NBSIR 78-1458

Projecting the Age Distribution of Full-Time Permanent Professional Staff at the National Bureau of Standards

L. S. Joel

Center for Applied Mathematics
National Bureau of Standards
Washington, D.C. 20234

March 1978

Final

Prepared for
Institute for Materials Research

NBSIR 78-1458

PROJECTING THE AGE DISTRIBUTION OF FULL-TIME PERMANENT PROFESSIONAL STAFF AT THE NATIONAL BUREAU OF STANDARDS

L. S. Joel

Center for Applied Mathematics
National Bureau of Standards
Washington, D C. 20234

March 1978

Final

Prepared for
Institute for Materials Research

U.S. DEPARTMENT OF COMMERCE, Juanita M. Kreps, *Secretary*

Dr. Sidney Harman, *Under Secretary*

Jordan J. Baruch, *Assistant Secretary for Science and Technology*

NATIONAL BUREAU OF STANDARDS, Ernest Ambler, *Director*

ACKNOWLEDGEMENT

This study was undertaken in response to suggestions from within the Institute for Materials Research (IMR) of the National Bureau of Standards, and was partially funded by that Institute.

The author wishes to acknowledge the enthusiastic and cogent assistance of members of the Personnel Division, the IMR Office, the Office of the Associate Director for Programs (OADP), the Office of Standard Reference Data (OSRD) and the Office of the Management and Organization Division. Particular thanks are due to I. Bartky of IMR and to E. Bunten and S. Newman of OADP for many insights, B. Conrad of Personnel and B. Molino of OSRD for help in retrieval and interpretation of personnel data, C. Johnson of the Applied Mathematics Division (AMD) for technical discussions relating to Markov models, to J. Lechner and P. B. Saunders of AMD who improved the exposition and finally to A. J. Goldman of AMD who furnished an appendix and was a general technical factotum. The residual flaws are all mine.

Lambert S. Joel

ABSTRACT

This report presents a simple mathematical model to project the age distribution of the full-time permanent professional (FTPP) staff of the National Bureau of Standards. The report includes a brief description of types of models currently in use for manpower analyses, discussion of the probable data requirements for reliable models, some staff profile information which supplements material in recent administrative reports, and the description of our model. The principal projection is that under "status quo" assumptions (FTPP staff size, age distribution of hires, and the separation rates from age cohorts remain constant), the FTPP staff will in about 25 years reach a steady-state age distribution with average age about 1/2 year higher than its current level (just over 42 yrs.). Comparison of this distribution with the present one shows moderate increases in the fractions of staff in ages 21-30 and 51-60, a fractional rise in turnover, and a general decrease in the groups 31-50 years old. A near term effect is the intensified loss of senior scientists.

Keywords: Data analysis; discrete Markov models; manpower planning models; simple Markov models

TABLE OF CONTENTS

Page

1. Introduction and Overview 1

   1.1 Summary 1

   1.2 Related Literature and Analysis Methods 2

   1.3 Study of Age Distributions 4

   1.4 Concluding Remarks & Possible Extensions 11

2. Data 15

3. The Model 20

   3.1 Model Concepts 20

   3.2 Discussion of Concepts 23

4. The Working Model and Its Output 30

5. Bibliography 40

Appendix A. The Model in Matrix Form 43

Appendix B. Analysis of the Markov Model 55

   B.1 Formulas for Steady State Quantities 55

   B.2 Attainability of "Target" Distributions 59

LIST OF TABLES AND FIGURES

Table 1. Hires & Separations, Jan. 1, 1970 - Jan. 31, 1976.............18

Table 2. Means and Median Ages of Hires & Separations.................18

Table 3. Relevant Data and Model Parameters...........................31

Table 4. Projected Changes in Age Class Sizes and Distribution.........36

Table 5. Rank Orders of Age Classes, 1975 - 1990......................38

Table 6. Selected Values from Tables 3 and 4..........................57

Figure 1. Age-Distribution Shapes for Hires & Separations.............19

Figure 2. Median Ages of Hires & Separations, at Years & Half-Years....27

Figure 3. FTPP Staff Age Distributions, 1970 - 1974...................32

Figure 4: Composition of FTP Staff by Occupational Category,
1964 - 1974...33

1. INTRODUCTION AND OVERVIEW

1.1 Summary

This report describes a simple model to project, under "status quo" assumptions, the age distribution of the full-time permanent professional (FTPP) staff of the National Bureau of Standards. That particular topic and the vehicle chosen for its analysis, a "Discrete Time Markov Model", may be regarded as paradigms for investigating the prospective usefulness of mathematical models in manpower planning at NBS. The report includes a cursory description of types of models currently in use for manpower analyses, discussion of the probable data requirements for reliable models, some staff profile information (derived from personnel tape records) which supplements material in recent administrative staff reports, and the description of our model and its output. Facets of the technical analysis are included in the main text when they are vital to the continuity of the discussion, but in general they are exiled to an appendix.

The salient output from the Markov model is that the FTPP staff will arrive at a steady state in about 25 years with an average age about 1/2 year higher than the current average age, moderate increases in the fractions of staff aged 21-30 and 51-60, a general decrease in groups 31-50 years old, and a small increase in turnover, <u>if</u>

1. The size of the FTPP staff remains constant.
2. The age distribution of separations and the age distribution of hires for FTPP hereafter remain constant at their mean levels during the period JAN 1970 - JAN 1976.

The model's outputs project accelerated loss of senior scientists during the next 10 years and a slight reduction in departures of professionals of ages 30-40.

Further specifics are given in Table 4 on p. 36.

Tabulations of the available data show that while mean and median ages of FTPP hires during 1970-1976 are high relative to conventional expectation (just under 36 and just under 34 yrs. respectively), they have held approximately level during this period and are consistent with mean and median hire ages of FTP technicians, wage board, administrative and clerical staff, all also high. This means that indicators of concentration of employees at increasing age levels are not peculiar to professionals but apply to the entire FTP staff, and may be characteristic of the entire federal establishment or conceivably the entire employed labor force of the United States.*

Analysis of the model supports the intuitively evident notion that the age distribution of staff can be controlled to the extent that the hire and separation age distributions can be controlled. The model (or minor variants) can be utilized to explore the effects of stipulated changes in these "input" distributions.

1.2 Related Literature and Analysis Methods

In general, models of manpower systems have been developed for the study of

(1) assignments to jobs and training, [12]

(2) perceived and undesired trends toward "top-heavy" rank structure in mature organizations** and

*In conversation with the author, a Bureau of Labor Statistics source opined that the Federal work force has aged subsequent to the employment boom of the 1960's, but that early retirements have had the opposite effect on the national employed labor force.

**Control of the grade structure of an organization, which depends on promotion paractices as well as on hires and separations, may serve as a method of indirect management of age distributions because of the typical correlations between rank and age, especially in systems in which seniority is a factor in promotion practice.

(3) departures one way or the other from target ratios between line and staff force sizes (or, alternatively, between professional and administrative force sizes). [1, 2, 3, 12]

Traditionally, little explicit attention has been paid to the first type of problem at NBS because at working levels this was regarded as a casual task of section chiefs which could usually be accomplished informally (and easily) using educational background as the prime criterion. Determination of job assignments using formal methods usually founders on the difficulty of defining reliable numerical "figures of merit" or scoring standards for anticipated performance. (Determining post hoc performance standards is hard enough!) Assignment problems are generally attacked with simple linear programming models (and a grain of salt), when a formal treatment is attempted. [12]

Normative (i.e., prescriptive) approaches to problems of types (2) and (3) are subject to the same kind of obstacle: it is difficult to define objective numerical criteria for target values. The models here tend to be either complicated mathematical optimization formulations with a heavy bias in favor of assumptions yielding reduction to (complicated) linear programs, or else models using the constructs of dynamic programming and control theory, derived from analogies of the manpower systems with physical engineering systems, and again characterized by ruthless assumptions of linearity to make the calculations tractable to known methods of solution [1, 12]

Four kinds of models, sometimes in combination, account for a great preponderance of the published studies in which analysts' sights are lowered, from prescription to description (and estimation, and prediction).

(A) Regression, or fitting curves to empirical data. Without prior hypotheses about the functional relationships between causes and effects, this kind of "model" amounts to fishing for "trends", i.e. linear (or sometimes log-linear) fits to data points.*

(B) "Atomic" or discrete "Monte Carlo" simulations. These involve attempts to reproduce numerically the actual behavior and decisions of individuals in an organization and to measure their aggregate consequences. Such models are usually very complicated and are characterized by extensive activity to develop inputs. Ironically, model types often classified under this rubric need not be atomic nor probabilistic, i.e. "Monte Carlo". For example, the well-known methods of J. Forrester [4] (customarily called "system dynamics" and widely used today by management analysts) involve deterministic representation of aggregate behavior.

(C) Renewal theory, primarily Integral Equation models. These have proven valuable in the study of biological population phenomena and have become popular for analysis of turnover rates, e.g., when responses to losses occurring at random points in time are of interest. [1, 12]

(D) Continuous time and Discrete time Markov process models. These are highly esteemed because they have an analysis-simplifying property: the future development of a Markov system depends on its current condition and not on its entire previous history. [1-3, 5-9, 15]

1.3 Study of Age Distributions

The previous list of three fundamental problem areas addressed by manpower analysis omits the one which is the principal topic of this report,

*"Trend fishing", despite the disdain implicit in such a designation, can be extremely useful in short term analysis.

viz. the tendency toward staff aging in organizations -- that is, the increasing concentration into higher age groups.*

The composition of an organization's working force, described in terms of its distributions into classes according to the members' characteristics: sex, age, education, rank, seniority, occupational specialty, etc., is a natural vehicle to carry measures of the organization's suitability for achieving its perceived missions. A central concern in manpower planning is the degree to which changes in composition over time can be predicted, or better, controlled, given the range of political, social and economic factors affecting them. [1, 11, 13]

At NBS, a current matter of management interest is the age distribution of the FTPP staff "component". The professional staff is perceptibly "aging"; the popular notion that professional productive output decreases with workers' ages is supported in research literature, and some investigators have found evidence that quality of output deteriorates as well. [10, 13]

Various plausible arguments have been advanced** to explain the increasing concentration of FTPP staff members in high age groups. Among these are:

(a) the termination of "blue sky" research programs that produced a young-professional recruitment explosion in the late forties and the fifties,

*Analyzing age distributions is conceptually equivalent to the analysis of rank distribution in an organizational hierarchy in which all promotions take place at regular intervals, promotions are always to the next higher rank, and no demotions can occur. While explicit models of aging are hard to find in the literature, studies of organizational rank hierarchies are common.

**Conversations between the author, and E. Bunten and S. Newman of the Office of the Associate Director for Programs.

(b) economic contractions that have diminished opportunities outside NBS for professional (indeed all) staff members, resulting in longer terms of service,

(c) a tendency for students to remain at universities longer than heretofore (for a multiplicity of reasons) before seeking first employment, and a concurrent tendency of project managers to seek "expert" recruits, both resulting in a high average age of hires; and finally

(d) the existence at NBS of certain programs involving long term research goals, leading to progressive refinement of results along narrow lines, producing a "life's work" attitude in which workers are reluctant to move to other activities or to relinquish career activities by retiring.

Accordingly, the primary object of the current study was analysis of personnel data to check and quantify the current impressions about shifts in the age distribution of FTPP, with particular emphasis on the rate at which this distribution would approach a possible steady state, given continuation of current short term trends in the components of change.

A most direct way to project the age distribution would be to trace over a period of years the number (or proportion) of staff members at each "age" (or in each of a set of age intervals) and to fit trendlines or other curves to these "time series" by regression techniques.

The results of the regressions would then be adjusted or "normalized" to conform to the assumption of constant total staff size in the case of sizes of the groups, or to the requirement of 100% totals if the regression is on age-group percentages of the total. In practice, it is desirable to group ages into at least 5-year intervals in order to reduce the number of series to be treated and to "smooth out" local variations, except in rare instances in which there are known factors distinguishing the progressions for finer division. [This might be true, for example, if we were considering all full time professional and subprofessionals including term appointees, as the population of interest. Then if the data were recorded on a fiscal year basis (i.e. staff status on July 1), and there were a recurring summer Quality High School Student employment program, it would make sense to distinguish 16-year olds or 16 and 17-year olds from, say, an "under 25" age interval.] Obviously, a cursory estimate of a trend in age can be obtained by examining and projecting a single indicator of the distributions, either the average or median age. This is conceptually equivalent to an extreme case in which the classes are aggregated into a single component.

