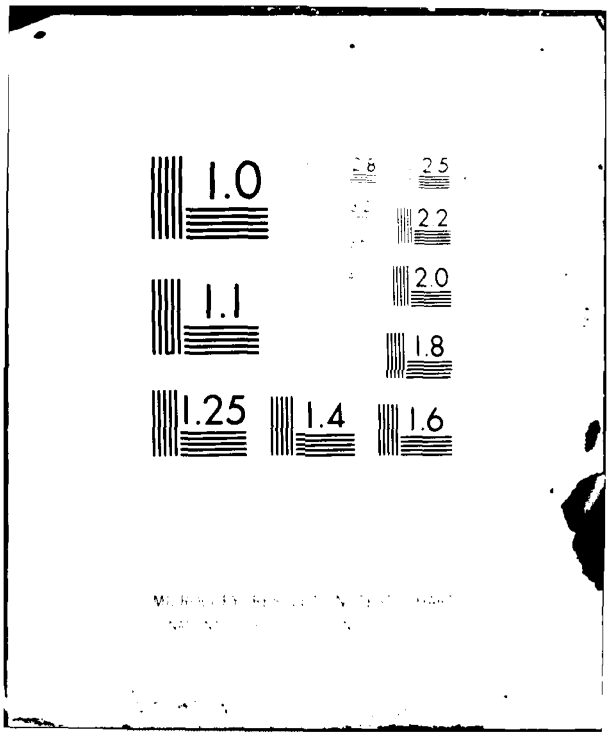
AD-A115 934 RAND CORP SANTA MONICA CA F/G 15/5
UNCERTAINTY IN PERSONNEL FORCE MODELING.(U)
APR 82 G J HALL, S C MOORE F49620-82-C-0018
UNCLASSIFIED RAND/N-1842-AF NL

1 OF 2
AD A
5934


1.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

A RAND NOTE

UNCERTAINTY IN PERSONNEL FORCE MODELING

Gaineford J. Hall, Jr.
S. Craig Moore

April 1982

N-1842-AF

Prepared For

The United States Air Force

Rand
SANTA MONICA, CA. 90406

The research reported here was sponsored by the Directorate of Operational Requirements, Deputy Chief of Staff/Research, Development, and Acquisition, Hq USAF, under Contract F49620-82-C-0018. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

Published by The Rand Corporation

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER N-1842-AF	2. GOVT ACCESSION NO. AD-A[illegible]934
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) Uncertainty in Personnel Force Modeling	5. TYPE OF REPORT & PERIOD COVERED Interim
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Gaineford J. Hall, Jr., S. Craig Moore	8. CONTRACT OR GRANT NUMBER(s) F49620-82-C-0018
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Rand Corporation 1700 Main Street Santa Monica, CA. 90406	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Requirements, Programs & Studies Group (AF/RDQM) Ofc, DCS/R&D and Acquisition Hq USAF, Washington, D.C. 20330	12. REPORT DATE April 1982
	13. NUMBER OF PAGES 107
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

No Restrictions

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Air Force Personnel
Enlisted Personnel
Force Structure Planning
Markov Processes
Mathematical Models
Management Planning

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See Reverse Side

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Most military personnel planning models are deterministic steady-state models. This Note examines the impact of various types of uncertainties on projections of force structures using a Markov flow model of the first-term force. In particular, it addresses the impact of uncertainties related to the supply of enlisted Air Force personnel (stay/leave decisions by or about individual airmen, the makeup of accession cohorts, retention rate estimation, and recruiting shortfalls) on force planning factors such as accession requirements, reenlistment requirements, and personnel costs. The analysis indicates that projections of many force characteristics can involve sizeable uncertainties. Individual stay/leave decisions comprise the largest source of this uncertainty. Another potentially larger contributor is uncertainty in the proportion of accession requirements that can actually be met. Uncertainties regarding estimates of flow rates, while important in projecting values for certain subsets of the force, appear to contribute little to uncertainty in overall force characteristics.

12

A RAND NOTE

UNCERTAINTY IN PERSONNEL FORCE MODELING

Gaineford J. Hall, Jr.
S. Craig Moore

April 1982

N-1842-AF

Prepared For The United States Air Force

Rand
SANTA MONICA, CA. 90406

PREFACE

This Note was prepared as part of Rand's "Enlisted Force Management" project. The project, which is part of the Project AIR FORCE Resource Management Program, is being performed for the Directorate of Personnel Plans, Headquarters, United States Air Force. The purpose of the project is to develop the specifications for an enlisted force management planning system to replace the Air Force's current system known as TOPCAP (Total Objective Plan for Career Airman Personnel).

All of the models in the TOPCAP system are deterministic--i.e., they ignore the uncertainty inherent in projecting both the demand (requirements) for manpower and the supply of personnel that will be available to meet the demand. Before undertaking serious development of new models, the authors carried out an investigation of the degree of uncertainty implicit in personnel flows. The investigation was to evaluate the need for incorporating uncertainty in the new models, and to consider alternative ways of doing so.

This Note focuses on uncertainty in the supply of personnel: stay/leave decisions of airmen, the composition of accession cohorts, retention rates, and recruiting shortfalls. It discusses the effect of uncertainty on the relationships between these variables and such work force characteristics as accession requirements, reenlistment requirements, and costs. The analytical tool used is a Markov chain model representing flows in the first-term enlisted work force.

This Note should be of interest to personnel planners in the Air Force and the other armed services, as well as to analysts developing models for use in analyzing manpower and personnel policies in both the public and the private sectors.

SUMMARY

The Air Force's TOPCAP system (Total Objective Plan for Career Airman Personnel) includes a number of models that describe the flow of people through the enlisted work force. These models are deterministic (e.g., they ignore the uncertainty implicit in personnel loss projections) and most of them are steady-state (i.e., they ignore the current enlisted personnel "inventory" and its evolutionary possibilities). In this Note we address the impact on work force structure uncertainty of various factors related to the "supply" of personnel: stay/leave decisions by or about individual airmen, the makeup of accession cohorts, retention rate estimation, and recruiting shortfalls. Specifically, we analyze the impact of uncertainty in these random quantities on work force characteristics such as accession requirements, reenlistment requirements, and personnel costs. Further, these relationships are examined over time. We also discuss methods for improving estimates of "flow rates" for personnel planning models and for estimating how these rates will change under altered personnel policies. The intent is to ascertain appropriate directions for extending Air Force personnel planning models, with particular regard to uncertainty, rather than to describe the actual amounts of uncertainty that exist or to demonstrate or evaluate different methods of dealing with uncertainty.

We evaluate the extent of uncertainty in projections of work force structures using an analytical Markov flow model that focuses on the first-term enlisted work force. Our analysis indicates that projections for many work force characteristics can involve sizeable uncertainties. Two-standard deviation confidence intervals often contain values differing 10 to 40 percent from corresponding expected (mean) values. Individual stay/leave decisions comprise the largest source of this uncertainty. Another potentially large contributor is uncertainty in the proportion of accession requirements that actually can be met. Uncertainties regarding the mix of people that can be accessed and regarding estimates of flow rates, while important in projecting

values for certain subsets of the work force, appear to contribute less to uncertainty in overall work force characteristics.

Since uncertainties in projecting the values of these characteristics can be substantial, there may also be substantial uncertainty in predicting the effects of policymakers' decisions. This leads to the question of assessment of risk--that is, the problem of determining how far off mean value calculations are likely to be or of determining the likelihood of certain undesirable events (e.g., unusually large accession quantities or required reenlistment rates). We conclude that if "protection" from undesirable events is important, it can be obtained by adding to deterministic flow models constraints determined using stochastic post-processors that could compute the approximate probabilities of certain events and/or of actual results differing from mean value estimates by specified amounts.

We also recommend that improved procedures be developed for estimating probabilistic parameters in personnel flow models--e.g., loss rates. Improved methods should provide consistent, interpretable, and parsimonious sets of parameters for estimating flow rates, they should incorporate time series data (in order to detect underlying trends), they should include "environmental" data such as occupational categories and corresponding civilian economic conditions, and they should admit to statistical goodness-of-fit procedures.

Finally, we recommend that recently-developed retention decision models (for the Air Force Officer Corps) be revised and extended to predict how the flow behaviors for the various categories of enlisted personnel will change if management "control" policies such as compensation, promotion opportunity, educational benefits, or retirement programs are changed.

ACKNOWLEDGMENTS

We gratefully acknowledge the help of Lieutenant Colonel Ronald Kerchner and Major James Hoskins, who recognized the potential importance of uncertainty in USAF personnel flow modeling and provided valuable advice in deciding exploratory modeling questions. We are grateful to our Rand colleagues Michael D. Miller, who established an important mathematical simplification, and Glenn Gotz and Cindy Williams, who contributed substantially through their reviews of the manuscript.

CONTENTS

FIGURES

TABLES

1. INTRODUCTION

Among the primary models employed in planning and programming for the enlisted component of the Air Force work force, many are deterministic models of personnel flow. These models are part of the TOPCAP System (Total Objective Plan for Career Airman Personnel) and include, for example, the Objective Force Model ("OBFOR"), the Airman Force Steady State Model ("the Static Model"), the Promotion Flow Model ("the Dynamic Model"), the Five Level Redistribution Program ("FLRP"), the Career Progression Group Model ("CPG Static"), and the Airman Skill Force Model ("ASKIF II"). Several of these models are static (i.e., steady-state); they represent the work force structure that should eventually develop if management policies, manpower requirements, retention behavior, upgrade and promotion rates, etc., remain unchanged. In addition, practically all of these models treat only the career portion of the enlisted work force--those individuals serving beyond their initial enlisted term of service (which usually lasts four years).

In discussing possibilities for the form and structure of extensions to the capabilities represented in the TOPCAP models, especially with regard to improving analysis capabilities focusing on the first-term work force, analysts at Rand and in the Air Force jointly agreed that uncertainty must be considered directly. Many of the inputs (e.g., retention rates) and virtually all of the outputs of such models (e.g., reenlistment rates and annual recruitment projections) are subject to uncertainty--often because they must be estimated using sample data and/or because they may depend on future decisions made by or about individual airmen. (Depending on the use of the information, stay/leave decisions made by many thousands or perhaps only a few airmen may be of interest.)

Uncertainty warrants special attention for two primary reasons, one technical and the other having both technical and decisionmaking ramifications. The first reason is that personnel flow models often employ nonlinear equations to relate random quantities. For example, "required" first-term reenlistment rates are nonlinear functions of requirements for career force entrants, accession quantities, and retention rates. The expected value of one random variable in such a relationship, unfortunately, cannot be found by replacing the other random variables with their expectations and solving the resulting equation--at least not in general. We determined to address this "nonlinearity of expectations" problem before beginning to develop further deterministic models that ignore it. The second reason for examining uncertainty relates to possible alternative forms for stating decision criteria and for specifying corresponding model objective and/or constraint functions. For example, we might wish to consider management options only if they provide confidence of certain events occurring--e.g., a 90% chance that a particular career progression group (CPG) would require a reenlistment rate no higher than 45% in 1983. Or we might want to consider a policy change only if it is likely to yield results substantially different from the current policy--e.g., if A_X and A_Y represent a particular CPG's accession requirement in 1983 under alternative policies X and Y, respectively, we may want to assure a 90% chance that A_X exceeds A_Y by, say, 5 percentage points. Alternatively, we may simply want to know the probability that our expected-value estimates will be off by particular amounts--e.g., what is the probability that our estimate of the number of people in the fourth year of service for 1984 is off by 10% or more? Assessment of uncertainty is obviously essential if information of this type is to be provided to decisionmakers, and underlying mathematical structure is clearly affected if "chance constraints" are to be incorporated in planning models.

Since these concerns about uncertainty arose during consideration of first-term planning models, much of this Note's discussion refers to aspects of the first-term work force. Nevertheless, the concepts, recursive equations, etc. can be generalized easily to include the career work force as well.

The next section reviews previous research regarding uncertainty in personnel flow models. We then address the sources of uncertainty in personnel flow models, develop analytical means for evaluating the extent of these uncertainties using a Markov model, and present sample descriptive results based on an implementation of the model that focuses on the first-term force. Identifying appropriate categories of enlisted personnel and the flow rates among the categories is critically important in personnel flow models. Section 4 addresses statistical methods for discerning categories of airmen whose retention behaviors (flow rates) differ and for predicting how their behavior would change under altered management policies (e.g., larger reenlistment bonuses, revised compensation tables, improved promotion opportunities, or altered retirement benefits). We conclude with recommendations for how uncertainty should be incorporated in future personnel flow models and for statistical/behavioral modeling work (related to personnel "supply") that should be accomplished and incorporated in improved Air Force personnel planning and programming models.

2. RELEVANT PREVIOUS WORK

Personnel flow models that incorporate uncertainty typically are either Markov chain models, renewal models, or simulation models. Markov chain models are "push" models in the sense that flow occurs due to natural progression from one state (e.g., pay grade or length of service) to the next, and due to the influx of new personnel into the system. Renewal models are "pull" models driven by the need to fill vacancies. Simulation models can be either push or pull models or combinations of the two.

Markov chain models assume that from one observation to the next the net changes in state exhibit the Markov property. That is, the probability of changing from one state to another depends only on the current state, and not on how that state was reached. Such models describe how changes (transitions) occur between states from one time point to the next using transition probabilities or proportions. They generally are composed of three components (see, for example, Bartholomew and Forbes [5]):

- o A description of how flows take place within the system (specified by transition probabilities);
- o A description of how attrition occurs from the system (specified by attrition rates);
- o A specification of the number of recruits at each point in time and the allocation of recruits to different states or categories.

The two primary uses of a Markov chain model are prediction of future behavior of the system (assuming no change in the parameters) and control of the system through policy changes (e.g., by altering recruitment, changing promotion rates, or expanding or contracting certain categories). Thus the problem of control arises naturally as a consequence of prediction.

In the Markov chain model, the transition rates typically are fixed, while the numbers of people in different states (or categories) change over time. In contrast, the renewal model has fixed numbers of people in the different states, while the flows are allowed to vary. A renewal system model can possess awkward mathematical features, but simple renewal systems can be handled within the Markov chain framework (Bartholomew [4]). The central assumptions of the renewal model relate to wastage flow (i.e., losses, or departures from the work force) and how this flow depends on length of service. Since for the first-term enlisted force the state (category) sizes are not fixed, a renewal model seems less appropriate than a Markov chain model. The renewal model would be more appropriate for studying a long-term system, such as the career force, since vacancies play a larger role in career force management.

Renewal models seem inappropriate in our context for other reasons as well:

- First-term airmen move lock-step through their first term of obligation (their promotion and upgrade times vary little);
- They are promoted more on the basis of length of service and skill qualification than on the availability of vacancies (vacancies become more important at the higher enlisted grades and in officer grades);
- The primary Air Force policy controls applied to the first-term work force concern enlistment and reenlistment, and these depend mainly on length of service and overall numerical requirements rather than on vacancies;
- The main vacancy aspects of first-term work force modeling concern the needs for (a) career force entrants to sustain a desirable and stable career work force structure and (b) raw recruits to achieve an overall work force of specified size--and both of these can be incorporated, as we shall see, in a Markov model.

In the steady state, it is difficult to distinguish between Markov chain and renewal models from state size and flow rate data alone. This is because state sizes achieve equilibrium in the Markov chain model and flow rates achieve equilibrium in the renewal model. Thus, as described in Bartholomew [4], in the steady state either model performs equally well with regard to expected values. The distinction may become crucial, however, in modeling the transient behavior of the system. Moreover, the Markov chain model permits much easier evaluation of standard deviations (a measure of uncertainty) for the random quantities of interest.

A third method for incorporating uncertainty in personnel flow modeling employs simulation models (or Monte Carlo models, as they are widely termed). Simulation models are typically employed when computationally more efficient methods prove inadequate in representing the details of system operation. A recent example of a simulation approach to personnel flow modeling--indeed to Air Force work force modeling--is the Integrated Simulation Evaluation Model [14]. This model, however, has as its primary aim the prediction of central tendencies--expected numbers of accessions, promotions, transfers, etc.--rather than of variations around expectations. Simulation modelers quite often ignore such uncertainty, although there is a substantial literature regarding variance reduction techniques (see, e.g., Fishman [10]). In contrast, our aim is to examine the extent and sources of uncertainty or "spread" that are ignored in widely-employed deterministic personnel flow models. Further, the relative simplicity of the flows in typical personnel planning models makes the computation expense of full-fledged simulation techniques unnecessary. As will be seen later, however, we do resort to simulation to augment our analytic Markov model in one situation (in order to consider uncertainty in estimates of the transition and accession probabilities) because the analytic stochastic model simply becomes too complex. We find that combining analytic and simulation models to represent uncertainty achieves a desirable degree of economy in both analysis and computation time.

The bulk of the open literature on personnel flow models employs the Markov chain structure but ignores its implicit uncertainty. Instead, the focus is on expected values, and most treatments could more accurately be termed deterministic fractional flow models than Markov chain models. Grinold and Marshall [12], for example, in a longitudinal comparison of two cohorts of U.S. Marine Corps entrants, note a "significant divergence" in the numbers remaining after several years. Probabilistically, however, that divergence should not be thought uncommon--i.e., it could not be characterized as "statistically significant."

The most notable publications regarding uncertainty in personnel flow models are by Bartholomew (see, e.g., Refs. 1 through 6). Mainly, however, Bartholomew merely catalogs the various sources of uncertainty and suggests circumstances where they are likely to be important; he generally does not show how to evaluate the extent of uncertainty. But he does note ([5], p. 110) that stochastic variation can be quite large and its analysis difficult:

> the errors in forecasts are likely to be quite large--the variances of the predictions being of the same order as the predicted values themselves. On top of this there is a further source of error arising from the fact that in most applications some, at least, of the parameters have to be estimated.... This source can give rise to errors of a similar magnitude to the random error arising from the stochastic assumptions of the model. This takes no account of the uncertainties of yet another kind arising from changes in the parameters which may occur during the forecast period. The whole question of how to cope with uncertainty in manpower planning is a complex one....

Finally, we note that previous work is concerned more with long-term (steady-state) than near-term (dynamic) aspects of work force modeling. Air Force enlisted personnel management is conducted in a notably dynamic environment, however, so our analysis examines the nature and size of uncertainty in this setting.

3. A RUDIMENTARY STOCHASTIC PERSONNEL FLOW MODEL

To ascertain the significance of uncertainty in nonlinear relationships, its magnitude, and the contributions of its various sources, we have developed a basic Markovian flow model that represents a simple, first-term work force. The sources of uncertainty considered (all of which are ignored in current Air Force personnel planning models) are:

1. Attrition (Retention) Behavior -- uncertainty due to the fact that the work force consists of individuals, and the numbers of these individuals who elect to leave the service or whose service the Air Force elects to terminate cannot be known precisely in advance. To illustrate, a projected first-year loss rate really represents a probability that an individual recruit, chosen at random, will not complete his or her entire first year of obligated service.
2. Accession Mix -- uncertainty due to the fact that the proportion of new recruits possessing particular characteristics--e.g., designated according to educational background, sex, race, marital status, or mental aptitude (characteristics that may correlate with retention behavior, productivity, and/or cost)--cannot be known precisely in advance. For example, a valuable input to a first-term personnel flow model might be the fraction of recruits having at least a high school education; in fact, this fraction estimates the probability that a new recruit, chosen at random, will have completed high school.
3. Parameters -- uncertainty due to the fact that the values employed to represent retention and accession probabilities are themselves only estimates. Depending on the size and nature of the historical data sample used to estimate these probabilities, there may be considerable uncertainty

regarding their actual values. Further, prediction of the values of these parameters under altered management policies (e.g., revised compensation patterns or promotion opportunities) introduces additional uncertainty.

4. Costs -- uncertainty due to the fact that expenditures to support different individuals within the same general category may vary. For example, within a particular year of service and within the same occupational specialty, individuals' pay and benefits may differ because they hold different pay grades, have different family situations, experience different health problems, etc.

While there may be numerous other sources of uncertainty, important in some situations (e.g., differences in job or task performance capabilities among apparently similar individuals), these are not considered here.

We should note early that the model is analytically based. It employs time-recursive relationships among key work force characteristics; individual behavior is not simulated in a Monte Carlo sense. Rather, probabilistic group behavior is considered. We resort to simulation only when necessary--namely in addressing the contributions of uncertainty implicit in underlying parameter estimates.

Recent consensus also emphasizes the importance of dynamics in personnel flow models. Most current personnel models are static rather than dynamic; hence they can yield only steady-state results concerning composition of the work force, attrition, and associated costs. In contrast, dynamic models can provide these results along with information about how long it may take to achieve a steady-state (static) distribution and information about the behavior of the system along the way. Our model has a dynamic structure, permitting us to investigate how uncertainty varies with time.

The model developed here incorporates several simplifying assumptions, notably:

1. A Four-Year Term of Service. Although airmen may enlist for either four or six years and although other enlistment terms are certainly possible, we treat a four-year first-term enlistment because that is currently the primary mode and because our model is designed for exploratory rather than descriptive use.

2. Specialty-Specific Categorization. Because (1) we expect uncertainty to be more significant when smaller personnel groups are considered, (2) manpower requirements typically are specified for individual occupations, and (3) it is computationally simpler to ignore the crosstraining and direct assignment channels which can move individuals from one occupation to another, we proceed as if we are considering only a single occupation. [Note: as structured, the model certainly can be employed to represent larger personnel aggregations, but at the expense of accuracy in representing individual occupations.]

