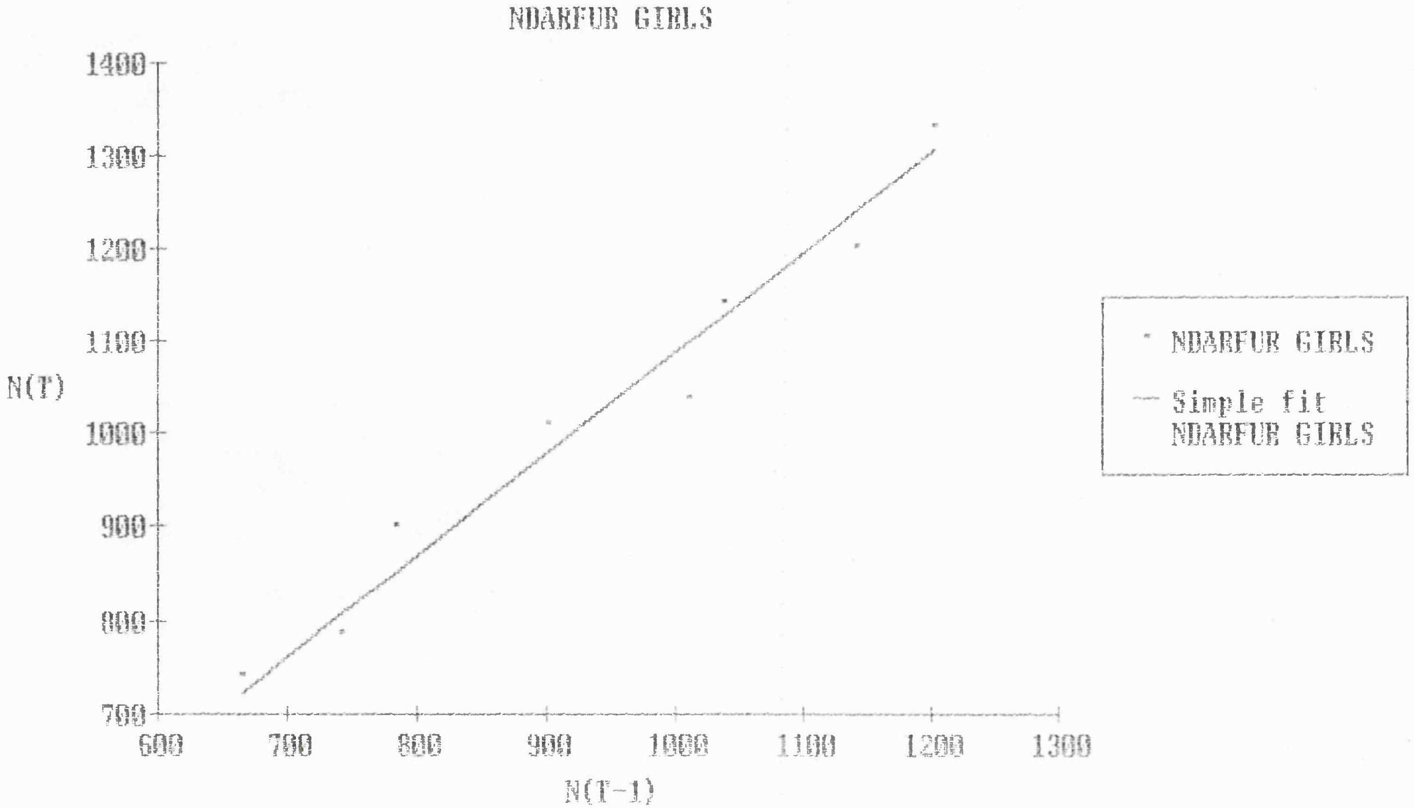

NDARFUR GIRLS
1400
1300
1200
1100
1000
900
800
700
N(T)
600
700
800
900
1000
1100
1200
1300
N(T-1)
NDARFUR GIRLS
Simple fit
NDARFUR GIRLS

Figure 3.6

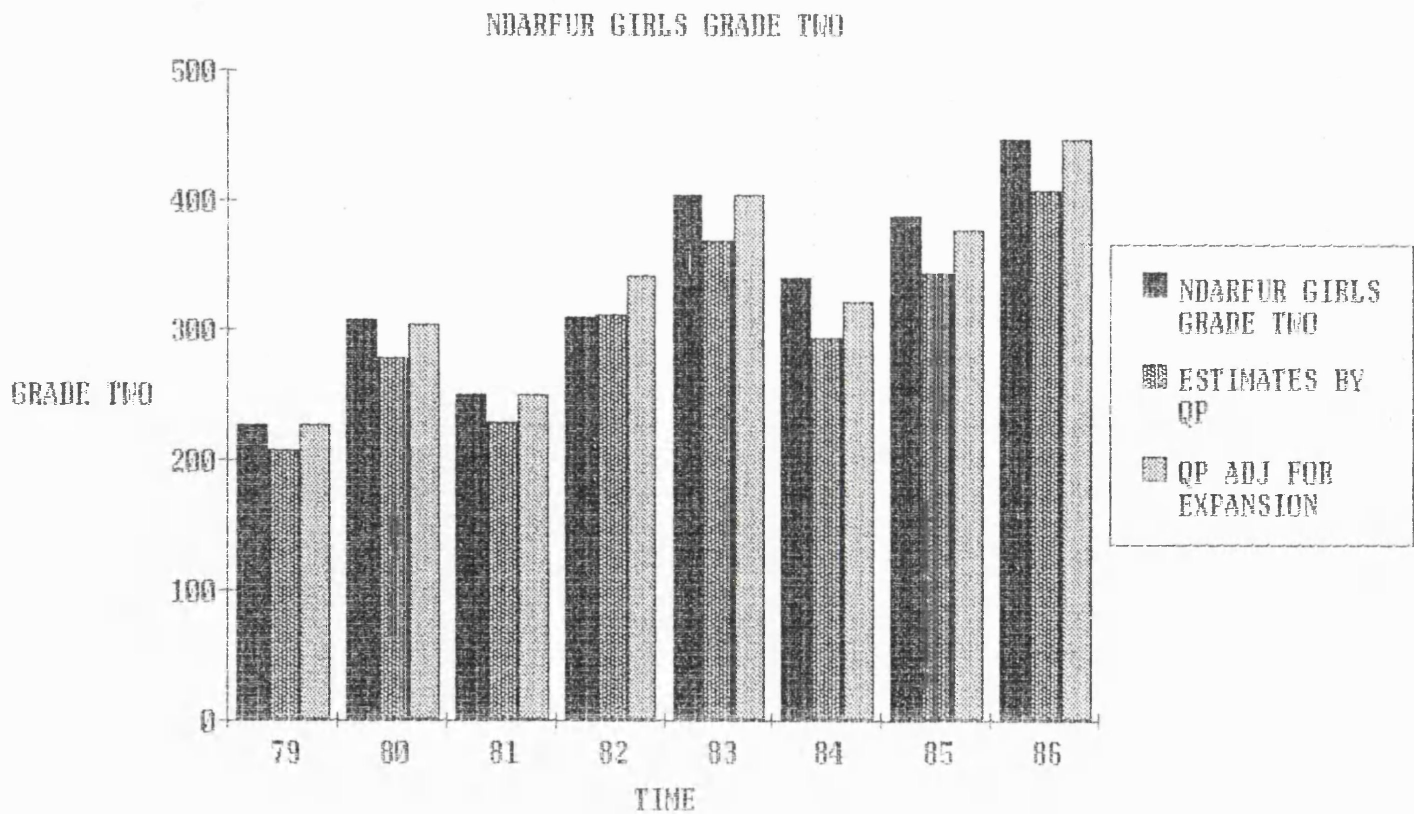

NDARFUR GIRLS GRADE TWO
GRADE TWO
500
400
300
200
100
0
79
80
81
82
83
84
85
86
TIME
NDARFUR GIRLS GRADE TWO
ESTIMATES BY QP
QP ADJ FOR EXPANSION

Figure 3.7

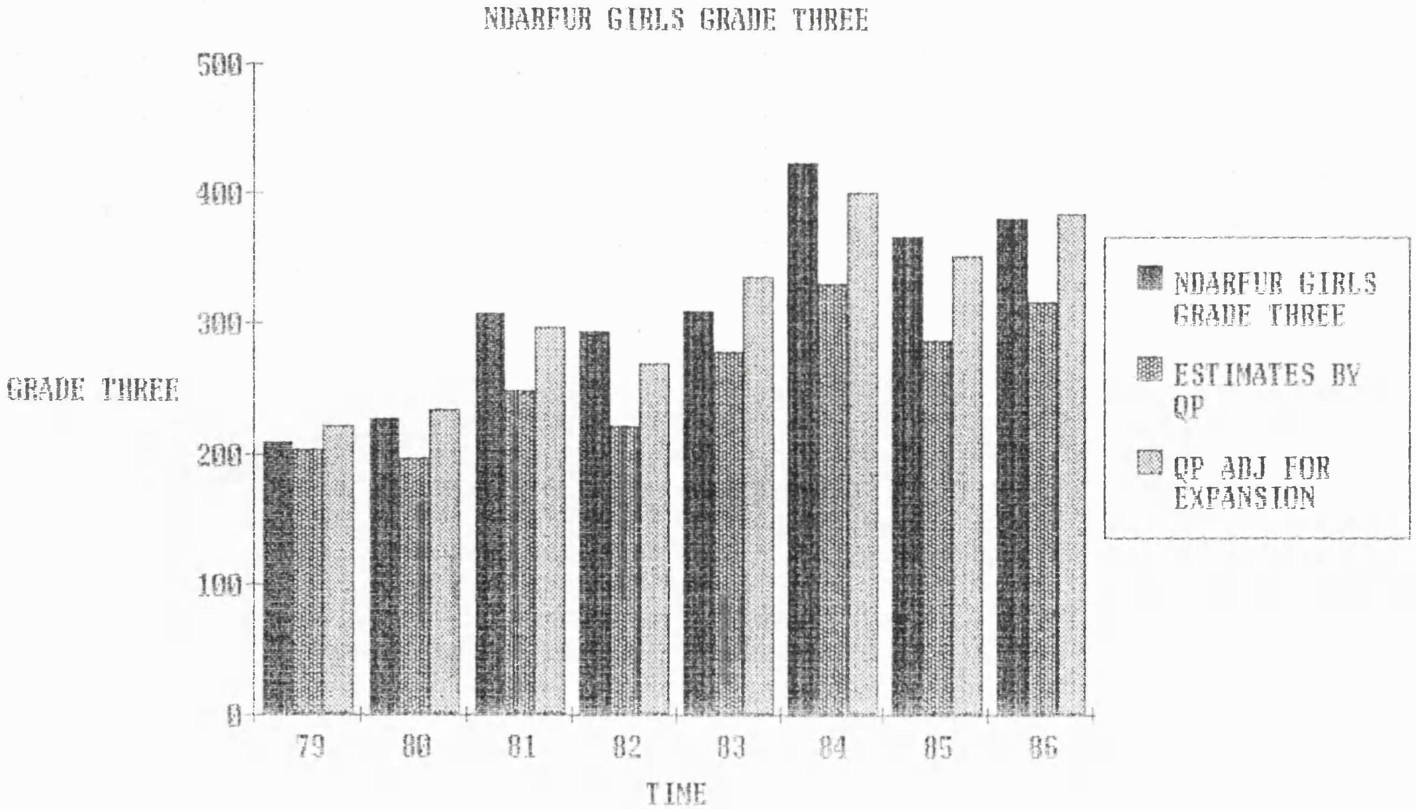

NDARFUR GIRLS GRADE THREE
500
400
300
200
100
0
GRADE THREE
79
80
81
82
83
84
85
86
TIME
NDARFUR GIRLS GRADE THREE
ESTIMATES BY QP
QP ADJ FOR EXPANSION