In the context of identification of a stable age distribution, the regression scheme has a minor drawback: it does not *ipso facto* furnish a model or hypothesis. When a "trendline" is fitted to data, the regression procedure finds the slope of that line which furnishes the best fit according to standard criteria. The "model" here is the (straight) line, and the regression determines the values of the parameters of the model, in this case a slope and an intercept. If regression should show very strong (non-horizontal) straight-line fits to the age distribution data, we would have to conclude that under present conditions no stable age distribution could be predicted. (A time series for average age lying along a rising/falling straight line would, if interpreted at face value, indicate an average age rising/falling indefinitely, precluding arrival at a stable age distribution.

Similarly, non level straight trendlines for separate age class sizes would say that some are rising, others falling, indefinitely, while if the data consisted of fractions of staff in the various classes, projections of such lines would be meaningless.)

In general, if observed data values are assumed to lie on a curve of specified form, regression fixes the best-fitting curve within the confines of that form. In searching for a steady state we would assume the time series for each age component to lie along a curve with a more or less horizontal asymptote, i.e. a curve that "levels off", but further pre-analysis or imaginative examination of plotted data points is needed before the regression procedure can be really useful.

The model we have chosen to employ instead, does not entail regression as such and therefore bypasses the need to deduce beforehand an appropriate shape for the path over time of the values of the age group sizes (or fractions) in order to arrive at a stable distribution. Our model is a routine application of the mathematical theory of discrete time Markov processes, which has widespread precedent for the analysis of manpower systems. [1, 2, 3, 7, 11, 12, 15]

For projecting basic age trends in a staff of fixed size, the model requires "only" separation rates by age and an age distribution of hires, both assumed to be constant. With these assumptions the model yields a stable age distribution for our FTPP staff and an estimate of how fast it is attained from the (known) current distribution.

In our case, the "constant" age distribution of hires and age-dependent rates were derived from averages of distributions for the years 1970-1976 as extracted from an NBS personnel History File data tape. These averages were informally judged representative, hence usable in the determination of the constants, on the grounds that the annual distributions (5-year age classes) for the six years did not "seem" to vary much. Of course, we have no assurance that they will not vary in an unknown manner in the future, since they depend partially on factors not under the control of management. The same argument applies to a lesser extent to the total size of the staff. (With regard to changes in total size, however, there are relatively simple extensions of the basic method which apply to expanding or contracting manpower systems.) [1, 3, 7, 12]

Markov models in the present context are subject to an inconvenient constraint that does not occur in the curve fitting approach described earlier. There the age intervals defining classes could be arbitrary, because class sizes were being projected independent of each other (except for the "consistent total" adjustment). Here, the class boundaries must be equally spaced because the structural framework is one of flows between classes, and the core of these flows is the <u>uniform</u> biological process of aging which could not be so represented if the age intervals varied.* [3]

A third methodology, that of system simulation models, could be used to study aging and <u>simultaneously</u> many other problems of importance in organizational structure and behavior. Usually synthesized using special computer languages, these models tend to be large and complicated and so

*In addition, if the length of the model's time periods is not an integral multiple of the class age interval, the calculations can produce nonsensical results due to "asynchrony"; an example appears on pp. 25-26, below.

none was considered as a practicable candidate for the present analysis. In contrast to "renewal models" (a class of continuous-time models primarily applied to questions of population growth and mentioned again here for the sake of completeness), simulation models represent philosophically a rationale so different from curve fitting and discrete stage Markov models that although they are far beyond the scope of the present project, they are mentioned here as possible tools for an expanded study. The earlier two methods are essentially schemes for data analysis in which the main questions are: "What happened?" "What will happen?" or "What is likely to happen?" but not "Why does it happen?" Simulations, in contrast, generally are intended to incorporate all the factors, behavioral, economic, political etc. that the analyst can identify and represent numerically. The resulting model is an attempt to represent realistically the interaction of the factors or forces defining a system and to generate outputs which can be studied as substitutes for the output of real (and less observable and controllable) systems. When such models are constructed too naively, and/or with totally inadequate baseline data, they are useless chimeras overfull of irrelevant details yet missing vital "components", and formulated with mathematical absurdities. When they are devised with even moderate intelligence and used with requisite caution, they can be gold mines of insight. Frequently, in fact, models developed by research analysts purely for the purpose of facilitating exposition of technical material to managers, have uncovered significant information about the modeled system (or organiation) previously known neither to analysts nor to management. [4]

1.4. Concluding Remarks & Possible Extensions

While the Markov age model begs the question of arriving at stable age distributions by assuming constant rates of change, it does tell us where we are heading given present trends. With moderate additional computational effort, the analysis can be applied, if desired, to staff components other than FTPP.

It may be of interest to examine, through such a model, the effects on staff age patterns of policy-induced changes in the hire and separation age distributions, such as incentives offered for early retirement, or changes in requirements for entry into service at various grades. These examples refer to actions mainly outside the scope of Bureau management. In the area of local policy aimed at such modification, promotion policies will affect voluntary separations, but of course some study would be required to estimate how "leaver" age distributions are related to promotion rules.* It would be relatively easy to estimate what modifications in hire age distributions would result from controlled numbers of appointments such as post-doctoral fellowships, or others in which there is a strong correlation between age and educational/experience requirements. This is one example of how planning for attainment of target age distributions in a full time staff (not exclusively permanent) is practicable with a Markov model as an analytical tool.

*A possible point of departure - compare leaver age distributions in organizations which tend to assign positions of responsibility to current staff members (in effect promoting them) when setting up mission task forces, to those in organizations which tend to hire task force leaders from outside.

Various other descriptions of manpower staff composition might be studied usefully with Markov models based on relatively accessible data:

(a) Trends in occupational mix and/or educational level of the professional staff as an indicator of shifts in the scientific mission of the Bureau.

(b) Trends in the mix of the major occupational categories, that is, the organizational breakdown into Administrative, Professional, Technician, Clerical and Wage Board employees, relevant to mission changes broader than mentioned in (a) above.

(c) Trends in fragmentation if any (here the distribution in question could be the fractions of staff or professional staff in units of various size classes).

(d) Trends in the educational level and number of skills represented in the total staff, to estimate growing (or diminishing) flexibility for organization of in-house task forces in response to externally imposed research programs.

(e) Trends in service life of equipment - this subject has no direct bearing on personnel questions, but we include it because of its suitability for Markov analysis and its putative interest to management. The emphasis would likely be not on durability but on avoiding technological obsolescence, with, by the way, a clear analogy to the level of continuing training in a personnel staff. [6]

If Markov (or other models) are to be used as guides for management control, and not just as predictive or projective mechanisms, then three important questions will demand consideration:

(a) Can optimal or desirable distributions of staff according to characteristics such as age be defined meaningfully? What kinds of costs are associated with policies intended to modify the distributions?

(b) To what extent are external (non-management) factors affecting distribution changes measurable or predictable?

(c) To what extent are distributions for different characteristics correlated? For instance if two MOUs at the Bureau had differing current and/or stable age distributions, would the difference be adventitious or would it be related to measurable structural or "personality" differences between the MOUs? Could these, in turn, result from differences in organizational mission?

These questions, unfortunately, fall into areas about which little useful information is contained in personnel records, so that study of them, presumably by simulation, would entail a substantial commitment of time and effort on data acquisition. For instance, a possible target criterion for an age distribution is that the mean or median age should not exceed some fixed value, k. When concern over rising ages is the primary motivation for analysis, such a ("single parameter") measure may be adequate although even this starkly simple measure begs the question of how the "critical" value should be determined. More generally one might try to formulate a staff member "personality profile" whose components are characteristics presumed to be age-dependent, e.g., enthusiasm, energy, currency of training, imagination, flexibility, prudence, breadth of professional contact, etc.

_If_ (1) these characteristics can be quantified, and if, (2) they can be accumulated over populations so that the sums of traits of several people is the set of traits of the group, and further if (3) "target" trait-mixes can be defined relative to organizational missions, then a linear programming "model" could determine desirable age distributions. The execution of the formal model is straightforward.* The major problems of judgment and data acquisition occur in determining, quantitatively, how the named characteristics relate to age, and then determining, again quantitatively, how the characteristics affect expected satisfaction of missions.

*Models which combine Markov processes with optimization procedures such as linear programming have received some attention recently as possible approaches to this kind of analysis. [7, 12]

2. DATA

The basic source of "hard" data for the analysis was a copy (edited for reasons of confidentiality) of a computer-tape compilation of personnel actions called the History File and maintained by the Personnel Division. The file contains an entry comprising 30 items (Date of Action, Name, SSN, Date of Birth, Educational Background, etc.) for each individual personnel action at NBS* (hire, promotion, in-grade promotion, change in occupational title, etc.) since Jan. 1, 1970. This file is augmented weekly at present.

From the approximately 70,000 records of actions which occurred in the period Jan. 1, 1970 - Jan. 31, 1976, the 10,598 referring to hires and separations** were extracted and subjected to statistical procedures aimed at (a) deriving approximate age distributions for hires and separations of FTPP staff as input to the Markov model, and (b) determining if these distributions are peculiar to the professional staff. Table 1 (p. 18) gives a breakdown of the hires and separations according to major occupational status for the total staff and its FTP subcategory. In light of the known stability of total-staff and occupational-component sizes there <u>appear</u> to be discrepancies in the table, particularly for full time permanents who show over 1 1/2 times as many separations as hires. These discrepancies occur in general because of unfilled vacancies at the end of the data period and, particularly as regards FTP, change of status of employees who become FTP after originally being hired into another tenure category. A continuously employed staff member whose status so changes will not have been recorded as a FTP hire, but upon separation will be counted among FTP separations.

*Including the Boulder Laboratories.

**There are also a few entries for calendar 1969. File data include 520 hires and 151 separations in undesignated major occupational categories.

Table 2 gives summary data on hire and separation ages of the five FTP staff occupational categories. Some cautious speculations can be advanced on the basis of the numbers in Table 2, though verification would require further evidence and/or analysis:

First, the age measures for hires in all groups appear "high", but perhaps only in comparison with those believed typical for a growing organization with rapidly shifting missions. Second, although the five occupational categories have distinctly characteristic hire and separation ages, it does not seem as if any group can be singled out as displaying age statistics inconsistent with those of the others.

For example, the approximate 3 year age difference between professionals and technicians is about what would be expected given the difference in educational requirements; possibly 4 to 5 years for early career appointments and diminishing later, as job experience becomes a significant factor. Mean and median hire ages for wage board employees fall between the values for professional workers and technicians, reflecting relatively demanding preparatory training in, e.g., trades and crafts. (We would expect the wage board hire ages to be closer to those for clerical employees if the data in the table were not restricted to FTP, on the grounds that "Wage Board" would then include a smaller proportion of those trades involving a degree of "craftsmanship" akin to "professionalism".) The relatively high age figure for Administrative hires is very likely a result of the fact that until fairly recently there was relatively little emphasis on development of managerial skills at NBS, so that in this category more than others appointees have tended to come to NBS in mid career, transferring from jobs elsewhere.

The very low age of clerical hires indicates that the "winds of change" for this group have not blown with significant effect. In the 1970's as in previous decades, hire data are dominated by large numbers of recent high school graduates, primarily women; turnover rates are high, as can be deduced from the separation age data. There is reason to believe that they will remain high even if gender-related events, i.e., marriage and child rearing, should no longer so frequently precipitate withdrawal from the employed labor force. This may result from the relative lack of specific <u>assignment</u>-related skill requirements. That is, since clerk-typists generally do not have <u>professional</u> affinities related to the missions of particular organizations, their effective mobility may be high.

We can make the same superficial analysis of the separation data. Except for the inversion in rank between the means and medians, for the wage board and administrative employees, the rankings of these numbers conform with plausible explanations. For example, technicians should be expected to remain in given career positions longer than professionals or "managers" (administrative staff), because their skill development and refinement on the job is most closely associated with assignment. Administrative personnel will tend to move more frequently than scientific professionals because position advancement is for managers a vital component of the <u>definition</u> of their profession, whereas for scientist professionals it is nominally incidental.

Finally, the shapes of the "in" and "out" age distributions for the groups, sketched from data histograms (fig. 1), seem consistent with the foregoing conjectures.