3. No Cross-Flow Among Categories. Again, although transitions such as changes in marital status, number of dependents, skill level, pay grade, etc., certainly occur for individuals within the first-term enlisted work force, we exclude them here in the interest of simplicity. The model can be extended to include such transitions in a fairly straightforward manner.

4. Fixed Work Force Size. Because our intent is to investigate the uncertainty inherent in personnel flow models and not the uncertainty implicit in (often fluctuating) manpower requirements, we treat the total size of the first-term work force being considered as constant. It would be a straightforward extension to allow this size to vary over time due to planned requirements changes. One extension we do incorporate later is the possibility of recruiting shortfalls; that is, although the required number of people may not change, the number having the proper qualifications who can actually be recruited may be inadequate to bring end-strength up to the desired level.

These seemingly restrictive assumptions clearly can be relaxed if it is desired to build a model embodying more of the detailed reality of possible personnel flows. But we believe the present model can provide the necessary insight into the magnitude and importance of uncertainty in personnel flow models.

Here we will focus on determining means and variances for important work force characteristics such as accession quantities, year-group sizes, required reenlistment rates, and costs. We focus on means because of the potential nonlinearity problem and because they are the conventional indicators of central tendency. Variances (and standard deviations), correspondingly, are the usual indicators of uncertainty or "spread," and are typically more tractable computationally than alternative measures of uncertainty. Ideally, of course, we should obtain entire probability distributions for the quantities of interest, but that seems neither necessary nor practical for this exploratory analysis.

The remainder of this section describes the basic model more explicitly; detailed mathematics are relegated to Appendix A. We begin by describing the model's basic inputs: attrition rates, accession mix, etc., and their probabilistic interpretations. Then follows a brief discussion of the uncertainty associated with important parameters in the model and how it is represented. The third subsection describes the model's computed outputs, and the last subsection presents example results and relevant observations and conclusions.

3.1. Model Inputs

In this section we give a brief overview of the model and describe its basic inputs. The fundamental model inputs, each described below, are:

- Subdivisions of the work force, characterized by year of service (YOS) i and category m.
- Work force size.
- Attrition rates.
- Accession mix.

- o Costs.
- o Planning horizon.

Subdivisions of the Work Force. For simplicity we assume that all airmen enlist for a four-year period. Thus, for any calendar year t, an airman belongs to year-of-service (YOS) i where $1 \leq i \leq 4$. Individual airmen can be categorized according to any number of characteristics such as education (e.g., high school graduate), race, marital status, mental aptitude test scores, pay grade, AFSC,† etc. Transitions between categories are not incorporated in the current model or computer program (although they could be included in a fairly straightforward extension). Thus, for current purposes, individual characteristics which may change over time (e.g., skill level, pay grade, or marital status) should not be considered as category-distinguishing characteristics. Within the model, airmen in each calendar year t are distinguished by their YOS i $(1 \leq i \leq 4)$ and category m $(1 \leq m \leq M)$. Notationally, we let $N_{im}(t)$ = number of airmen in YOS i and category m, for calendar year t. This number is generally a random variable.

Work Force Size. In this model it is assumed that the work force is kept constant at size N. Thus, we assume initially that the Air Force can enlist as many airmen as necessary to keep its force size fixed. N is a variable whose value is chosen by the decisionmaker. For each calendar year t, we have

$$N = \sum_{m=1}^{M} \sum_{i=1}^{4} N_{im}(t) .$$

We also treat in our examples and in Appendix A a case where the work force size is a random variable--in particular, we admit the possibility of recruiting shortfalls.

Attrition Rates. Attrition is treated by supposing that each individual in YOS i and category m stays in the service another year with probability p_{im} (so that the probability of attrition

---

†Air Force Specialty Code, basically an occupational designator.

between YOS i and $i + 1$ is $1 - p_{im}$). The individual stay/leave decisions, whether made by the Air Force or the airman, are assumed to occur independently of one another. Thus, for a particular personnel group (say YOS i and category m), the number remaining from one year to the next has the following conditional binomial probability distribution:

$$P(N_{i+1,m}(t + 1) = k \mid N_{i,m}(t) = n) = \binom{n}{k} p_{im}^{k}(1 - p_{im})^{n-k} .$$

Accession Mix. In treating accessions into the first year of service, the model assumes that we first observe the total number of people leaving the force from each year of service and each category. Hence, the number of people needed to enter the first YOS in calendar year $t + 1$ to keep the force size fixed at N is

$$L(t + 1) = N - \sum_{j=2}^{4} \sum_{m=1}^{M} N_{jm}(t + 1) .$$

Equivalently, enough airmen must be enlisted to ensure that

$$\sum_{m=1}^{M} N_{1m}(t + 1) = L(t + 1) .$$

Now each $N_{1m}(t + 1)$, $1 \leq m \leq M$ is a random variable whose distribution must be determined. We consider two methods for modeling these random variables:

- Fixed proportion model. Here it is assumed that $N_{1m}(t + 1)$ is a fixed fraction of $L(t + 1)$. Let $\pi_1, \ldots, \pi_M$ be positive numbers such that $\sum_{m=1}^{M} \pi_m = 1$. (These are the accession mix parameters). If $L(t + 1)$ takes the value

k , this approach assumes $N_{1m}(t + 1) = \pi_m k$.

- o <u>Multinomial model</u>. In this perhaps more realistic model, we take the parameters $\pi_1, \ldots, \pi_M$ as representing probabilities. In particular, we assume that $N_{11}(t + 1), \ldots, N_{1M}(t + 1)$ are jointly multinomially distributed with parameters $L(t + 1)$ and $\pi_1, \ldots, \pi_M$. The conditional probability density function is

$$P\,(N_{11}(t + 1) = n_1, \ldots, N_{1M}(t + 1) = n_M | L(t + 1) = k)$$

$$\equiv f(n_1, \ldots, n_M) = \frac{k!}{n_1!\ n_2!\ \ldots\ n_M!}\ \pi_1^{n_1} \cdots \pi_M^{n_M}$$

where $n_1, \ldots, n_M$ are nonnegative integers such that

$$\sum_{m=1}^{M} n_m = k\ .$$

Thus, <u>given</u> $L(t + 1)$, the variable $N_{1m}(t + 1)$ is not random using the fixed proportion model, but it is random using the multinomial model. However, since $L(t + 1)$ itself is random, $N_{1m}(t + 1)$ is actually random in both models.

<u>Costs</u>. Since the expense of maintaining the work force is of obvious interest, cost values are included as model inputs. A cost $C_{im}(t)$ is associated with each YOS i and category m during planning year t. $C_{im}(t)$ is the yearly cost for one airman with characteristics (i, m) . Since in fact, different airmen in the same class (i, m) may be compensated differently, depending on marital status, pay grade, etc., our model treats $C_{im}(t)$ as a random variable. The question of uncertainty in costs is treated in greater detail in the next section.

<u>Planning Horizon</u>. Since the model is dynamic as well as stochastic, the length of the planning horizon is an input value. Thus, the model may be used to answer questions concerning enlistment quantities, reenlistment rates, costs, etc., five years from now,

ten years from now, etc. If no policy or behavioral changes occurred, the long-run solutions would eventually converge to the steady-state (equilibrium) answers.

3.2. Treatment of Uncertainty in Input Parameters

We now address the question of modeling the uncertainty in several of the parameters just mentioned: the retention rates (the p_{im}'s), accession mix (the π_m's) and costs (the $C_{im}(t)$'s). The values of these parameters are uncertain because they can be estimated only from historical data, and historical estimates themselves possess some innate variability. One method of treating uncertainty in these parameters is explained more fully in Appendix A. There we hypothesize that the parameter estimates are based on one year's observation of a work force of size N, the same as we assume for future work force sizes. (This is convenient for exploratory computational purposes and is consistent with Air Force use of the most recent year's data for estimating future retention rates, etc.) We denote the number of people from this historical data set in year of service i and category m as n_{im}. The estimate $\hat{p}_{im}$ of the retention probability p_{im} has a variance inversely proportional to n_{im}. The estimate $\hat{\pi}_m$ of the accession mix parameter π_m has variance inversely proportional to $\sum_k n_{1k}$.

For our computer code, the estimate of p_{im} has been modeled as if it had a normal distribution with mean p_{im} and variance $p_{im}(1-p_{im})/n_{im}$. Future refinements of the model should consider any dependence among the estimates of the retention rates and should model the distribution of their estimates more precisely. (The normal approximation is generally good for large n_{im}, but not for small n_{im}.) The assumption that the estimates of p_{im} and $p_{j\ell}$ for $m \neq \ell$ are independent is probably fairly reasonable, since events concerning individuals in different categories are likely to be independent of one another. The assumption that the estimates of p_{im} and $p_{i+1,m}$ are independent is perhaps less reasonable, and this assumption merits further investigation. (This is beyond the scope of this study, but can be considered within the context of the statistical behavioral models discussed in Section 4.)

The estimates of $\pi_1,\ldots,\pi_M$ cannot be treated as if they were independent, since $\pi_1 + \ldots + \pi_M = 1$. For reasons described in Appendix A, the distribution of the estimates of $(\pi_1,\ldots,\pi_M)$ can be modeled as a Dirichlet distribution.[†] In this case, the variance of the estimate of π_m is inversely proportional to $\sum_k n_{1k}$ and the correlation between the estimates of π_m and π_ℓ, $m \neq \ell$, is negative.

There also can be considerable uncertainty in the costs $C_{im}(t)$ associated with the different classes (i, m). This is because an individual selected at random from class (i, m) may be payed considerably more (or less) than another individual from the same class, depending on the pay grade each holds, the number of dependents each has, etc. Our model assumes that the variances $\tau_{im,im} = \mathrm{Var}\,[C_{im}(t)]$ and the covariances $\tau_{im,j\ell} = \mathrm{Cov}\,[C_{im}(t), C_{j\ell}(t)]$ have the form

$$\tau_{im,im} = g(i)^2$$

$$\tau_{im,j\ell} = \rho g(i)\, g(j)$$

where $g(\cdot)$ is some suitably chosen positive function, and $0 \leq \rho \leq 1$. Thus the correlation between two classes (i,m) and (j,ℓ) is ρ. If uncertainty in cost turns out to be important, more work will be required to determine a more realistic formulation for the variances and covariances of its components.

[†]The random variables $X_1,\ldots, X_M$ have a Dirichlet distribution with parameters $\alpha_1 > 0,\ldots, \alpha_M > 0$ if their joint density has the form

$$f(x_1,\ldots, x_M) = \frac{\Gamma(\alpha_1 + \ldots + \alpha_M)}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_M)} x_1^{\alpha_1 - 1} \ldots x_M^{\alpha_M - 1}$$

where

$\Gamma(\cdot)$ is the gamma function, each $x_m > 0$, and $\sum_{m=1}^{M} x_m = 1$.

3.3. Computational Outputs

The primary computational outputs of our model are:

1. The means (expected values) and covariances of the numbers of individuals in the various classes (i, m) for each year t (these include accession quantities); i.e., $E(N_{im}(t))$ and $\mathrm{Cov}\,[N_{im}(t), N_{j\ell}(t)]$.
2. The mean value and standard deviation of the number of accessions for each year t ; i.e.,

$$E\left[\sum_{m=1}^{M} N_{1m}(t)\right] \quad \text{and} \quad \left\{\mathrm{Var}\left[\sum_{m=1}^{M} N_{1m}(t)\right]\right\}^{1/2} .^{\dagger}$$

3. The mean value and standard deviation of the required overall reenlistment rate for each year t. We shall assume that the desired reenlistment quantity, say R, is known. This value may represent the number of people required to enter the career portion of the work force in order to achieve some personnel structure there (e.g., the "objective" work force structure identified using OBFOR for a particular CPG). For convenience, we take the reenlistment target R as proportional to the fixed first-term force size N; i.e., we use $R = cN$ or $R_m = c_m N$, respectively, depending on whether we are considering an overall or a category-specific reenlistment rate. Consequently the required reenlistment rate is defined as the quotient of the reenlistment target R (fixed) and the total size of the fourth-year group (random), i.e., $\omega_t = cN/\sum_{m=1}^{M} N_{4m}(t)$, and we wish to compute the mean $E\omega_t$ and standard deviation $\{\mathrm{Var}\ \omega_t\}^{1/2}$ of the required reenlistment rate.

---

† For a random variable X, we will use both EX and E(X) for the mean of X, and both Var X and Var(X) for its variance.

4. The mean value and standard deviation of the reenlistment rate for each category m, for each year t; i.e.,

$$E\,\omega_{mt} \quad \text{and} \quad \{\mathrm{Var}\,(\omega_{mt})\}^{1/2} ,$$

where

$$\omega_{mt} = c_m\, N/N_{4m}(t) .$$

5. The mean value and standard deviation of the cost of the first-term force for each year t; i.e.,

$$E\,C(t) \quad \text{and} \quad \{\mathrm{Var}\,(C(t))\}^{1/2} ,$$

where

$$C(t) = \sum_{i=1}^{4} \sum_{m=1}^{M} C_{im}(t)\, N_{im}(t) .$$

The model can evaluate these quantities for the cases where the p_{im}'s are fixed or random, the π_m's are fixed (both the fixed proportion model and the multinomial model) or random, and the costs $C_{im}(t)$ are fixed or random.

3.4. Results and Observations

In this section we present results, conclusions, and recommendations reached from exercising a computerized version of this model--using a range of assumptions and input parameters.

3.4.1. Parameter Inputs

For exploratory purposes, we use total force sizes of $N = 100$, 500 and 1,000 and a planning horizon T of 10 years. For illustration, we assume that personnel are classified according to two characteristics (e.g., A = educational background and B = sex), each characteristic having two levels or values (e.g., nongraduate or high-school graduate and male or female, respectively). Thus personnel are subdivided into M=4 categories or cells. Table 1 presents the list of accession mix parameters.

Table 1

ACCESSION MIX PARAMETERS

	B_1	B_2
A_1	0.40 (π_1)	0.05 (π_2)
A_2	0.45 (π_3)	0.10 (π_4)

Thus π_1 = 0.40, π_2 = 0.05, π_3 = 0.45, π_4 = 0.10 -- i.e., 10% of new accessions have the second "value" for characteristic A and the second "value" for characteristic B. (Throughout, the data we employ are only indicative. The intent is to explore the potential importance of uncertainty, not to measure it precisely.)

The retention probabilities p_{im} are listed in Table 2. To illustrate, p_{32} = .95 reflects an estimate that an average of 95% of the personnel in Category 2 who complete 3 years of service will complete their 4th year of service.

Table 2

RETENTION PROBABILITIES

	Category			
YOS	1	2	3	4
1	0.900	0.920	0.880	0.850
2	0.920	0.950	0.910	0.880
3	0.940	0.950	0.940	0.900

As explained earlier (Section 3.2) when the accession mix parameters and the retention probabilities are treated as random quantities (assumed to be estimated from historical data), their variances are assumed to be proportional to $1/(\sum_k n_{1k})$ and $1/n_{im}$, respectively, where n_{im} is the number of people in YOS i and category m in the historical sample.

Specifically, recall that the retention probability for the i^{th} YOS and cell m is modeled as a normal random variable with given mean p_{im} and variance $p_{im}(1 - p_{im}) / n_{im}$, and that the accession parameters are modeled as a Dirichlet distribution, so that the accession parameter for cell m has mean π_m and variance $\pi_m(1-\pi_m) / (\sum_k n_{1k} + 1)$. Since (for realism) we wish n_{im} to vary proportionately to the work force size N (and for convenience and ease of computation), we introduce proportionality parameters γ_{im}, defined as $\gamma_{im} = n_{im}/N$, or $n_{im} = \gamma_{im}N$.

The fixed values of γ_{im} are displayed in Table 3. Thus, for example, if N = 100, the number of people n_{23} (in the historical data set) in the second YOS and third category is about 12 (so the variance of p_{23} is proportional to 1/12). For N = 1,000, this number is about 124 (so the variance of p_{23} is proportional to 1/124).

We have characterized the uncertainty in the accession mix parameters and the retention rate estimates as depending on N, the size of

Table 3

PROPORTIONALITY PARAMETERS

	Category			
YOS	1	2	3	4
1	0.1290	0.0161	0.1452	0.0323
2	0.0903	0.0181	0.1239	0.0258
3	0.1016	0.0068	0.0903	0.0271
4	0.0890	0.0155	0.0735	0.0155

the work force being considered. For convenience, the initial work force structure N_{im} (0) is also directly proportional to N. To begin the computation we must input the initial force $\{ N_{im} (0) \}$. We set $N_{im} (0) = N\delta_{im}$, where the relative mix parameters δ_{im} are listed in Table 4. For example, if N = 1,000, then initially there are 100 people in Category 3 in their second year of service.

Table 4

RELATIVE MIX PARAMETERS

YOS	Category 1	Category 2	Category 3	Category 4
1	0.130	0.020	0.120	0.040
2	0.090	0.030	0.100	0.030
3	0.100	0.020	0.080	0.040
4	0.080	0.030	0.070	0.020

The expected costs EC_{im} are listed in Table 5.

Table 5

EXPECTED COSTS

YOS	Category 1	Category 2	Category 3	Category 4
1	7,400	7,200	7,400	7,300
2	8,200	8,000	8,200	8,000
3	9,100	8,900	9,100	8,900
4	9,600	9,500	9,600	9,500

Here we assume, only for simplicity, that costs do not depend on time t.

The standard deviations of the costs for each category are displayed in Table 6.

Table 6

STANDARD DEVIATIONS OF COSTS

YOS i	Standard Deviation
1	600
2	550
3	700
4	800

To obtain the covariances between costs for different categories and years of service as described in Section 3.2, we set $\rho = 0.80$.

Using these example data, we turn our attention to our two primary concerns, the "nonlinearity of expectations" problem and the associated uncertainty in computed estimates of (random) quantities. In our analysis we examine five different cases corresponding to increasing levels of uncertainty in the model:

Case 0: The completely deterministic case with no uncertain parameters, no uncertainty in stay/leave decisions, and no uncertainty in the distribution of accessions among the m categories.

Case 1: Uncertainty in stay/leave decisions (characterized by known p's), and proportional accessions with known π's.

Case 2: Same as Case 1, but assuming multinomial accessions with known π's.

Case 3: Same as Case 2, but with estimated (random) p's and π's.

Case 4: Same as Case 3, but with random L* (accession shortfalls are allowed).

In Case 4, the actual number of accessions L^* is assumed to be a random variable with mean α L and variance β L, where L is the required number of accessions to keep the total force size fixed at N (see Appendix A for details).[+] In the limit (as the time horizon grows) the expected value of the total force size is slightly higher than α N. In our analysis we chose $\alpha = \beta = 0.90$.

3.4.2. Nonlinearity in Means

Current personnel force flow models typically involve nonlinear equations that relate descriptive random variables (such as accessions, attrition rates, reenlistments and costs), but use only mean values of these random variables in computations and completely ignore the associated uncertainties in the results. One of the primary goals of our research has been to determine the effects of uncertainties in these variables on the computed outputs of such models.

We have found that for the limited (but representative) parameter values we have employed, the "nonlinearity in expectations" problem is not serious. The computed means vary little among the four cases examined. It can be seen from the equations in Appendix A that except for reenlistment rates, there is no difference in means between cases 1 and 2. (Required reenlistment rates are an exceptional case since they are inverses of random quantities. Reenlistment rates will be discussed in greater detail below.) The reason is that the computations of the mean involve the same values of π in the fixed-proportion and multinomial cases and do not involve the variances. Case 0, the deterministic case, will have the same mean as cases 1 and 2 (again, except for reenlistments). Moreover, it can be shown from results in Appendix A that the fourth case is "linear"--that is, any change in the mean is due to the "shortfall" parameter α and not to a nonlinear equation. Thus only in case 3 (randomness in retention and mix parameters) is there a

---

[+]We assume that $0 \le L^* \le L$. Further, as can be seen from the equations in Appendix A, the results only depend on the mean and variance of L^*, and not on the form of its probability density function.

"nonlinearity in expectations" problem. But for accessions, year-group size and costs, the means vary little between case 2 and case 3. As can be seen from the tables in Appendix B, the difference in means (for N = 1,000) is generally much less than 1%. Even for N = 100, the difference in means for accessions between case 2 and case 3 is at most 2%-4%. For fourth-year group size (still N=100), the difference for the total group and also the large cell is 2%-4%; for the small cell it is at most 16%. However, since the expected cell size is only about 1.2, this 16% difference is uninteresting. For costs, the difference is generally 0.2% or less. It can happen for these three quantities that the mean may either increase or decrease between case 2 and case 3. As is expected, there is always a decrease in mean between case 3 and case 4.

It is also of interest to examine the quotients of computed means for the different work force sizes--e.g., comparing the accessions mean for N = 1,000 with that for N = 100. For all computed quantities (accessions, fourth-year-group size, and cost) we found that the ratio of the computed mean for N = 1,000 to that for N = 100 was always 10/1 for case 1 and case 2 (as is expected). However, for cases 3 and 4, the ratio is not generally 10/1, but may vary from 9.75 (accessions into a large cell) to 11.82 (for the fourth-year group in a small cell). This is due to the fact that the variances of the p's and π's are taken as inversely proportional to the work force size and not due to any underlying nonlinearity of expectations.