Figure 3.8

FIGURE 4.1A (from equation 1)

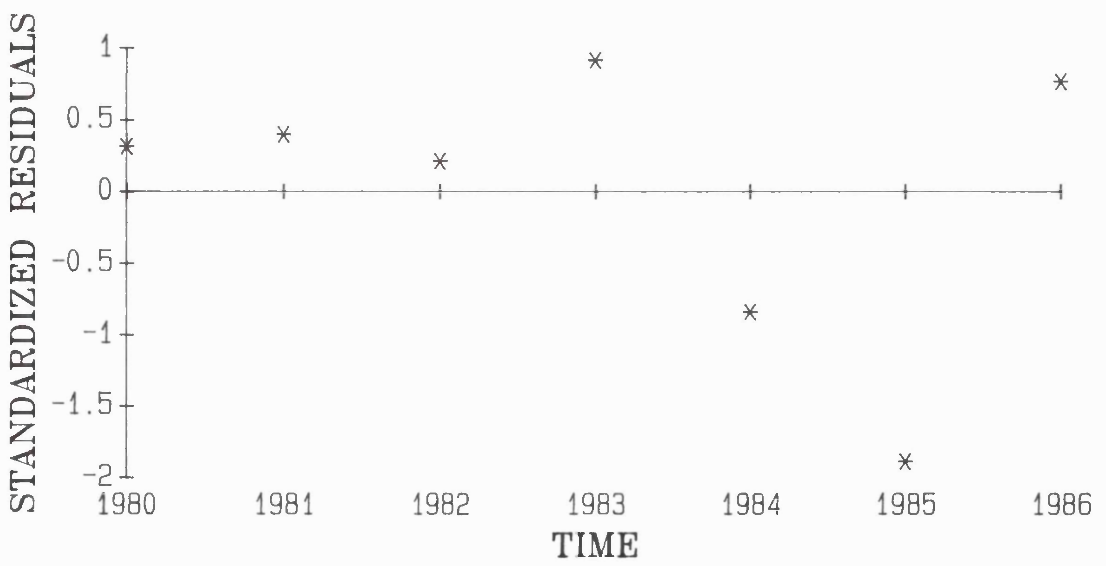
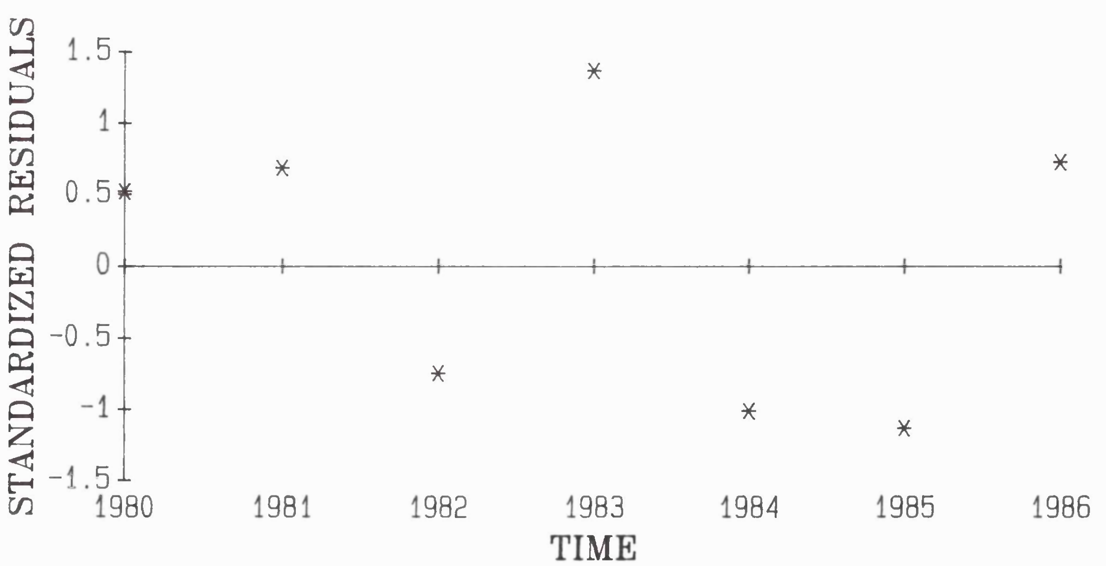


FIGURE 4.1B (from equation 2)



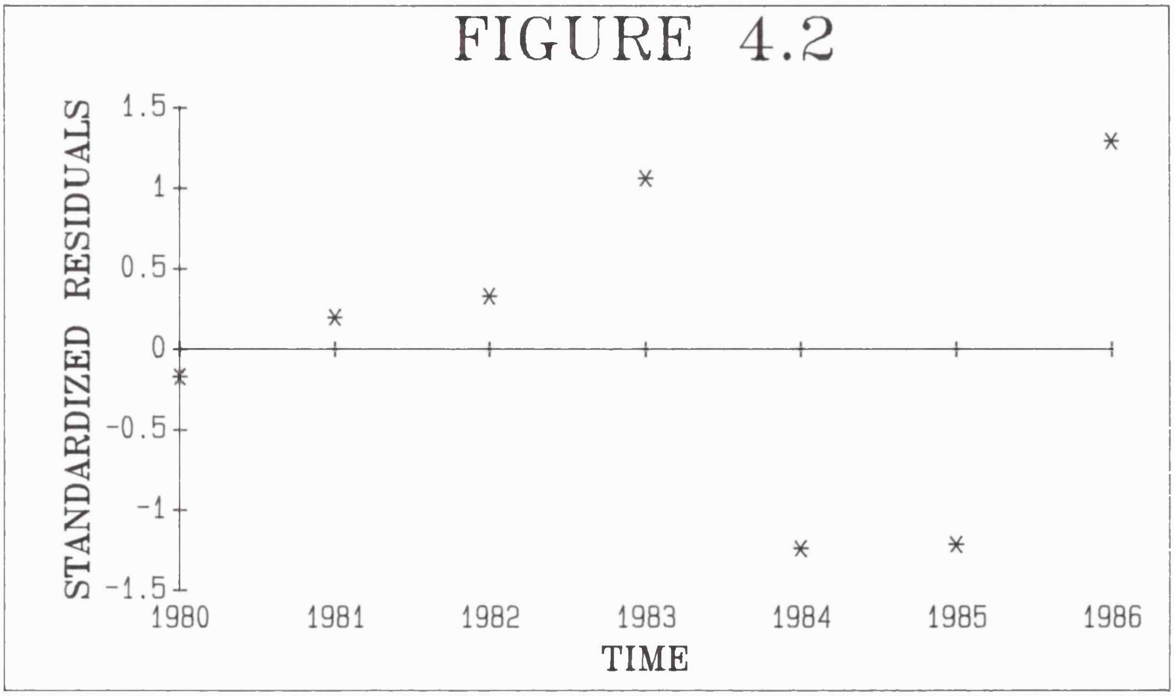


FIGURE 4.2
STANDARDIZED RESIDUALS
1.5
1
0.5
0
-0.5
-1
-1.5
1980
1981
1982
1983
1984
1985
1986
TIME

FIGURE 5.1

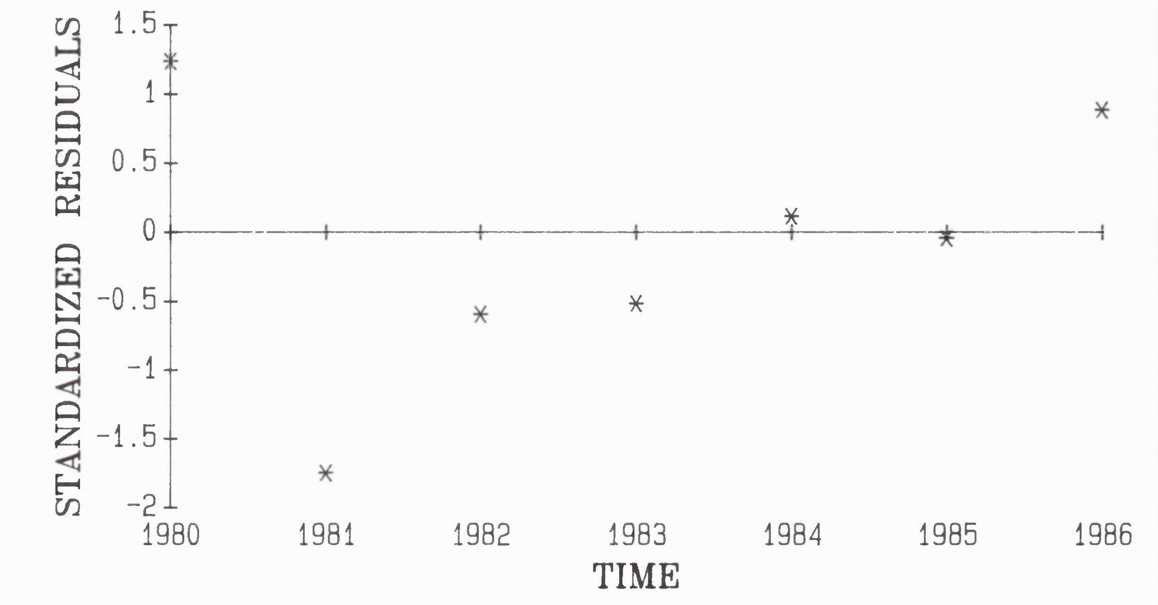
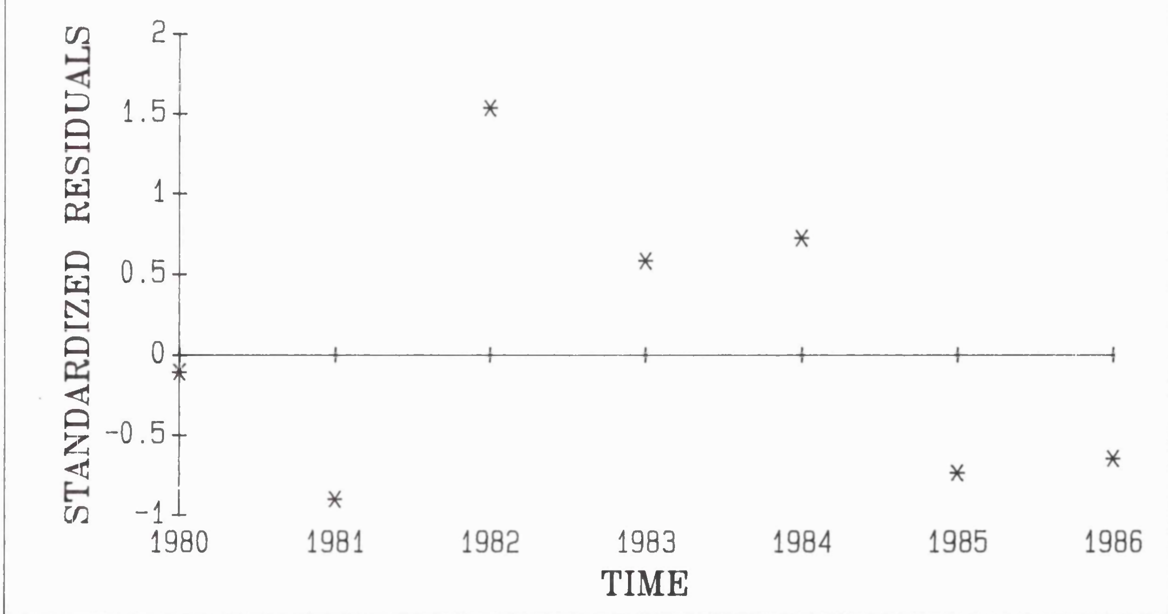