TABLE 1

OCCUPATIONAL CATEGORY	TOTAL HIRES	TOTAL SEPARATIONS	FULL TIME PERMANENTS HIRES	FULL TIME PERMANENTS SEPARATIONS
Professional	1371	1389	347	538
Technicians	577	652	59	206
Wage Board	1330	1195	170	272
Administrative	272	404	123	180
Clerical	1368	1369	379	462
Totals	4918*	5009*	1078	1658

Hires and Separations for the period Jan. 1, 1970 - Jan. 31, 1976. (From the History File)

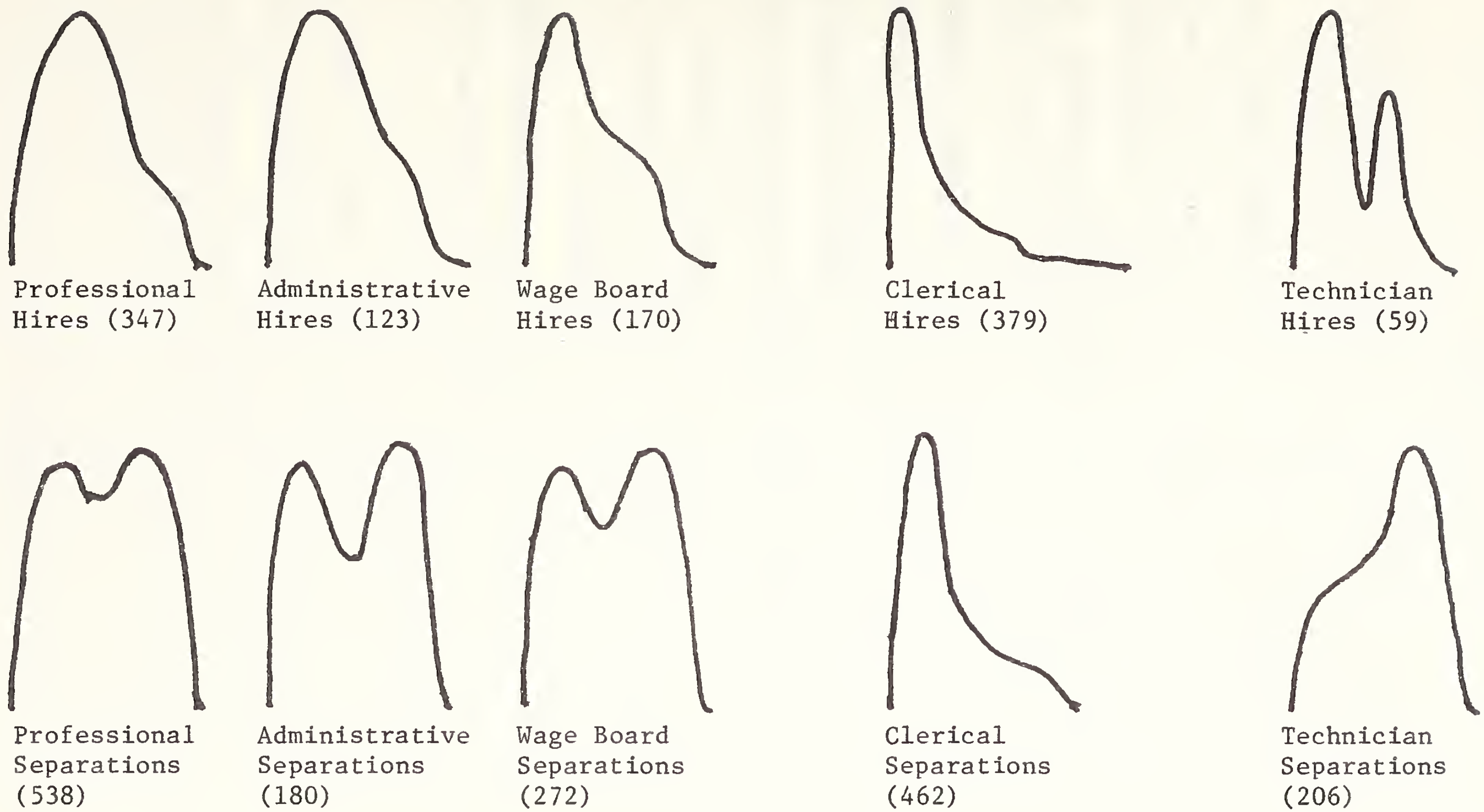
TABLE 2

OCCUPATIONAL CATEGORY	HIRES MEANS	HIRES MEDIANS	SEPARATIONS MEANS	SEPARATIONS MEDIANS
Professionals	35.9	33.8	45.5	45.7
Technicians	32.5	30.1	47.8	52.5
Wage Board	33.4	30.9	43.5	45.6
Administrative	36.9	37.6	44.5	45.0
Clerical	26.6	23.3	33.8	27.0

Mean and Median Ages of Hires and Separations of Full Time Permanent Employees for the period Jan. 1, 1970 - Jan. 31, 1976. By occupational category. (From the History File)

*Hires and separations = 9927. Inclusion of the 671 unlabelled hires and separations mentioned in the footnote on p. 15 gives a total of 10,598.

FIGURE 1



Curve shapes of the age distributions of hires and separations (Jan. 1, 1970- Jan. 31, 1976) for the primary personnel categories of NBS full time permanent (FTP) staff, grouped according to perceived similarities. The staff component label for each curve is followed by the number of hires or separations. Note that the curve for clerical hires (but not separations) could be viewed as a highly skewed version of those for Professional, Administrative and Wage Board employees (and hence similar). At this level of detail, the curves for technicians appear to be distinct from the others. (From History File Data.)

3. THE MODEL

3.1 Model Concepts

We present a simple age distribution analysis to clarify the form and manner of operation of a Markov model, and its level of realism. Suppose the size of the working staff of an organization is assumed to be constant, that is, all losses (retirements, etc.) are made up by recruitment, but new hires are limited by "slot" availability. The age distribution of the staff is assumed to be "recorded" every 25 years. Suppose also, that the staff members' ages may range from 21 to 70 (i.e., there is a minimum age required for employment and retirement is mandatory at 70 years of age), and we break down the staff into only two classes: 21 to 45-year olds and 46 to 70-year olds.* With these ground rules we would say that the staff is aging if the fraction in the second class increases over time. We are interested in the conditions which will produce a stable age configuration - one in which the sizes of the two classes no longer change. Now we make just three more assumptions:

(1) When vacancies occur, the replacements are split between the two age classes in a constant ratio, so that say, 1/2 of new hires are 45 or younger and 1/2 are older.

(2) The separation rates for the classes are constants. This means that in each 25-year period 1/2, say, of the first class will leave the staff. We do not differentiate among reasons for leaving, so that for the second class it suffices to say that all will leave,

*Demographers call such classes "age cohorts."

during a 25-year period, because regardless of what fraction leaves prior to retirement, all those who are members at the start of the period will by its end reach the age of mandatory retirement.

(3) Separations and hires will be treated as if they all take place simultaneously at the start of each 25-year "accounting" period.

Thus in any 25-year period 1/2 the members of the first class leave and the other 1/2 all age into the second class; all members of the second class leave the system.

Regular repetition of this accounting sequence, with the new class sizes replacing the old after each step, defines a discrete time Markov process or model. The necessary elements are: a fixed time step (in the example, 25 years), a set of classes or "states" (here, the two age classes) and constant coefficients relating the state values (the class sizes) after a time step to their values before it. [8, 9] *

In equations:

$$n_1(t+25) = 1/2(1/2\, n_1(t) + n_2(t)) = 1/4\, n_1(t) + 1/2\, n_2(t)$$

$$n_2(t+25) = 1/2\, n_1(t) + 1/2(1/2\, n_1(t) + n_2(t)) = 3/4\, n_1(t) + 1/2\, n_2(t),$$

where

t is the current or old year,

$t+25$ is the forward or new year,

$n_1(t)$ is the current size of the first class,

$1/2\, n_1(t) + n_2(t)$ is the total number of leavers during the 25-year

*Strictly speaking, according to the theory of Markov processes, which deals with random phenonema of a certain type, the coefficients are probabilities and the rates are thus rates "on the average". In applications, however, that is in analysis of "real" systems using Markov theory, the distinction between random and deterministic processes is usually ignored. The computational characteristics of these systems do not depend on the labels attached to the coefficients, e.g. "constant rates" or "constant probabilities".

period [cf.(2) above],

$n_2(t)$ is the current size of the second class,

$n_1(t+25)$ is the new size of the first class,

$n_2(t+25)$ is the new size of the second class.

Now if we operate this process on an organizational staff of 1000 members, half of them currently 45 years old or younger and half older -- that is, $n_1(1975) = 500$ and $n_2(1975) = 500$ -- then the result is

$$n_1(2000) = 1/4\ n_1(1975) + 1/2\ n_2(1975) = 1/4 \cdot 500 + 1/2 \cdot 500 = 375,$$

$$n_2(2000) = 3/4\ n_1(1975) + 1/2\ n_2(1975) = 3/4 \cdot 500 + 1/2 \cdot 500 = 625;$$

$$n_1(2025) = 1/4\ n_1(2000) + 1/2\ n_2(2000) = 1/4 \cdot 375 + 1/2 \cdot 625 = 406.25,$$

$$n_2(2025) = 3/4\ n_1(2000) + 1/2\ n_2(2000) = 3/4 \cdot 375 + 1/2 \cdot 625 = 593.75;$$

$$n_1(2050) = 398.4375 \quad , \quad n_2(2050) = 601.5625;$$

$$n_1(2075) = 400.390625 \quad , \quad n_2(2075) = 599.609375;$$

$$n_1(2100) = 399.9023438 \quad , \quad n_2(2100) = 600.0976563;$$

$$n_1(2125) = 400.0244141 \quad , \quad n_2(2125) = 599.9755859.$$

The staff (that is, the model) has a stable distribution with 400 persons in the 20-45 year old class and 600 from 46-70. To all intents and purposes, that distribution is reached in 100 years or 4 cycles of the process (3 if we aren't fussy).

3.2 Discussion of Concepts

(1) The example has 2 classes and a 25-year step interval. Our model for FTPP used 10 age classes with a five-year step interval, and with moderately small additional effort in preparing data, input could be disaggregated to 50 classes (e.g., ages 21, 22, ..., 70) and an annual time step. The "mechanics" of the process would be the same in all cases.

(2) With rare exceptions*, aging models of the type studied here will converge to stable distributions. The stable distribution, by the way, is determined entirely by the model parameters (the equation coefficients, i.e. the "transition probabilities"). That is, all initial age distributions of a given sized population will converge to the same stable distribution (but possibly at differing rates). If, for instance, in our example we had assumed the initial distribution

$$n_1(1975) = 1000 \quad , \qquad n_2(1975) = 0 \quad ,$$

we would have calculated the sequence

$n_1(2000) = 250$, $\qquad n_2(2000) = 750$;

$n_1(2025) = 437.5$, $\qquad n_2(2025) = 562.5$;

$n_1(2050) = 390.625$, $\qquad n_2(2025) = 609.375$;

$n_1(2075) = 402.34375$, $\qquad n_2(2050) = 597.65625$;

$n_1(2100) = 399.4140625$, $\qquad n_2(2100) = 600.5859375$;

$n_1(2125) = 400.1464844$, $\qquad n_2(2125) = 599.8535156$;

requiring two additional cycles to reach the (same) stable distribution to the nearest integer. (Compare the values for year 2125 above to those for year 2075 in the previous example.)

*Involving zero hire and/or separation rates for a large subset of the age classes. In the example, for instance, if there were no separations from the first class and no hires into the second, the two class sizes would interchange endlessly from the start, and thus not converge.

We chose the numerical values of the example's model parameters for arithmetic simplicity. In general, the hire fractions (in the example, 1/2 for each class) must merely be non negative and add up to one, while the separation fractions (in the example, 1/2 for the first class and 1 for the second) must merely be non negative.

The coefficients of the model, which in the example were calculated from the aging of staff members and the hire and separation fractions, are similarly free. That is, since the coefficients of any given class, in total, "account" for the "contribution" of that class to the classes in the next step, they must merely not be negative or greater than one in value, and must sum to one for a system with a constant total.* (In the example we had 1/4 and 3/4 as the coefficients of n_1 in the two equations; 1/2 and 1/2 as the coefficients of n_2.)

The non-dependence of the shape of a Markov model's stable distribution on the starting distribution furnishes a crude yardstick for measuring the validity of "present is like the long-term past" assumptions in a given application.

Specifically, suppose we have estimated the transition probabilities (coefficients) from recent data and we wish to assume that they have <u>been</u> unchanged for some extended period of time previously. This would imply that our application represents a Markov process which has <u>already been operating for some time</u>, and thus our "starting" <u>current</u> distribution should not be drastically different from the stable distribution which represents the expected outcome of the process and which we can calculate from continued application of the model as in the example.

*One of the reasons these coefficients are called "transition probabilities" in Markov theory is that they have these probability-like properties.

On the other hand, we might be led to assume or deduce that the process defined by our coefficients represents a substantial difference from the one previously "in effect", that is, that the organization's staff is affected by a different set of forces now than previously. In this case our suspicions would be raised by a stable distribution that resembled the current one too closely.*

(3) In the sample calculations above, the Markov process operated on a distribution described by the numbers of staff members in each class. These numbers should be thought of as "expected values" after each step. Rounding them to whole numbers to preserve token realism would be improper because it is equivalent to making small changes in coefficients which must be constants to satisfy the mathematical definition of the Markov process. The "psychological" problem of interpreting the stepwise outputs as age-class population sizes can be avoided by stating the distributions in fractional or percentage terms. Thus in the example we would have used $n_1(1975) = .5$ or $n_1(1975) = 50.0\%$ instead of $n_1(1975) = 500$, leading to stable values of .4 or 40 %, respectively.