The mean required reenlistment rate is the expectation of the inverse of a random variable. The function 1/x is more curved when x is closer to zero. Thus if the mean of the fourth-year-group size is small, we could expect the mean required reenlistment rate to be affected by the "nonlinearity problem." But the variance of the fourth-year-group size also is potentially important. The mean reenlistment rates can be approximated by a Taylor series expansion of the function 1/x. As shown in Appendix A, the approximation involves the use of both the mean and variance of the fourth-year-group size (either in total or by cell). Since the variance of the fourth-year-group size increases over all four cases (as described below), and since the fourth year-group mean decreases between case 3 and case 4, the mean required reenlistment rate steadily increases from case 1 to case 4. However, this increase in mean is so slight as to be almost negligible, with the exceptions of case 4 and/or small cell size.

For N = 1,000, the change in mean reenlistment rate for the total group and for the large cell is at most 2% across the 5 cases; for the small cell it is at most 14%. For N = 100, the total reenlistment rate mean varies at most 3%; for the large cell it varies at most 10%; the computed mean reenlistment rate for the small cell exceeds 1, because of the Taylor approximation, and hence is not meaningful.

In summary, our analysis shows that the "nonlinearity in means" question is generally no real problem, and that using expected values in nonlinear equations to relate various quantities such as accession quantities, reenlistment rates, fourth-year-group size and cost is a safe practice, except when N (and consequently some expected cell sizes) is small.

3.4.3. Increase in Uncertainty

Although the degree of nonlinearity now appears of little practical concern, the magnitude of uncertainty can be substantial--at least warranting explicit consideration. As shown in the next section, actual values of various random quantities of interest can be quite different from their expected values. This undercuts confidence in mean-value predictions of accession quantities, reenlistment rates, etc.

To assess the magnitude of uncertainty associated with our outputs, we use the coefficient of variation (CV)--the standard deviation of a quantity divided by its mean. Notationally, this is $CV = \sigma/\mu$. Thus we may write the mean "plus or minus" two standard deviations as $\mu \pm 2\sigma = \mu(1 \pm 2CV)$. As can be seen from this formula, all we need do to assess the magnitude of uncertainty is to compare the CV (or $2 \cdot CV$) with 1.

Appendix B consolidates important outputs from numerous model runs. For all four stochastic cases and for each year in the 10-year planning horizon, the means, standard deviations, and coefficients of variation are tabulated for total accessions, fourth-year-group size, required reenlistment rate, and annual cost. For illustration, the first three measures are also tabulated for two subsets of the work force: the largest and smallest "cells" (e.g., male high-school graduates and female nongraduates, respectively). Generally, the CVs increase yearly, they typically approach their limiting values after about four years (primarily because the model assumes a four-year term of service), they increase from case 1 to case 4, they are substantially larger when N is smaller, and they are larger for subsets of the work force than in total.

For total accessions, we found that the CV increases slightly over the first three cases, but that for case 4 it is double the value of case 1. For case 4, the CV (10th year) is 0.316 for N=100 and 0.103 for N=1000. The large increases are from case 0 to case 1 (the simple addition of binomial choices) and from case 2 to case 3 (randomness in the p's and π's). For the large cell (cell 3) the CV doubles from case 1 to case 3 and increases moderately for case 4. All four sources of uncertainty contribute significantly. The results for the small cell (cell 2) are even more dramatic. Here we find that in some cases we get CVs greater than one, for N=100.

Figure 1 displays, as a function of work force size, the mean plus or minus twice the standard deviation for total accessions and for accessions into the largest and smallest personnel categories (cells). The graph is based on case 3 (uncertainty in attrition behavior, accession mix, and in estimation of the p's and π's). Note that uncertainty is substantially larger, relatively, for smaller subsets of the work force.

For total fourth-year-group size, the CV generally increases by a factor of 2 to 5 from year 1 to year 10. It increases slightly over the first three cases, then nearly doubles for case 4. For case 4, the CV ranges from 0.392 for N = 100 to 0.104 for N = 1000. The large

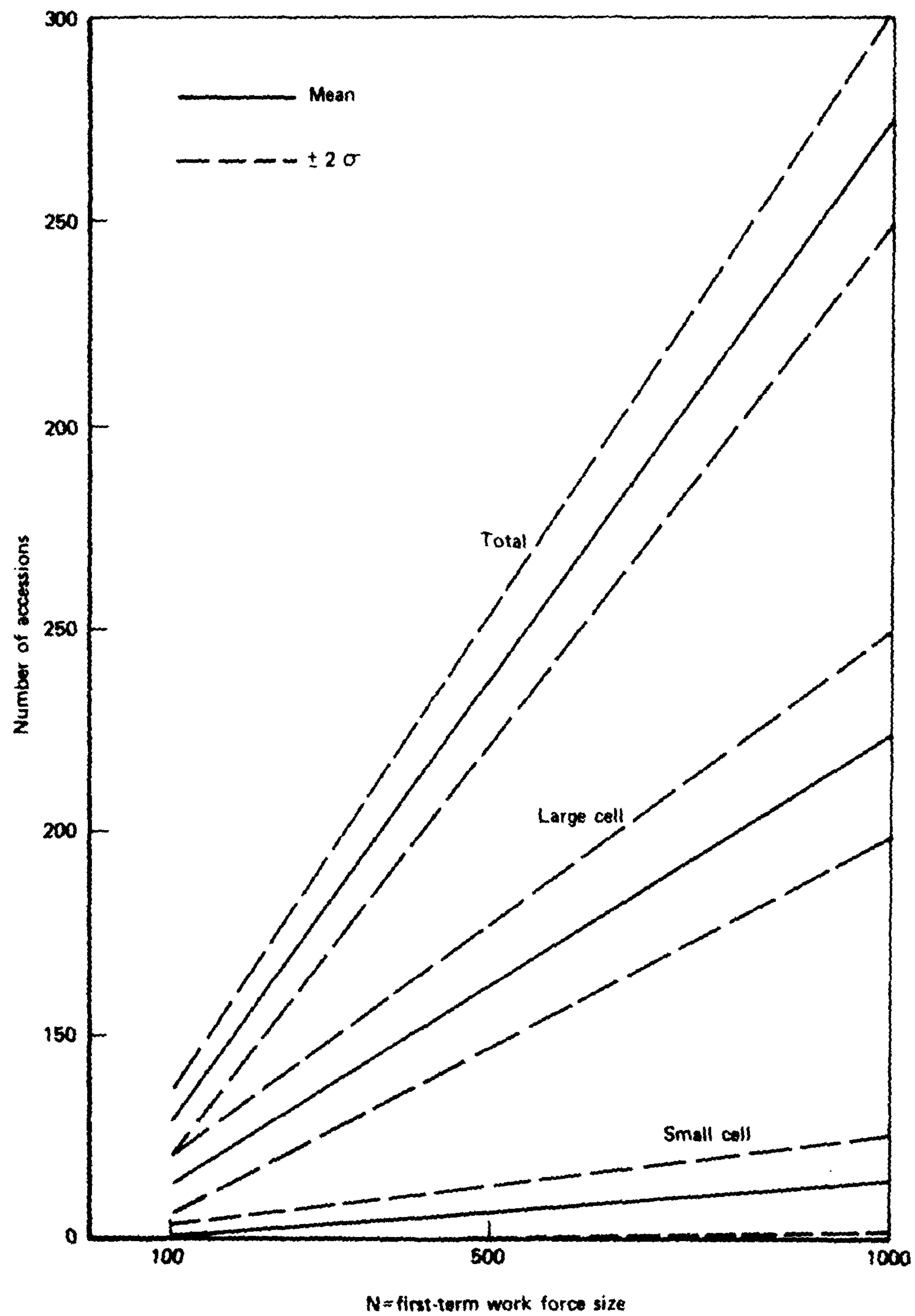


Fig. 1 — Expectations and uncertainties in accession quantities
(fifth year in the planning horizon)

increases are from case 0 to case 1 and from case 3 to case 4. The CV for the fourth-year group in cell 3 (the large cell) increases substantially over all four cases. It increases by a factor of more than 2 between case 1 and case 4. In case 4 the CV ranges from 0.460 for N = 100 to 0.148 for N = 1000. For the small cell the CV increases by a factor of three from case 1 to case 4. It is greater than 1 (for N=100) for both cases 3 and 4.

For total required reenlistment rates, the CV increases only slightly for the first three cases, then nearly doubles for the fourth case. In case 4, the CV varies from 0.297 for N = 100 to 0.103 for N = 1000. Here, again, the large increases are due to case 1 and case 4. For the large cell, each case contributes significantly to the increase in CV. For the fourth case, CV ranges in value from 0.379 for N = 100 to 0.145 for N = 1000. For the small cell, the CV again increases significantly for each case. For case 1, the CV is 0.346 at N = 100 and is 0.125 at N = 1000. For Case 4, it is 0.482 and 0.344, respectively.

Figure 2 displays a confidence band for the overall required reenlistment rate as a function of the work force size N for the fifth year in the planning horizon (again for case 3). Recall that this rate is uncertain because it depends on the random number of people who actually complete four years of service; the desired (required) reenlistment quantity is held fixed. Note that the confidence band widens considerably when a small work force is considered but remains fairly stable even for a fairly large first-term work force. (Recall again that these work force sizes are representative for many Air Force occupational specialties.)

As may be seen from Appendix B, the coefficient of variation for the overall total cost was only about 7% (0.07). Since the input standard deviations of the individual costs $C_{im}(t)$ were on the order of about 7% of the means (the C_{im}'s), this would indicate that the cost results are dominated by these inputs. To verify this, we doubled the input standard deviations of cost and found that the CV of the overall total cost approximately doubled to 0.14. Thus, cost uncertainties are sensitive to the ascribed values of these standard

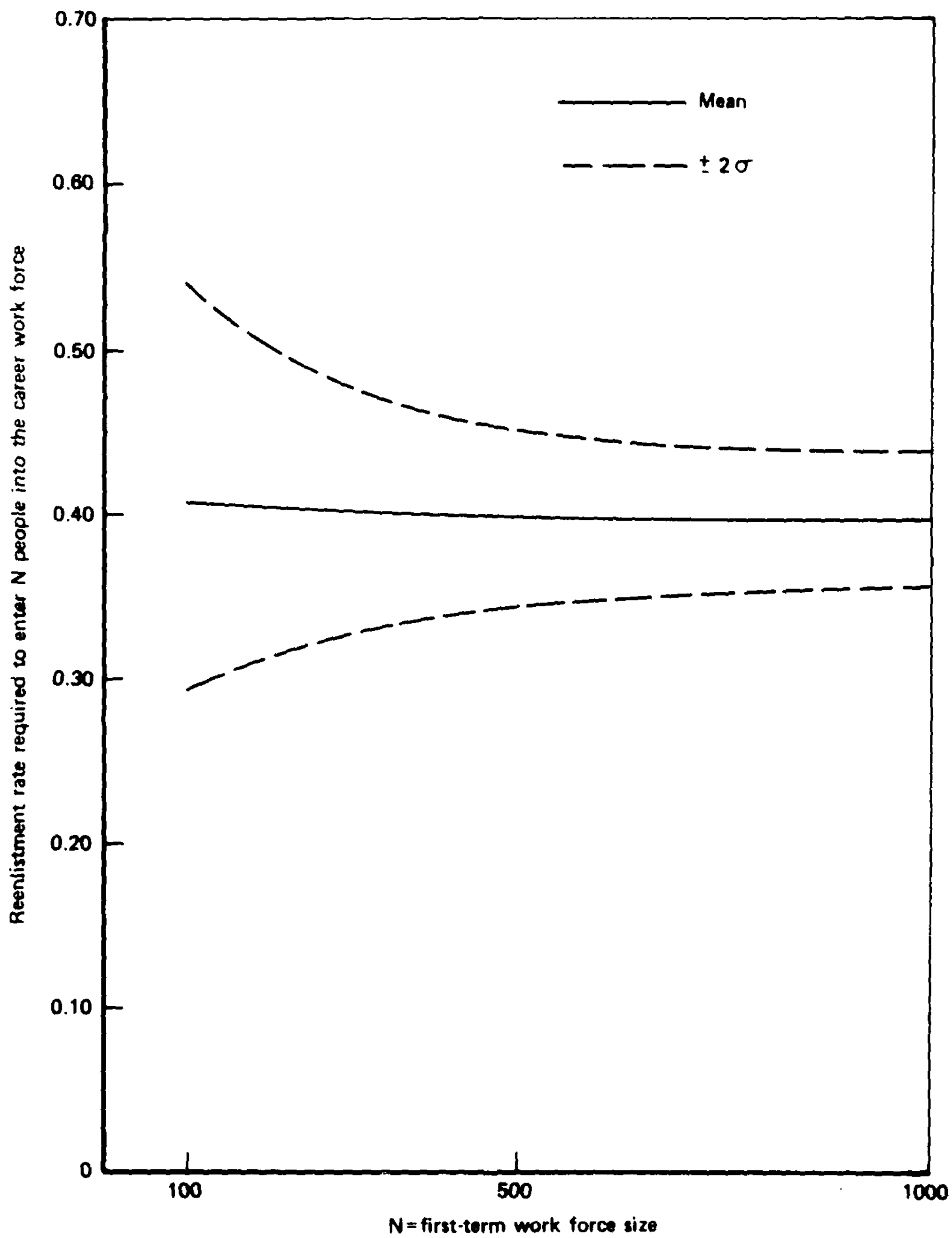


Fig. 2 – A confidence band for required reenlistment rate
(fifth year in the planning horizon, Case 3)

deviations, and, consequently, their accurate prediction requires good estimates of these standard deviations. Such estimates could be obtained in practice by using random samples from the Uniform Airman Record (UAR).

Our findings regarding the increase in uncertainty due to the successive sources in cases 1-4 are typified graphically in Fig. 3. This display shows the coefficient of variation in each case for total accessions, required reenlistment rate, and cost. Uncertainty increases most notably with smaller work force sizes and with the assumption in case 4 of enlistment shortfalls.

Finally, Figs. 4 and 5 reflect the dynamics involved in our model. Figure 4 is a plot of total accessions versus time for the ten-year horizon, and Fig. 5 is a plot of total required reenlistment rate versus time. Again, both figures are based on case 3. The variation over time apparent in these graphs is due to the fact that the initial work force configuration is not the mean equilibrium configuration. This "zigzagging" dampens over time, and the expectations of these quantities eventually would converge to stable equilibrium values.

3.4.4. Assessment of Risk

These results indicate that uncertainties in projecting the values of several work force characteristics can be substantial. We are naturally led to wonder about the probabilities of certain events occurring. For example, what is the probability that more than X people will have to be recruited in 1984 in order to maintain a first-term work force of a specified size, or what is the probability that a reenlistment rate higher than 40% will be required in 1987 in order to enter Y people into an occupation's career work force? Using the means and variances identified with our stochastic flow model, we can approximate these probabilities. We do this by using the method of moments to estimate the parameters of probability distributions that approximate those of the subject random variables. For example, since $N_4(t) = \sum_m N_{4m}(t)$, the number of people in the fourth year of

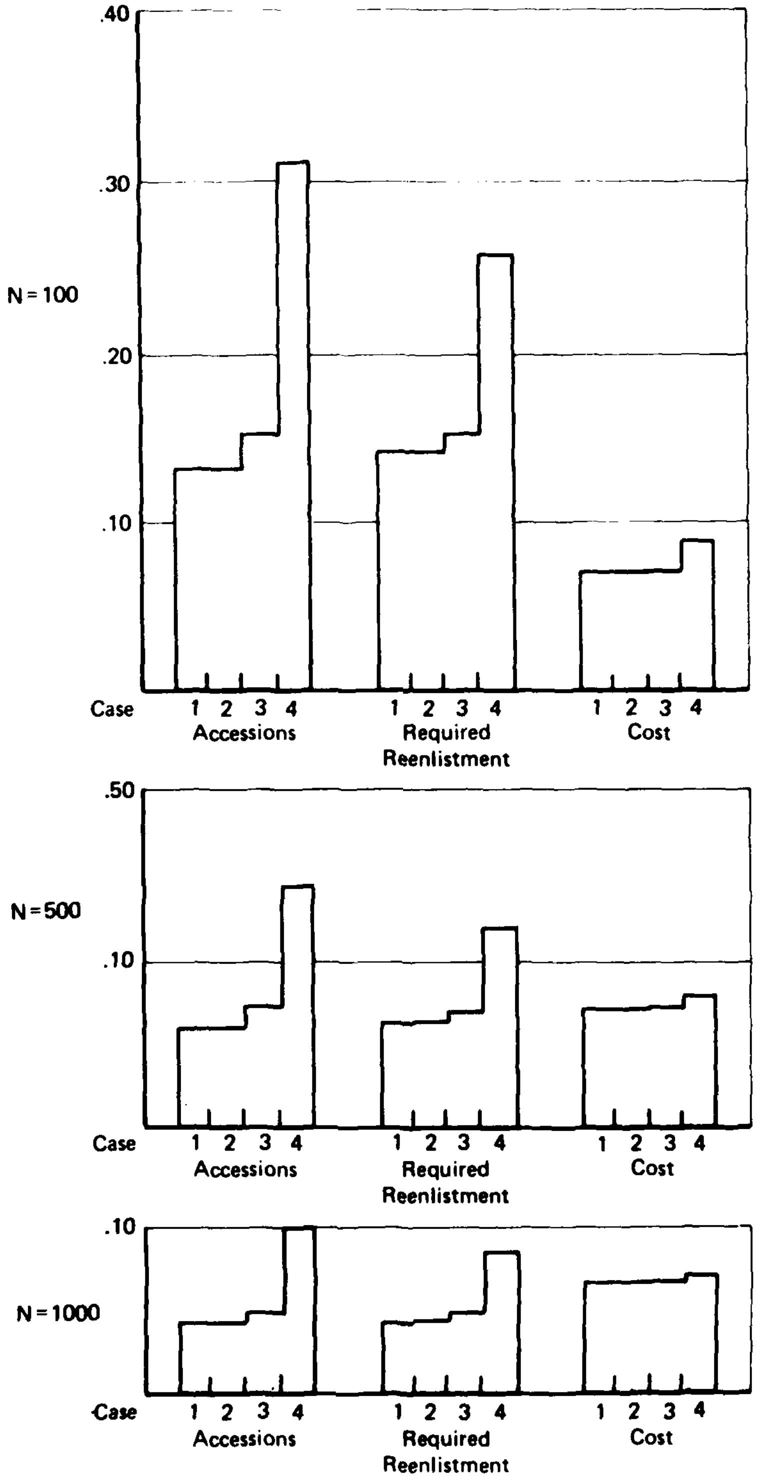


Fig. 3—Coefficients of variation (fifth year)

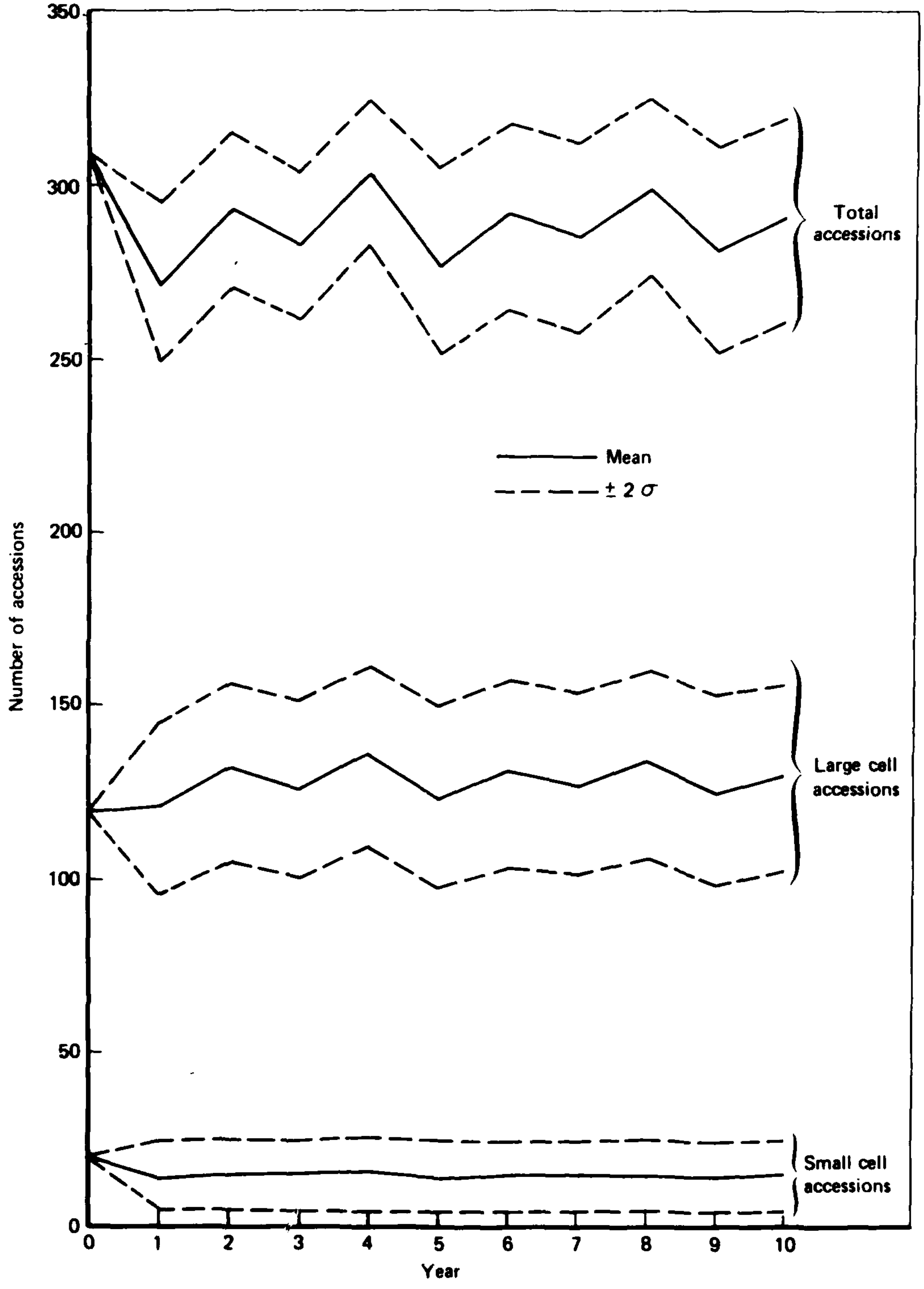


Fig. 4—Accessions versus time (N = 1,000, Case 3)