FIGURE 5.2



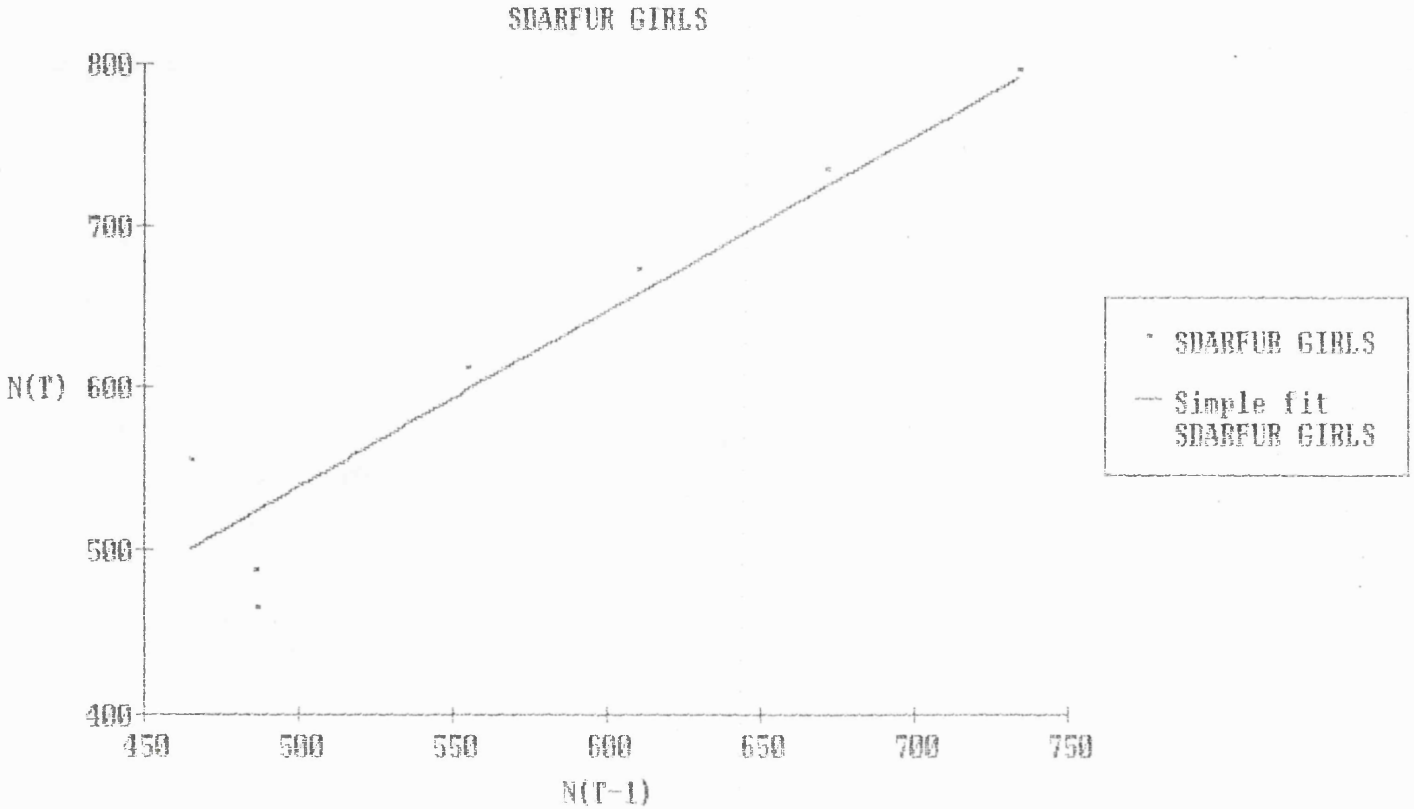
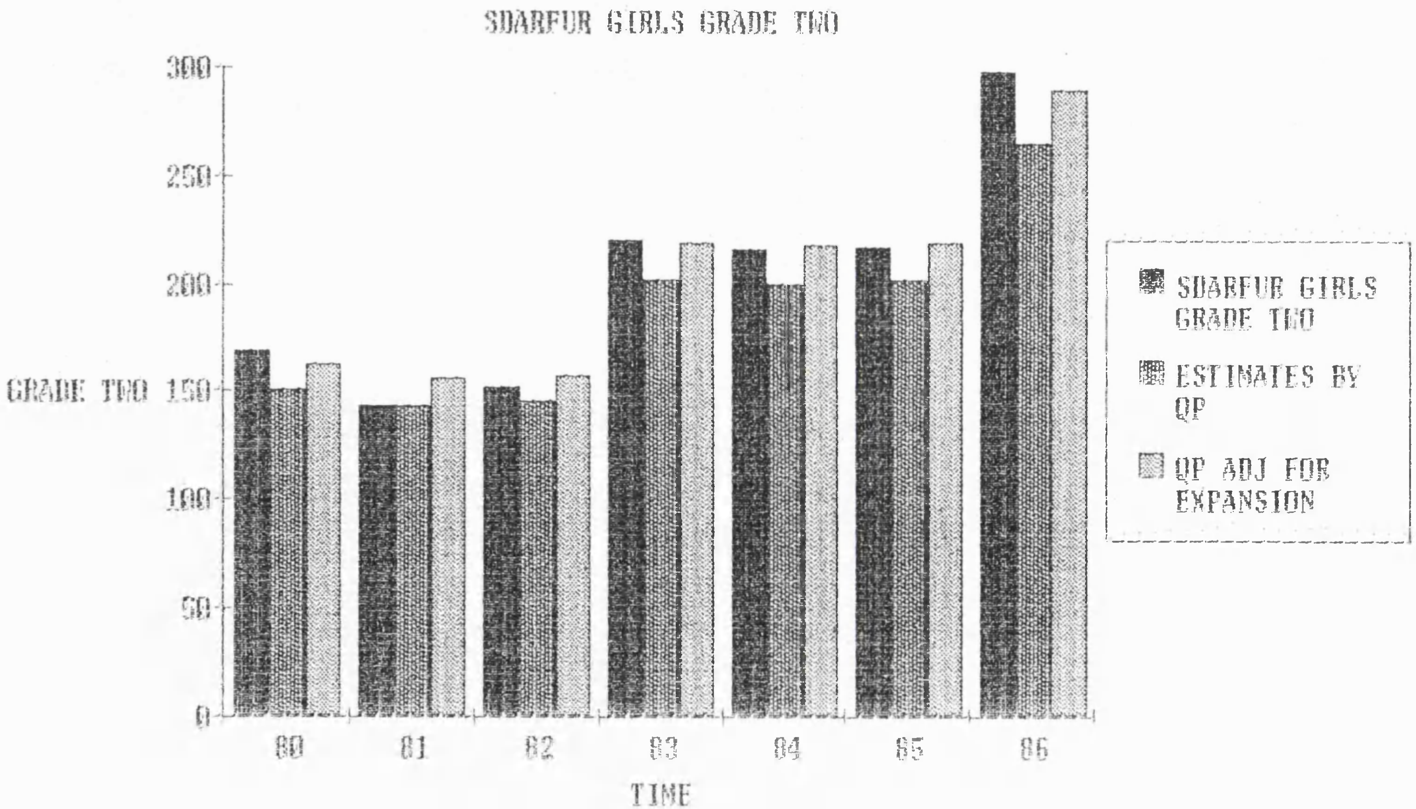
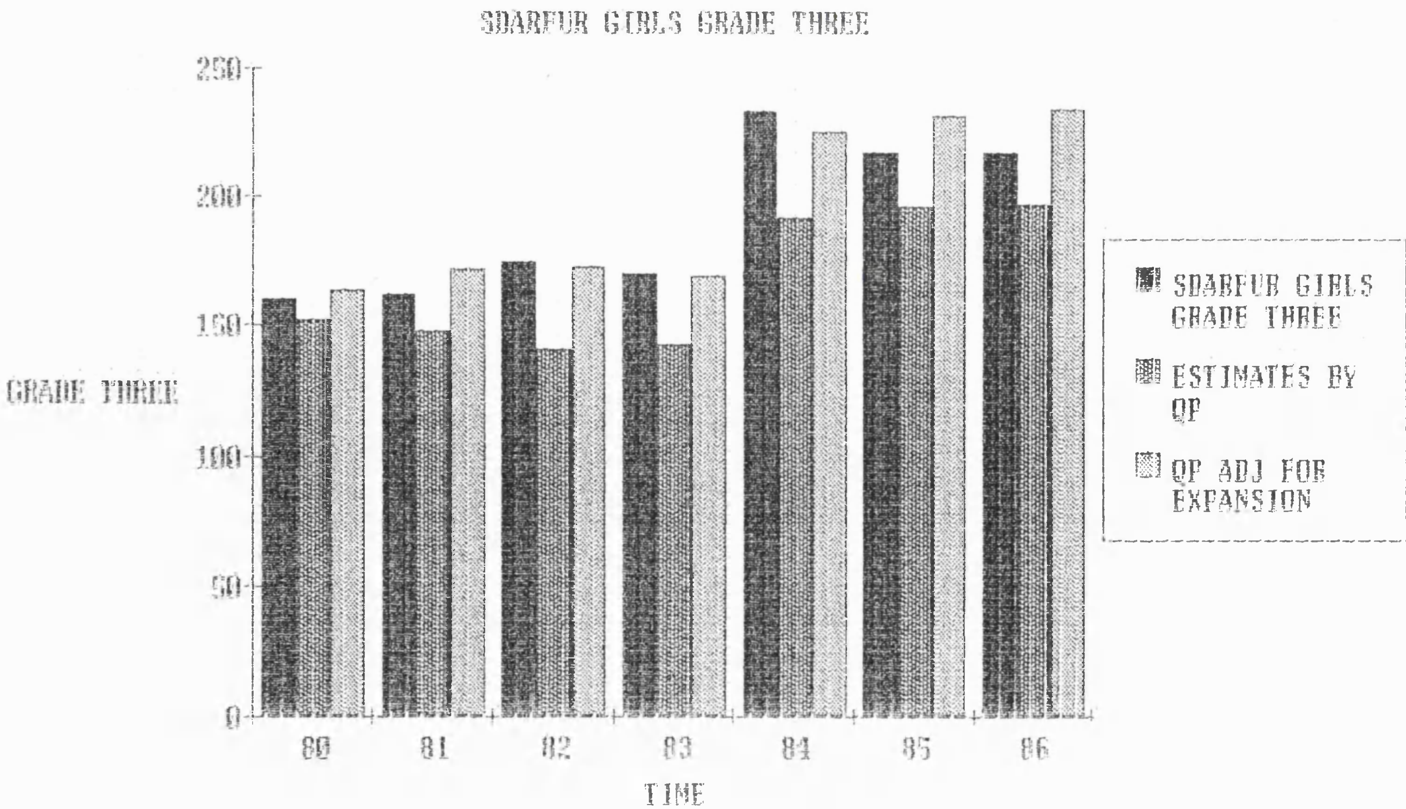