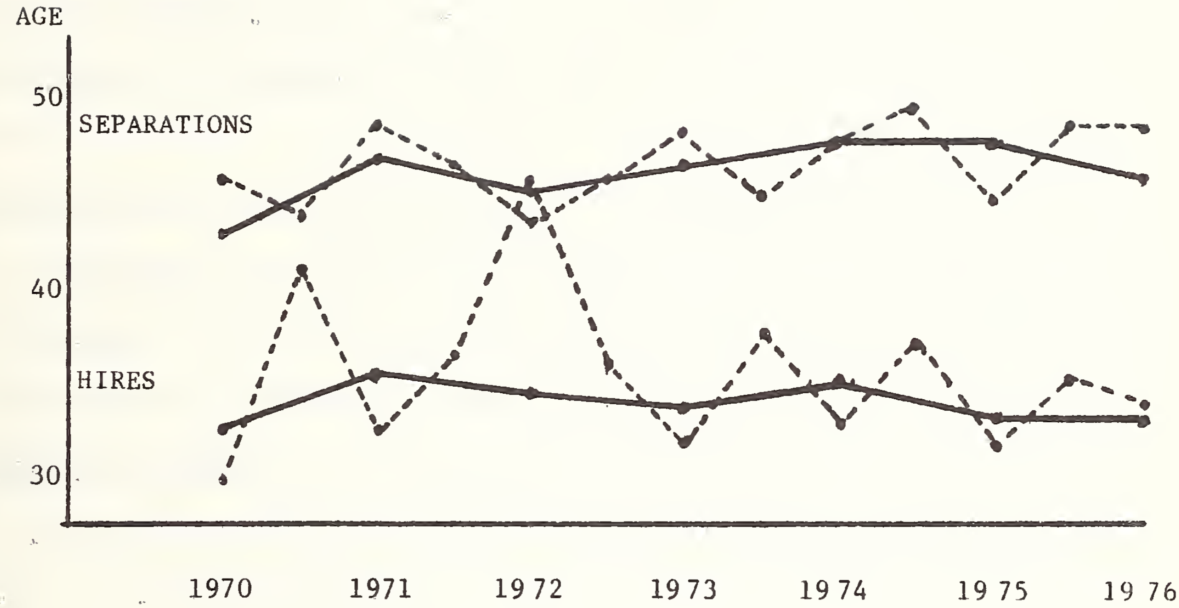
(4) The reason for requiring that the process step time match the class age interval can best be explained with an example: If we wished to pursue the changes in the distribution into two classes more closely by examining this change every five years instead of every 25, we would need 5-year constant rates in place of 25. Thus we would first have to assume that the age distribution within the classes be sufficiently uniform so that each five years exactly 1/5 of those from 21-45 who didn't leave the

*"Drastically different" and "too closely" are matters of judgment. In practice simple statistical procedures such as the Chi-square test are pulled from the shelf and applied.

organization could age into the 46-70 year old class, and that similarly exactly 1/5 of the members of the second class would age into retirement. The main difficulty begins at the end of the second five year period, because a "Markov process has no memory" so that the people who passed into the second class five years ago are "diffused" in it and 1/5 of them will now be "promoted" to age 70 -- that is, the ages of some of our staff members would increase by over 25 years in a 10 year period!

(5) We specified in our example that the transitions were to be treated as occurring simultaneously at the starts of time periods. This was done to avoid a situation in which for instance, during a single period a 23-year old staff member leaves the staff, is replaced by a 29-year old, who in turn leaves and is replaced by a 22-year old, leading to counting-ambiguities. (The turnover is two, but we see only one new face at the end of the period.) This artificiality is not a defect peculiar to this type of model, but is one that is shared by all sorts of accounting procedures unless they are continuous. A case in point is the administrative tabulation of the total size of the NBS FTPP staff, which has been quite constant for six years when recorded at calendar-year ends but shows fluctuations on a July 1 basis. This sort of variation leads to a basic dilemma in modeling. Discrete, that is, fixed time-step treatment of continuous processes such as aging frequently produces paradoxical results. A more realistic representation of the flow of people in and out of NBS would take into account the relatively random nature of the "departure time" of leavers, and time lags between separations and hires with their possible consequences. Such mathematical models exist but they are complicated and entail difficult (i.e., costly) data demands. It is possible that two different sets of constant rates

FIGURE 2



Median ages of FTPP hires and separations recorded annually (solid lines) and semiannually (dotted lines) in the period Jan. 1970 - Jan. 1976. (From History File data) could be observed depending on the choice of starting points for the regular intervals; fiscal year end and calendar year end, say, might result in two different Markov processes "existing in tandem." This is one of the few cases in which one can usefully introduce a single formal model which actually permits variation in the rates. Such models, with cyclic variation of the parameters, lead to a pair (or more) of alternating "stable" distributions. [1, 8]

How the stable distribution will vary, as a result of differing hire distributions and/or leaver rates, is interesting as an approach to studying the control of the age distribution, and also as a means of measuring the effect of error in estimating the "true" hire and leaver parameters. The question of such variations (i.e., sensitivity analysis of the steady state) will not be treated in any depth in this report, because specifying suitable "scenarios" of such input changes is a complicated problem beyond the scope of the present introductory study. Even the present two-class example is sufficiently "rich" that with (effectively) only two quantities subject to modification (the fraction of hires into and the fraction leaving the first class), it is not immediately obvious how to program a sequence of meaningful changes. We append an example to show that in this case, at any rate, moderately small changes in the parameters lead to small changes in results:

If the hire fraction into the first class is changed from .5 to .48 and the leaver fraction from the first class is also changes from .5 to .48, the coefficients (1/4, 3/4) of n_1 become (.2304, .7696) and the coefficients (1/2, 1/2) of n_2 become (.48, .52). These lead to a stable fractional distribution of (.384, .616) instead of (.400, .600).

We close this section with some opinions concerning the appropriate level of detail for a predictive model of the NBS staff age distribution useful for planning purposes. Obviously, the 25-year chunks of time of the example can't give us much meaningful information - too much goes on in the organization during such long intervals, and confining the description of the distribution to a ratio of "young" to "old" (the two classes) presents a very fuzzy picture. Moreover, since just four cycles of the process represent 100 years of real time (over 30 years longer than NBS has existed), we would be straining credulity too far to expect the change rates to hold constant over such a period. On the other hand, a slight amount of fuzziness is desirable, if it militates against the temptation to regard model outputs as gospel because they "look" very realistic. This more or less rules out "atomic" monthly or even quarterly models. An annual model appears to be an attractive alternative, but we believe that the data at hand are not suitable for the determination of representative (constant) annual hire and separation rates by one-year age intervals. The five-year level of detail in the model used as the basis for this report, was adopted primarily because of the ready availability of five-year summary and profile data in the "Employment Trends" personnel report [14], but we think that it also affords a reasonable compromise between an overly crude representation and one for which smoothing out input-data fluctuations is too difficult. Setting questions of convenience aside, computations with a time period and age class interval of about three years appear to offer the greatest likelihood of reliable projections given the now familiar assumptions of constancy in the sets of (age-dependent) separation and hire rates.

4. THE WORKING MODEL AND ITS OUTPUT

The Markov model which was operated to project the age distribution of the FTPP staff is conceptually the same as the sample model of section 3.1 except that it has 10 classes representing 5-year age intervals, and a 5-year step size. The assumptions too, are conceptually the same, except that they are applied to observed data (Table 3), instead of "made up" numbers. Thus the model's distribution of hires into age classes was determined by calculating the corresponding fractions of the 347 FTPP hires recorded in the History File between Jan. 1, 1970 and Jan. 31, 1976.

The separation fractions could not similarly be read directly from the History File data, because this calculation would give only the distribution into age classes of the total group of leavers in the 6 1/12 year period of recording, whereas what was required was the fraction of each age class that was separated during a 5-year period. These fractions were calculated by first estimating a representative number of leavers in each class during a five year period by multiplying the number of leavers from the History File by .82 (the approximate ratio between 5 and 6 1/12), and then dividing by the size of the class. The class sizes themselves were derived from "informal" averages (i.e. judged rather than calculated) of (a) the percentage breakdowns of FTPP staff into 5-year cohorts for the years 1970-74 and (b) the breakdown of the FTP staff by sizes of occupational category components given as charts in [14] and reproduced here as figs. 3 and 4. It was necessary to "massage" the separation data to avoid recording a separation rate greater than 100% in the highest age class,

TABLE 3

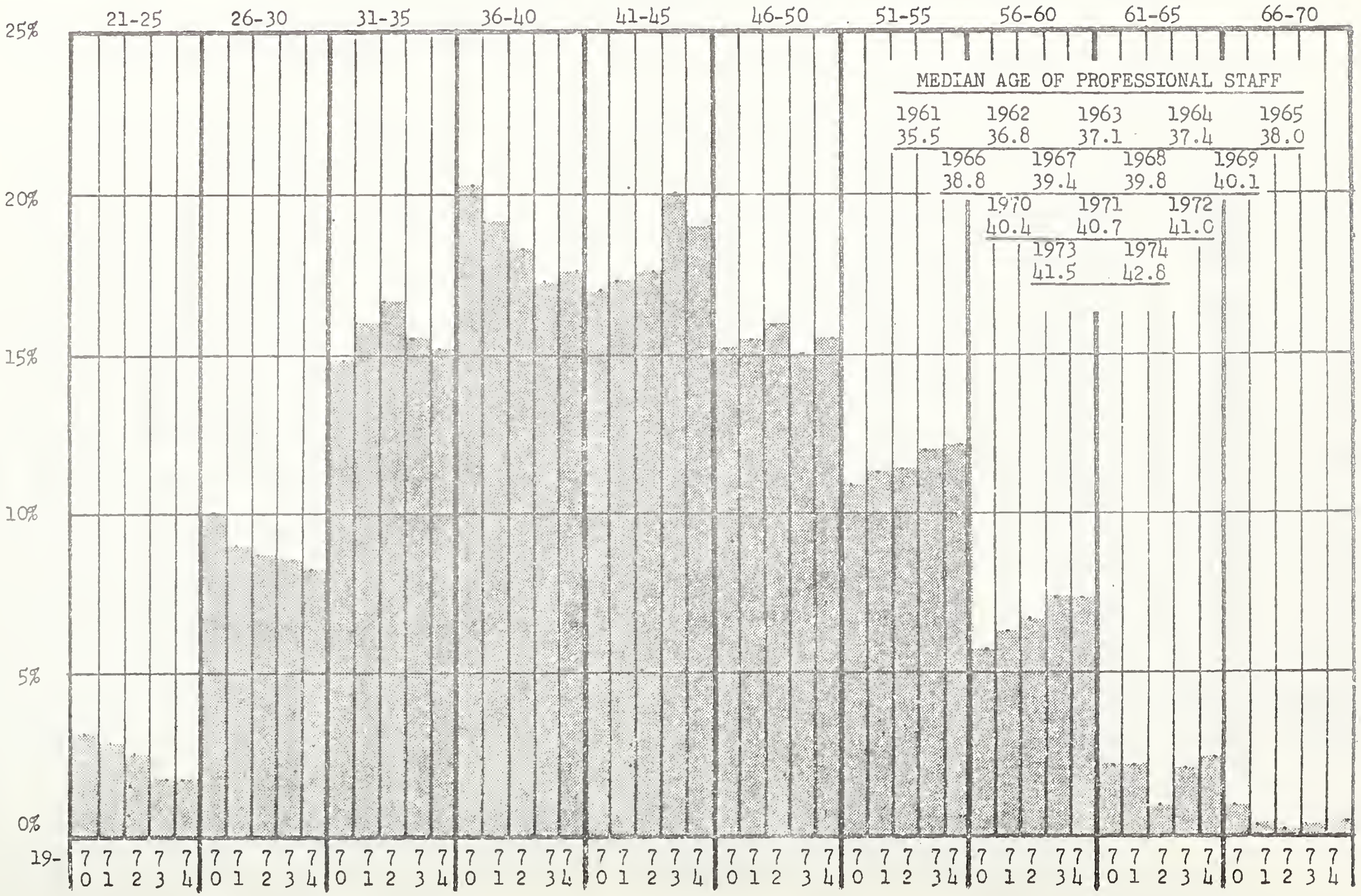
1	2	3	4	5	6	7	8	9	10
CLASS	AGE	STAFF CLASS SIZE	STAFF CLASS FRACTION	HIRE COUNTS	SEPARATION COUNTS	(ADJUSTED) 5-YEAR SEP. COUNTS	5-YEAR SEPARATION RATES	5-YEAR "AGE" RATE	HIRE DISTRIBUTION
1	21-25	42	.030	39	32	26	.6190	.3810	.1124
2	26-30	126	.090	92	61	51.5	.4087	.5913	.2651
3	31-35	217	.155	74	66	56.5	.2604	.7396	.2133
4	36-40	252	.180	52	57	48	.1905	.8095	.1498
5	41-45	238	.170	27	56	47	.1975	.8025	.0778
6	46-50	217	.155	30	62	52.5	.2419	.7581	.0865
7	51-55	175	.125	17	61	51.5	.2943	.7057	.0490
8	56-60	91	.065	13	78	67	.7362	.2638	.0375
9	61-65	35	.025	3	42	34	.9714	.0286	.0086
10	66-70	7	.005	0	23	7	1.0000	.0000	.0000
TOTAL		1400	1.000	347	538	441			