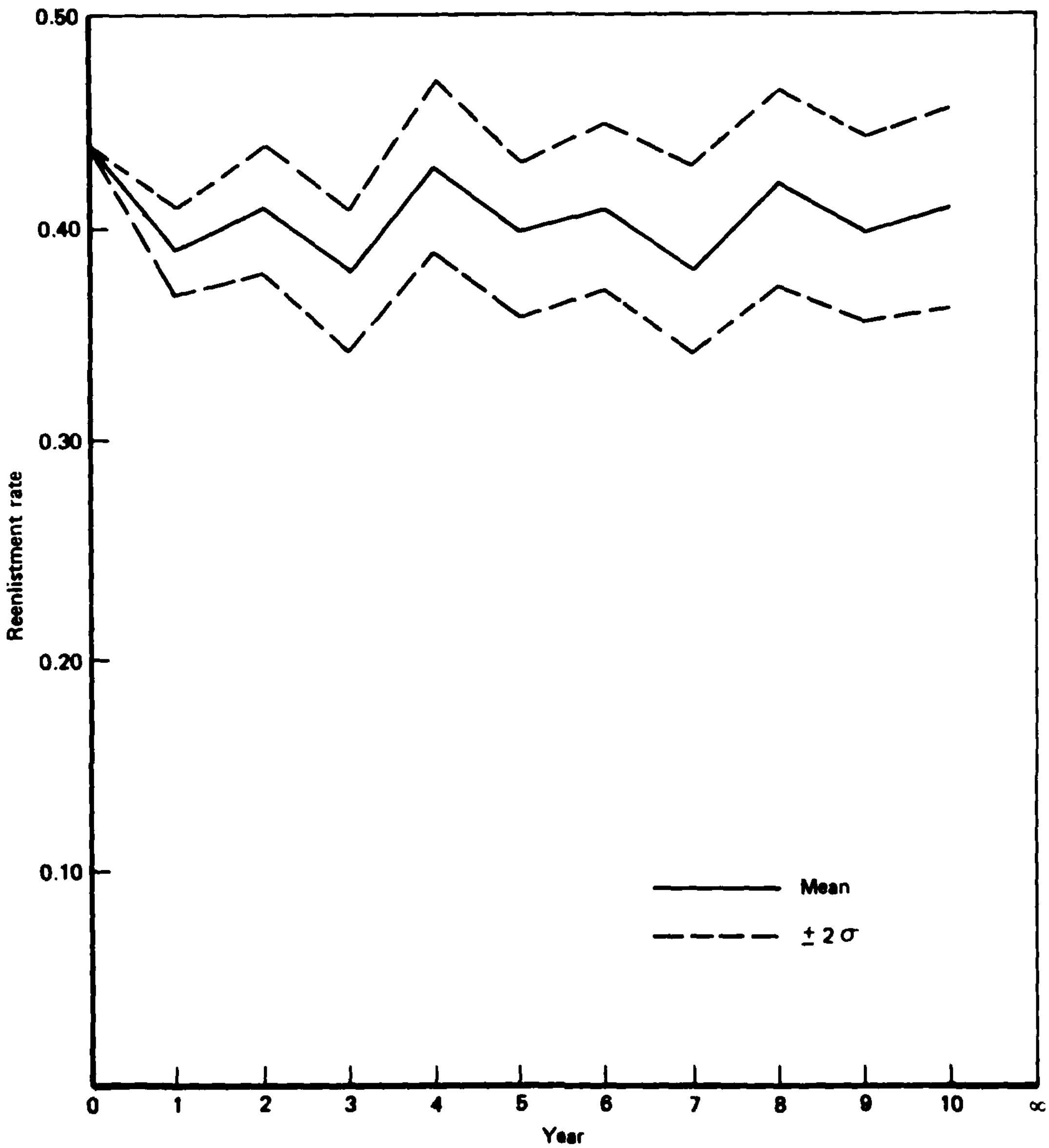


Fig. 5 – Required reenlistment rate versus time (N = 1,000, Case 3)

service in year t, is determined as the result of a large number of decisions by or about individual airmen, we can expect that it follows a normal probability distribution approximately. As an illustration, consider $N_4(t)$ in our example case with N = 100, multinomial accessions, and uncertainty in the estimates of π_m and p_{im} (i.e., case 3). In this case $E[N_4(5)] = 21.8$ and $\sqrt{Var[N_4(5)]} = 3.3$. If we assume that $N_4(5)$ has a normal probability distribution, then we can determine the (approximate) probability that, say, a reenlistment rate higher than 40% is required if 9 people are to enter the career force in year 6:

$$P(9/N_4(5) > .4) = P(N_4(5) < 9/.4 = 22.5)$$
$$\approx P(Z < (22.5 - 21.8)/3.3 = .212) \approx 0.584,$$

where Z is the standard normal random variable.

In this example, suppose that a required reenlistment rate as high as 40% is something to be avoided, for example, because it may require a reenlistment bonus. In this case (and in general) we may ask what is the probability that our expected-value estimates will be off by particular amounts--e.g., what is the probability that our estimate of $N_4(5)$ $(E[N_4(5)]$ is off by 10% or more? Figure 6 provides a ready means for determining such error probabilities, providing the random variable of interest can be assumed to follow a normal distribution approximately. In this example, the coefficient of variation of $N_4(5)$ is about 0.15 ($3.3/21.8 \approx 0.15$), and the graph indicates a probability of about 0.48 that the actual value of $N_4(5)$ will differ from $E[N_4(5)]$ by at least 10%. (Note: 0.48 represents an interpolation between the CV = .10 and CV = .20 curves plotted in Figure 6.) Naturally, the smaller the percentage error we consider, the higher its corresponding probability. But the smaller the coefficient of variation for the subject random variable, the lower the probability of error. The dimensionless nature of the graph in Figure 6 permits its use for estimating the

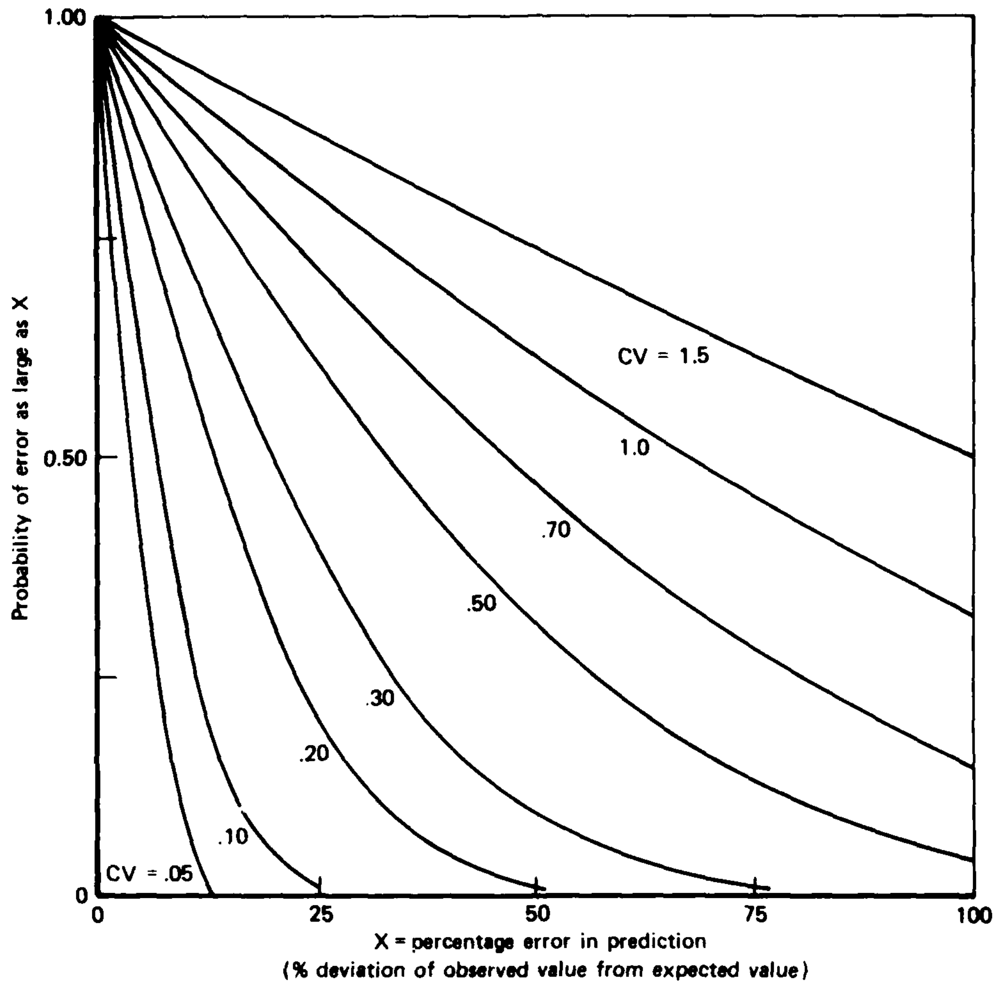


Fig. 6 — Probability of prediction error using expected values

error potential implicit in any of the mean values calculated by our flow model--as long as the corresponding random variables follow approximately normal distributions.

It would be comparatively easy to include this kind of risk-assessment capability in a post-processor for deterministic personnel flow models. That is, in addition to calculating the standard deviations (and more accurate expected values) for work force characteristics, a stochastic post-processor could compute the approximate probabilities of certain events and/or of actual results differing from mean-value estimates by specified amounts.

As indicated in this study's introduction, Air Force analysts and planners might desire to operate personnel flow models in such a way that they preclude management actions which admit unacceptable risks. For example, they may wish to establish recruiting levels which give high probabilities that the numbers of people subsequently available for reenlistment will be sufficient to make a reenlistment bonus unnecessary. How difficult would it be to construct flow models that could identify options providing protection from risk? To examine the complexity involved in the necessary calculations, let us use the example mentioned; that is, we seek to determine some number, A, of people that should be recruited to assure a probability of at least b that a reenlistment rate no higher than r will be required to reenlist c of these people for the career force. More specifically, we want to find the minimum value a such that

$$P(R = c/Y \leq r \mid A = a) \geq b,$$

where R is the (random) required reenlistment rate of interest, and Y is the number of people remaining after four years of service of those a who are recruited initially. Since Y has a conditional binomial distribution (conditioned on a) in case the first-term annual retention probabilities are known with certainty, then the normal approximation to the binomial distribution can be employed to obtain†

$$a = \frac{2pc/r + z_b^2 p(1-p) + \sqrt{[2pc/r + z_b^2 p(1-p)]^2 - 4\, p^2 c^2/r^2}}{2\, p^2},$$

where $p = \sum_m \pi_m p_{1m} p_{2m} p_{3m}$ represents the probability that an individual recruit makes it through his initial four-year obligation, and z_b satisfies $P(Z \geq z_b) = b$. For example, if c = 9, p = .75, r = .4, and

†See Appendix C for the derivation.

$b = .90$, we find $a = 34.33$ or, rounding up to an integer, 35. This contrasts with the 30 people (determined as $9/[(.4)(.75)] = 30$) we would obtain from a deterministic treatment assuming a reenlistment rate of .4 could be attained. Thus, in this case, "insurance" against having to offer a reenlistment bonus costs about 17% in added accessions and overall (expected) first-term work force size and costs. This example is representative of a work force whose first-term component contains about 100 airmen. For a first-term work force of about 1000 airmen, we can change c, say to 87, and leave the other parameters fixed. The result is $a = 302$, representing about 4% "safety stock" over the 291 that would be indicated by deterministic assumptions.

The situation becomes somewhat more complicated if, realistically, the retention fraction p is not known precisely--as is the case when the π_m's and p_{im}'s are treated as random--because then the random variable Y does not have a conditional binomial distribution. As a simple illustration, suppose p assumes the value .75 with probability .50 and the values .60 and .90 with equal probabilities of .25. In this case the expected retention rate is still .75, but its standard deviation is about .11 and its coefficient of variation about .14. It is fairly straightforward, but tedious, to ascertain in this case that $a = 39$ if $c = 9$ and that $a = 367$ if $c = 87$. In both cases, additional accessions exceed the corresponding deterministically determined quantities by over 25%.

In reality the actual distribution of the value of p is very much more complex than the simple one used here. In principle it can be determined for a subset of the work force--e.g., a CPG--by examining the distributions of its determinants, a set of π_m's and p_{im}'s. But in practice this would be very difficult, and the mechanics of an algorithm to perform the kind of calculations accomplished above would be quite complex in the presence of an involved distribution for p. Hence, we recommend that such capabilities not be attempted in personnel flow models. In case "protection" from undesirable events is important, however, it can be obtained by adding constraints to a deterministic flow model run in conjunction with a stochastic post-processor. For example, the deterministic flow model could contain constraints providing lower limits for the numbers of people of particular types which should be recruited for each year in the planning horizon. If the results provided by the deterministic model do not provide the desired probabilities for particular events--say, reenlistment rates below specified limits--then the lower limits on accessions could be increased and the deterministic model rerun. This process could continue, with constraint values being increased or decreased, until acceptable probabilities are achieved. These constraint adjustments could be made either interactively, with program users observing intermediate results and changing parameter values, or in logical "loops" which would adjust constraints using specified rules and rerun the deterministic model and post-processor until the results meet a priori specifications. We recommend these iterative approaches because they are analytically simple and computationally practical. Descriptive personnel flow models typically execute in very short times, and they can be rerun with different constraints very economically. Incorporating the necessary probabilistic computations in the basic flow model itself, while possible in principle, would add immensely to its complexity and computation time and would make its initial development and testing much more difficult and time-consuming.

4. PREDICTING PERSONNEL FLOW RATES

The flow model described in the previous section, and indeed each of the flow models in current use by the Air Force, assumes that the work force is already partitioned into subsets or categories whose retention rates differ. Further, the means (and in our stochastic model, the variances) of these flow rates are critical inputs to the descriptive models. Of course there are many reasons why it is important to distinguish personnel categories in flow models -- e.g., behavioral differences (of primary interest is retention behavior), productivity differences, cost differences, and availability differences (i.e., differences in the numbers of people in the enlistment-eligible population with particular attributes and differences in their propensities for joining the Air Force). In this section we focus on ways to distinguish categories of people whose behaviors differ and to characterize those differences. These fundamental categories and associated flow rates will continue to constitute critical inputs for personnel flow models, whether the models are deterministic or stochastic. In either case, the identification of the categories and corresponding rates remains a statistical problem. This section begins with a description of the important statistical issues relative to this problem and proceeds with brief discussions of two complementary statistical modeling approaches we believe will provide the requisite capabilities for handling these issues.

4.1. Statistical Estimation Issues

Whatever the techniques used to examine retention behavior, we believe they should meet three important criteria:

- <u>Statistical accuracy</u>. They should provide accurate predictions of retention rates, the precision implicit in their estimates should be characterized, and they should admit to convenient tests of hypotheses (particularly goodness-of-fit tests).
- <u>Logical consistency</u>. They should provide interpretable relationships between variables which predict retention rates and

the corresponding predictions, and their stability should be assessed (i.e., the regular presence and importance of the identified characteristics in predicting flow rates).

o Environmental robustness. They should be able to predict retention behavior under altered personnel management policies such as revised compensation tables, promotion opportunities, and/or retirement programs.

The primary retention rate estimation technique in current use by the Air Force is the Automatic Interaction Detector (AID), a method which partitions its data sample iteratively using the explanatory characteristic that provides the maximum decrease in overall mean squared prediction error. Its users have found this method usually adequate for predicting overall retention rates, but have noted substantial errors when subsets of the work force are considered. AID is employed within a special Air Force information system, the Airman Loss Probability System (ALPS), to provide flow rate estimates for numerous personnel planning and programming models. ALPS has the capability to bypass the AID partitioning/estimating routines for subsets of the work force, and this is frequently done for the first-term component of the work force. For this component a simple set of predictive characteristics is input to ALPS and flow rates are calculated for the corresponding categories. Another estimation procedure involving trend extrapolation also is used occasionally for first-term retention prediction.

These estimation/prediction procedures have limitations with respect to all three statistical criteria cited above. They seem to suffer least from lack of predictive accuracy -- at least in the aggregate, as already noted -- although problems in this area have led to recent revisions in the way the AID-identified categories and corresponding rate estimates are employed. But the system apparently has no capabilities for characterizing the precision implicit in the rates it identifies or for subjecting them or their underlying structure to goodness-of-fit tests, although "validation" runs listing comparisons between predicted and actual retention quantities are made regularly. Another structural limitation, at least in the basic AID logic, is an inability to consider possible time trends in the sample data. Regarding the logical consistency criterion, we note that the categories of airmen identified by the AID logic are not always the same. That is, some predictive characteristics appear to influence retention behavior more during some time periods and less during others. In fact this may be characteristic of an observation made by Doyle and Fenwick [9]: The sequential AID logic can "find" explanatory power (in characteristics) where it doesn't exist, and miss it where it does. A further logical shortcoming of AID is that it does not permit systematic study of possible interactions among predictive characteristics. (For example, educational background, mental aptitude, sex, and race may interact in subtle ways which would contribute to understanding retention behavior and possibly point toward useful personnel management policy revisions.) Finally, with respect to the environmental robustness criterion, current methods provide no real capability to predict retention behavior under revised management policies.

In the remainder of this section we discuss briefly two improvements which can enhance considerably the Air Force's capabilities for flow rate estimation: (1) application of log-linear models for behavioral category identification and rate estimation, and (2) development of a sequential decisionmaking model for prediction of flow rates under altered management policies. As we will see, use of log-linear models should provide a sound logical and statistical foundation for rate estimation in the absence of policy change and should identify distinct categories of personnel for which the sequential decisionmaking models should be employed separately. The sequential decisionmaking model can be based on the model developed by Gotz and McCall [11] for prediction of officer personnel retention.

4.2. Category Identification and Rate Estimation in the Absence of Policy Change

We propose the use of log-linear models to establish the relationship between flow rates--e.g., attrition, extension-of-obligation, and reenlistment--and various explanatory (predictor) categorical characteristics, such as mental category, educational background, training, job category, history of experience in the work force, etc. Our purpose is to ensure statistical soundness in inferences being drawn from the available data and to lay a solid foundation for the development of a sequential decisionmaking model to be used in predicting behavioral changes in attrition (and other) rates due to possible changes in Air Force policy.

To develop appropriate structures for flow models, we need to know what characteristics generally distinguish personnel categories and how those characteristics interact. In our view the most reasonable approach to identifying and analyzing these characteristics employs log-linear models for discrete multivariate data. This method is based on sound statistical theory, its results submit readily to tests of significance, and it possesses a number of other advantages mentioned in the following brief discussion of the analysis approach.

Typically, in log-linear analysis, we have a sample set of size N, and each data point falls within one of several categories. For example, in a study of attrition rates, we may have a group of enlisted airmen categorized according to characteristics such as marital status, race, mental category, educational level, skill level, geographical origin, etc. and according to whether they stay in or leave the service in the observed time period.

To illustrate, suppose each datum is classified by the values of three different discrete variables (characteristics) labeled A, B, and C. Assume that

A has I levels (values)
B has J levels (values)
C has K levels (values)

For example, A might represent retention behavior (two values: stay or leave) during some time period, B might represent mental category (say, using the four major values), and C might represent race (three values: white, black, other). Thus there are $2 \times 4 \times 3 = 24$ "elementary" cells. Of course, we may also have variables D, E, F, ..., and so on.

Generally, we are restricted to those variables maintained in the Uniform Airman Record (UAR) for information on individuals in the Air Force.

Assuming that the data set of size N represents the outcomes of some stochastic experiment, let

$$p_{ijk} = \text{probability of an observation } \textit{falling in cell } (i,\ j,\ k)$$

and

$$\begin{aligned} m_{ijk} &= \text{expected count for cell } (i,\ j,\ k) \\ &= N\, p_{ijk}. \end{aligned}$$

The log-linear model is obtained by writing the natural logarithm of the expected cell counts as a linear combination of terms which represent "effects" due to the characteristics A, B, and C, and to their various combinations (*i.e., their interaction effects*). Formally, by analogy with analysis of variance, we write

$$\begin{aligned} \log m_{ijk} = u &+ u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} \\ &+ u_{23(jk)} + u_{123(ijk)}, \end{aligned}$$

where the variables u *represent the linear contributions* of the various combinations of the characteristics A, B, and C to the logarithm of m_{ijk}--hence the name "log-linear models."[*]

Thus, the probabilities of interest are found by computing

$$p_{ijk} = \exp\left(u + u_{1(i)} + u_{2(j)} + \cdots + u_{123(ijk)}\right)/N$$

and then taking the proper summations.

For example, using the definitions of the three categories given above,

---

[*]The symbol + used as a subscript denotes summation over all values of the corresponding index.

$$p_{1++} = \sum_j \sum_k p_{1jk}$$

= Prob (a person chosen at random leaves)

and

$$p_{1|j} = p_{1j+}/p_{+j+}$$

$$= \{\sum_k p_{1jk}\}/\{\sum_i \sum_k p_{ijk}\}$$

= Prob (a person with characteristic j chosen at random leaves).