SDARFUR GIRLS
800
700
600
500
400
N(T)
450
500
550
600
650
700
750
N(T-1)
SDARFUR GIRLS
Simple fit
SDARFUR GIRLS

Figure 5.3


SDARFUR GIRLS GRADE TWO
GRADE TWO
300
250
200
150
100
50
0
80
81
82
83
84
85
86
TIME
SDARFUR GIRLS GRADE TWO
ESTIMATES BY QP
QP ADJ FOR EXPANSION

5.14


SDARFUR GIRLS GRADE THREE
250
200
150
100
50
0
GRADE THREE
80
81
82
83
84
85
86
TIME
SDARFUR GIRLS GRADE THREE
ESTIMATES BY QP
QP ADJ FOR EXPANSION

APPENDIX C

Figures for Chapter V

PROBABILITY OF MOVING FROM GRADE 1 TO 2

Figure VI

```
nt   Midpoint
 0     .6725 |
 0     .6800 |
 0     .6875 |
 0     .6950 |
 0     .7025 |
 0     .7100 |
 1     .7175 |*****
 0     .7250 |
 0     .7325 |.
 1     .7400 |:****
 1     .7475 |:****
 0     .7550 | .
 1     .7625 |*:***
 1     .7700 |**:**
 0     .7775 |    .
 0     .7850 |    .
 0     .7925 |     .
 1     .8000 |*****  .
 1     .8075 |*****   .
 1     .8150 |*****     .
 1     .8225 |*****      .
 0     .8300 |             .
 8     .8375 |**************:*************************
 3     .8450 |***************.
 4     .8525 |*****************:**
 2     .8600 |**********          .
 5     .8675 |********************:****
 6     .8750 |*********************:********
 4     .8825 |********************   .
 6     .8900 |***********************:******
 4     .8975 |********************    .
 4     .9050 |********************    .
 5     .9125 |**********************:**
 4     .9200 |********************   .
 1     .9275 |*****                 .
 7     .9350 |*******************:***************
 4     .9425 |******************:*
 2     .9500 |**********        .
 4     .9575 |***************:****
 2     .9650 |**********     .
 3     .9725 |***********:***
 2     .9800 |**********.
 3     .9875 |********:******
 0     .9950 |       .
 8    1.0025 |*****:***********************************
 0    1.0100 |     .
 0    1.0175 |    .
 0    1.0250 |   .
 0    1.0325 |   .
 0    1.0400 |  .
 0    1.0475 |.
             I....+....I....+....I....+....I....+....I....+....I
             0         2         4         6         8         10
                      Histogram Frequency

      .900      Std Err        .006     Median          .899
     1.000      Std Dev        .063     Variance        .004
s     .080      S E Kurt       .478     Skewness       -.412
w     .241      Range          .279     Minimum         .721
     1.000      Sum          89.982
```

FIG. V2

```
   PROBABILITY OF REMAINING IN GRADE 2
unt   Midpoint
 0     -.0075  |.
 0      .0000  |.
 0      .0075  | .
 0      .0150  | .
 0      .0225  |  .
 0      .0300  |   .
 1      .0375  |****:
 2      .0450  |*****:****
 4      .0525  |******:*************
 3      .0600  |********:******
 2      .0675  |*********:
 3      .0750  |***********:***
 3      .0825  |*************:*
 2      .0900  |**********    .
 5      .0975  |****************:********
 5      .1050  |******************:******
 3      .1125  |***************    .
 2      .1200  |**********          .
 3      .1275  |***************      .
 4      .1350  |********************   .
 6      .1425  |***********************:******
 3      .1500  |***************       .
 6      .1575  |***********************:******
 6      .1650  |**********************:*******
 8      .1725  |**********************:*****************
 2      .1800  |**********         .
 3      .1875  |***************   .
 2      .1950  |**********      .
 7      .2025  |****************:******************
 3      .2100  |**************:
 3      .2175  |************:**
 0      .2250  |         .
 2      .2325  |*********:
 1      .2400  |*****  .
 1      .2475  |***** .
 0      .2550  |      .
 1      .2625  |***:*
 1      .2700  |***:*
 0      .2775  |  .
 0      .2850  | .
 1      .2925  |*:***
 0      .3000  |.
 0      .3075  |.
 1      .3150  |:****
 1      .3225  |*****
 0      .3300  |
 0      .3375  |
 0      .3450  |
 0      .3525  |
 0      .3600  |
 0      .3675  |
               I....+....I....+....I....+....I....+....I....+....I
               0         2         4         6         8         10
                         Histogram Frequency

          .149      Std Err        .006      Median        .149
          .175      Std Dev        .062      Variance      .004
is       -.014      S E Kurt       .478      Skewness      .372
ew        .241      Range          .285      Minimum       .037
n         .322      Sum          14.897

ases       100      Missing Cases     0
```

Fig V3

```
   PROBABILITY OF MOVING FROM GRADE 2 TO 3
nt   Midpoint
 0      .5275 |
 0      .5350 |
 1      .5425 |*****
 0      .5500 |
 1      .5575 |*****
 0      .5650 |
 0      .5725 |
 0      .5800 |
 0      .5875 |
 0      .5950 |
 0      .6025 |.
 0      .6100 |.
 1      .6175 |:****
 0      .6250 |.
 0      .6325 | .
 1      .6400 |*:***
 2      .6475 |**:*******
 0      .6550 |    .
 0      .6625 |    .
 1      .6700 |****:
 0      .6775 |      .
 2      .6850 |*******:**
 3      .6925 |********:******
 0      .7000 |          .
 1      .7075 |*****       .
 2      .7150 |**********    .
 0      .7225 |               .
 4      .7300 |****************:***
 2      .7375 |**********         .
 1      .7450 |*****               .
 3      .7525 |***************       .
 7      .7600 |*********************:**************
 6      .7675 |*********************:*********
 2      .7750 |**********            .
 4      .7825 |********************  .
 6      .7900 |**********************:*******
 7      .7975 |*********************:**************
 3      .8050 |***************       .
 5      .8125 |*******************:*****
 5      .8200 |******************:******
 6      .8275 |*****************:************
 8      .8350 |***************:************************
 4      .8425 |**************:*****
 4      .8500 |************:*******
 1      .8575 |*****       .
 2      .8650 |*********:
 3      .8725 |********:******
 1      .8800 |***** .
 1      .8875 |*****.
 0      .8950 |    .
 0      .9025 |   .
               I....+....I....+....I....+....I....+....I....+....I
               0         2         4         6         8        10
                         Histogram Frequency

          .782      Std Err        .007      Median        .794
          .832      Std Dev        .066      Variance      .004
ls       2.065      S E Kurt       .478      Skewness    -1.239
ew        .241      Range          .344      Minimum       .541
n         .885      Sum          78.196

Cases      100      Missing Cases     0
```

Fig V4

```
   PROBABILITY OF REMAINING IN GRADE 3
unt   Midpoint
 0         .01 |   .
 0         .02 |    .
 0         .03 |     .
 0         .04 |      .
 2         .05 |*******:**
 0         .06 |         .
 2         .07 |**********.
 3         .08 |************:**
 7         .09 |**************:********************
 3         .10 |***************  .
 8         .11 |******************:*********************
 6         .12 |********************:*********
 4         .13 |********************  .
 7         .14 |***********************:***********
 5         .15 |************************:
 9         .16 |*************************:*******************
 4         .17 |********************     .
 3         .18 |***************          .
 6         .19 |************************:*****
 8         .20 |***********************:****************
 4         .21 |********************  .
 3         .22 |***************     .
 1         .23 |*****             .
 3         .24 |*************** .
 0         .25 |              .
 0         .26 |            .
 1         .27 |*****     .
 3         .28 |*********:*****
 1         .29 |*****  .
 0         .30 |     .
 1         .31 |****:
 0         .32 |   .
 1         .33 |**:**
 2         .34 |*:********
 0         .35 | .
 0         .36 |.
 1         .37 |:****
 0         .38 |.
 0         .39 |
 0         .40 |
 0         .41 |
 0         .42 |
 0         .43 |
 1         .44 |*****
 0         .45 |
 1         .46 |*****
 0         .47 |
 0         .48 |
 0         .49 |
 0         .50 |
 0         .51 |
               I....+....I....+....I....+....I....+....I....+....I
               0         2         4         6         8        10
                         Histogram Frequency
```