NOTES:
COL 3, 4 From charts in [14]
COL 5, 6 History File data (6 years +1 month)
COL 7 .82 x COL 6 entries adjusted to maintain total and match last row value to that in Col. 3.
COL 8 COL 7/COL 3
COL 9 1 - Entry in COL 8
COL 10 History File data

RELEVANT FTPP DATA AND MODEL PARAMETERS

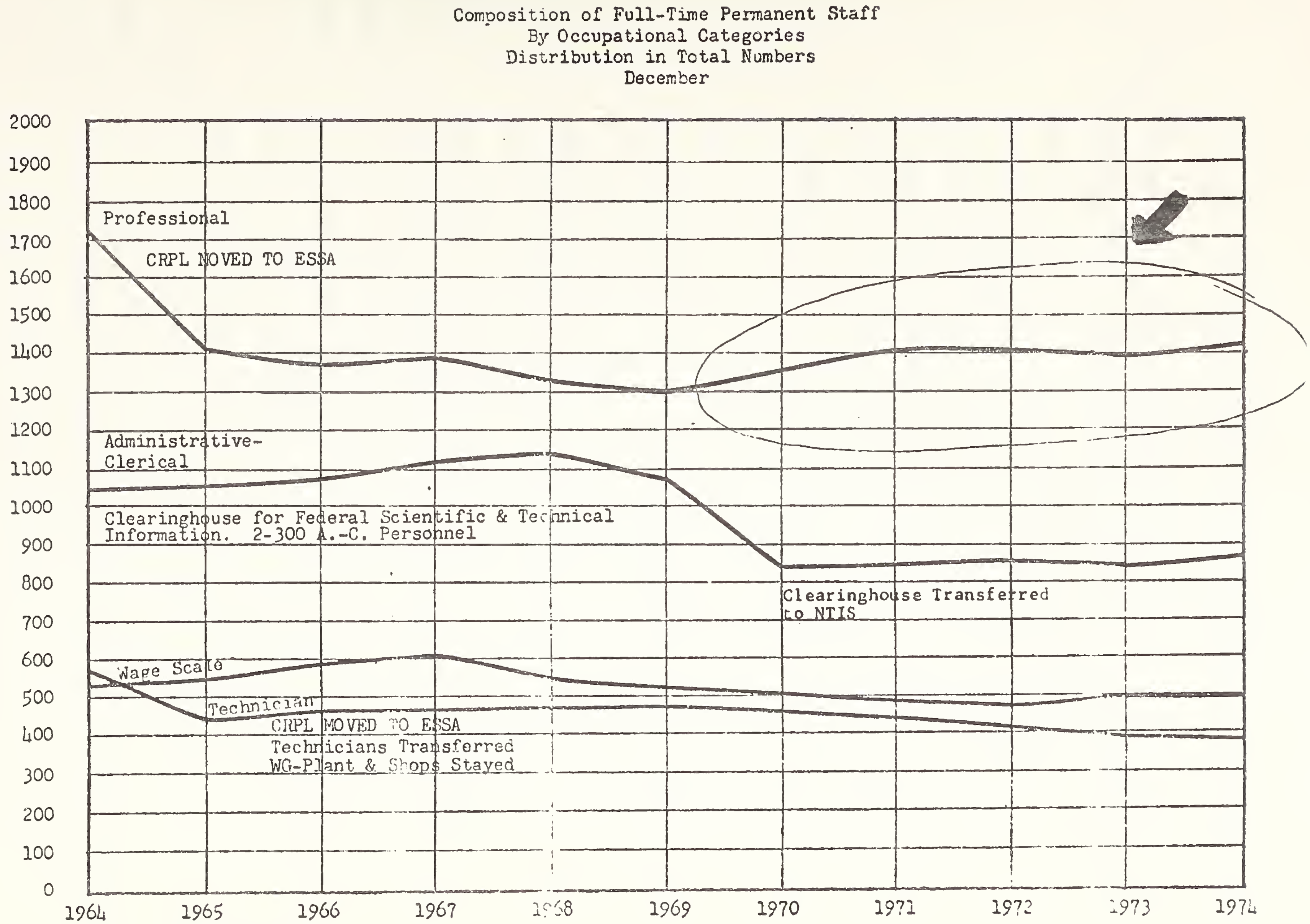
FIG. 3 (REPRODUCED FROM [14])

Full-Time Permanent Professional Staff
Percentage Distribution by Age
December



Source: NBS-023, NBS-025 & NBS-075

FIG. 4 (REPRODUCED FROM [4]) SHOWING APPROXIMATELY CONSTANT FTPP STAFF SIZE 1970-74



Source: NBS-020 & NBS-012

the estimated leavers in 5 years having exceeded the size of the class as read from the chart. (We have previously touched on some of the likely causes for such a discrepancy: unfilled slots, twice filled slots and charges of occupational status.)

These details of the crude estimation of hire and leaver rates (assumed representative and constant) have been included to emphasize the provisional nature of the projections from the Markov model. We remind the reader again (cf. p. 23) that Markov models almost invariably will produce ultimately stable distributions for the kind of system treated here and that with some additional expenditure, more precise estimates of the coefficients can be derived from the recorded data of the past 6 or 7 years. The real catch is that current knowledge gives very little idea to what extent the forces shaping the hire and separation distributions may vary in the future (and indeed, how they have varied prior to the recent past). The projections of current trends resulting from the present exercise are valuable as a reference point, but if they are to be succeeded by attempts to make accurate predictions, then we believe that improvements are more likely to derive from analysis of causes of change than from quantitatively refining the procedures we used for the estimation of parameters. We give three plausible areas for such analysis:

(1) possible relationships over time between the age distributions of various components of the nation's population and the age distribution of hires in a group of interest, e.g. FTP or FTPP at NBS (the notion here is that the composition of an

available "pool" may determine the composition of the recruited force);

(2) relationships between length-of-service distributions and rank (grade) structures or promotion policies. These might then be combined with hire-age distributions to estimate the distribution of leaver ages.

(3) A study aimed at amplifying the data base by broadening the subject population through redefinition. On pp. 2, 16 above we intimated a suspicion that other components of NBS staff might be equivalent to FTPP with respect to evolving patterns in age distribution. If this suspicion were to be substantiated, the other group or groups could then be amalgamated with FTPP for the purpose of analysis, furnishing an expanded set of observations (from data files) from which to calculate presumably more reliable parameters for a Markov model. On the other hand, careful redefinition of the subject population could result in a restriction of the data set if, for instance, it turned out that persons whose status changes after entry to the staff constituted a small but variable fraction with substantially distinct aging characteristics. Even in this case, the reduced data base might afford improvement if it contained fewer "outliers" to the average trends.

Briefly stated, the Markov model outputs (Table 4) show the staff age distribution becoming stable (to many decimal places) in 31 cycles (155 years!) but in practical measures in about 5 cycles or 25 years, with an increase of less than 1 year in average age.*

*The staff class sizes have been rounded to whole values (integers) for display only. They were carried in the calculations to "full" computer significance.

TABLE 4

CLASS NUMBER

		1	2	3	4	5	6	7	8	9	10		AVERAGE CLASS SIZE CHANGE IN 5 YEARS	AVERAGE AGE
	YEAR	21-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60	61-65	66-70	TOTAL		
(A)	1975	42	126	217	252	238	217	175	91	35	7	1400		42.3
	1980	50	133	169	227	238	229	186	140	28	1	1400	17.41	43.0
	1990	54	147	186	203	194	217	199	156	43	1	1400	7.8	43.2
	2005	55	150	193	216	212	209	179	141	44	1	1400	2.8	42.8
	2020	54	148	191	212	210	211	184	147	42	1	1400	0.5	42.8
	2130	54	148	191	213	210	210	183	147	43	1	1400	0.0	42.8

AGE DISTRIBUTION OF FTPP FOR 6 SELECTED STAGES OF THE MARKOV MODEL

(B)	1975	3.0	9.0	15.5	18.0	17.0	15.5	12.5	6.5	2.5	0.5	100.0%
	2130	3.8	10.6	13.6	15.2	15.0	15.0	13.1	10.5	3.1	0.1	100.0%

INITIAL AND STABLE % AGE DISTRIBUTIONS OF FTPP STAFF

											AVERAGE MAGNITUDE
(C)	12	22	-27	-39	-28	-7	8	56	8	-6	21.3
	28.6	17.5	-12.4	-5.5	-11.8	-3.2	4.6	61.5	22.9	-85.7	26.41%

CHANGE IN AGE CLASS SIZE FROM INITIAL TO STABLE DISTRIBUTION

% CHANGE IN CLASS SIZE FROM INITIAL TO STABLE DISTRIBUTION

The relatively large shifts in the early periods indicate that the Markov process deduced from the information in the data files is probably different from one which might have produced the starting (1975) age distribution (cf. p. 24). This conforms with a generally-held perception that NBS' current and recent state differs distinctively from that during a preceding period of growth. It also suggests, among other things, that in addition to examining the (relatively long term) steady state distributions of the Markov model, a collateral attempt to extrapolate historical changes and intervals of changes in hire or separation patterns might be useful in order to estimate parameters for periodic "restart" of the model.

The general tenor of the shifts is a moderate increase in the fractions of staff in ages 21 to 30 and 51 to 60, and a general decrease throughout the middle aged groups 31-50 years old. The model's turnover (the number of staff members who can be expected to be separated in a time step) varies as the process progresses, even though the total staff size remains fixed and the "loss intensities" (the separation rates relative to the age-class sizes) are constant. This variation occurs because these rates are applied to class sizes which vary over time. The stable distribution has a turnover of 482, that is, 32.9% for a 5-year period or 6.8% annually. This represents a slight increase over the starting value of 441 (31.5% per period, 6.3% annually) which was derived from the History File data and which is consistent with information graphed on p. 20 of [14].

Variations in the losses from individual classes could be disconcerting from the point of view of management if the model is realized in flesh and blood terms, particularly since our particular model presents a system which shifts drastically in the early stages as noted above. The number of 56-60 year olds increases in one period by over 50% (from 91 to 140) to a value close to its size in the stable distribution (147). Separations from this age-group, currently numbering about 13 annually, would occur at almost double that rate* (about 21 annually) during the next ten years. This points to a fairly heavy exodus of those staff members now of ages 46-55 who constitute the bulk of senior scientific leadership at NBS. To a somewhat lesser extent the "transfer out" of the "young PhD" class (aged 31-40 at separation but now under 30) will decrease during the same period from approximately 21 annually to 18.

	AGE CLASS				
YEAR	31-35	36-40	41-45	46-50	51-55
1975	3.5	1	2	3.5	5
1980	5	3	1	2	4
1985	5	4	2	1	3
1990	5	2	4	1	3
1995	4	1	2	3	5

TABLE 5. RANK ACCORDING TO CLASS SIZE (5 LARGEST CLASSES) FROM COMPUTER OUTPUT OF MODEL

A notion related to the possible cause for sudden losses of large numbers of critical senior personnel is the possibility of an "age bulge" propagating wavelike through the staff over time, as a result of massive hires of young scientists in the 1950's and early 1960's. We therefore offer table 5 showing progressive class size ranks, but leave the risk of interpreting it to the reader.

*Note that this is a rate in people/year, not to be confused with the "separation rate" which is a constant fraction of the class size. In fact the numbers cited are calculated by multiplying the separation rate for the class (.7362) by the class size (see tables 3 & 4).

As we have implied earlier, the desireability of any particular age distributions and strategies for their attainment or avoidance are matters which cannot be settled by Markov models (although appendix B includes discussion of the relationships among the hire, separation and stable staff age distributions). The principal benefit of the Markov model is that it describes the resolution of current trends instead of merely designating their initial directions. In addition, it can be used to determine the resolution of specified modifications in these trends.

5. BIBLIOGRAPHY

1. Bartholemew, David J., Stochastic Models for Social Processes, (New York: John Wiley & Sons, 1967).

2. Davies, G. S., "Consistent Recruitment in a Graded Manpower System," Management Science, Vol. 22, No. 11, pp. 1215-1220, (July 1976).