The u_{12} and other pairwise u-terms are the two-factor effects; u_{123} is the three-way interaction term. If the u_{12}, u_{23}, u_{13} and u_{123} terms are all zero, the three variables are mutually independent. If $u_{123} = 0$ but the others are not, we have all two-way interactions present but no three-way interaction.

Maximum likelihood is the method employed to fit these models. In some situations exact closed-form solutions can be obtained. Generally, however, iterative proportional fitting methods must be employed. Computer programs for this purpose are available.

When the model is saturated (no u-terms are taken as zero), we have as many parameters to estimate as there are "elementary cells"; otherwise, we may have far fewer parameters to estimate. The choice of variables to be included in log-linear models and the examination of the fit must be made carefully, since in "near-saturated" models there may be many u-terms to estimate.

In the initial stages of analyses, it is wise to fit only the simplest of models, models with no more than two-factor interaction effects. There are several reasons for this.

1. We can obtain cell estimates for every cell in a sparse array; fitting unsaturated models gives estimates for elementary cells that

have positive probabilities but no sample observations. For example, a particular sample may have no black, female airmen in mental category II in a certain specialty, and yet the probability of such an occurrence may not be zero.

2. Models with two-factor effects yield elementary cell estimates that are more stable than observed cell counts. Successively higher-order terms can be regarded as deviations from the average value of related lower-order terms, and so models with the higher-order terms removed are useful in describing the gross structure of a data array. Such models describe general trends and hence can be regarded as "smoothing" devices.

3. Simple models facilitate the detection of outliers. The detection of sporadic cells that are unduly large may be of importance. For example, it will be desirable to determine which combinations of variable categories give an excessive number of leavers from the work force.

As an example, one may find that married personnel with a good educational background and a high skill level have a higher attrition rate because of the interaction of these characteristics.

After initially fitting the model with two-factor interactions only, the model can be extended (if necessary) to include higher-order interaction terms. It is also possible that some two-factor interactions could be dropped from the model. We should always seek to develop as simple a model as possible that is still consistent with the data, since generally it is much easier to interpret the parameters of a simple model than of a more complex one. Additionally, a model with fewer parameters may improve the precision of the parameter estimates.

Log-linear models have the additional capability of using the natural ordering of categories. In our example, A and C (retention behavior and race) are not ordered, whereas B (mental category) is naturally ordered. The natural ordering can be used by assigning ordered scores to the various levels (values) of the ordered categories (characteristics). This is useful in reducing the number of general (higher-order) interaction u-terms and aids in developing understandable, interpretable and effective models. For example, in a two-way table with variables A (retention behavior) and B (mental category), a general log-linear model would have the form

$$\log m_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} .$$

However, since B has ordered levels, it may be preferable to examine the model in which scores v_1, v_2, v_3, v_4 are assigned to mental categories I, II, III, and IV, so that

$$\log m_{ij} = u + u_{1(i)} + u_{2(j)} + (v_j - \bar{v})u'_{1(i)}$$

where $\bar{v}$ is the average of the v's. Such a model has fewer parameters to estimate, and adds only a few extra degrees of freedom to the no-interaction model.

We may also use log-linear models to analyze discrete multivariate data forming a series through time--i.e., a Markov chain. We may wish to analyze trends in attrition rates as they change over time. Log-linear models can easily be adapted to this type of problem, whereas other modes of analysis (such as AID) do not generally lend themselves to such an investigation.

Logit regression models are a special type of log-linear model. By treating certain marginal totals as fixed, we may rewrite a log-linear model for the variables A, B, and C as

$$\log \left(\frac{m_{1jk}}{m_{0jk}} \right) = w + w_{2(j)} + w_{3(k)} + w_{23(jk)} .$$

This results from the fact that the conditional probability of attrition given characteristics (j, k) can be written

$$p(1|j, k) = \frac{m_{1jk}}{m_{+jk}} .$$

Hence,

$$p(0|j, k) = 1 - p(1|j, k) = \frac{m_{0jk}}{m_{+jk}}$$

and

$$\log\left(\frac{p(1|j,k)}{p(0|j,k)}\right) = \log\left(\frac{m_{1jk}}{m_{0jk}}\right)$$

$$= \log m_{1jk} - \log m_{0jk} \; .$$

We may use logit regression models where the explanatory variables are continuous and/or discrete. Further, we may relate attrition rates to various economic variables such as inflation rates, the cost of living index, joblessness, etc., to develop a model of attrition rates dependent on both personal variables (in the UAR) and economic indexes varying over time.

Additionally, it should be noted that log-linear models and logit regression models can be extended so that the dependent variable A is multinomial (or polytomous)--i.e., A may have more than two outcomes, such as leave, extend or reenlist.

In all the above models, careful attention must be paid to the analysis of residuals and various "goodness-of-fit" criteria to detect any serious model inadequacy. Well-established methods exist for examining the fit of log-linear models to actual data. Alternative methods, like AID, generally ignore the question of model fit; they provide simple, untested point estimates.

4.3. Behavioral Response to Policy Changes

As mentioned above, log-linear models include logit regression models as a special case, and logit models commonly have been used to estimate stay/leave behavioral alteration in response to policy or environmental change. However, logit models do not implicitly represent decisionmaking by individual airmen.

Airmen's decisions occur primarily near reenlistment points, and these decisions are subject to some influence through personnel management policies such as bonus levels and promotion rates. Gotz and McCall [11] have developed a sequential decisionmaking model of stay/leave behavior for Air Force officers which offers two key advantages over alternative retention rate estimation procedures:

- History dependence. The dynamic programming model of Gotz and McCall demonstrates that retention rates of Air Force officers depend both on prospective future financial returns to remaining in the military and on past occurrences. Their analysis shows, for example, that ordinary regression models can overpredict retention rates for years beyond the offer of a bonus. These models ignore the fact that some individuals may have stayed in service only to obtain the bonus; hence, their post-bonus retention rates should be expected to be lower. The important extension in the Gotz-McCall model which allows such behavior to be predicted is explicit incorporation of a term representing permanent differences in individuals' tastes for the military. (Of course the distribution of these tastes must be estimated empirically.)
- Structure which incorporates management policies directly. Personnel policies affecting individuals' income streams (i.e., expected military versus civilian incomes with differences depending on compensation tables, promotion opportunities, retirement pay, and other financial benefits) are represented explicitly in the underlying sequential decisionmaking model.

We believe this sequential modeling concept should be developed further and generalized for application to enlisted retention modeling. Key differences between the Gotz-McCall model and the sequential model for enlisted personnel may include:

- Multiple years between decision points. Although every year of service sees some airmen leave the Air Force--e.g., due to health problems, personal emergencies, or unsatisfactory behavior or job performance--most airmen face continuation decisions at four-year intervals, the usual enlistment or reenlistment obligation. This contrasts with an officer's more frequent opportunities to make stay/leave decisions, and it necessitates structural differences in a sequential model representing enlisted personnel decisionmaking.

o <u>Extension of obligation beyond normal enlistment terms</u>. Near the end of an enlistment term, an airman judged by the Air Force to be acceptable for continued service has a third option in addition to reenlisting or separating: extension. That is, the airman can extend his or her current term of service somewhat and delay the stay/leave decision. This option needs explicit representation in a behavioral model based on decisionmaking timing and options for enlisted personnel.

o <u>Transient differences in "taste" for the military</u>. Enlisted people typically enter the service much younger, less educated, less experienced, and with less forethought than officer personnel. Hence they cannot be expected to have as stable an affinity for the Air Force. Their impressions of service life may be much more influenced by their induction, training, assignment, and initial work experiences than are those of officers. The Air Force is often the first full-time, long-term job for enlisted people, and they really don't know what to expect either from their employer or from themselves in their newfound responsibilities and independence. Thus, the "taste" term to be employed in a sequential decisionmaking model of their behavior may need to be generalized so that, at least during the early years of service, it can follow different distributions. Alternatively, the annual "disturbance" factor represented in the Gotz-McCall approach might be allowed to play a larger role until airmen have sufficient time and experience to stabilize their impressions of military life.

o <u>Crosstraining</u>. In contrast to officers, enlisted personnel receive virtually all of their job-related training directly from the Air Force; essentially they are "given" occupations by their employer. Sometimes, when occupations become overmanned, a condition for continued employment can be that an individual must change occupations. This usually necessitates a period of crosstraining, whether formal or informal, and can result in changed working conditions, a different set of possible assignment locations, altered advancement potential, different opportunities in civilian life, etc. Such a change is usually more drastic and may be less expected than corresponding changes experienced by officer personnel; hence it may need explicit representation in an enlisted decisionmaking model.

The key linkages between the log-linear modeling approach, which focuses on identification of categories of personnel whose retention behaviors differ, and the sequential decisionmaking modeling approach, which focuses on how behavior will change under altered management policies, are the representations of tastes for the military and of transient disturbances affecting continuation decisions. We expect that the same general model structure can permit estimation of personnel flow rate changes under altered policies regardless of behavioral category, but the different categories will require different model parameters. Thus, the modeling approaches are complementary: the log-linear model provides an initial "filter" to separate and identify behaviorally distinct categories of personnel, and the sequential decisionmaking model predicts how each category's retention behavior will change if management policies are changed.

5. CONCLUSIONS AND RECOMMENDATIONS

We have ascertained that projections for many work force characteristics can incorporate sizeable uncertainties: "two-sigma" confidence intervals often contain values differing 10%-40% from corresponding expectations. Thus, especially when smaller segments of the work force are considered, substantial deviations from expected-value projections should be expected fairly frequently. From our very limited computational experience, it appears that the largest contributor to this uncertainty is usually the simple uncertainty in individual stay/leave behavior (regardless of whether the individual or the Air Force makes the determination). Another potentially large contributor is uncertainty in the proportion of accession requirements which actually can be met. Uncertainties regarding the mix of people that can be accessed and regarding estimates of flow rates, while they can be important in projecting values for certain subsets of the work force (e.g., the number of minority, male, and high-school graduates in a particular CPG who will be eligible for reenlistment five years hence), appear to contribute less to uncertainty in overall work force characteristics (e.g., the <u>total</u> number of people in a CPG who will be eligible for reenlistment five years hence).

While we see that uncertainties can be substantial, we find that one of our original concerns, the "nonlinearity of expectations" problem, is not too important. That is, at least for the limited (but representative) parameter values we have employed, the nonlinear equations which relate descriptive random variables (the random variables representing enlisted work force characteristics such as accessions, attrition, reenlistments, and costs) also hold approximately when the random variables are replaced by their expectations. Thus we can expect deterministic personnel flow models, which typically employ the best available estimates of expected values within systems of nonlinear equations, to yield fairly accurate predictions of the remaining expectations. Hence, as long as our primary interests are in expected values, we needn't make the substantial extra effort to develop models which include uncertainty explicitly.

When our interests shift to risk aversion, things get much more complicated. Since we have not evaluated the entire distribution of the work force "state," we cannot make precise statements about the probabilities for joint events (e.g., of needing reenlistment rates no higher than 40% in 1981-1984). But this work does enable us to determine the approximate probabilities of individual events (e.g., the probability that a reenlistment rate no higher than 40% will be required in 1983 or the probability that accessions in 1982 need not exceed 250 people in a particular CPG). These approximations are obtained by employing the calculated means and variances of the corresponding random variables to estimate the parameters of a specified probability distribution, usually the normal distribution. Since these means and variances are determined in the course of the model's execution, it is a simple matter to determine the approximate probabilities upon run completion. But if execution is to be affected by limits on these probabilities (e.g., access enough people in 1982 to provide 90% confidence that in 1986 a reenlistment rate no higher than, say, 40% will be required to provide some fixed number of career entrants within a CPG), the ongoing calculations become exceedingly complex.

Thus we recommend that personnel flow models not be encumbered with these intricate calculations during their basic operations; i.e., they should continue to be constructed as deterministic. However, uncertainty can be significant enough that we recommend that such models have appended stochastic "post-processors" which evaluate associated means, variances, and approximate event probabilities using the methods we have developed here. This will provide ready assessments of the extent of uncertainty in model projections. Further, if risk aversion is important, such post-processors can be employed to identify constraints which should be incorporated in the deterministic flow models. This could be accomplished either interactively, involving personnel policy analysts in changing inputs and exercising the flow models iteratively, or directly, by specifying in advance desired probabilities for certain events and imbedding the flow models in "logical loops" which would alter and rerun them until acceptable results are achieved.

Regarding the stochastic inputs necessary for personnel flow models (e.g., upgrade, reenlist t, and loss rates), we believe improved estimation procedures should be developed. These methods should provide consistent, interpretable, and parsimonious sets of parameters for estimating flow rates, they should incorporate time series data (in order to detect and project underlying trends), they should include "environmental" data such as occupational categories and corresponding civilian economic conditions, and they should admit to statistical goodness-of-fit procedures so that their accuracy can be assessed systematically. Uncertainty in these rate estimates is particularly important since, as we have seen, it can contribute significantly to uncertainty in related work force characteristics. We recommend that the first step in investigating these flow rates be to employ log-linear models (with logit models as a special case) to identify categories of enlisted personnel whose flow behaviors differ--e.g., subdividing the work force according to occupational subsets, educational backgrounds, mental aptitudes, marital status, etc., as appropriate. The second step should be to develop a sequential decisionmaking model which will predict how the flow behaviors for the various categories of personnel will change if management "control" policies such as compensation, promotion opportunity, educational benefits, or retirement programs are changed. These estimates of revised behavior under altered policies are obviously crucial if flow models are to be useful in evaluating and/or selecting improved personnel management policies. This step also should result in estimates of flow rates and their inherent uncertainty, again because of the potential importance of this uncertainty when the flow rates are employed in descriptive flow models.

We are convinced that substantially improved personnel flow models can be developed for Air Force use in developing and evaluating alternative personnel management policies--especially if smaller segments of the work force are to be examined (1) simultaneously with the total work force and (2) dynamically. Ideally, these models would be constructed to achieve directed results, for example, by using optimization techniques to pursue user-determined objectives. We have concluded that these models should be developed as deterministic flow models; but uncertainty is sufficiently important that such models should include stochastic post-processors to evaluate the degree of uncertainty implicit in identified results. In addition, careful attention must be given to estimation of important flow rates which ordinarily are model inputs--particularly to altered flow rates which may apply under different management policies and to the categories of personnel for which these behaviors apply.

Appendix A

DEVELOPMENT OF THE STOCHASTIC FLOW MODEL

A.1. Introduction

This appendix gives a detailed description of the stochastic (dynamic) flow model. As described in Section 3.1, the inputs to the model are:

- $N_{im}(0)$, the initial number of people in YOS i and category m
- Fixed force size, N
- The retention rates p_{im} = probability that a person in class (i,m) (YOS i, category m) will flow into class (i + 1, m) in the next time period. Thus, $1-p_{im}$ are the attrition rates. (The retention rates may depend on time t; i.e., the model considers the values $p_{im}(t)$).
- Accession mix $\pi_1, \ldots, \pi_M$.
- Costs, $C_{im}(t)$.
- Planning horizon, T.

Since each individual in class (i,m) stays in the service with probability p_{im}, it is clear that

$$N_{i+1,m}(t+1) = \text{number of individuals in YOS } i+1, \text{ class } m \text{ in year } t+1$$

$$= \sum_{\lambda=1}^{N_{im}(t)} X_\lambda$$

where X_1, X_2, ... are independent and identically distributed Bernoulli random variables with probability of success p_{im}; i.e.,

$$P\ (X_\lambda = 1) = 1 - P\ (X_\lambda = 0) = p_{im}.$$

The flow of people through the work force can be illustrated as follows:

$$\begin{tikzcd}
 & & \text{YOS=1} & 2 & 3 & 4 \\
 & t-1 & N_{1m}(t-1) \arrow[dr] & N_{2m}(t-1) \arrow[dr] & N_{3m}(t-1) \arrow[dr] & N_{4m}(t-1) \\
\text{Time} & t & N_{1m}(t) \arrow[dr] & N_{2m}(t) \arrow[dr] & N_{3m}(t) \arrow[dr] & N_{4m}(t) \\
 & t+1 & N_{1m}(t+1) & N_{2m}(t+1) & N_{3m}(t+1) & N_{4m}(t+1)
\end{tikzcd}$$

In our model it is convenient to assume the process starts at time $t = 0$ and runs until $t = T$ (the planning horizon). The values of $N_{im}(0)$ are given (fixed), and the process is allowed to evolve.

Fundamental questions to which we require answers are:

1. How do we compute $E\,(N_{im}(t) \mid \underset{\sim}{N}\,(t-1))$, the conditional expectation of $N_{im}(t)$ given the values

$$\underset{\sim}{N}\,(t-1) = \{N_{j\ell}(t-1) : 1 \leq j \leq 4,\ 1 \leq \ell \leq M\} ?$$

2. How do we compute $\mathrm{Cov}\,[N_{im}(t),\ N_{j\ell}(t) \mid \underset{\sim}{N}(t-1)]$, the conditional covariance?
3. How do we then compute $EN_{im}(t)$ and $\mathrm{Cov}\,[N_{im}(t),\ N_{j\ell}(t)]$, the unconditional means and covariances?
4. Are there closed-form analytic solutions to the above questions, or must we perform simulation runs (Monte Carlo runs) to obtain the answers?
5. What happens when $T \to \infty$? Is there a long-run, steady-state (equilibrium) distribution for $\{N_{im}(t)\}$? If so, can it be characterized?

In answer to the fourth question, if the p_{im}'s and π_m's are fixed parameters (not random), then we can obtain simple analytic expressions for the above-mentioned quantities. If, however, either the p_{im}'s or the π_m's are treated as random, then we must perform simulation runs to obtain the answers--repeatedly sampling from the distribution of the p_{im}'s and the π_m's.

Since the π_m's may be fixed (hence the proportional or multinomial model) or random, the p_{im}'s may be fixed or random, and the $C_{im}(t)$'s may be fixed or random, there are essentially

3 x 2 x 2 = 12 different cases that can be treated. We treat here only the 3 most interesting cases:

1. $\underset{\sim}{p}$ and $\underset{\sim}{\pi}$ fixed, $\underset{\sim}{\pi}$ proportional, $\underset{\sim}{C}$ random

2. $\underset{\sim}{p}$ and $\underset{\sim}{\pi}$ fixed, $\underset{\sim}{\pi}$ multinomial, $\underset{\sim}{C}$ random

3. $\underset{\sim}{p}$ and $\underset{\sim}{\pi}$ random, $\underset{\sim}{\pi}$ multinomial, $\underset{\sim}{C}$ random

The first three of the above questions are answered in Section A.2. Section A.3 discusses the modeling of parameter uncertainty, and Sections A.4 and A.5 discuss incorporation of parameter uncertainty in the flow model (as well as the answer to the fourth question). Section A.6 derives approximations to the mean and variance of reenlistment rates, and the last section answers the fifth question concerning the existence of a long-run, steady-state distribution for the flow model.

A.2. Dynamic Equations for the Means and Covariances of System State Contents

In this section, we develop expressions to compute iteratively the means and covariances of $N_{im}(t+1)$ as functions of $N_{j\ell}(t)$; i.e., we compute the conditional means and covariances

$$E\,(N_{im}(t+1) \mid \underset{\sim}{N}\,(t))$$

$$\mathrm{Cov}\,[N_{im}(t+1),\ N_{j\ell}(t+1) \mid \underset{\sim}{N}\,(t)]$$

where $\underset{\sim}{N}\,(t) = (N_{im}(t))$. We also develop the unconditional means and covariances at time t + 1 as functions of the unconditional means and covariances at time t. Hence, if the process starts at the initial values $N_{im}^{0} = N_{im}(0)$ at time t = 0, we can trace its evolution as t grows.

Up until now, we have assumed that the accession mix π and the transition probabilities $p = (p_{im})$ do not depend on t. However, to remind ourselves that a change in policy or behavior at time t can affect both π and p, we show the dependence of π and p on t by writing

$$\pi(t) = (\pi_m(t)) \text{ and } p(t) = (p_{im}(t)).$$

Throughout this section, neither $\pi(t)$ nor $p(t)$ is random.

We develop these results for two importantly different cases: in the first case the work force size is held constant at N, in the second it is allowed to fall below N (i.e., the possibility of recruiting shortfalls is introduced). We will treat both the fixed proportional case and the multinomial case for $\pi(t)$.

A.2.1. Fixed Work Force Size

First, to obtain the conditional means and covariances, let L = $L(t) = N - \sum_{j\geq 2} \sum_{\ell} N_{j\ell}(t)$ = number of accessions required for the planning year t.

(i). **Fixed proportional case.** We have $N_{1m}(t) = \pi_m(t)L$, so $E(N_{1m}(t) \mid L) = \pi_m(t)L$ and $\mathrm{Var}\,(N_{1m}(t) \mid L) = 0$.