Fig F1

```
    PROBABILITY OF MOVING FROM GRADE 1 TO 2
unt    Midpoint
  1       .749 |**
  0       .754 |
  2       .759 |:****
  0       .764 |.
  1       .769 |:*
  1       .774 |:*
  0       .779 |.
  0       .784 |.
  0       .789 |.
  0       .794 | .
  1       .799 |*:
  0       .804 | .
  3       .809 |*:*****
  0       .814 |  .
  1       .819 |**.
  3       .824 |**:****
  2       .829 |**:**
  1       .834 |**  .
  0       .839 |    .
  1       .844 |**  .
  3       .849 |****:**
  0       .854 |     .
  2       .859 |****:
  2       .864 |*****.
  4       .869 |*****:****
  3       .874 |*****:*
  2       .879 |*****  .
  3       .884 |******:
  2       .889 |*****  .
  2       .894 |*****  .
  2       .899 |*****  .
  1       .904 |**     .
  3       .909 |******:
  5       .914 |******:*****
  1       .919 |**     .
  1       .924 |**     .
  2       .929 |*****  .
  3       .934 |******:
  2       .939 |*****  .
  1       .944 |**     .
  3       .949 |*****:*
  5       .954 |*****:******
  4       .959 |*****:****
  3       .964 |****:**
  3       .969 |****:**
  1       .974 |**   .
  2       .979 |***:*
  2       .984 |***:*
  1       .989 |**  .
  3       .994 |**:****
 12       .999 |**:***************************
               I....+....I....+....I....+....I....+....I....+....I
               0         4         8         12        16        20
                          Histogram Frequency

          .912     Std Err        .007     Median        .915
         1.000     Std Dev        .067     Variance      .004
s        -.557     S E Kurt       .478     Skewness     -.495
ew        .241     Range          .251     Minimum       .749
         1.000     Sum          91.160
```

Fig F2

```
      PROBABILITY OF REMAINING IN GRADE 2
ount     Midpoint
  0       -.0175     .
  0       -.0100  |  .
  0       -.0025  |    .
  0        .0050  |    .
  0        .0125  |     .
  0        .0200  |      .
  1        .0275  |*****.
  2        .0350  |******:***
  4        .0425  |********:***********
  5        .0500  |*********:**************
  3        .0575  |**********:****
  2        .0650  |**********     .
  3        .0725  |*************:*
  6        .0800  |**************:*************
  8        .0875  |***************:**********************
  3        .0950  |***************    .
  3        .1025  |***************      .
  3        .1100  |***************        .
  3        .1175  |***************          .
  5        .1250  |********************:***
  4        .1325  |********************  .
  4        .1400  |********************  .
  4        .1475  |********************  .
  0        .1550  |                         .
  3        .1625  |***************         .
  4        .1700  |******************:*
  7        .1775  |*****************:*****************
  3        .1850  |***************  .
  1        .1925  |*****          .
  2        .2000  |**********   .
  1        .2075  |*****      .
  0        .2150  |           .
  4        .2225  |*********:**********
  3        .2300  |********:******
  2        .2375  |******:***
  0        .2450  |      .
  0        .2525  |     .
  2        .2600  |***:******
  0        .2675  |   .
  0        .2750  |   .
  0        .2825  | .
  3        .2900  |*:*************
  1        .2975  |:****
  0        .3050  |.
  1        .3125  |:****
  0        .3200  |.
  0        .3275  |
  0        .3350  |
  0        .3425  |
  0        .3500  |
  0        .3575  |
                  I....+....I....+....I....+....I....+....I....+....I
                  0         2         4         6         8         10
```

Fig F3

```
   PROBABILITY OF MOVING FROM GRADE 2 TO 3
unt    Midpoint
 1      .5575 |*****
 1      .5650 |*****
 0      .5725 |
 1      .5800 |:****
 0      .5875 |.
 0      .5950 |.
 0      .6025 |.
 0      .6100 |.
 0      .6175 | .
 0      .6250 | .
 0      .6325 | .
 1      .6400 |**:**
 0      .6475 |   .
 1      .6550 |***:*
 0      .6625 |    .
 2      .6700 |****:*****
 1      .6775 |*****.
 2      .6850 |******:***
 2      .6925 |*******:**
 1      .7000 |*****    .
 4      .7075 |*********:**********
 3      .7150 |**********:****
 0      .7225 |            .
 2      .7300 |**********   .
 3      .7375 |*************:*
 3      .7450 |**************:
 0      .7525 |               .
 3      .7600 |***************.
 1      .7675 |*****           .
 4      .7750 |*****************:**
 3      .7825 |***************  .
 4      .7900 |*****************:**
 2      .7975 |**********        .
 2      .8050 |**********        .
 6      .8125 |*****************:************
 5      .8200 |*****************:*******
 3      .8275 |***************  .
 5      .8350 |****************:********
 3      .8425 |***************.
 2      .8500 |**********     .
 3      .8575 |*************:*
 8      .8650 |************:***************************
 0      .8725 |           .
 1      .8800 |*****     .
 4      .8875 |*********:**********
 3      .8950 |********:******
 4      .9025 |*******:************
 2      .9100 |******:***
 2      .9175 |*****:****
 1      .9250 |****:
 1      .9325 |****:
               I....+....I....+....I....+....I....+....I....+....I
               0         2         4         6         8        10
                         Histogram Frequency

        .798     Std Err        .008     Median        .812
        .863     Std Dev        .082     Variance      .007
sis     .235     S E Kurt       .478     Skewness     -.697
ew      .241     Range          .379     Minimum       .556
m       .935     Sum          79.804

Cases    100     Missing Cases     0
```

Fig. F4

```
   PROBABILITY OF REMAINING IN GRADE 3

nt     Midpoint
 0       -.04   |   .
 0       -.03   |   .
 0       -.02   |     .
 0       -.01   |       .
 4        .00   |*****:**************
 0        .01   |           .
 2        .02   |*******:**
 5        .03   |********:****************
 3        .04   |**********:****
 2        .05   |**********   .
 2        .06   |**********     .
 3        .07   |**************:
 5        .08   |***************:*********
 5        .09   |****************:********
 5        .10   |*****************:*******
 4        .11   |******************:*
 5        .12   |*******************:*****
 4        .13   |********************.
 6        .14   |********************:*********
 4        .15   |********************.
 2        .16   |**********          .
 3        .17   |***************     .
 3        .18   |***************    .
 4        .19   |******************:*
 2        .20   |**********        .
 1        .21   |*****          .
 3        .22   |***************.
 3        .23   |**************:
 1        .24   |*****         .
 4        .25   |***********:********
 3        .26   |**********:****
 3        .27   |********:******
 2        .28   |*******:**
 1        .29   |*****  .
 0        .30   |      .
 1        .31   |****:
 1        .32   |***:*
 1        .33   |**:**
 0        .34   |   .
 0        .35   |  .
 0        .36   |  .
 0        .37   |.
 0        .38   |.
 .        .39   |:****
 0        .40   |.
 1        .41   |*****
 0        .42   |
 1        .43   |*****
 0        .44   |
 0        .45   |
 0        .46   |
                I....+....I....+....I....+....I....+....I....+....I
                0         2         4         6         8         10
                          Histogram Frequency

           .151      Std Err        .009      Median       .139
           .000      Std Dev        .094      Variance     .009
is         .155      S E Kurt       .478      Skewness     .635
ew         .241      Range          .426      Minimum      .000
m          .426      Sum          15.064

Cases      100       Missing Cases     0
```

Figure S1

```
PROBABILITY OF MOVING FROM GRADE 1 TO 2
  Midpoint
     .6525 |
     .6600 |
     .6675 |
     .6750 |
     .6825 |**
     .6900 |
     .6975 |
     .7050 |
     .7125 |**
     .7200 |
     .7275 |.
     .7350 |:*
     .7425 |.
     .7500 |:****
     .7575 |.
     .7650 |*:
     .7725 | .
     .7800 | .
     .7875 |**.
     .7950 |  .
     .8025 |** .
     .8100 |   .
     .8175 |****:
     .8250 |****:
     .8325 |*****:******
     .8400 |******:***
     .8475 |******:********
     .8550 |*******.
     .8625 |*****  .
     .8700 |*****   .
     .8775 |*****   .
     .8850 |*****   .
     .8925 |*********:
     .9000 |*******  .
     .9075 |*********:
     .9150 |*******  .
     .9225 |*******  .
     .9300 |******* .
     .9375 |******* .
     .9450 |        .
     .9525 |*******.
     .9600 |*******:**
     .9675 |***** .
     .9750 |******:
     .9825 |*****:****
     .9900 |*****:****
     .9975 |****:*******************************************
    1.0050 |   .
    1.0125 |   .
    1.0200 |  .
    1.0275 |  .
           I....+....I....+....I....+....I....+....I....+....I
           0         4         8         12        16        20
                          Histogram Frequency

  .909      Std Err        .008      Median        .914
 1.000      Std Dev        .077      Variance      .006
 -.157      S E Kurt       .478      Skewness     -.617
  .241      Range          .321      Minimum       .679
 1.000      Sum          90.897
```

Figure S2

```
   PROBABILITY OF REMAINING IN GRADE 2
unt   Midpoint
 1     -.0025 |***:*
 0      .0050 |    .
 1      .0125 |*****.
 0      .0200 |      .
 3      .0275 |******:********
 5      .0350 |*******:*****************
 3      .0425 |********:******
 6      .0500 |*********:********************
 4      .0575 |***********:********
 4      .0650 |************:*******
 3      .0725 |*************:*
 2      .0800 |**********    .
 2      .0875 |**********     .
 0      .0950 |               .
 4      .1025 |****************:***
 4      .1100 |*****************:**
 5      .1175 |******************:******
 1      .1250 |*****           .
 3      .1325 |***************  .
 3      .1400 |***************  .
 4      .1475 |******************:*
 2      .1550 |**********        .
 4      .1625 |*****************:**
 4      .1700 |****************:***
 2      .1775 |**********       .
 2      .1850 |**********      .
 5      .1925 |**************:**********
 4      .2000 |*************:******
 5      .2075 |************:************
 3      .2150 |**********:****
 3      .2225 |*********:*****
 0      .2300 |         .
 0      .2375 |        .
 0      .2450 |       .
 2      .2525 |*****:****
 0      .2600 |     .
 0      .2675 |     .
 0      .2750 |    .
 1      .2825 |**:**
 0      .2900 |   .
 1      .2975 |*:***
 0      .3050 |  .
 1      .3125 |*:***
 0      .3200 |.
 1      .3275 |:****
 1      .3350 |:****
 0      .3425 |.
 0      .3500 |
 0      .3575 |
 1      .3650 |*****
 0      .3725 |
              I....+....I....+....I....+....I....+....I....+....I
              0         2         4         6         8         10
                        Histogram Frequency

         .136      Std Err         .008     Median          .132
         .167      Std Dev         .079     Variance        .006
e       -.006      S E Kurt        .478     Skewness        .664
```