3. Feichtinger, Gustav and Mehlmann, Alexander, "The Recruitment Trajectory Corresponding to Particular Stock Sequences in Markovian Person-Flow Models," Mathematics of Operations Research, Vol. 1, No. 2, pp. 175-184, (May 1976).

4. Forrester, J. W., Industrial Dynamics, (MIT Press, 1961).

5. Freese, Ralph and Johnson, Charles R., "Comments on the Discrete Matrix Model of Population Dynamics," Journal of Research of NBS, Vol. 78B, No. 2, pp. 73-78, (Apr.-June 1974).

6. Gornick, F. and Passaglia, E., A Dynamic Model for a Population of Durable Goods, (Draft of a Report from IMR of NBS, 1976).

7. Jaquette, D. L. and Nelson, G. R., "Optimal Wage Rates and Force Composition in Military Manpower Planning," Management Science, Vol. 23, No. 3, pp. 257-266, (November 1976).

8. Kemeny, John C. and Snell, J. Laurie, Finite Markov Chains, (Van Nostrand, Princeton, NJ 1960).

9. Pearl, Martin, Matrix Theory and Finite Mathematics, (McGraw Hill, NY 1973).

10. Simonson, E. and Weiser, P. (eds.), Psychological Aspects & Physiological Correlates of Work and Fatigue, (Charles C. Thomas, Springfield, IL 1976).

11. Schinnar, Arie P., Eco-demographic Planning and Forecasting (to appear).

12. Smith, A. P. (ed.), Models of Manpower Systems - Proceedings of a NATO Conference, (American Elsevier, NY 1971).

13. Stahl, Michael J., "Organizational Variables Associated With Innovation and Productivity in Research and Development," - Paper #FP2.12, 1976 Joint National Meeting of The Institute for Management Science/Operations Research Society of America, (Phila., PA., April 2, 1976).

14. Unsigned Report, "Employment Trends 1964-1974," National Bureau of Standards, Office of the Associate Director for Administration, Personnel Division.

15. Young, A. and Almond G., "Predicting Distributions of Staff," Computer Journal Vol. 3, pp. 246-50 (1961).

PREFACE TO APPENDICES

Appendix A is a restatement of the material in Sections 3 and 4 of the report, using matrix notation, augmented by a brief discussion of some basic properties of discrete time Markov Models. Appendix B is an analysis of the algebraic relationships among the quantities that define the Markov transition probabilities (namely, the Hire and Separation age distributions) and the output of the model (the staff's stable age distribution).

APPENDIX A: THE MODEL IN MATRIX FORM

The Markov Model traces the progressive changes in the age distribution of staff over a sequence of time steps, through linear equations whose coefficients are rate constants (called transition probabilities or transition coefficients), which relate the components of the distribution after each time step to the components of the previous distribution.

According to the customary classifications of Markov models ("Markov Chains" or "Markov Processes"), the model of this report is a "Finite State Discrete Time Simple Markov Chain with Stationary Probabilities": the set of states (age classes) is finite, the process operates (is recorded) at regular time periods rather than continuously, and the equations that project the classes one period forward require only the coefficients and the population's current status. (A "simple" model (in the preceding sense) is also called a "first order" model. If, instead, information about both the <u>preceding</u> period <u>and</u> the <u>current</u> period were needed to project forward, the model would be described as second order; a chain whose equations include data for k periods including the current one to project the next is called a k-th order Markov chain.) Finally, the coefficients, as stated previously, are constant.

Most elementary texts simply use the term Markov Chain for the kind of model described here, and we adopt the abbreviation Markov Model from this point on. For an introduction to Markov models see references [8, 9] of the report or <u>Introduction to Finite Mathematics</u> by the authors of [8]. A readable survey of more general Markov models is contained in "Statistical Methods in Markov Chains" by Patrick Billingsley, <u>The Annals of Mathematical Statistics</u>, Vol. 32, 1961, pp. 12-40.

We have a staff of n members divided into m classes according to age, with the number of members in the i-th age class denoted by $n_i(\tau)$, where τ is a discrete variable representing time steps or stage intervals over which the age distribution will be traced.

The values $n_i(\tau)$ define (row) vectors $N(\tau)$, $N(0)$ being the initial distribution, with $n_i(\tau) \geq 0$, $i = 1,2...m$, $\tau = 0, 1, 2, ...$, and

(1) $$\sum_{i=1}^{m} n_i(\tau) = n \quad \text{for } \tau = 0, 1, 2, ...$$

The constant transition matrix $P = (p_{ij})$ with

$$p_{ij} \geq 0, \ \sum_{j=1}^{m} p_{ij} = 1 \quad \text{for } i = 1, 2, ..., m,$$

determines the stage vectors, of expected age-class sizes, $N(\tau)$, through

(2) $N(\tau) = N(\tau-1)P = N(0)P^{\tau}$

for $\tau = 1, 2, ...$ where P^{τ} is the τ-th power of P.

Frequently, it is convenient to replace the distribution vector $N(\tau)$ with a proportional distribution vector $F(\tau) = (f_1(\tau), f_2(\tau), ..., f_m(\tau))$ where $f_i(\tau)$ is the fraction of the staff in class i at time τ. Thus $f_i(\tau) = n_i(\tau)/n$, with

(1a) $$\sum_{i=1}^{m} f_i(\tau) = 1 \quad \text{for } \tau = 0, 1, 2, ...$$

and, analogously to (2),

(2a) $F(\tau) = F(\tau-1)P = F(0)P^{\tau}$.

The distinction is essentially semantic, but the representation in terms of $F(\tau)$ bypasses the necessity of rounding the expected values $n_i(\tau)$ to integers for display of intermediate outputs of the model, to avoid confusion in interpretation of non-integer class sizes.

To determine the entries of P, we first label two artificial "classes" <u>outside</u> the staff structure to represent a source of hires into the classes 1, 2, ..., m, and a sink for separations from these classes. (Note that they do not appear in the age distribution.) We assign these two artificial classes, respectively, the indices 0 and m+1. Thus $p_{i,m+1}$ (for i=1, 2, ..., m) is the fraction of the i-th class that will be separated, and p_{oj} (for j=1, 2, ..., m) the fraction of hires into the j-th class, in a time period. For mnemonic convenience we will now relabel the quantities $p_{i,m+1}$ as s_i and p_{oj} as h_j. Then, with a_{ij} (for i, j=1, 2, ..., m) the fraction of the i-th class that will actually age to the j-th class, in a time period.

(3) $p_{ij} = a_{ij} + s_i h_j \quad i,j = 1, 2, \ldots, m.$

In our aging model

a_{ij} will be zero for $j \neq i+1$ and

$a_{i,i+1}$ will be $1-s_i$

The relationship (3) rests on replacing the total number of separatees in a period by hires to maintain the constant staff size n. Thus the products $s_i h_j$ (j = 1, 2, ..., m) distribute the replacements for the i-th class separatees among the age classes--clearly, $\sum_{j=1}^{m} h_j = 1$ by definition.

The remarkable and important fact about Markov models is that if the matrix P, or merely some power of P, is positive, i.e., has no zero entries (such a matrix is said to be regular or in algebraic terminology primitive), then the process defined by the model is guaranteed to converge as τ increases.

In other words, the sequence N(0), N(1) ... approaches a unique vector $N(\infty)=N(\infty)P$. Moreover, this "solution" vector does not depend on the starting distribution N(0). Thus writing $N(\infty)$ as N we obtain

(4) $N = NP = N(0)P^{\infty}$

(5) $P^{\infty}P = PP^{\infty} = P^{\infty}$

i.e., the powers of P converge.

Equation (4) is a linear equation. Therefore it is valid for arbitrary scaling of the vector N(0). Thus for any real number α and $\tilde{N}(0) = \alpha N(0)$,

(4a) $\tilde{N} = \alpha N = \tilde{N}(0)P^{\infty}$, as well.

Hence the substitution of a fractional distribution for the distribution of class sizes as in equation 2a could be made before or after the solution of the model with equivalent results. ($\alpha = 1/n$.) It turns out that the rows of P^{∞} are all equal to the stable vector in normalized form, i.e. in terms of our model, a stable fractional age distribution vector. This means that while the starting distribution might affect the speed of numerical convergence to a particular number of decimal places, the final distribution itself is determined entirely by the transition matrix. Transition matrices for our age analysis can be expected to be regular except for very unlikely hire and separation distributions and thus, for us, the FTPP staff does approach a stable age distribution determined solely by the separation rates and hire age distribution assuming they are constant. Moreover, whatever the staff age distribution at any point in time, altering the separation rates and hire age distribution to "new" constant values will define a "new Markov process at that point, allowing us to determine in advance the direction of change in the staff age distribution if the hire and separation distributions can be controlled by management policy. Appendix B develops algebraic relationships among the staff distribution, the hire and separation distributions and the turnover rate.

The specifications of the modeled system were as follows:

(1) The FTPP staff was divided into 10 classes defined by 5-year age intervals from 20 to 69. The initial distribution was derived from reference [14].

(2) The total FTPP staff size was assumed to remain constant at 1400 persons (from [14]).

(3) Age dependent separation rates and a hire age distribution according to 5-year age intervals were derived from the History Data File and assumed to remain constant. All hires and separations are assumed to occur at the beginning of a time step. (In calculating the averaged hire and separation distributions we did not discard data records outside the model age limits, but instead included those under 20 in the 20-24 year old class and those 70 and older in the 65-69 year old class.)

(4) The time step or "stage interval" for the model is the same as the class age interval, 5 years.

The numerical values for the components of the vectors (h_j) and (s_i) for the 5-year class interval FTPP model were determined as follows:

The FTPP hires for the entire 6-year period (Jan. 1970 - Jan. 1976) covered by the History File were grouped by 5-year classes of age at hire, and the fractions of the total that fell into each class were assumed to define a representative hire distribution, i.e. these fractions were considered to define the values for (h_j).

Treated similarly, the FTPP separations from the History File gave an age distribution of all leavers from the staff during the same period. Call these fractions g_i (i=1, ...m). The product of g_i by t, the average 5-year turnover, yields the representative number of "departures" from the i-th class. Call these quantities d_i. Thus

$$g_i t = d_i \ , \ i = 1, 2, \ldots, 10$$

and

$$s_i = d_i/n_i(0) \ , \ i = 1, 2, \ldots, 10$$

where as before the $n_i(0)$ are the initial sizes of the age classes, the average number in each class during (1970-1976).

Two extreme cases for the separations can be analyzed fairly easily.

(1) If $s_i = 1$ for all classes, i.e. turnover is total in each period, then intuitively, the age distribution should be determined solely by hires. This is indeed true of the model, for the transition matrix becomes

$$P = (p_{ij}) = (h_j),$$

which since $\sum_1^m h_j = 1$, has the property

$$P = P^2 = \ldots = P^\infty.$$

It follows that

$$N(1) = N(2) = \ldots = N = (h_1 n, h_2 n, \ldots, h_m n).$$

(2) If there are no separations except for mandatory retirement, then

$s_i = 0$ for all classes except $s_m = 1$. P now becomes

$$p_{i,\ i+1} = 1 \quad i < m$$

$$p_{ij} = 0 \quad i < m;\ j \neq i+1$$

$$p_{mj} = h_j \quad j = 1, 2, \ldots, m.$$

Or in tableau:

$$P = \begin{bmatrix} 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 1 \\ h_1 & h_2 & h_3 & \cdot & \cdot & \cdot & h_m \end{bmatrix}$$

We see that the equation $N = NP$, which determines the stable vector $N = N(\infty)$, takes the form

$$(n_1, n_2, \ldots, n_m) = (h_1 n_m\ ,\ n_1 + h_2 n_m, \ldots, n_{m-1} + h_m\, n_m)$$

or equivalently

$$n_1 = h_1 n_m$$

$$n_2 = n_1 + h_2 n_m = h_1 n_m + h_2 n_m = (h_1 + h_2) n_m$$

$$\cdot$$
$$\cdot$$
$$\cdot$$

$$n_m = n_{m-1} + h_m n_m = \ldots = (h_1 + \ldots + h_m) n_m$$

If $n_m = 0$, $N = (0, 0, \ldots, 0)$. Otherwise, we use the notation of equation 4a and a device suggested by that equation, to scale so that $n_m = 1$ and obtain

$$\tilde{n}_1 = h_1$$

$$\tilde{n}_2 = h_1 + h_2$$

$$\vdots$$

$$\tilde{n}_m = \sum_1^m h_i$$

Normalizing, i.e. rescaling through division by the sum $\sum_1^m ih_i$, we arrive at a fractional distribution $f_1, f_2, \ldots, f_m$ (following the notation of equation (2)). The values n_i are determined by $n_i = nf_i$ where as usual, n is the size of the staff.