(ii). Multinomial case. Given L, the vector $(N_{11}(t),\ldots,N_{1M}(t))$ has a multinomial distribution, i.e., $(N_{11}(t),\ldots, N_{1M}(t)) \sim \mathcal{M}(L, \pi_1(t),\ldots, \pi_M(t))$. Hence, $E\,(N_{1m}(t) \mid L) = \pi_m(t)\, L$ as before, but

$$\mathrm{Var}\,(N_{1m}(t) \mid L) = \pi_m(t)(1-\pi_m(t))\, L$$

and

$$\mathrm{Cov}\,(N_{1m}(t), N_{1\ell}(t) \mid L) = -\pi_m(t)\, \pi_\ell(t)\, L\,.$$

To keep the notation consistent, we assume that the transition $N_{j\ell}(t) \to N_{j+1,\ell}(t+1)$ is determined by $p_{j\ell}(t)$, and the distribution of $(N_{11}(t+1), \ldots, N_{1M}(t+1))$ is determined by $\underset{\sim}{\pi}(t)$. In our derivation, we make extensive use of the fact that, given $\underset{\sim}{N}(t)$,

$$N_{i+1,m}(t+1) = \sum_{\lambda=1}^{N_{im}(t)} X_\lambda,$$

where X_λ are independent, identically distributed (i.i.d.), having a binomial distribution with parameters 1 and $p_{im}(t)$.

a. <u>The Conditional Means</u>. For $t \geq 0$ and $1 \leq i \leq 3$, i.e., for the categories of continuing personnel, we have

$$\begin{aligned} E(N_{i+1,m}(t+1) \mid \underset{\sim}{N}(t)) &= E\left(\sum_{\lambda=1}^{N_{im}(t)} X_\lambda \mid \underset{\sim}{N}(t)\right) \\ &= p_{im}(t)\, N_{im}(t). \end{aligned}$$

The conditional means for the accession categories are the same whether we consider the fixed proportion or the multinomial case. We have

$$\begin{aligned} E(N_{1m}(t+1) \mid \underset{\sim}{N}(t)) &= \pi_m(t)\, E\,(L(t+1) \mid \underset{\sim}{N}(t)) \\ &= \pi_m(t)\, [N - \sum_{j \geq 2} \sum_{\ell} E(N_{j\ell}(t+1) \mid \underset{\sim}{N}(t))] \\ &= \pi_m(t) \left\{ N - \sum_{j \geq 2} \sum_{\ell} p_{j-1,\ell}(t)\, N_{j-1,\ell}(t) \right\} \end{aligned}$$

b. <u>The Conditional Covariances</u>. For the continuing categories of personnel, again for $1 \leq i \leq 3$ and $t \geq 0$, we have

$$\mathrm{Var}\ (N_{i+1,m}(t+1)\ |\ \underset{\sim}{N}(t)) = \mathrm{Var}\ (\sum_{\lambda=1}^{N_{im}(t)} X_\lambda\ |\ \underset{\sim}{N}(t))$$

$$= p_{im}(t)\ [1-p_{im}(t)]\ N_{im}(t)$$

The variances of the accession quantities $N_{1m}(t)$ depend on whether we treat the fixed proportion or the multinomial case. For brevity we treat the two cases simultaneously. The difference is that an extra term enters in the multinomial case; we handle this by introducing the indicator variable I as follows:

$$I = \begin{cases} +1 \text{ if in multinomial case} \\ \ \ 0 \text{ if in fixed proportion case} \end{cases}$$

We then have (for the first equation, see DeGroot [8]),

$$\mathrm{Var}\ (N_{1m}(t+1)\ |\ \underset{\sim}{N}(t)) = \mathrm{Var}\ (E(N_{1m}(t+1)\ |\ L(t+1),\ \underset{\sim}{N}(t))\ |\ \underset{\sim}{N}(t))$$

$$+\ E(\mathrm{Var}(N_{1m}(t+1)\ |\ L(t+1),\ \underset{\sim}{N}(t))|\ \underset{\sim}{N}(t))$$

$$= \mathrm{Var}\ (\pi_m(t)\ L(t+1)\ |\ \underset{\sim}{N}(t))$$

$$+\ I\ E(\pi_m(t)[1-\pi_m(t)]L(t+1)\ |\ \underset{\sim}{N}(t))$$

$$= \pi_m^{\ 2}(t)\ \mathrm{Var}\ (N - \sum_{j\geq 2} \sum_{\ell} N_{j\ell}(t+1)\ |\ \underset{\sim}{N}(t))$$

$$+\ I\ \pi_m(t)\ [1-\pi_m(t)]\ \{N - \sum_{j\geq 2} \sum_{\ell} p_{j-1,\ell}(t)\ N_{j-1,\ell}(t)\}$$

$$= \pi_m^{\ 2}(t)\ \sum_{j\geq 2} \sum_{\ell} \mathrm{Var}(N_{j\ell}(t+1)\ |\ \underset{\sim}{N}(t))$$

$$+\ I\ \pi_m(t)\ [1-\pi_m(t)]\ \{N - \sum_{j\geq 2} \sum_{\ell} p_{j-1,\ell}(t)\ N_{j-1,\ell}(t)\}$$

because, as we show below, if $(j,\ell) \neq (k,n)$ and $j \geq 2$, $k \geq 2$ then $Cov(N_{j\ell}(t+1), N_{kn}(t+1) \mid \underset{\sim}{N}(t)) = 0$. The equation continues as

$$Var\ (N_{1m}(t+1) \mid \underset{\sim}{N}(t)) = \pi_m^2(t) \sum_{j\geq 2} \sum_{\ell} p_{j-1,\ell}(t)\ [1-p_{j-1,\ell}(t)]\ N_{j-1,\ell}(t)$$

$$+ I \cdot \pi_m(t)\ [1-\pi_m(t)]\ \{N - \sum_{j\geq 2} \sum_{\ell} p_{j-1,\ell}(t)\ N_{j-1,\ell}(t)\}$$

To obtain the covariances, we first consider the cases where $(i,m) \neq (j,\ell)$ and $i \geq 2$, $j \geq 2$. Then if $\{X_\lambda\}$ are i.i.d. $b(1,p_{im}(t))$ and $\{X_\mu'\}$ are i.i.d. $b(1, p_{j\ell}(t))$,† and if $\{X_\lambda\}$ and $\{X_\mu'\}$ are independent systems, we get

$$Cov(N_{im}(t+1), N_{j\ell}(t+1) \mid \underset{\sim}{N}(t))$$

$$= Cov\left(\sum_{\lambda=1}^{N_{i-1,m}(t)} X_\lambda, \sum_{\mu=1}^{N_{j-1,\ell}(t)} X_\mu' \mid \underset{\sim}{N}(t)\right)$$

$= 0$, since X_λ, X_μ' are independent.

Now, in case $i = 1$ and $j \geq 2$, we have‡

$$Cov(N_{1m}(t+1), N_{j\ell}(t+1) \mid \underset{\sim}{N}(t))$$

$$= E\ [Cov(N_{1m}(t+1), N_{j\ell}(t+1) \mid N_{j\ell}(t+1), \underset{\sim}{N}(t)) \mid \underset{\sim}{N}(t)]$$

---

†That is, the random variables are independent and identically distributed (i.i.d.), having a binomial distribution with parameters $n = 1$ and probability of success $p_{j\ell}(t)$.

‡See De Groot [8] for the relevant conditional result which says that for random variables X, Y and Z,

$$Cov(X,Y) = E[Cov(X,Y|Z)] + Cov[E(X|Z), E(Y|Z)].$$

$$+ \text{Cov}\ [E(N_{1m}(t+1) \mid N_{j\ell}(t+1), \underset{\sim}{N}(t)), E(N_{j\ell}(t+1) \mid$$

$$N_{j\ell}(t+1), \underset{\sim}{N}(t)) \mid \underset{\sim}{N}(t)]$$

$$= 0 + \text{Cov}\ [\pi_m(t)\ E(L(t+1) \mid N_{j\ell}(t+1), \underset{\sim}{N}(t)), N_{j\ell}(t+1) \mid \underset{\sim}{N}(t)]$$

$$= \pi_m(t)\ \text{Cov}\ [N - \sum_{\substack{k \geq 2, n \\ (k,n) \neq (j,\ell)}} \sum p_{k-1,n}(t)\ N_{k-1,n}(t) - N_{j\ell}(t+1), N_{j\ell}(t+1) \mid \underset{\sim}{N}(t)]$$

$$= - \pi_m(t)\ \text{Var}(N_{j\ell}(t+1) \mid \underset{\sim}{N}(t))$$

$$= - \pi_m(t)\ p_{j-1,\ell}(t)\ [1-p_{j-1,\ell}(t)]\ N_{j-1,\ell}(t)\ .$$

In case $i = j = 1$ and $m \neq \ell$, we have

$$\text{Cov}(N_{1m}(t+1), N_{1\ell}(t+1) \mid \underset{\sim}{N}(t))$$

$$= E\ [\text{Cov}(N_{1m}(t+1), N_{1\ell}(t+1) \mid L(t+1), \underset{\sim}{N}(t)) \mid \underset{\sim}{N}(t)]$$

$$+ \text{Cov}[E(N_{1m}(t+1) \mid L(t+1), \underset{\sim}{N}(t)), E(N_{1\ell}(t+1) \mid L(t+1), \underset{\sim}{N}(t)) \mid \underset{\sim}{N}(t)]$$

$$= I \cdot E(-\pi_m(t)\ \pi_\ell(t)\ L(t+1) \mid \underset{\sim}{N}(t))$$

$$+ \text{Cov}(\pi_m(t)\ L(t+1), \pi_\ell(t)\ L(t+1) \mid \underset{\sim}{N}(t))$$

$$= - I \cdot \pi_m(t)\ \pi_\ell(t)\ E(L(t+1) \mid \underset{\sim}{N}(t))$$

$$+ \pi_m(t)\ \pi_\ell(t)\ \text{Var}(L(t+1) \mid \underset{\sim}{N}(t))$$

$$= - I \cdot \pi_m(t)\, \pi_\ell(t) \{ \sum_{k\geq 2} \sum_n p_{k-1,n}(t)\, N_{k-1,n}(t) \}$$

$$+ \pi_m(t)\, \pi_\ell(t) \{ \sum_{k\geq 2} \sum_n p_{k-1,n}(t)\, [1-p_{k-1,n}(t)]\, N_{k-1,n}(t) \}$$

In summary, if we let $a(t) = N - \sum_{k\geq 2} \sum_n p_{k-1,n}(t)\, N_{k-1,n}(t)$

$$= E\{L(t+1) | \underset{\sim}{N}(t)\}$$

and

$$b(t) = \sum_{k\geq 2} \sum_n p_{k-1,n}(t)\, [1-p_{k-1,n}(t)]\, N_{k-1,n}(t)$$

$$= \mathrm{Var}[L(t+1) | \underset{\sim}{N}(t)],$$

then for $i \geq 2,\ j \geq 2,\ m \neq \ell$,

$$E(N_{im}(t+1) \mid \underset{\sim}{N}(t)) = p_{i-1,m}(t)\, N_{i-1,m}(t)$$

$$E(N_{1m}(t+1) \mid \underset{\sim}{N}(t)) = \pi_m(t)\, a(t)$$

$$\mathrm{Var}(N_{im}(t+1) \mid \underset{\sim}{N}(t)) = p_{i-1,m}(t)\, [1 - p_{i-1,m}(t)]\, N_{i-1,m}(t)$$

$$\mathrm{Var}(N_{1m}(t+1) \mid \underset{\sim}{N}(t)) = I \cdot \pi_m(t)\, [1 - \pi_m(t)]\, a(t) + \pi_m^2(t)\, b(t)$$

$$\mathrm{Cov}(N_{im}(t+1), N_{j\ell}(t+1) \mid \underset{\sim}{N}(t)) = 0$$

$$\mathrm{Cov}(N_{1m}(t+1), N_{j\ell}(t+1) \mid \underset{\sim}{N}(t)) = - \pi_m(t)\, p_{j-1,\ell}(t)\, [1-p_{j-1,\ell}(t)]\, N_{j-1,\ell}(t)$$

$$\mathrm{Cov}(N_{1m}(t+1), N_{1\ell}(t+1) \mid \underset{\sim}{N}(t)) = \pi_m(t)\, \pi_\ell(t)\, \{-I\, a(t) + b(t)\}$$

Thus the only differences between the fixed proportion case and the multinomial case are the variances and covariances involving the first year of service.

c. <u>The Unconditional Means and Covariances</u>. Next, we derive the unconditional means and covariances of $\underset{\sim}{N}(t+1)$ as functions of the means and covariances of $\underset{\sim}{N}(t)$. From the preceding equations, we have

$$E(N_{im}(t+1)) = p_{i-1,m}(t)\; E(N_{i-1,m}(t))$$

$$E(N_{1m}(t+1)) = \pi_m(t)\; \{N - \sum_{k\geq 2} \sum_{n} p_{k-1,n}(t)\; E(N_{k-1,n}(t))\}$$

$$\text{Var } N_{im}(t+1) = E\;[\text{Var}(N_{im}\;(t+1) \mid \underset{\sim}{N}\;(t))] + \text{Var}\;[E(N_{im}(t+1) \mid \underset{\sim}{N}(t))]$$

$$= p_{i-1,m}(t)\;[1-p_{i-1,m}(t)]\; E\; N_{i-1,m}(t) + p^2_{i-1,m}(t)\; \text{Var}\; N_{i-1,m}(t)$$

$$\text{Var } N_{1m}(t+1) = E\;[\text{Var}(N_{1m}(t+1) \mid \underset{\sim}{N}(t))] + \text{Var}\;[E(N_{1m}(t+1) \mid \underset{\sim}{N}(t))]$$

$$= I \cdot \pi_m(t)\;[1-\pi_m(t)]\;\{N - \sum_{k\geq 2} \sum_{n} p_{k-1,n}(t)\; E\; N_{k-1,n}(t)\}$$

$$+ \pi^2_m(t)\;\{\sum_{k\geq 2} \sum_{n} p_{k-1,n}(t)\;[1-p_{k-1,n}(t)]\; E\; N_{k-1,n}(t)\}$$

$$+ \pi^2_m(t)\;\{\sum_{\substack{k\geq 2\\ j\geq 2}} \sum_{n,\ell} p_{k-1,n}(t)\; p_{j-1,\ell}(t)\; \text{Cov}(N_{k-1,n}(t),\; N_{j-1,\ell}(t))\}$$

To compute the unconditional covariances, we again use the fact that $\text{Cov}(X,Y) = E[\text{Cov}(X,Y|Z)] + \text{Cov}[E(X|Z),\; E(Y|Z)]$ to obtain:

---

† For a random variable X, we use the notation EX or E(X) for the mean and Var X or Var(X) for the variance.

$$\text{Cov}\,(N_{im}(t+1),\ N_{j\ell}(t+1)) = p_{i-1,m}(t)\ p_{j-1,\ell}(t)\ \text{Cov}(N_{i-1,m}(t),\ N_{j-1,\ell}(t))$$

$$\text{Cov}\,(N_{1m}(t+1),\ N_{j\ell}(t+1)) = -\pi_m(t)\ p_{j-1,\ell}(t)\ [1-p_{j-1,\ell}(t)]\ E\ N_{j-1,\ell}(t)$$

$$-\pi_m(t)\ p_{j-1,\ell}(t)\ \sum_{k\geq 2}\sum_n p_{k-1,n}(t)\ \text{Cov}(N_{k-1,n}(t),N_{j-1,\ell}(t))$$

$$\text{Cov}\,(N_{1m}(t+1),\ N_{1\ell}(t+1)) = \pi_m(t)\ \pi_\ell(t)\ \{-I\ [N - \sum_{k\geq 2}\sum_n p_{k-1,n}(t)\ E\ N_{k-1,n}(t)]$$

$$+ \sum_{k\geq 2}\sum_n p_{k-1,n}(t)\ [1-p_{k-1,n}(t)]\ E\ N_{k-1,n}(t)\}$$

$$+ \pi_m(t)\pi_\ell(t)\{\sum_{\substack{k\geq 2\\ j\geq 2}}\ \sum_{n,\ell} p_{k-1,n}(t)p_{j-1,\ell}(t)\text{Cov}(N_{k-1,n}(t),$$

$$N_{j-1,\ell}(t))\}\ .$$

A.2.2. Random Work Force Size

Until now, we have assumed that we may recruit as many people as necessary to keep the total size constant. We now assume that there may be shortfalls, so that the recruiting quota is not always met.

Specifically, in our earlier notation, if we need to recruit a total of

$$L = L(t+1) = N - \sum_{j\geq 2}\sum_\ell N_{j\ell}\ (t+1)$$

recruits, to keep the force size at N, we assume that the actual number L^* of people recruited is a random variable whose distribution depends on L. To introduce this type of uncertainty, we shall assume that both the mean and variance are linear functions of L; i.e.,

$$\alpha \cdot L = E(L^* \mid L) = \text{conditional mean}$$

$$\beta \cdot L = \text{Var}(L^* \mid L) = \text{conditional variance.}$$

For example, if the distribution of L^*, given L is binomial with parameters L and q, written $(L^* \mid L) \sim b(L, q)$, then $\alpha = q$ and $\beta = q(1 - q)$. If $\alpha = 1$ and $\beta = 0$, then $L^* \equiv L$, and we recruit as many as are needed.

We shall assume that p and π are nonrandom, and that we have the multinomial case for accessions.* Note that, given L^*, the distribution of $N_{11}(t + 1), \ldots, N_{1M}(t + 1)$ is $\mathcal{M}(L^*, \pi_1, \ldots, \pi_M)$, that is, multinomial with parameters L^* and $\pi_1, \ldots, \pi_M$. As before, we first obtain the conditional means and covariances, and then the unconditional means and covariances.

a. <u>The Conditional Means</u>. For $1 \leq i < 3$ and $t \geq 0$, we have, for the categories of continuing personnel,

$$E(N_{i+1,m}(t + 1) \mid N(t)) = p_{im}(t)\, N_{im}(t) \;.$$

For the accession categories, we have

$$E(N_{1m}(t + 1) \mid N(t)) = E(E(N_{1m}(t + 1) \mid L, N(t)) \mid N(t))$$

$$= E(E(E(N_{1m}(t + 1) \mid L^*, L, N(t)) \mid L, N(t)) \mid N(t))$$

---

*The fixed proportional case for accessions can also be easily treated.

$$= E(E(\pi_m(t)\, L^* \mid L, \underset{\sim}{N}(t)) \mid \underset{\sim}{N}(t))$$

$$= E(\pi_m(t)\, E(L^* \mid L) \mid \underset{\sim}{N}(t))$$

$$= \pi_m(t)\, E(\alpha L \mid \underset{\sim}{N}(t))$$

$$= \pi_m(t)\, \alpha \{N - \sum_{j\geq 2} \sum_{\ell} p_{j-1,\ell}(t)\, N_{j-1,\ell}(t)\} .$$

b. <u>The Conditional Covariances</u>. For the continuing categories of personnel, again for $1 \leq i \leq 3$ and $t \geq 0$, we have

$$\mathrm{Var}(N_{i+1,m}(t+1) \mid \underset{\sim}{N}(t)) = p_{im}(t)\,[1 - p_{im}(t)]\, N_{im}(t) .$$

For the variances of the accession quantities, we obtain

$$\mathrm{Var}(N_{1m}(t+1) \mid \underset{\sim}{N}(t)) = \mathrm{Var}[E(N_{1m}(t+1) \mid L, \underset{\sim}{N}(t)) \mid \underset{\sim}{N}(t)]$$

$$+ E[\mathrm{Var}(N_{1m}(t+1) \mid L, \underset{\sim}{N}(t)) \mid \underset{\sim}{N}(t)]$$

$$= \mathrm{Var}[E(E(N_{1m}(t+1) \mid L^*, L, \underset{\sim}{N}(t)) \mid L, \underset{\sim}{N}(t)) \mid \underset{\sim}{N}(t)]$$

$$+ E\{E(\mathrm{Var}(N_{1m}(t+1) \mid L^*, L, \underset{\sim}{N}(t)) \mid L, \underset{\sim}{N}(t))$$

$$+ \mathrm{Var}(E(N_{1m}(t+1) \mid L^*, L, \underset{\sim}{N}(t)) \mid L, \underset{\sim}{N}(t)) \mid \underset{\sim}{N}(t)\}$$

$$= \mathrm{Var}[E(\pi_m(t)\, L^* \mid L, \underset{\sim}{N}(t)) \mid \underset{\sim}{N}(t)]$$

$$+ E\,\{E(\pi_m(t)\,[1 - \pi_m(t)]\, L^* \mid L, \underset{\sim}{N}(t))$$

$$+ \mathrm{Var}(\pi_m(t)\, L^* \mid L, \underset{\sim}{N}(t)) \mid \underset{\sim}{N}(t)\}$$

$$= \mathrm{Var}[\pi_m(t)\ \alpha L \mid N(t)]$$

$$+ E\{\pi_m(t)\ [1 - \pi_m(t)]\ \alpha L + \pi_m^2(t)\ \beta L \mid N(t)\}$$

$$= \pi_m^2(t)\ \alpha^2\ \mathrm{Var}\ [L \mid N(t)]$$

$$+ \{\pi_m(t)\ [1 - \pi_m(t)]\ \alpha + \pi_m^2(t)\ \beta\}\ E(L \mid N(t))$$

$$= \pi_m^2(t)\ \alpha^2 \sum_{j \geq 2} \sum_{\ell} p_{j-1,\ell}(t)\ [1 - p_{j-1,\ell}(t)]\ N_{j-1,\ell}(t)$$

$$+ \{\pi_m(t)\ [1 - \pi_m(t)]\ \alpha + \pi_m^2(t)\ \beta\}\ \{N - \sum_{j \geq 2} \sum_{\ell} p_{j-1,\ell}(t)\ N_{j-1,\ell}(t)\}\ .$$

To obtain the covariance, we first consider the cases where $(i, m) \neq (j, \ell)$ and $i \geq 2$, $j \geq 2$.