PROBABILITY OF MOVING FROM GRADE 2 TO 3

Figure S3

```
  Midpoint
       .50 |
       .51 |
       .52 |
       .53 |
       .54 |.
       .55 |:*********
       .56 |:****
       .57 |:*********
       .58 | .
       .59 |*:********
       .60 |  .
       .61 |***:*
       .62 |    .
       .63 |*****.
       .64 |      .
       .65 |*****  .
       .66 |        .
       .67 |          .
       .68 |********** .
       .69 |*************:***********
       .70 |**********    .
       .71 |**********      .
       .72 |*****            .
       .73 |**********        .
       .74 |*******************:*****
       .75 |********************.
       .76 |**********           .
       .77 |*********************:***
       .78 |********************  .
       .79 |*********************:********
       .80 |*********************:******************
       .81 |********************:****
       .82 |*******************:*****
       .83 |***************   .
       .84 |*****************:*****************
       .85 |***************.
       .86 |**********    .
       .87 |************:**********************
       .88 |********** .
       .89 |*********:
       .90 |********:*
       .91 |       .
       .92 |*****:*********
       .93 |    .
       .94 |   .
       .95 |  .
       .96 |**:**
       .97 | .
       .98 | .
       .99 |.
      1.00 |.
           I....+....I....+....I....+....I....+....I....+....I
           0         2         4         6         8        10
                        Histogram Frequency
```

.778	Std Err	.009	Median	.795
.845	Std Dev	.089	Variance	.008
.421	S E Kurt	.478	Skewness	-.767
.241	Range	.418	Minimum	.545
.963	Sum	77.848		

es 100 Missing Cases 0

PROBABILITY OF REMAINING IN GRADE 3

Figure S4

```
unt   Midpoint
  0      -.03  |     .
  0      -.02  |     .
  0      -.01  |       .
  2      -.00  |****:*****
  2       .01  |****:*****
  0       .02  |          .
  1       .03  |*****      .
  1       .04  |*****       .
  3       .05  |********:******
  3       .06  |*********:*****
  2       .07  |**********    .
  3       .08  |************:**
  5       .09  |*************:***********
  4       .10  |**************:*****
  5       .11  |***************:*********
  5       .12  |****************:********
  5       .13  |*****************:*******
  3       .14  |***************    .
  1       .15  |*****              .
  7       .16  |******************:****************
  4       .17  |*******************:
  9       .18  |******************:**************************
  2       .19  |**********         .
  4       .20  |******************:*
  1       .21  |*****              .
  2       .22  |**********         .
  3       .23  |***************  .
  3       .24  |***************.
  1       .25  |*****           .
  2       .26  |**********     .
  1       .27  |*****         .
  1       .28  |*****       .
  3       .29  |*********:*****
  2       .30  |********:*
  0       .31  |        .
  0       .32  |       .
  0       .33  |      .
  1       .34  |****:
  1       .35  |***:*
  0       .36  |    .
  1       .37  |**:**
  1       .38  |**:**
  1       .39  |*:***
  1       .40  |*:***
  1       .41  |:****
  1       .42  |:****
  0       .43  |.
  0       .44  |.
  2       .45  |:*********
  0       .46  |
  0       .47  |
               I....+....I....+....I....+....I....+....I....+....I
               0         2         4         6         8         10
                         Histogram Frequency
```