Thus the fractions of total hires no older than the age thresholds for the age classes are proportional to the class sizes in the stable age distribution when there are no separations except from the oldest class. This "top heavy" structure from the model thus confirms intuition as in the preceding example.

A "pure" hierarchical age model in which there are no losses (separations) except from the highest age group, i.e., mandatory retirements, and in which all hires are into the lowest age group, is an example of a Markov model which does not converge to a stationary state because its transition matrix is not regular. The transition matrix is, in fact, the cycle matrix

$$\begin{bmatrix} 0 & 1 & 0 & 0 & . & . & . & 0 \\ 0 & 0 & 1 & 0 & . & . & . & 0 \\ 0 & 0 & 0 & 1 & . & . & . & 0 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & . & . & . & 1 \\ 1 & 0 & 0 & 0 & . & . & . & 0 \end{bmatrix}$$

Successive powers shift the columns one position to the right with the last column moving to the first; i.e., the columns cycle so that all powers of the matrix contain zero elements. The action of this matrix on a distribution vector permutes its elements cyclically one place in each time step. If the starting configuration happens to have equal numbers in all ages, the organization will appear to be age stable, but any other (non uniform) configuration will cycle endlessly.

Thus with a staff of 150 stratified into five age groups with ages 20-29, 30-39, 40-49, 50-59, and 60-69, mandatory retirement at 70, all hires (replacements) at age 20 and a 10-year time step, if there are 30 individuals in each group initially there will always be 30 in each group, 30 recruits at age 20 replacing the 30 retirees from the high group at each step. If, instead, the staff were initially distributed into the groups in numbers

20-29	30-39	40-49	50-59	60-69
10	20	30	40	50

the staff distribution would be successively

50	10	20	30	40
40	50	10	20	30, etc.

completely recycling each 5 periods. This example tacitly illuminates an important consideration in age models which is sometimes overlooked in applications, particularly when the transition coefficients are regarded as probabilities, to wit: the time step size must match the interval defining the age groups. If not, then serious distortions may occur, as the following simple example demonstrates.

Consider a constant-sized personnel staff stratified by 10-year age brackets: 20-29, 30-39, 40-49, 50-59, 60-69, with mandatory retirement at age 70, replacements at age 20 and no other losses. For a simple Markov transition process with 10-year time-step, the transition matrix is the matrix

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

That is, in each 10 year period the fraction of any 10-year age class that ages into the next class is 1, or in terms of probability, a person in any class at time τ will be in the next higher class with probability 1 at time $\tau+10$ years.

To represent an <u>annual</u> aging process, i.e., to use a one-year time step with the same structural description (5 classes determined by age decades), we might expect the matrix

$$\begin{bmatrix} .9 & .1 & 0 & 0 & 0 \\ 0 & .9 & .1 & 0 & 0 \\ 0 & 0 & .9 & .1 & 0 \\ 0 & 0 & 0 & .9 & .1 \\ .1 & 0 & 0 & 0 & .9 \end{bmatrix}$$

to furnish an annual aging model equivalent to the previous one. Here the row entries purport to show that on the average, annually, 1/10 of the individuals in a (10-year) class age into the next higher class while 9/10 "stay put." We see that this requires first, an additional assumption, namely, that <u>within</u> the classes, the members are divided into equal sized subclasses corresponding to the 10 years in the decade of ages, and indeed, because this process is to be iterated, that the condition must persist over time if constant transition rates are to be meaningful. An unpleasant inconsistency then results: After one year 1/10 of some age class with 200 members (i.e., 20 individuals), progresses into the next class. Then, as _this_ class is homogeneously distributed, the tenth that ages into the next class after another period, includes two of the new members so that in expectation, these two have aged over ten years in two! Moreover, this new model is no longer cyclic but has a regular matrix: the fifth power of the matrix has no zeros and so the process yields eventually a stable distribution with any initial vector. Interestingly, the stable distribution is the uniform distribution, which we recall to be the stable vector of the cyclic model, the difference, being, of course, that for stability in the cyclic system, the _initial_ distribution had to be the stable one.

In formulating the erroneous* model above, we chose an annual time step for convenience. Let us look now at some time steps larger than the class interval. A time step of twenty years would induce the transition matrix

*This age model can be defined as erroneous in terms of Markov theory in that the stipulated structure with an annual time step is in fact not really a Markov process because it has "memory" which is being ignored in the representation. That is, the individuals going from, say, the 30-39 year age class to the 40-49 year age class are in reality 40 years old, but the model "forgets" this in distributing their ages evenly through the 40-49 interval. The result is a Markov model of something, but not of the intended "age" system.

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

which satisfies intuition by cycling the distribution entries two places each step instead of one as in the base model. More generally: the matrices representing time steps which are multiples of ten are powers of the "original" or ten-year matrix. Following the procedure of the "erroneous model", y-year time step with y between zero and ten, would result in a matrix of the form

$$\begin{bmatrix} 1-\alpha & \alpha & 0 & 0 & 0 \\ 0 & 1-\alpha & \alpha & 0 & 0 \\ 0 & 0 & 1-\alpha & \alpha & 0 \\ 0 & 0 & 0 & 1-\alpha & \alpha \\ \alpha & 0 & 0 & 0 & 1-\alpha \end{bmatrix}$$

in which $\alpha = y/10$. $\alpha = 1$ corresponds to the base model of the example. $\alpha = 0$ gives a model for which every age distribution is stationary (the identity matrix I). For $0<\alpha<1$ all models yield the uniform distribution asymptotically.* If y is greater than 50 years, the model has substantially no meaning because it gives a rate matrix identical to that for y - 50 years.

The "smear" phenomenon exhibited by these models would not necessarily distort reality in situations revolving about grade or rank (instead of age), when the coefficients might represent actual promotion policies. Also, age systems with "proper" time steps, with losses (separations) at rates $1-\alpha$ and with hires for direct replacement could be represented by such models.

*It is easy to see that any such matrix is regular, for with $0<\alpha<1$ each successive power will add positive entries to a new circulant diagonal. The (n - 1)-st power, i.e. the result of n - 2 matrix multiplications, will contain no zeros.

APPENDIX B: ANALYSIS OF THE MARKOV MODEL*

B.1 Formulas for Steady-State Quantities

A principal output of the model is the limiting or "steady-state" age distribution of FTPP staff. In Section 3, we showed how that distribution could be calculated numerically by repeated application of the model's transition equations, beginning with an initial age distribution. In this self-contained technical appendix, we will give an algebraic treatment, leading to explicit formulas for the limiting distribution (and some related quantities) in terms of the model's input data. Relations among these inputs, necessary for a consistent model, will be identified. Finally, an analysis will be given of what limiting age distributions are attainable if some of the model inputs are fixed but others are adjustable.

The steady-state age distribution will be described by a vector $\vec{F} = (f_1, \ldots, f_m)$, whose components

f_i = limiting fraction of FTPP staff lying in i-th age group

are nonnegative and satisfy

(1) $\Sigma_1^m f_i = 1.$

An auxiliary model output, which will also figure in the analysis, is

t = limiting turnover rate.

Other model outputs, which might for example be of interest should the issue of age-discrimination arise, are the quantities

λ_i = limiting fraction of staff-leavers lying in i-th age group;

*By A. J. Goldman, Applied Mathematics Division.

these quantities form a vector $\vec{\Lambda} = (\lambda_1,\ldots,\lambda_m)$ whose components λ_i are nonnegative and satisfy

(2) $\Sigma_1^m \lambda_i = 1.$

Note that t and $\vec{\Lambda}$ are defined only in terms of the steady state; a turnover rate or a leaver-age distribution constant over time is <u>not</u> assumed.

The inputs to the model can be described by two vectors. The first one, $\vec{H} = (h_1,\ldots,h_m)$, gives the age distribution of new hires; its components

h_i = fraction of new hires lying in i-th age group

are nonnegative and sum to 1. The second vector, $\vec{s} = (s_1,\ldots,s_m)$, has as components the separation rates

s_i = fraction of i-th age group separated during a time period,

which satisfy the conditions $0 \leq s_i \leq 1$. Because of the mandatory retirement age, $s_m = 1$.

It follows from these definitions that

(3) $t = \Sigma_1^m s_i f_i$

and

(4) $\lambda_i t = s_i f_i.$

Equation (4) can be used to calculate $\vec{\Lambda}$, once $\vec{F}$ and t are determined. The numerical values of these various quantities, for the particular case of the NBS FTPP staff, are given in Table 6.

We will first develop formulas for $\vec{F}$ (and t) in terms of $\vec{H}$ and $\vec{S}$. For this purpose, note that in the steady state a fraction t of the total FTPP staff leaves each period; since the total staff size is assumed constant, the same fraction t of the next period's staff consists of new hires,

		MODEL PARAMETERS		INITIAL VS. STEADY STATE ("FINAL") DISTRIBUTIONS							
				TURNOVER				FTPP STAFF			
				NUMBERS		PERCENTAGES		NUMBERS (n_i)		PERCENTAGES	
CLASS	AGES	DISTRIBUTION OF HIRES (h_i)	SEPARATION RATES (s_i)	INIT.	FINAL	INIT.	FINAL	INIT.	FINAL	INIT.	FINAL
1	21-25	.1124	.6190	26	33	5.9	6.9	42	54	3.0	3.8
2	26-30	.2651	.4087	51.5	60	11.7	12.5	126	148	9.0	10.6
3	31-35	.2133	.2604	56.5	50	12.8	10.4	217	191	15.5	13.6
4	36-40	.1498	.1905	48	41	10.9	8.5	252	213	18.0	15.2
5	41-45	.0778	.1975	47	41	10.6	8.5	238	210	17.0	15.0
6	46-50	.0865	.2419	52.5	51	11.9	10.6	217	210	15.5	15.0
7	51-55	.0490	.2943	51.5	54	11.7	11.2	175	183	12.5	13.1
8	56-60	.0375	.7362	67	108	15.2	22.5	91	147	6.5	10.5
9	61-65	.0086	.9714	34	42	7.7	8.7	35	43	2.5	3.1
10	66-70	.0000	1.0000	7	1	1.6	0.2	7	1	0.5	0.1
TOTAL		1.0000		441	481	100.0	100.0	1400	1400	100.0	100.0
			5-year Turnover (%)	31.5	34.4		Average Age	42.3	42.8		
			Annual Turnover (%)	6.3	6.9						

TABLE 6. MODEL PARAMETERS, STARTING VALUES AND STEADY STATE OUTPUTS

and so a fraction $h_i t$ consists of new hires into the i-th age group. Since the time period's length coincides with the span of each group, such new hires are the only source of personnel for the first (i.e., youngest) age group, and so that group's balance equation is

(5) $f_1 = h_1 t$.

For j>1, the j-th group consists not only of new hires but also of those staff members who spent the previous period in the (j-1)-st group and were not separated; thus the balance equation is

(6) $f_j = h_j t + (1-s_{j-1}) f_{j-1} \qquad (j>1)$.

It is convenient to define

$$s'_i = 1 - s_i,$$

so that (6) can be written

(7) $f_j = h_j t + s'_{j-1} f_{j-1} \qquad (j>1)$,

and also to define

(8) $\sigma_{ij} = \Pi_{k=i}^{j-1} s'_k \qquad (\sigma_{ij} = 1 \text{ for } i \geq j)$.

Then the solution to the recurrence (7), with (5) as initial condition, is (by mathematical induction)

(9) $f_j = t \, \Sigma_{i=1}^{j} \sigma_{ij} h_i$.

To determine t we sum (9) over j and apply (1), obtaining

$$1 = \Sigma_1^m f_j = t \, \Sigma_{j=1}^m \Sigma_{i=1}^j \sigma_{ij} h_i$$

$$= t \, \Sigma_{i=1}^m h_i \Sigma_{j=i}^m \sigma_{ij}.$$

In the double sum, the coefficient of each h_i contains the positive summand σ_{ii} and so is strictly positive; since at least one h_i is strictly positive (they sum to 1), the same is true of the double sum itself and we may write

(10) $t = 1/\Sigma^m_{i=1} h_i \, \Sigma^m_{j=i} \sigma_{ij} > 0.$

It now follows from (8) and (9) that $f_j = 0$ if and only if (a) $h_j=0$ and (b) for each i with $i < j$ and $h_i > 0$, there is a k with $i \leq k < j$ such that $s_k=1$. For example, $f_1=0$ if and only if $h_1=0$ (here (b) is satisfied vacuously).

From (4) and (9) we have

(11) $\lambda_j . = s_j(f_j/t) = s_j \Sigma^j_{i=1} \sigma_{ij} h_i.$

B.2 Attainability of "Target" Distributions

We turn now to a simplest of the questions arising in connection with using the model for "control" purposes rather than just predictive ones. That question is: given one of the two vectors $(\vec{H}, \vec{S})$ as fixed, and the other as freely adjustable, which steady-state age distributions $\vec{F}$ can be attained? In other words, we regard $\vec{F}$ and one of the pair $(\vec{H}, \vec{S})$ as given, and ask whether (and which) choice of the other member of the pair will make $\vec{F}$ the solution of the balance equations. This is of course a highly idealized situation, since in practice the parameters would be neither precisely fixed nor freely adjustable, and since dynamic as well as limiting system behavior would be of interest; more realistic "scenarios" may be formulated and analyzed in subsequent work.