$$\mathrm{Cov}(N_{im}(t + 1),\ N_{j\ell}(t + 1) \mid N(t)) = 0\ .$$

Now, if $i = 1$ and $j \geq 2$, we have

$$\mathrm{Cov}(N_{1m}(t + 1),\ N_{j\ell}(t + 1) \mid N(t))$$

$$= \mathrm{Cov}\ [E(N_{1m}(t + 1) \mid N_{j\ell}(t + 1),\ N(t))\ ,$$

$$E(N_{j\ell}(t + 1) \mid N_{j\ell}(t + 1), N(t)) \mid N(t)]$$

$$= \mathrm{Cov}\ [\pi_m(t)\ \alpha\ E(L \mid N_{j\ell}(t + 1),\ N(t)),\ N_{j\ell}(t + 1) \mid N(t)]$$

$$= -\pi_m(t)\ \alpha\ p_{j-1,\ell}(t)\ [1 - p_{j-1,\ell}(t)]\ N_{j-1,\ell}(t)\ .$$

In case $i = j = 1$ and $m \neq \ell$, we get

$$\mathrm{Cov}(N_{1m}(t+1), N_{1\ell}(t+1) \mid N(t))$$

$$= \mathrm{Cov}\,[E(N_{1m}(t+1) \mid L, N(t)), E(N_{1\ell}(t+1) \mid L, N(t))]$$

$$+ E\,\{\mathrm{Cov}(N_{1m}(t+1), N_{1\ell}(t+1) \mid L, N(t)) \mid N(t)\}$$

$$= \mathrm{Cov}\,[\pi_m(t)\,\alpha L, \pi_\ell(t)\,\alpha L \mid N(t)]$$

$$+ E\,\{E(\mathrm{Cov}(N_{1m}(t+1), N_{1\ell}(t+1) \mid L^*, L, N(t)) \mid L, N(t))$$

$$+ \mathrm{Cov}(E(N_{1m}(t+1) \mid L^*, L, N(t)), E(N_{1\ell}(t+1) \mid L^*, L, N(t))$$

$$\mid L, N(t)) \mid N(t)\}$$

$$= \pi_m(t)\,\pi_\ell(t)\,\alpha^2\,\mathrm{Var}(L \mid \underset{\sim}{N}(t))$$

$$+ E\,\{E(-\pi_m(t)\,\pi_\ell(t)\,L^* \mid L, \underset{\sim}{N}(t))$$

$$+ \mathrm{Cov}(\pi_m(t)\,L^*, \pi_\ell(t)\,L^* \mid L, \underset{\sim}{N}(t) \mid \underset{\sim}{N}(t))\}$$

$$= \pi_m(t)\,\pi_\ell(t)\,\alpha^2\,\mathrm{Var}(L \mid \underset{\sim}{N}(t))$$

$$+ E\,\{-\pi_m(t)\,\pi_\ell(t)\,\alpha L + \pi_m(t)\,\pi_\ell(t)\,\beta L \mid \underset{\sim}{N}(t)\}$$

$$= \pi_m(t)\,\pi_\ell(t)\,\alpha^2 \sum_{j\geq 2}\sum_\ell p_{j-1,\ell}(t)\,[1 - p_{j-1,\ell}(t)]\,N_{j-1,\ell}(t)$$

$$+ \pi_m(t)\,\pi_\ell(t)\,(\beta - \alpha)\,\{N - \sum_{j\geq 2}\sum_\ell p_{j-1,\ell}(t)\,N_{j-1,\ell}(t)\}\,.$$

In summary, if $a(t)$ and $b(t)$ represent the (conditional on $\underset{\sim}{N}(t)$) mean and variance of the number of recruits required in year t+1 to achieve the target end-strength of N--i.e., if

$$a(t) = N - \sum_{k \geq 2} \sum_{n} p_{k-1,n}(t)\ N_{k-1,n}(t)$$

$$b(t) = \sum_{k \geq 2} \sum_{n} p_{k-1,n}(t)\ [1 - p_{k-1,n}(t)]\ N_{k-1,n}(t)$$

then for $i \geq 2$, $j \geq 2$, and $n \neq \ell$,

$$E(N_{im}(t + 1) \mid \underset{\sim}{N}(t)) = p_{i-1,m}(t)\ N_{i-1,m}(t)$$

$$E(N_{1m}(t + 1) \mid \underset{\sim}{N}(t)) = \pi_{m}(t)\ \alpha\ a(t)$$

$$\mathrm{Var}\ (N_{im}(t + 1) \mid \underset{\sim}{N}(t)) = p_{i-1,m}(t)\ [1 - p_{i-1,m}(t)]\ N_{i-1,m}(t)$$

$$\mathrm{Var}(N_{1m}(t + 1) \mid \underset{\sim}{N}(t)) = \pi_{m}^{2}(t)\ \alpha^{2}\ b(t)$$

$$+ \{\pi_{m}(t)\ [1 - \pi_{m}(t)]\ \alpha + \pi_{m}^{2}(t)\ \beta\}\ a(t)$$

$$\mathrm{Cov}(N_{im}(t + 1), N_{j\ell}(t + 1) \mid \underset{\sim}{N}(t)) = 0$$

$$\mathrm{Cov}(N_{1m}(t + 1), N_{j\ell}(t + 1) \mid \underset{\sim}{N}(t)) = -\pi_{m}(t)\ \alpha\ p_{j-1,\ell}(t)$$

$$[1 - p_{j-1,\ell}(t)]\ N_{j-1,\ell}(t)$$

$$\mathrm{Cov}(N_{1m}(t + 1), N_{1\ell}(t + 1) \mid \underset{\sim}{N}(t)) = \pi_{m}(t)\ \pi_{\ell}(t)\ \alpha^{2}\ b(t)$$

$$+ \pi_{m}(t)\pi_{\ell}(t)\ (\beta - \alpha)\ a(t).$$

c. <u>The Unconditional Means</u>. For $1 \leq i \leq 3$ and $t \geq 0$, we have

$$E\,N_{i+1,m}(t+1) = p_{im}(t)\,E\,N_{im}(t) \;.$$

For the case $i = 1$ and $t \geq 0$,

$$E\,N_{1m}(t+1) = \pi_m(t)\,\alpha\,\{N - \sum_{j \geq 2} \sum_{\ell} p_{j-1,\ell}(t)\,E\,N_{j-1,\ell}(t)\} \;.$$

d. <u>The Unconditional Covariances</u>. For the continuing categories of personnel, again for $1 \leq i \leq 3$ and $t \geq 0$, we have

$$\begin{aligned}
\mathrm{Var}(N_{im}(t+1)) &= E(\mathrm{Var}(N_{im}(t+1) \mid \underset{\sim}{N}(t))) \\
&\quad + \mathrm{Var}(E(N_{im}(t+1) \mid \underset{\sim}{N}(t))) \\
&= p_{i-1,m}(t)\,[1 - p_{i-1,m}(t)]\,E\,N_{i-1,m}(t) \\
&\quad + p^2_{i-1,m}(t)\,\mathrm{Var}\,N_{i-1,m}(t) \;.
\end{aligned}$$

For the accession quantities,

$$\begin{aligned}
\mathrm{Var}\,N_{1m}(t+1) &= E(\mathrm{Var}(N_{1m}(t+1) \mid \underset{\sim}{N}(t))) \\
&\quad + \mathrm{Var}(E(N_{1m}(t+1) \mid \underset{\sim}{N}(t))) \\
&= \pi^2(t)\,\alpha^2\;E\,b(t) + \{\pi_m(t)\,[1 - \pi_m(t)]\,\alpha + \pi_m^2(t)\,\alpha^2\,E a(t) \\
&\quad + \pi_m^2(t)\,\alpha^2\,\mathrm{Var}\,a(t) \;.
\end{aligned}$$

To obtain covariances, first assume $(i, m) \neq (j, \ell)$ and $i \geq 2$, $j \geq 2$. Then

$$\begin{aligned}
Cov(N_{im}(t+1), N_{j\ell}(t+1)) &= E(Cov(N_{im}(t+1), N_{j\ell}(t+1) \mid \underset{\sim}{N}(t))) \\
&\quad + Cov(E(N_{im}(t+1) \mid \underset{\sim}{N}(t)), E(N_{j\ell}(t+1) \mid \underset{\sim}{N}(t))) \\
&= 0 + Cov(p_{i-1,m}(t)\, N_{i-1,m}(t), p_{j-1,\ell}(t)\, N_{j-1,\ell}(t)) \\
&= p_{i-1,m}(t)\, p_{j-1,\ell}(t)\, Cov(N_{i-1,m}(t), N_{j-1,\ell}(t)).
\end{aligned}$$

If $i = 1$ and $j \geq 2$, we have

$$\begin{aligned}
Cov(N_{1m}(t+1), N_{j\ell}(t+1)) &= E(Cov(N_{1m}(t+1), N_{j\ell}(t+1) \mid \underset{\sim}{N}(t))) \\
&\quad + Cov(E(N_{1m}(t+1) \mid \underset{\sim}{N}(t)), E(N_{j\ell}(t+1) \mid \underset{\sim}{N}(t))) \\
&= -\pi_m(t)\, \alpha\, p_{j-1,\ell}(t)\, [1 - p_{j-1,\ell}(t)]\, E\, N_{j-1,\ell}(t) \\
&\quad + Cov(\pi_m(t)\, \alpha\, a(t), p_{j-1,\ell}(t)\, N_{j-1,\ell}(t)) \\
&= -\pi_m(t)\, \alpha\, p_{j-1,\ell}(t)\, [1 - p_{j-1,\ell}(t)]\, EN_{j-1,\ell}(t) \\
&\quad + \alpha\, \pi_m(t)\, p_{j-1,\ell}(t)\, Cov(a(t), N_{j-1,\ell}(t)) .
\end{aligned}$$

Finally, for $i = j = 1$ and $m \neq \ell$,

$$\begin{aligned}
Cov(N_{1m}(t+1), N_{1\ell}(t+1) \mid \underset{\sim}{N}(t)) &= E\, (Cov(N_{1m}(t+1), N_{1\ell}(t+1) \mid \underset{\sim}{N}(t))) \\
&\quad + Cov(E(N_{1m}(t+1) \mid \underset{\sim}{N}(t)), E(N_{1\ell}(t+1) \mid \underset{\sim}{N}(t))) \\
&= \pi_m(t)\, \pi_\ell(t)\, [\alpha^2\, Eb(t) + (\beta - \alpha)\, E\, a(t)] \\
&\quad + Cov(\pi_m(t)\, \alpha\, a(t), \pi_\ell(t)\, \alpha\, a(t))
\end{aligned}$$

$$= \pi_m(t)\ \pi_\ell(t)\ [\alpha^2\ E\ b(t) + (\beta - \alpha)\ E\ a(t)]$$

$$+ \pi_m(t)\ \pi_\ell(t)\ \alpha^2\ \mathrm{Var}\ a(t)\ .$$

A.3. Modeling of Parameter Uncertainty

Recall that the retention rates $(p_{im}(t))$ and the accession mix fractions $(\pi_m(t))$ are inputs. To consider uncertainty in these parameters in our analysis, we have ignored the possibility of their time dependence and supposed they have been estimated from actual data--as would be the case if the AID* techniques were employed. Thus the estimate $\hat{p}_{im}$ of p_{im} is treated as a normal random variable with mean p_{im} and variance $p_{im}(1-p_{im})/n_{im}$, where n_{im} is the number of individuals in YOS i and category m whose stay/leave behavior is observed in the sample. Although $0 \le p_{im} \le 1$, we are assuming that n_{im} is sufficiently large so that the normal distribution approximates the binomial distribution adequately.

To model uncertainty in the π_m's, we will require that $\sum_m \hat{\pi}_m = 1$. Since each "estimate" $\hat{\pi}_m$ is also required to lie in the interval [0,1], a useful probability distribution to model the joint distribution of $(\hat{\pi}_1,\ldots,\hat{\pi}_M)$ is the Dirichlet distribution. (See De Groot [8].) If $\alpha_1,\ldots,\alpha_M$ (all nonnegative) denote the parameters of this distribution, its density is defined by

$$f(x_1,\ldots,x_M) = \frac{\Gamma(\alpha_o)}{\prod_{m=1}^{M} \Gamma(\alpha_m)}\ x_1^{\alpha_1 - 1} \cdots x_M^{\alpha_M - 1}$$

---

*Automatic Interaction Detector, the primary statistical approach used by the Air Force for discerning behavioral categories and estimating corresponding flow rates.

where $\alpha_o = \sum_m \alpha_m$ and $x_1,\ldots,x_M$ are all nonnegative, and $\sum_m x_m = 1$. Moreover, if $(X_1,\ldots,X_M)$ have a Dirichlet distribution, then it can be shown that

$$EX_m = \alpha_m/\alpha_o$$

$$\mathrm{Var}\ X_m = \frac{\alpha_m(\alpha_o-\alpha_m)}{\alpha_o^2\ (\alpha_o+1)}$$

$$\mathrm{Cov}\ (X_m,X_\ell) = -\frac{\alpha_m\alpha_\ell}{\alpha_o^2(\alpha_o+1)}\ ,\ m \neq \ell\quad .$$

For convenience in our computer program, we chose (n_{1m}) such that $n_{1m}/\sum_\ell n_{1\ell} = \pi_m$, for each m, and let $\alpha_m = n_{1m}$. (This gives estimates for the Dirichlet distribution parameters consistent with those obtained from using the most recent year's observed accession mix as data.) Consequently, $(\hat{\pi}_1,\ldots,\hat{\pi}_M)$ has a Dirichlet distribution with

$$E\ \hat{\pi}_m = \pi_m = n_{1m}/\sum_k n_{1k}$$

$$\mathrm{Var}\ (\hat{\pi}_m) = \frac{n_{1m}(\sum_k n_{1k}-n_{1m})}{(\sum_k n_{1k})^2(\sum_k n_{1k}+1)}$$

$$= \frac{\pi_m(1-\pi_m)}{\sum_k n_{1k}+1}$$

$$\simeq \frac{\pi_m(1-\pi_m)}{\sum_k n_{1k}}$$

$$\mathrm{Cov}\,(\hat{\pi}_m, \hat{\pi}_\ell) = -\frac{n_{1m}\, n_{1\ell}}{(\sum_k n_{1k})^2 (\sum_k n_{1k}+1)}$$

$$= -\frac{\pi_m \pi_\ell}{\sum_k n_{1k}+1}$$

$$\simeq -\frac{\pi_m \pi_\ell}{\sum_k n_{1k}}\,, \quad m \neq \ell\,.$$

Thus the first and second moments of $(\hat{\pi}_1, \ldots, \hat{\pi}_M)$ are approximately those of estimators of the parameters of a multinomial distribution, with sample size $\sum_k n_{1k}$.

The above approach provides an approximate feel for the order of magnitude of the variance of the p_{im}'s and the π_m's if they are estimated cross-sectionally using a recent work force of size N.

A.4. Incorporating Parameter Uncertainty in the Flow Model

Section A.2 described the formulas used to obtain E $N_{im}(t)$ and Cov $(N_{im}(t), N_{j\ell}(t))$ for fixed values of p_{im} and π_m. Now we denote these quantities by

$$E(N_{im}(t)\,|\,\underset{\sim}{p}, \underset{\sim}{\pi}) \text{ and } \mathrm{Cov}\,(N_{im}(t), N_{j\ell}(t)\,|\,\underset{\sim}{p}, \underset{\sim}{\pi})$$

to indicate their dependence on $\underset{\sim}{p}$ and $\underset{\sim}{\pi}$. Since there is uncertainty associated with estimating the true values of $\underset{\sim}{p}$ and $\underset{\sim}{\pi}$, both the means and the covariances vary depending on the estimated values of $\underset{\sim}{p}$ and $\underset{\sim}{\pi}$.

The mean $E\, N_{im}(t)$ is obtained by averaging $E(N_{im}(t) \mid \underset{\sim}{p}, \underset{\sim}{\pi})$ over the distribution of $\underset{\sim}{p}$ and $\underset{\sim}{\pi}$; i.e.,

$$E\, N_{im}(t) = E[E(N_{im}(t) \mid \underset{\sim}{p}, \underset{\sim}{\pi})]$$

and Cov $(N_{im}(t), N_{j\ell}(t))$ is obtained from:

$$Var(N_{im}(t)) = E[Var(N_{im}(t) \mid \underset{\sim}{p}, \underset{\sim}{\pi})] + Var[E(N_{im}(t) \mid \underset{\sim}{p}, \underset{\sim}{\pi})]$$

$$Cov[N_{im}(t), N_{j\ell}(t)] = E(Cov[N_{im}(t), N_{j\ell}(t) \mid \underset{\sim}{p}, \underset{\sim}{\pi}])$$

$$+ Cov[E(N_{im}(t) \mid \underset{\sim}{p}, \underset{\sim}{\pi}), E(N_{j\ell}(t) \mid \underset{\sim}{p}, \underset{\sim}{\pi})]$$

These quantities are nonlinear functions of $\underset{\sim}{p}$ and $\underset{\sim}{\pi}$, so it is difficult to obtain (even approximate) analytical expressions for these quantities, based on the distributions of $\underset{\sim}{p}$ and $\underset{\sim}{\pi}$. To evaluate these (and other) quantities of interest, we employ Monte Carlo simulations using the distributions of $\underset{\sim}{p}$ and $\underset{\sim}{\pi}$. We generally performed 400 replications in our simulations.

A.5. A Brief Look at Cost Uncertainty

Our model assumes that the cost $C_{im}(t)$ associated with each individual in the class (i, m) in year t is also random. The total (random) cost for the work force in calendar year t is

$$C(t) = \sum_{i,m} C_{im}(t)\, N_{im}(t).$$

Given $N(t)$, the (conditional) expected cost for year t is

$$E(C(t) \mid N(t)) = \sum_{i,m} EC_{im}(t)\, N_{im}(t)$$

and the (conditional) variance is

$$\mathrm{Var}(C(t)\,|\,\underset{\sim}{N}(t)) = \sum_{i,j,m,\ell} \mathrm{Cov}[C_{im}(t)\, N_{im}(t),\ C_{j\ell}(t)\, N_{j\ell}(t) \mid p,\ n]$$

$$= \sum_{i,j,m,\ell} N_{im}(t) N_{j\ell}(t) \mathrm{Cov}[C_{im}(t),\ C_{j\ell}(t)].$$

Hence the unconditional expected cost is

$$EC(t) = \sum_{i,m} EC_{im}(t)\, EN_{im}(t)$$

and the unconditional variance is

$$\mathrm{Var}\ C(t) = E[\mathrm{Var}(C(t)\,|\,\underset{\sim}{N}(t))] + \mathrm{Var}[E(C(t)\,|\,\underset{\sim}{N}(t))]$$

$$= \sum_{i,j,m,\ell} E(N_{im}(t) N_{j\ell}(t)) \mathrm{Cov}[C_{im}(t),\ C_{j\ell}(t)]$$

$$+ \sum_{i,j,m,\ell} EC_{im}(t) \cdot EC_{j\ell}(t) \mathrm{Cov}[N_{im}(t),\ N_{j\ell}(t)]$$

$$= \sum_{i,j,m,\ell} \mathrm{Cov}[C_{im}(t),\ C_{j\ell}(t)]\ \{\mathrm{Cov}[N_{im}(t),\ N_{j\ell}(t)] + EN_{im}(t) EN_{j\ell}(t)\}$$

$$+ \sum_{i,j,m,\ell} \mathrm{Cov}[N_{im}(t),\ N_{j\ell}(t)]\ EC_{im}(t)\ EC_{j\ell}(t).$$

This shows how the variance of the total cost depends on the means and covariance of $\underset{\sim}{N}(t)$ and $\underset{\sim}{C}(t)$.

A.6. Approximation of Means and Variances for Required Reenlistment Rates

Reenlistment rates are computed for the first-term work force as a whole, and also by category. Let $0 < c < 1$ (where c may depend on the category m), N be force size, and the random variable $X(t)$ be either the number in the entire fourth year group ($\sum_m N_{4m}(t)$) or the number in the fourth year group for cell m ($N_{4m}(t)$). The expected required reenlistment rate and the variance of the required reenlistment rate are

$$E(cN/X(t)) \text{ and } \mathrm{Var}\ (cN/X(t)) \quad .$$

We wish to obtain approximations to these quantities by means of the first and second moments of $X(t)$. Theoretically, since $X(t)$ is a discrete random variable with positive probability of being equal to zero, the above quantities are not finite. However, in practice, we know that our model is only an approximation to reality, and that the likelihood that $X(t) = 0$ is quite small. Therefore in our calculations we assume that $P(X(t) = 0) = 0$ (i.e., we truncate the distribution of $X(t)$ away from zero).

Let g be a continuously differentiable function. Let Y be any random variable, and let $\mu = E\ Y$ be its mean value. Consider the Taylor series expansion:

$$g(Y) \approx g(\mu) + g'(\mu)[Y-\mu] + \frac{g''(\mu)}{2}[Y-\mu]^2 + \text{remainder.}$$

Assuming the remainder term can be ignored, we find

$$\begin{aligned} E\ g(Y) &\approx g(\mu) + g'(\mu)[\mu-\mu] + \frac{g''(\mu)}{2} E(Y-\mu)^2 \\ &\approx g(\mu) + \frac{g''(\mu)}{2} \mathrm{Var}\ Y \quad . \end{aligned}$$

Moreover, by ignoring the second and higher order terms in the Taylor series expansion, we have

$$\begin{aligned} \operatorname{Var} g(Y) &\approx \operatorname{Var}\,(g(\mu) + g'(\mu)(Y-\mu)) \\ &= \operatorname{Var}\,(g'(\mu)(Y-\mu)) \\ &= g'(\mu)^2 \operatorname{Var} Y \quad . \end{aligned}$$

Now let $g(y) = 1/y$, and let $Y = X/cN$. The higher order terms of the Taylor expansion can be dropped if they are $O(N^{-\gamma})$, where $\gamma > 1$.† We know from empirical results and from the limiting stationary distribution (see Section A.7 below) that

$$\operatorname{Var} Y = O(N^{-1}) \quad ,$$

because

$$\begin{aligned} E\,X &\propto N \\ \operatorname{Var} X &\propto N \quad . \end{aligned}$$

Since $\operatorname{Var} Y = O(1/N)$, the above approximations using the Taylor expansion (together with truncating X) is justified (Bickel and Doksum [7]).