APPENDIX D

Computer Programmes

APPENDIX D

```
          PROGRAM SIMUL
    C
    C     THIS PROGRAM SIMULATES  THE RANDOM  BEHAVIOUR OF A 3-GRADE
    C     SYSTEM IN WHICH THE TOTAL SIZE (DENOTED BELOW BY U(T))IS
    C     ALSO EXPANDING AT A RANDOM RATE.
    C     THE RANDOM NUMBER GENERATOR USED IS RN55 WHICH IS A DOUBLE
    C     PRECISION FUNCTION THAT RETURNS A RANDOM NUMBER X FROM THE
    C     UNIFORM DISTRIBUTION (0,1).
    C     THE PROGRAM PRINTS THE NUMBER OF INDIVIDUALS IN EACH GRADE
    C     AT TIME T,THE TOTAL SIZE OF THE SYSTEM AND THE RANDOM RATE
    C     OF EXPANSION.  IT ALSO PRINTS THE PROPORTION OF INDIVIDUALS
    C     IN EACH GRADE.
    C
          IMPLICIT DOUBLE PRECISION (A-H),(O-Z)
          INTEGER N
          INTEGER IFAIL
          DIMENSION X(23),Y(23),RESULT(20)
          DIMENSION O(0:50),U(50),SUMN(2:3,0:50),H(50),R(9000)
          DIMENSION M(2:3,9000),S(2:3,9000)
          DOUBLE PRECISION RN55,FAC1,FAC2
          ISEED=INT(SECNDS(0.0))
          CALL RNSD(ISEED)
    C     RN55 IS INITIALIZED BY THIS CALL TO THE SUBROUTINE RNSD.
          READ *,O(0),SUMN(2,0),SUMN(3,0)
    C     THESE ARE RESPECTIVELY THE INITIAL SIZES OF GRADES 1,2,3.
          READ *,P12,P22,P23,P33
    C     THESE ARE THE TRANSITION PROBABILITIES OF MOVEMENTS BETWEEN
    C     THE GRADES. NO REPETITION IS ASSUMED IN GRADE 1.
          READ *,FAC1,FAC2
    C     THE SYSTEM IS ASSUMED TO EXPAND RANDOMLY BETWEEN A GIVEN
    C     RANGE SAY ALPHA 1 TO ALPHA 2. FAC1 = ALPHA 2 - ALPHA 1 AND
    C     FAC2 = ALPHA 1.
          W1= 1.0 - P12
          W2= 1.0 - P22 - P23
          W3= 1.0 - P33
          U(0)= O(0) + SUMN(2,0) + SUMN(3,0)
          READ *,T
          READ *,NLOOP
    C     IN THE PROCEDURE BELOW S(R,I) TAKES A VALUE OF 1 WHEN AN
    C     INDIVIDUAL KNOWN TO BE IN GRADE R-1 AT TIME T-1 (R=2,3)
    C     MOVES TO GRADE R AT TIME T.
    C     M(R,I) TAKES A VALUE OF 1 WHEN AN INDIVIDUAL KNOWN TO BE
    C     IN GRADE R AT TIME T-1 REMAINS IN GRADE R AT TIME T.
          DO 200 LOOP=1,NLOOP
          WRITE(6,201)LOOP
    201   FORMAT(3X,'THIS IS SIMULATION NUMBER ',I3)
          DO 100 J=1,T
                SUMN(2,J)=0
                SUMN(3,J)=0
                DO 50 I=1,O(J-1)
                      R(I)=RN55()
                            IF (R(I).LT.P12) THEN
                            S(2,I)=1
```

```
                        SUMN(2,J)=SUMN(2,J) + S(2,I)
                        END IF
50          CONTINUE
            DO 60 K=1,SUMN(2,J-1)
                  R(K)=RN55()
                        IF (R(K).LT.P22) THEN
                        M(2,K)=1
                        SUMN(2,J)=SUMN(2,J) + M(2,K)
                        END IF
                        IF(R(K).GE.P22.AND.R(K).LE.(P22+P23)) THEN
                        S(3,K)=1
                        SUMN(3,J)=SUMN(3,J) +S(3,K)
                        END IF
60          CONTINUE
            DO 70 L=1,SUMN(3,J-1)
                  R(L)=RN55()
                        IF (R(L).LT.P33) THEN
                        M(3,L)=1
                        SUMN(3,J)=SUMN(3,J) + M(3,L)
                        END IF
70          CONTINUE
            H(J)=(RN55()*FAC1 + FAC2
C           H(J) IS THE RANDOM NUMBER WHICH REPRESENTS THE RATE OF
C           EXPANSION.  IT IS RESCALED HERE IN ORDER TO LIE
C           BETWEEN THE PRE-DEFINED UPPER AND LOWER LIMITS
            U(J)=H(J)*U(J-1) + U(J-1)
            O(J)=U(J)-SUMN(2,J)-SUMN(3,J)
100   CONTINUE
      DO 300 J=O,T,4
            WRITE(6,211)
211         FORMAT(1HO,T3,'GRADES')
            WRITE(6,212)J
212         FORMAT(1H+,T18,'T='I3)
            WRITE(6,213)J+1
213         FORMAT(1H+,T30,'T='I3)
            WRITE(6,214)J+2
214         FORMAT(1H+,T42,'T='I3)
            WRITE(6,215)J+3
215         FORMAT(1H+,T53,'T='I3)
            WRITE(6,316)1,O(J),O(J+1),O(J+2),O(J+3)
316         FORMAT(1HO,I5,T13,4F12.2)
      WRITE(6,317)2,SUMN(2,J),SUMN(2,J+1),SUMN(2,J+2),SUMN(2,J+3)
317         FORMAT(1H ,I5,T13,4F12.2)
      WRITE(6,317)3,SUMN(3,J),SUMN(3,J+1),SUMN(3,J+2),SUMN(3,J+3)
            WRITE(6,218)
218         FORMAT(1H ,T5,'U(T)')
            WRITE(6,319) U(J),U(J+1),U(J+2),U(J+3)
319         FORMAT(1H+,T13,4F12.3)
            WRITE(6,220)
220         FORMAT(1H ,T5,'A(T)')
            WRITE(6,319) H(J),H(J+1),H(J+2),H(J+3)
300   CONTINUE
      DO 10 J=1,11
      X(J)=U(J-1)
      Y(J)=U(J)
10    CONTINUE
```

```
      IFAIL=0
      CALL G02CBF(11,X,Y,RESULT,IFAIL)
      WRITE(6,33)RESULT(6)
33    FORMAT(3X,'ESTIMATE OF 1+A USING LEAST SQUARES= 'F7.5)
      OPEN(UNIT=25,FILE='LP.DAT',STATUS='NEW',
      4CARRIAGECONTROL='LIST')
      WRITE(UNIT=25,FMT=2)
2     FORMAT('SLPMODELS SIMUL DAT;')
      WRITE(UNIT=25,FMT=3)
3     FORMAT('SINDICESS I,J,T;')
      WRITE(UNIT=25,FMT=4)
4     FORMAT('SDATAIDSS Y(1:J:3,1:T:12);')
      WRITE(UNIT=25,FMT=6)
6     FORMAT('SVARIDSS Z(1:J:3,2:T:12) : REALNONNEG,')
      WRITE(UNIT=25,FMT=7)
7     FORMAT(9X,'P(1:I:3,1:J:3) : REALNONNEG;')
      WRITE(UNIT=25,FMT=8)
8     FORMAT('SROWSS FUNC: -SIGMA(J,1,3,
      6SIGMA(T,2,12,2*Y(J,T)*Z(J,T))) SMINS,')
      WRITE(UNIT=25,FMT=9)1.0/RESULT(6)
9     FORMAT(5X,'ZD(1:J:3,2:T:12) : Z(J,T)SEQS
      2SIGMA(I,1,3,Y(I,T-1)*',F7.4,'*P(I,J)),')
      WRITE(UNIT=25,FMT=12)
12    FORMAT(5X,'PS(1:I:3) : SIGMA(J,1,3,P(I,J)) SLEQS 1,')
      WRITE(UNIT=25,FMT=23)
23    FORMAT(5X,'ND(2:I:3) : P(I,I-1) SEQS 0,')
      WRITE(UNIT=25,FMT=14)
14    FORMAT(5X,'EX(3:I:3) : P(I,I-2) SEQS 0,')
      WRITE(UNIT=25,FMT=15)
15    FORMAT(5X,'RP(1:I:1) = P(I,I) SEQS 0,')
      WRITE(UNIT=25,FMT=16)
16    FORMAT(5X,'JM(1:I:1) : P(I,I+2) SEQS 0;')
      WRITE(UNIT=25,FMT=17)
17    FORMAT('SEOMS')
```

```
      WRITE(UNIT=25,FMT=18)
18    FORMAT('SLPOATAS SIMUL DAT;')
      WRITE(UNIT=25,FMT=19)(O(J)/U(J),J=0,8)
19    FORMAT(3X,'Y(1,T) +',F5.4,' +',F5.4,' +',F5.4,' +',
      3F5.4,' +',F5.4,' +',F5.4,' +',F5.4,' +',F5.4,' +',F5.4)
      WRITE(UNIT=25,FMT=21)(O(J)/U(J),J=9,11)
21    FORMAT(10X,'+',F5.4,' +',F5.4,' +',F5.4,',')
      WRITE(UNIT=25,FMT=24)(SUMN(2,J)/U(J),J=0,8)
24    FORMAT(3X,'Y(2,T) +'.F5.4,' +',F5.4,' +',F5.4,' +',
      6F5.4,' +',F5.4,' +',F5.4,' +',F5.4,' +',F5.4,' +',F5.4)
      WRITE(UNIT=25,FMT=25)(SUMN(2,J)/U(J),J=9,11)
25    FORMAT(10X,'+',F5.4,' +',F5.4,' +',F5.4,',')
      WRITE(UNIT=25,FMT=27)(SUMN(3,J)/U(J),J=0,8
27    FORMAT(3X,'Y(3,T) +',F5.4,' +',F5.4,' +',F5.4,' +',
      2F5.4,' +',F5.4,' +',F5.4,' +',F5.4,' +',F5.4,' +',F5.4)
      WRITE(UNIT=25,FMT=28)(SUMN(3,J)/U(J),J=9,11)
28    FORMAT(10X,'+',F5.4,' +',F5.4,' +',F5.4,';')
      WRITE(UNIT=25,FMT=31)
31    FORMAT('SEODS')
      WRITE(UNIT=25,FMT=32)
32    FORMAT('"PICTURE","A","P",/')
      CLOSE(UNIT=25,STATUS='KEEP')
200   CONTINUE
      STOP
      END
```

```
      PROGRAM PREDICT
C
C     THIS PROGRAM CALCULATES THE EXPECTED STRUCTURE OF A
C     3-GRADE SYSTEM WHICH IS EXPANDING.  ESTIMATES OF THE
C     EXPANSION RATE AND THE TRANSITION PROBABILITIES MUST
C     BE READ INTO THE PROGRAM.  THE PROGRAM ALSO GIVES FOR
C     EACH TIME PERIOD, THE VARIANCES AND COVARIANCES OF THE
C     PREDICTIONS.  COVIJ(T) DENOTES THE COVARIANCE BETWEEN
C     THE PREDICTED VALUES N(I,T) and N(J,T).
C
      REAL N1, N2, N3
      DIMENSION N1(0:28,100),N2(0:28,100), N3 (0:28,100),P12(100)
      DIMENSION P22(100),P23(100),P33(100),A(100),U(0:28,100)
      DIMENSION COV11(0:28,100),COV33(0:28,100),COV12(0:28,100)
      DIMENSION
      COV22(0:28,100),COV23(0:28,100),W1(100),W2(100),W3(100)
      DIMENSION
      SN1(28),SN2(28),SN3(28),SU(28),SCOV33(28),COV13(0:28,100)
      DIMENSION
      SCOV11(28),SCOV12(28),SCOV13(28),SCOV22(28),SCOV23(28)
      READ *, N1(0,1),N2(0,1),N3(0,1)
      READ *,TS
      DO 6 J=1,TS
            N1(0,J)=N1(0,1)
            N2(0,J)=N2(0,1)
            N3(0,J)=N3(0,1)
            U(0,J)=N1(0,J) + N2(0,J) + N3(0,J)
            COV11(0,J)=0
            COV12(0,J)=0
            COV13(0,J)=0
            COV22(0,J)=0
            COV23(0,J)=0
            COV33(0,J)=0
6     CONTINUE
      DO 66 J=1,TS
            READ *,P12(J),P22(J),P23(J),P33(J),A(J)
66    CONTINUE
      DO 88 L=1,28,9
            SN1(L)=0
            SN2(L)=0
            SN3(L)=0
            SU(L)=0
            SCOV11(L)=0
            SCOV12(L)=0
            SCOV13(L)=0
            SCOV22(L)=0
            SCOV23(L)=0
            SCOV33(L)=0
            DO 77 J=1,TS
                  W1(J)=1.0 - P12(J)
                  W2(J)=1.0 - P22(J) - P23(J)
                  W3(J)=1.0 - P33(J)
                  DO 50 I=1,2,88
                  U(I,J)=A(J)*U(I-1,J) + U(I-1,J)
                  N3(I,J)=P33(J)*N3(I-1,J) + P23(J)*N2(I-1,J)
```

```
                        N2(I,J)=P22(J)*N2(I-1,J) + P12(J)*N1(I-1,J)
                        N1(I,J)=U(I,J)-N2(I,J)-N3(I,J)
      COV11(I,J)=COV11(I-1,J)*W1(J)*W1(J)+COV12(I-1,J)*W1(J)*W2(J)*2.0 +
            2COV13(I-1,J)*W1(J)*W3(J)*2.0 + COV22(I-1,J)*W2(J)*W2(J) +
            7COV23(I-1,J)*2.0*W3(J)*W2(J) +
            3COV33(I-1,J)*W3(J)*W3(J) + (W1(J)-W1(J)*W1(J))*N1(I-1,J) +
            4(W2(J)-W2(J)*W2(J))*N2(I-1,J)+(W3(J)-W3(J)*W3(J))*N3(I-1,J)
      COV12(I,J)=COV11(I-1,J)*W1(J)*P12(J)COV12(I-1,J)*W1(J)*P22(J) +
            5COV12(I-1,J)*W2(J)*P12(J) + COV22(I-1,J)*W2(J)*P22(J) +
            2COV13(I-1,J)*W3(J)*P12(J) + COV23(I-1,J)*W3(J)*P22(J) -
            6W1(J)*P12(J)*N1(I-1,J) - W2(J)*P22(J)*N2(I-1,J)
      COV13(I,J)=COV12(I-1,J)*W1(J)*P23(J)COV13(I-1,J)*W1(J)*P33(J) +
            7COV22(I-1,J)*W2(J)*P23(J) + COV23(I-1,J)*W2(J)*P33(J) +
            2COV23(I-1,J)*W3(J)*P23(J) + COV33(I-1,J)*W3(J)*P33(J) -
            9W2(J)*P23(J)*N2(I-1,J) - W3(J)*P33(J)*N3(I-1,J)
COV22(I,J)=COV11(I-1,J)*P12(J)*P12(J)+COV12(I-1,J)*P12(J)*P22(J)*2.0
9+COV22(I-1,J)*P22(J)*P22(J)+(P12(J)-P12(J)*P12(J))*N1(I-1,J) +
            1(P22(J)-P22(J)*P22(J))*N2(I-1,J)
            COV23(I,J)=COV12(I-1,J)*P12(J)*P23(J)
            +COV13(I-1,J)*P12(J)*P33(J) +
            2COV22(I-1,J)*P22(J)*P23(J) + COV23(I-1,J)*P22(J)*P33(J) -
            3P22(J)*P23(J)*N2(I-1,J)
            COV33(I,J)=COV22(I-1,J)*P23(J)*P23(J)
            +2.0*COV23(I-1,J)*P33(J)*P23(J)
            4+ COV33(I-1,J)*P33(J)*P33(J) +
            P33(J)-P33(J)*P33(J))*N3(I-1,J)+
            5(P23(J)-P23(J)*P23(J))*N2(I-1,J)
      50    CONTINUE
            SN1(L)=SN1(L) + N1(L,J)
            SN2(L)=SN2(L) + N2(L,J)
            SN3(L)=SN3(L) + N3(L,J)
            SU(L)=SU(L) + U(L,J)
            SCOV11(L)=SCOV11(L) + COV11(L,J)
            SCOV12(L)=SCOV12(L) + COV12(L,J)
            SCOV13(L)=SCOV13(L) + COV13(L,J)
            SCOV22(L)=SCOV22(L) + COV22(L,J)
            SCOV23(L)=SCOV23(L) + COV23(L,J)
            SCOV33(L)=SCOV33(L) + COV33(L,J)
      77    CONTINUE
            PRINT *,'T=',L
            PRINT *,SCOV11(L)/TS,SCOV12(L)/TS,SCOV13(L)/TS
            PRINT *,'               ',SCOV22(L)/TS,SCOV23(L)/TS
            PRINT *,'                             ,SCOV33(L)/TS
      88    CONTINUE
            WRITE(6,211)
      211   FORMAT (1HO,T3,'GRADES')
            WRITE (6,212)1
      212   FORMAT (1H+,T18,'T='12)
            WRITE (6,213) 10
      213   FORMAT (1H+,T30,'T='12)
            WRITE(6,214) 19
      214   FORMAT (1H+,T42,'T='12)
            WRITE (6,215) 28
      215   FORMAT (1H+, T54,'T='12)
            WRITE(6,216)1,SN1(1)/TS,SN1(10)/TS,SN1(19)/TS,SN1(28)/TS
      216   FORMAT(1HO,15,T13,4F12.2)
```

```
      WRITE(6,217)2,SN2(1)/TS,SN2(10)/TS,SN2(19)/TS,SN2(28)/TS
217   FORMAT(1H ,I5,T13,4F12.2)
      WRITE(6,217)3,SN3(1)/TS,SN3(10)/TS,SN3(19)/TS,SN3(28)/TS
      WRITE(6,218)
218   FORMAT(1H ,T5,'U(T)')
      WRITE(6,219)SU(1)/TS,SU(10)/TS,SU(19)/TS,SU(28)/TS
219   FORMAT(1H+,T13,4F12.2)
      STOP
      END
```

REFERENCES

Abdellauoi, G.M. (1985). 'Stochastic control in manpower planning'. Ph.D. Thesis, University of London.