Suppose first that $\vec{F}$ and $\vec{S}$ are given. From (3) and (10) it follows that

(12) $s_i f_i > 0$ for at least one i

is a first compatibility condition on the pair $(\vec{F}, \vec{S})$. By (5) and (6),

(13) $h_1 = f_1/t \;, \quad h_j = (f_j - s'_{j-1} f_{j-1})/t \quad (j > 1),$

so that $\vec{H}$ is determined uniquely. The requirement that its components sum to 1 is satisfied automatically, since (13) and (3) give

$$\Sigma_1^m h_j = [f_1 + \Sigma_{j=2}^m (f_j - s'_{j-1} f_{j-1})]/t$$

$$= [(f_1 + \Sigma_2^m f_j) - \Sigma_{i=1}^{m-1} s'_i f_i]/t$$

$$= [\Sigma_1^m f_i - \Sigma_1^m s'_i f_i]/t \qquad \text{(since } s'_m = 0\text{)}$$

$$= [\Sigma_1^m (1-s'_i) f_i]/t = [\Sigma_1^m s_i f_i]/t = 1.$$

The requirement that $h_j \geq 0$ leads, by (13), to the further compatibility condition

(14) $f_j \geq s'_{j-1} f_{j-1}$

on the pair $(\vec{F}, \vec{S})$. $\vec{\lambda}$ is uniquely determined by (11).

Next, suppose $\vec{F}$ and $\vec{H}$ are given. It will prove convenient to introduce

(15) $J = \min \{j\colon h_j > 0\}$.

It follows from (5) and (6) that

(16) $f_j = 0$ for $j < J$,

and thus also that

(17) $t = f_J/h_J$,

so that (10) gives

(18) $f_J > 0$;

(16) and (18) are compatibility conditions on the pair $(\vec{F}, \vec{H})$. The choice s_m=1 is required, and for $j \leq m$, (7) yields

(19) $s'_{j-1} = (f_j - h_j t)/f_{j-1}$ (if $f_{j-1} > 0$),

so that s_i is uniquely determined if $f_i > 0$; if $f_i = 0$, then the value of s_i is immaterial for a steady-state analysis. Since $0 \leq s_i \leq 1$, (19) yields

$$0 \leq f_j - h_j t \leq f_{j-1}$$

as a further compatibility condition on $(\vec{F}, \vec{H})$; with the aid of (17), this reads

(20) $h_j f_J \leq h_J f_j \leq h_J f_{j-1} + h_j f_J$ (if $f_{j-1} > 0$).

The remaining compatibility condition, also obtained from (5) or (7) and (17), is

(21) $h_J f_j = h_j f_J$ (if $j=1$ or $f_{j-1} = 0$).

A somewhat less natural pair of problems involves stipulating a particular choice of $\vec{\Lambda}$, rather than $\vec{F}$, as the "target" to be attained. In these problems we regard $\vec{\Lambda}$ and one of the pair $(\vec{H}, \vec{S})$ as given, and ask whether (and which) choice of the other member of the pair will "realize" $\vec{\Lambda}$ in the steady state.

Suppose first that $\vec{\Lambda}$ and $\vec{S}$ are given. The compatibility condition

(22) $\lambda_i = 0$ if $s_i = 0$

is an immediate consequence of (11). For what follows, it is convenient to introduce the cumulative sums

(23) $\Lambda_j = \Sigma_1^j \lambda_i$, $H_j = \Sigma_1^j h_i$.

Using (4), we can rewrite (6) as

(24) $f_j = (h_j - \lambda_{j-1}) + f_{j-1}$ ($j>1$).

This recursion, with (5) as initial condition, has as its solution

(25) $f_j = (H_j - \Lambda_{j-1})$ ($\Lambda_o = 0$),

leading *via* (4) to

(26) $\lambda_j = s_j(H_j - \Lambda_{j-1})$.

Now let

(27) $j(1) < j(2) < \ldots < j(\mu) = m$

consist of those $j \varepsilon \{1,2,\ldots,m\}$ for which $s_j > 0$; here $j(\mu) = m$ because $s_m = 1$. It follows from (26) that

$$H_{j(\nu)} = \Lambda_{j(\nu)-1} + \lambda_{j(\nu)}/s_{j(\nu)} \qquad (\nu = 1,2,\ldots,\mu),$$

which by (22) can be written

$$(28) \quad H_{j(\nu)} = \Lambda_{j(\nu-1)} + \lambda_{j(\nu)}/s_{j(\nu)} \qquad (j(0) = 0).$$

These values automatically obey the first of the conditions

$$H_m = 1 \quad , \quad H_{j(\nu+1)} \geq H_{j(\nu)}$$

which are necessary for the H_j's to arise from an age distribution $\vec{H}$, while the second of these requirements leads via (28) to the further compatibility condition

$$(29) \quad \lambda_{j(\nu+1)}/s_{j(\nu+1)} \geq \lambda_{j(\nu)} s'_{j(\nu)}/s_{j(\nu)}$$

on the pair $(\vec{\lambda}, \vec{s})$.

If all $s_j > 0$ (i.e., $\mu = m$ in (27)) then (28) determines a unique $\vec{H}$. If $\mu < m$, then the sets

$$\{h_j : \ j(\nu) < j \leq j(\nu+1)\} \qquad (\nu = 0,1,\ldots,\mu-1)$$

are only determined by (28) "up to" their respective sums, so that there are multiple choices for $\vec{H}$, leading via (25) to different (in general) solutions $\vec{F}$ of the balance equations. In the context of (25), an interesting expression for t is obtained by summing over j and applying (1). This yields

$$\begin{aligned}
1 &= t(\Sigma_1^m H_j - \Sigma_1^{m-1} \Lambda_j) \\
&= t(\Sigma_{j=1}^m \Sigma_{i=1}^j h_i - \Sigma_{j=1}^{m-1} \Sigma_{i=1}^j \lambda_i) \\
&= t(\Sigma_{j=1}^m h_i \Sigma_{j=i}^m 1 - \Sigma_{i=1}^{m-1} \lambda_i \Sigma_{j=i}^{m-1} 1) \\
&= t[\Sigma_{i=1}^m h_i (m+1-i) - \Sigma_{i=1}^{m-1} \lambda_i (m-i)] \\
&= t[(m+1) - \Sigma_1^m ih_i - m(1-\lambda_m) + \Sigma_1^{m-1} i\lambda_i] \\
&= t[1 + A_\lambda - A_h],
\end{aligned}$$

where

(30) $A_\lambda = \Sigma_1^m i\lambda_i \quad , \qquad A_h = \Sigma_1^m ih_i$

are the average steady-state leaving age and the average hiring age, respectively, in age-group span units. Thus

(31) $t = 1/(1 + A_\lambda - A_h)$.

Next suppose that $\vec{\Lambda}$ and $\vec{H}$ are given. Then (25) and (31) determine $\vec{F}$ uniquely. We claim that

(32) $H_j \geq \Lambda_j \qquad (j=1,2,\ldots,m)$

is a necessary compatibility condition on the pair $(\vec{\Lambda}, \vec{H})$. To prove this, first note that if $t = 0$ then (5) and (24) would imply that all f_j=0, contradicting (1). Therefore $t>0$, and so the condition $f_j \geq 0$ applied to (25) yields $H_j \geq \Lambda_{j-1}$. If strict inequality holds in this relation, then (26) uniquely determines s_j as

(33) $s_j = \lambda_j/(H_j - \Lambda_{j-1}) \qquad (\text{if } H_j > \Lambda_{j-1})$

and so the requirement $s_j \leq 1$ leads to (32). If equality holds in the relation then s_j is not determined but its value is immaterial for the steady-state analysis since f_j=0 by (25); this implies λ_j=0 by (4), so that (32) holds as an equality. Thus we have the further compatibility condition

(34) $H_j = \Lambda_{j-1}$ implies $\lambda_j = 0$.

As the final step in this Appendix, we assume that both $\vec{F}$ and $\vec{\Lambda}$ are given. Interestingly, in this case, it is also possible to stipulate a "target" value of t, with $0 < t \leq 1$. It follows from (5) that

(35) $h_1 = f_1/t \quad ,$

and from (24) that

(36) $h_{j+1} = \lambda_j + (f_{j+1} - f_j)/t \quad (j < m),$

so that $\vec{H}$ is uniquely determined. The requirement $h_{j+1} \geq 0$ leads via (36) to the compatibility condition

(37) $\lambda_j t + (f_{j+1} - f_j) \geq 0 \quad (j < m)$

on $(\vec{f}, \vec{\lambda}, t)$. It follows from (35) and (36) that

$$H_m = h_1 + \Sigma_1^{m-1} h_{j+1} = (f_1/t) + \Lambda_{m-1} + [\Sigma_1^{m-1}(f_{j+1} - f_j)]/t$$

$$= (f_1/t) + (1 - \lambda_m) + (f_m - f_1)/t$$

$$= 1 - (\lambda_m - f_m/t) = 1 \quad ,$$

as desired, where the further compatibility condition

(38) $f_m = \lambda_m t$

follows from (4) since $s_m = 1$. Requirement (32), for $i = 1$, imposes via (35) the compatibility condition

(39) $f_i \geq \lambda_i t,$

for $i = 1$, while the calculation from (35) and (36) that

$$H_{i+1} = h_1 + \Sigma_1^i h_{j+1}$$

$$= (f_1/t) + \Lambda_i + [\Sigma_1^i (f_{j+1} - f_j)]/t$$

$$= (f_1/t) + (\Lambda_{i+1} - \lambda_{i+1}) + (f_{i+1} - f_1)/t$$

$$= \Lambda_{i+1} + (f_{i+1} - \lambda_{i+1} t)/t$$

shows that (39) is also the necessary and sufficient condition for (32) when $i > 1$. Given (39), the same calculation shows that (34) will be satisfied. The separation rates s_i are determined to the same extent and in the same manner as at the end of the previous paragraph; requirement (29) turns out to be equivalent to (37).

NBS-114A (REV. 11-77)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET	1. PUBLICATION OR REPORT NO. NBSIR 78-1458	2. Gov't Accession No.	3. Recipient's Accession No.

4. TITLE AND SUBTITLE Projecting the Age Distribution of Full-Time Permanent Professional Staff at the National Bureau of Standards	5. Publication Date March 1978 6. Performing Organization Code NBSIR 78-1458
7. AUTHOR(S) L. S. Joel	8. Performing Organ. Report No.
9. PERFORMING ORGANIZATION NAME AND ADDRESS NATIONAL BUREAU OF STANDARDS DEPARTMENT OF COMMERCE WASHINGTON, D.C. 20234	10. Project/Task/Work Unit No. 2050151 (7120151) 11. Contract/Grant No.
12. Sponsoring Organization Name and Complete Address *(Street, City, State, ZIP)* Same as Item 9	13. Type of Report & Period Covered Final 14. Sponsoring Agency Code

15. SUPPLEMENTARY NOTES

16. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)*

This report presents a simple mathematical model to project the age distribution of the full-time permanent professional (FTPP) staff of the National Bureau of Standards. The report includes a brief description of types of models currently in use for manpower analyses, discussion of the probable data requirements for reliable models, some staff profile information which supplements material in recent administrative reports, and the description of our model. The principal projection is that under "status quo" assumptions, (FTPP staff size, age distribution of hires, and the separation rates from age cohorts remain constant) the FTPP staff will in about 25 years reach a steady-state age distribution with average age about 1/2 year higher than its current level (just over 42 yrs.). Comparison of this distribution with the present one shows moderate increases in the fractions of staff in ages 21-30 and 51-60, a fractional rise in turnover, and a general decrease in the groups 31-50 years old. A near term effect is the intensified loss of senior scientists.

17. KEY WORDS *(six to twelve entries; alphabetical order; capitalize only the first letter of the first key word unless a proper name; separated by semicolons)*

Data analysis; discrete Markov models; manpower planning models; simple Markov models

18. AVAILABILITY [X] Unlimited	19. SECURITY CLASS (THIS REPORT)	21. NO. OF PAGES
[ ] For Official Distribution. Do Not Release to NTIS	UNCLASSIFIED	70
[ ] Order From Sup. of Doc., U.S. Government Printing Office Washington, D.C. 20402, SD Stock No. SN003-003	20. SECURITY CLASS (THIS PAGE)	22. Price
[X] Order From National Technical Information Service (NTIS) Springfield, Virginia 22151	UNCLASSIFIED	$5.25