Abdellauoi, G.M. (1987). 'Probability of maintaining or attaining a structure in one step'. *J.Appl.Prob.*, 24.

Ali, M.H. (1986). 'Manpower Planning and development potentials: test case of Sudan'. Ph.D. Thesis, University of London.

Anderson, T.W., and L.A. Goodman (1957). 'Statistical inference about Markov chains'. *Ann.Math.Statist.*, 28, 89-109.

Armitage, P.H., C.S. Smith, and P. Alper (1969). *Decision Models for Educational Planning*, Allen Lane, The Penguin Press, London.

Armitage, P.H., C.M. Phillips, and J. Davies (1970). 'Towards a model of the upper secondary school system'. *J.R.Statist.Soc.*, A133, 166-205.

Balinsky, W., and A. Reisman (1972). 'Some manpower planning models based on levels of educational attainment'. *Man.Sci.*, 18, B691-B705.

Balinsky, W., and A. Reisman (1973). 'A taxonomy of manpower-educational planning models'. *Socio-econ.Plann.Sc.*, 7, 13-18.

Bartholomew, D.J. (1971). 'The statistical approach to manpower planning'. *Statistician*, 20, 3-26.

Bartholomew, D.J. (1976). 'Renewal theory models in manpower planning'. *Symposium Proceedings Series No. 8. The Inst. of Math. and its Applications*. 153, 57-73.

Bartholomew, D.J. (1977a). 'The analysis of data arising from stochastic processes'. Chapter 5 in C.A. O'Muircheartaigh and C. Payne(Eds.) *The Analysis of Survey Data*, Vol.II, John Wiley, Chichester.

Bartholomew, D.J. (1977b). 'Maintaining a grade or age structure in a stochastic enviroment'. *Adv.Appl.Prob*,9,1-17.

Bartholomew, D.J. and A. F. Forbes (1979). *Statistical Techniques for Manpower Planning*, John Wiley, Chichester.

Bartholomew, D.J. (1982). *Stochastic Models for Social Processes*,3rd Ed., John Wiley, Chichester.

Bartholomew, D.J. (1983). 'Some recent developments in social statistics'. *International.Stat.*Revue, Vol.51.

Bartholomew, D.J. (1984). 'Recent developments in nonlinear stochastic modelling of social processes'. *Canadian J. of Statis.*12,No.1,39-52.

Bhat, U.N. (1984). *Elements of Applied Stochastic Processes*, 2nd Ed., Wiley-Interscience, John Wiley & Sons, New York.

Britney, R.R. (1975). 'Forecasting educational enrollments: comparison of a Markov chain and dircuitless flow netweork model'. *Socio-Econ.Plan.Sci.*,9,53-60.

Clough, D.J. and W.P. MacReynolds (1966). 'State transition model of an educational sytem incorporating a constraint theory of supply and demand'. *Ontario J. of Educational Research*,9,1-18.

Collins, L.(1974). 'Estimating Morkov transition probabilities from micro-unit data'. *Appl.Statist.*,23,355-370.

Conlisk, J. (1976). 'Interactive Markov chains'. *J.Math.Sodiology*, **4**,157-185.

Conlisk, J. (1978). 'A stability theorem for an interactive Markov chain'. *J.Math.Sociology*,**6**,163-168.

De Stavola, B. (1988). 'Testing departures from time homogeneity in multistate Markov processes'. *Appl.Statist.*,**37**,242-250.

Davies, G.S. (1973). 'Structural control in a graded manpower system'. *Man.Sci.*,**20**,76-84.

Davies, G.S. (1982). 'Control of grade sizes in a partially stochastic Markov manpower model'. *J.Applied Prob.*,**19**,439-443.

Davies, G.S. (1983). 'A note on the geometric/probabilistic relationship in a Markov manpower model'. *J.Applied Prob.*,**20**,423-428.

Davies, G.S. (1985). 'A note on a continuous-time Markov manpower model'. *J.Applied Prob.*,**22**,932-938.

Forbes, A.F. (1971). 'Promotions and recruitment policies for the control of quasi-stationary hierachial systems'. In A.R. Smith (Ed.)(1971) *Models for Manpower Systems*, English University Press, London, 401-414.

Frydman, H., J.G. Kallberg and D. Kao (1985). 'Testing the adequacy of Markov chain and mover-stayer models as representations of credit behaviour'. *Operations Research*,**33**,1203-1214.

Gani, J. (1963). 'Formulae for projecting enrollments and degrees awarded in universities'. *J.R.Statist.Soc.*,**A126**,400-409.

Gupta, B.R. (1986). 'Stochastic models of brand choice behaviour'.

Ph.D. Thesis, University of London.

Haigh, J. (1983). 'Maintability of manpower structres; counterexamples, results and conjectures'. *J.Applied Prob.*,**20**.

Hassani, H. (1980). 'Semi-Markov models in manpower planning'. Ph.D. Thesis, University of London.

Janssen, J. (1986)(Ed.). *Semi-Markov models: Theory and Applications*.

Kalamatianou, A.G. (1984). 'Generalized Markovian manpower models: theory and applications'. Ph.D. Thesis, University of London.

Kalamatianou, A.G. (1988). 'A model for responding to promotion blockages'. *J.Applied Prob.*,**25**,268-278.

Kalbfleisch, J.D. and J.F. Lawless (1984). 'Least-squares estimation of transition probabilities from aggregate data'. *Canadian J.of Statist.*,**12**,169-182.

Kamat, A.R. (1968). 'Estimating wastage in a course of education'. *Sankhya*,**B30**,5-12.

Kelton, C.M. (1981). 'Estimation of time-independent Markov processes with aggregate data: a comparison of techniques'. *Econometrica*,**49**,517-518.

Kelton, C.M. and D.W. Kelton (1982). 'Advertising and intra-industry brand shift in the U.S. brewing industry'. *J.of Industrial Economics*,**30**,293-303.

Kelton, C.M. and D.W. Kelton (1984). 'Hypothesis tests for Markov process models estimated form aggregate frequency data'. *J. of the American Statist.Assoc.*,**79**,923-928.

Lawless, J.F. and D.L. McLeish (1984). 'The information in aggregate data in Markov chains'. *Biometrika*,**71**,419-430.

Leeflang, P.S. (1974). *Mathematical Models in Marketing*, H.E. Stenfert Kroese B.B., Leiden.

Leeson, G.W. (1980). 'A projection model for hierarchial manpower systems'. *J.Operat.Res.Soc.*,**31**,247-256.

Massy, W., D. Montgomery and D. Morrison (1970). *Stochastic Models of Buying Behavior*. Cambridge, Mass.: The MIT Press.

MacRae, E.C. (1977). 'Estimation of time-varying Markov processes with aggregate data'. *Econometrica*,**45**,183-198.

Moya-Angeler, J. (1976). 'A model with shortages of places for educational and manpower systems'. *Omega*,**4**,719-730.

McCarthy, C. and T. Ryan (1977). 'Estimates of voter transition probabilities from the British general elections of 1974'. *J.R.Statist.Soc.*,**A140**,Part 1.

McClean, S.I. (1980). 'A semi-Markov model for a multi-grade population with Poisson recruitment'. *J.Applied Prob.*,**17**,846-852.

McLeish, D.L. (1984). 'Estimation for aggregate models: The aggregate Markov chain'. *Canadian J. of Statist.*,**12**,265-282.

Plewis, I. (1981). 'Using longitudinal data to model teachers' ratings of classroom behaviour as a dynamic process'. *J.Educ.Statist.*,**6**,237- 255.

Rosenqvist, G. (1986). 'Micro and Macro data in statistical inference on Markov chains'. Ph.D. Thesis, Swedish School of Economics and Business Administration.

Sales, P. (1971). 'The validity of the Markov chain model for a branch of the Civil Service'. *Statistician*, 20,85-110.

Seal, H.L. (1945). 'The mathematics of a population composed of k stationary strata each recruited from the stratulm below and supported at the lowest level by a uniform annual number of entrants'. *Biometrika*, 33,226-230.

Sherif, F. and M.E. Thompson (1980). 'A brand share model for Canadian cigarettes'. *Canadian J. of Statist.*, 8,205-215.

Stone, R. (1965). 'A model of the educational system'. *Minervia*, 3,172-186.

Stone, R. (1972). 'A Markovian educational model and other examples linking social behaviour to the economy'. *J.R.Statist.Soc.*, A135,511-543.

Thompson, M.E. (1981). 'Estimation of the parameters of a semi-Markov process for censored records'. *Adv.Applied.Prob.*, 13,804-825.

Thonstad, T. (1969). *Education and Manpower: Theoretical Models and Empirical Applications*, Oliver and Boyd, Edinburgh and London.

Thorburn, D. (1982). 'Estimating transition matrices from aggregate stock data'. Department of Mathematical Statistics, University of Lund and Lund Institute of Technology.

Tuma, N.B., M.T. Hannan and L.P. Groenveld (1979). 'Dynamic analysis of event histories'. *American J. of Sociology*, 84,820-854.

Tziafetas, G. (1980). 'Micro and macro Bayesian estimation of population distributions.' *Biometrical Journal*, 22,583-591.

Vajda, S. (1947). 'The stratified semi-stationary population'.

Biometrika,**34**,243-254.

Vajda, S. (1975). 'Mathematical aspects of manpower planning'. *Operational Research Quart.*,**26**,527-542.

Van der Plas (1983). 'On the estimation of the parameters of Markov probability models using macro data'. *Annals of Statist.*,**11**,78-85.

Vassiliou, P.C. (1981). 'On the limiting behavior of a non-homogenous Markov chain model in manpower systems'.*Biometrika*,**68**,557-561.

Vassiliou, P.C. (1982). 'On the limiting behavior of a non-homogenous Markovian manpower model with independent Poisson input'. *J.Applied Prob.*,**19**,433-438.

Vassiliou, P.C. and I. Gerontidis (1985). 'Variances and covariances of the grade sizes in manpower systems'. *J.Applied Prob.*,**22**.

Wegner, T. (1985). 'Discrete simulation modelling for corporate manpower planning'. *Omega*,**13**.

Woodward, M. (1983). 'The limiting properties of population distributions with particular application to manpower planning'.*J.Applied Prob.*,**20**,19-30.

Young, A., and G. Almond (1961). 'Predicting distributions of staff'.*Comp.J.*,**3**,246-250.

Young, A., and P.C. Vassiliou (1974). 'A non-linear model on the promotion of staff'. *J.R.Statist.Soc.*,**A137**,584-595.