Note that

$$\begin{aligned} g'(y) &= -1/y^2 \quad , \\ g''(y) &= 2/y^3 \quad . \end{aligned}$$

Thus we obtain

$$\begin{aligned} E(cN/X) &\approx \frac{cN}{EX} + \frac{(cN)^3}{(EX)^3} \operatorname{Var}(X/cN) \\ &= \frac{cN}{EX} + \frac{cN}{(EX)^3} \operatorname{Var} X \end{aligned}$$

---

†Recall that a function g(N) is $O(N^{-\gamma})$ if $\lim_{N\to\infty} N^{\gamma} g(N)$ remains bounded.

$$= \frac{cN}{EX}\left[1 + \frac{\text{Var } X}{(EX)^2}\right]$$

and

$$\text{Var}(cN/X) \simeq \frac{(cN)^4}{(EX)^4}\text{Var}(X/cN)$$

$$= \frac{(cN)^2}{(EX)^4}\text{ Var } X.$$

In the limiting (stationary) case (see next section), $X(t)$ is binomial with parameters N and ξ, where $\xi = \xi_{4m}$ if $X = N_{4m}$ or $\xi = \sum_m \xi_{4m}$ if $X = \sum_m N_{4m}$. In this case, $EX = \xi N$ and $\text{Var } X = \xi(1 - \xi)N$. Thus

$$E(cN/X) \simeq \frac{c}{\xi}\left[1 + \frac{(1 - \xi)}{\xi N}\right]$$

$$\text{Var}(cN/X) \simeq \frac{c^2(1 - \xi)}{\xi^3 N}$$

A.7. The Limiting Stationary Distribution of the System

In this section we develop the limiting distribution of the system $\{N_{im}(t): 1 \leq i \leq 4, 1 \leq m \leq M\}$ as $t \to \infty$. We assume that the ξ_{im}'s and the π_m's are fixed, and that the π_m's are parameters for the multinomial distribution.

During the development of the work, the limiting stationary case was actually solved before the equations of Section A.2 were developed, and it subsequently was used as a check on the results from Section A.2. We feel the results of the limiting case are sufficiently interesting and useful to be recorded here.

Under the above assumptions, the state $(N_{im}(t))$ (a 4 x M matrix) of the system at time t moves to the state $(N_{im}(t+1))$ as follows:

(A7.1) 1. $N_{i+1,m}(t+1) = \sum_{\ell=1}^{N_{im}(t)} X_\ell$ where X_ℓ are independent,

identically distributed binomial random variables with parameters 1 and p_{im}, for $1 \leq i \leq 3$.

2. Given $L = N - \sum_{j=2}^{4} \sum_{\ell=1}^{M} N_{j\ell}(t+1)$, the accession vector

$(N_{11}(t+1), \ldots, N_{1M}(t+1))$ has a multinomial distribution with parameters L and $\pi_1, \ldots, \pi_M$.

Thus, the transition from the state at time t to the state at time t + 1 depends only on the state at time t. Hence $(N_{im}(t))$, t = 0, 1, 2, 3, ..., is a finite Markov chain (indeed, besides having a finite state space, the chain is aperiodic and irreducible) with state space

$$\mathcal{S} = \{(n_{im}) : \text{ each integer } n_{im} \geq 0, \ \sum_{i,m} n_{im} = N\}.$$

Thus, from the theory of Markov chains, $N_{im}(t)$ has a stationary limiting distribution $Q(\cdot)$ on $\mathcal{S}$ (Parzen [13]). That is, if

$$\underset{\sim}{N}(t) = (N_{im}(t))_{\substack{1 \leq i \leq 4 \\ 1 \leq m \leq M}}$$

$$\underset{\sim}{n} = (n_{im})_{\substack{1 \le i \le 4 \\ 1 \le m \le M}}$$

$$\underset{\sim}{n}' = (n'_{im})_{\substack{1 \le i \le 4 \\ 1 \le m \le M}}$$

and if

$$P_{\underset{\sim}{n},n'} = P(\underset{\sim}{N}(t+1) = n' \mid \underset{\sim}{N}(t) = \underset{\sim}{n})$$

there is a distribution $Q(n)$ on $\mathcal{N}$ such that

$$\text{(A7.2)} \qquad Q(\underset{\sim}{n}') = \sum_{\underset{\sim}{n}} P_{n,n'} \, Q(n)$$

where the sum is over all n such that $P_{n,n'} > 0$. Moreover, Q is unique (Parzen [13]).

Also,

$$\lim_{t \to \infty} P(N(t) = \underset{\sim}{n}) = Q(\underset{\sim}{n}).$$

The problem then is to determine Q; i.e., how Q relates to the parameters p_{im} and τ_m.

First, from (A7.1) and from Section A.2, we see that

$$\text{(A7.3)} \quad E(N_{i+1,m}(t+1) \mid N_{im}(t)) = p_{im} N_{im}(t)$$

and

$$E(N_{1m}(t+1) \mid L) = \tau_m L .$$

Thus

(A7.4) $$EN_{i+1,m}(t+1) = p_{im}\, EN_{im}(t)$$

and

$$EN_{1m}(t+1) = \pi_m\, EL$$

where expectation $E(\cdot)$ is taken with respect to the stationary distribution Q.

Let $\xi^*_{im} = \lim_{t\to\infty} EN_{im}(t)$. We have

$EN_{2m}(t+1) = p_{1m}\, EN_{1m}(t)$, and in the limit $\xi^*_{2m} = p_{1m}\, \xi^*_{1m}$;

$EN_{3m}(t+1) = p_{2m}\, EN_{2m}(t) = p_{2m}\, p_{1m}\, EN_{1m}(t-1)$, and in the limit

$$\xi^*_{3m} = p_{2m}\, p_{1m}\, \xi^*_{1m} ;$$

and similarly

$E_{4m}(t+1) = p_{3m} p_{2m} p_{1m} \xi_{1m}(t-2)$, and in the limit $\xi^*_{4m} = p_{3m} p_{2m} p_{1m} \xi^*_{1m}$.

Furthermore,

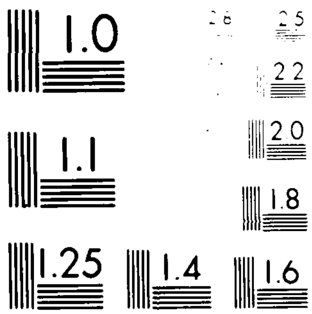
$$EN_{1m}(t+1) = \pi_m\, EL = \pi_m\, E\{N - \sum_{j=2}^{4} \sum_{\ell} N_{j\ell}(t)\}$$

$$= \pi_m\{N - \sum_{j=2}^{4} \sum_{\ell} E\, N_{j\ell}(t)\}$$

AD-A115 934   RAND CORP SANTA MONICA CA   F/G 15/5
UNCERTAINTY IN PERSONNEL FORCE MODELING.(U)
APR 82 G J HALL, S C MOORE   F49620-82-C-0018
UNCLASSIFIED RAND/N-1842-AF   NL

2 of 2
AD A
115934

END
DATE
FILMED
07-82
DTIC


1.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

and in the limit

$$\xi^*_{1m} = \pi_m\{N - \sum_{j=2}^{4} \sum_{\ell=1}^{M} \xi^*_{1\ell}\, p_{1\ell} \cdots p_{j-1,\ell}\}.$$

Thus since the p_{1m}'s and π_m's are known, we have a system of M linear equations in M unknowns (the ξ^*_{1m}'s):

$$\xi^*_{1m} + \pi_m \sum_{j=2}^{4} \sum_{\ell=1}^{M} \xi^*_{1\ell}\, p_{1\ell} \cdots p_{j-1,\ell} = \pi_m N, \quad 1 \leq m \leq M$$

or

$$\text{(A7.5)} \quad \xi^*_{1m} + \pi_m \sum_{\ell=1}^{M} \xi^*_{1\ell} \sum_{j=2}^{4} p_{1\ell} \cdots p_{j-1,\ell} = \pi_m N, \quad 1 \leq m \leq M.$$

In the remainder of this section we show how to solve these simultaneous equations for the expected values ξ^*_{1m} and then obtain the steady-state probability distribution $Q(\underset{\sim}{n})$, a multinomial distribution with parameters N and $\xi_{im} = \xi^*_{im}/N$.

First, if

$$b_\ell = \sum_{j=2}^{4} p_{1\ell} \cdots p_{j-1,\ell}, \quad 1 \leq \ell \leq M,$$

then (A7.5) can be written as

$$\xi^*_{1m} + \pi_m \sum_{\ell=1}^{M} \xi^*_{1\ell}\, b_\ell = \pi_m N, \quad 1 \leq m \leq M .$$

Define

$$B = \begin{bmatrix} \pi_1 b_1 & \pi_1 b_2 & \cdots & \pi_1 b_M \\ \pi_2 b_1 & \pi_2 b_2 & \cdots & \pi_2 b_M \\ \vdots & & \cdots & \\ \pi_M b_1 & \pi_M b_2 & \cdots & \pi_M b_M \end{bmatrix},$$

$\underset{\sim}{\xi}_1^{*T} = (\xi_{11}^*, \ldots, \xi_{1M}^*)$ and $\underset{\sim}{\pi}^T = (\pi_1, \ldots, \pi_M)$. Then we have

$$(I + B)\, \underset{\sim}{\xi}_1^* = N \cdot \underset{\sim}{\pi} .$$

Hence, solving for $\underset{\sim}{\xi}_1^*$, we obtain

$$\underset{\sim}{\xi}_1^* = N(I + B)^{-1}\, \underset{\sim}{\pi} .$$

<u>LEMMA</u>.† Let $s = tr(B) = \sum_{\ell=1}^{M} \pi_\ell b_\ell$.
Then

$$(I + B)^{-1} = I - (s + 1)^{-1} B .$$

<u>Proof (outline)</u>

It is easy to show that the ij^{th} element of B^2, $(B^2)_{ij}$, is equal to sB_{ij}. Hence $B^2 = sB$. Therefore

$$B^2 + B = (s + 1) B,$$

---

†We are indebted to Michael D. Miller, a Rand Corporation colleague, for this lemma.

so

$$(s+1)B - B - B^2 + (s+1)I = (s+1)I .$$

Thus

$$(B+I)\,((s+1)I - B) = (s+1)I,$$

or

$$(B+I)^{-1} = I - (s+1)^{-1} B .$$

$\boxtimes$

Consequently, we have

$$\underset{\sim}{\xi}{}^{*}_{1} = N(I - (s+1)^{-1} B)\, \underset{\sim}{\pi} .$$

Thus

$$\xi^{*}_{1m} = N \pi_m - \frac{N}{s+1} \pi_m \sum_{\ell} \pi_{\ell} b ,$$

$$= N \left[\pi_m - \frac{\pi_m s}{s+1} \right]$$

$$= N \frac{\pi_m}{s+1} .$$

Hence for all m, as $t \to \infty$, we have the following straightforward expressions for evaluating the steady-state expectations ξ^{*}_{im} :

$$E\, N_{1m}(t) \to N \cdot \pi_m/(s+1) = \xi^{*}_{1m}$$

and

$$E\, N_{im}(t) \to N \cdot p_{1m} \cdots p_{i-1,m}\, \pi_m/(s+1) = \xi^{*}_{im} ,\ i \geq 2 .$$

To obtain the entire probability distribution for the steady-state Markov chain, we define

$$\alpha_{1m} = \pi_m$$

and

$$\alpha_{im} = p_{1m} \cdots p_{i-1,m}\, \pi_m, \quad i \geq 2,$$

so that

$$\alpha_{im} = \alpha_{i-1,m}\, p_{i-1,m} \text{ for } i \geq 2.$$

Now let

$$\xi_{im} = \alpha_{im} / \sum_{j,\ell} \alpha_{j\ell},$$

so that $\xi_{im} = \xi_{i-1,m}\, p_{i-1,m}$. Then $\sum_{j,\ell} \alpha_{j\ell} = 1 + s$, and the above shows that $EN_{im}(t) = N\xi_{im}$. Note that each $\xi_{im} \geq 0$ and $\sum_{i,m} \xi_{im} = 1$. Consequently, since $\sum_{i,m} N_{im}(t) \equiv N$, one would guess that the stationary distribution of $(N_{im}(t))$ is multinomial with parameters N and (ξ_{im}), the latter representing "cell probabilities." This is in fact the case.

THEOREM. Let $\xi(\underset{\sim}{n})$ denote the multinomial distribution $\mathcal{M}(N, (\xi_{im}))$ on $\mathcal{S}$. Then $\xi = Q$.

Proof (outline)

We need to show that $\xi(\underset{\sim}{n})$ has the steady-state property that $\xi(\underset{\sim}{n}') = \sum_{\underset{\sim}{n}} P_{\underset{\sim}{n},\underset{\sim}{n}'}\, \xi(\underset{\sim}{n})$, for all $\underset{\sim}{n}, \underset{\sim}{n}' \in \mathcal{S}$. Since the steady-state distribution $Q(\underset{\sim}{n})$ is unique, we will then have $Q(\underset{\sim}{n}) = \xi(\underset{\sim}{n})$.

Now

$$\xi(\underset{\sim}{n}) = \frac{N!}{\prod_{i,m} n_{im}!} \prod_{i,m} \xi_{im}^{n_{im}} \quad \text{and} \quad \xi(\underset{\sim}{n}') = \frac{N!}{\prod_{i,m} n'_{im}!} \prod_{i,m} \xi_{im}^{n'_{im}}$$

Note that in order for $P_{\underset{\sim}{n},\underset{\sim}{n}'}$ to be positive, we must have

$n_{jm} \geq n'_{j+1,m}$, $1 \leq m \leq M$, $1 \leq j \leq 3$. Moreover,

$$\text{(A7.6)} \quad P_{\underset{\sim}{n},\underset{\sim}{n}'} = \underbrace{\prod_{m=1}^{M} \prod_{j=1}^{3} \binom{n_{jm}}{n'_{j+1,m}} p_{jm}^{n'_{j+1,m}} (1 - p_{jm})^{n_{jm}-n'_{j+1,m}}}_{\text{from fact that attritions are Bernoulli trials}} \; .$$

$$\cdot \left.\frac{(\sum_m n'_{1m})!}{\prod_m n'_{1m}} \prod_m \pi_m^{n'_{1m}} \right\} \text{from fact that accession mix is multinomial}$$

Also recall that $\xi_{j+1,m} = p_{jm} \xi_{jm}$.

Given $\underset{\sim}{n}'$, (A7.6), we get

$$\sum_{\underset{\sim}{n}} \xi(\underset{\sim}{n}) P_{\underset{\sim}{n},\underset{\sim}{n}'} = \sum_{\underset{\sim}{n}} \left(\frac{N!}{\prod_m \prod_i n_{im}!} \prod_{m=1}^{M} \prod_{k=1}^{4} \xi_{km}^{n_{km}} \right)$$

$$\left(\prod_m \prod_{j=1}^{3} \frac{n_{jm}!}{n'_{j+1,m}!(n_{jm}-n'_{j+1,m})!} \right.$$

$$\cdot \; p_{jm}^{n'_{j+1,m}} (1-p_{jm})^{n_{jm}-n'_{j+1,m}}$$

$$\left. \cdot \; \frac{(\sum_m n'_{1m})!}{\prod_m n'_{1m}} \prod_m \pi_m^{n'_{1m}} \right)$$

$$= \sum_{\underset{\sim}{n}} \frac{N!}{\prod_m \prod_i n_{im}!} \quad \frac{(\sum_m n'_{1m})!}{\prod_m n'_{1m}!}$$

$$\cdot \prod_{m=1}^{M} \left[\prod_{k=1}^{4} \xi_{km}^{n_{km}} \prod_{j=1}^{3} \frac{n_{jm}!}{n'_{j+1,m}!(n_{jm}-n'_{j+1,m})!} \right.$$

$$\left. \cdot\, p_{jm}^{n'_{j+1,m}} (1-p_{jm})^{n_{jm}-n'_{j+1,m}} \cdot \pi_m^{n'_{1m}} \right] .$$

For each m,

$$\xi_{1m}^{n_{1m}} \cdot \xi_{2m}^{n_{2m}} \cdots \xi_{4m}^{n_{4m}} = \xi_{1m}^{n_{1m}-n'_{2m}} \cdot \xi_{1m}^{n'_{2m}} \xi_{2m}^{n_{2m}-n'_{3m}}$$

$$\cdot\, \xi_{2m}^{n'_{3m}} \cdots \xi_{3m}^{n_{3m}-n'_{4m}} \cdot \xi_{3m}^{n'_{4m}} \xi_{4m}^{n_{4m}} .$$

Since $p_{jm}\,\xi_{jm} = \xi_{j+1,m}$, we have

$$p_{jm}^{n'_{j+1,m}} \xi_{jm}^{n'_{j+1,m}} = \xi_{j+1,m}^{n'_{j+1,m}} ,$$

hence we obtain

$$\text{(A7.7)} \qquad \frac{N!}{\prod_m \prod_{j=1}^{4} n'_{jm}!} \sum_{\underset{\sim}{n}} \frac{(\sum_m n'_{jm})!}{\prod_m n_{4m}!} \left[\prod_m \prod_{j=1}^{3} \frac{1}{(n_{jm}-n'_{j+1,m})!} \right]$$

$$\cdot\, \left(\prod_m \prod_{k=2}^{4} \xi_{km}^{n'_{km}}\right) \left(\prod_m \prod_{j=1}^{3} \xi_{jm}^{n_{jm}-n'_{j+1,m}}\right)$$

$$\cdot \left(\prod_m \prod_{j=1}^{3} (1-p_{jm})^{n_{jm}-n'_{j+1,m}}\right) \prod_m \xi_{4m}^{n_{4m}} \pi_m^{n'_{1m}}$$

$$= \frac{N!}{\prod_m \prod_{j=1}^{4} n'_{jm}!} \prod_m \prod_{k=2}^{4} \xi_{km}^{n'_{km}} \sum_{\underset{\sim}{n}} \frac{(\Sigma n'_{1m})!}{\prod_m n_{4m}!}$$

$$\cdot \prod_m \prod_{j=1}^{3} \frac{1}{(n_{jm}-n'_{j+1,m})!} \prod_m \prod_{j=1}^{3} (\xi_{jm} (1-p_{jm}))^{n_{jm}-n'_{j+1,m}}$$

$$\cdot \prod_m \xi_{4m}^{n_{4m}} \pi_n^{n'_{1m}} .$$

Now let $\ell_{jm} = n_{jm} - n'_{j+1,m}$, $1 \le j \le 3$, $1 \le m \le M$, so

$$\sum_{j,m} \ell_{jm} = \sum_m \left(\sum_{j=1}^{3} n_{jm} - \sum_{k=2}^{4} n'_{km}\right) = (N - \sum_m n_{4m}) - (N - \sum_m n'_{1m})$$

$$= \sum_m n'_{1m} - \sum_m n_{4m} .$$

Also note that

$$\xi_{1m} (1-p_{1m}) + \xi_{2m} (1-p_{2m}) + (\xi_{3m} (1-p_{3m})) + \xi_{4m} = \xi_{1m},$$

hence

$$\sum_m \sum_{j=1}^{3} \xi_{jm} (1-p_{jm}) = (\xi_{11}+\ldots+\xi_{1M}) = \frac{1}{1+s} ,$$

since

$$\xi_{jm} = \frac{\pi_m}{1+s} \text{ and } \sum_m \pi_m = 1 .$$

Thus the last sum in (A7.7) becomes

$$\prod_m \pi_m^{n'_{1m}} \sum_{\underset{\sim}{\ell}} \frac{(\sum_m n'_{1m})!}{\prod_m \prod_{j=1}^{3} \ell_{jm}! \cdot \prod_m n_{4m}!} \prod_m \prod_{j=1}^{3} (\xi_{jm} (1-\rho_{jm}))^{n_{jm}-n'_{j+1,m}}$$

$$\cdot \prod_m \xi_{4m}^{n_{4m}} .$$

But this sum is the multinomial expansion of

$$(\sum_m \sum_{j=1}^{3} \xi_{jm} (1-\rho_{jm}) + \sum_m \xi_{4m})^{\sum_m n'_{1m}} = (\frac{1}{1+s})^{\sum_m n'_{1m}} .$$

Thus we get

$$\prod_m \left(\frac{\pi_m}{1+s}\right)^{n'_{1m}} = \prod_m \xi_{1m}^{n'_{1m}} .$$

Hence we obtain

$$\frac{N!}{\prod_m \prod_1^4 n'_{\underset{\sim}{j}m}} \prod_m \prod_{k=2}^{4} \xi_{km}^{n'_{km}} \cdot \prod_m \xi_{1m}^{n'_{1m}} = \underset{\sim}{\xi}(\underset{\sim}{n}') .$$

Thus the proof is complete.

$\boxtimes$

Appendix B

TABLED COMPUTATIONS OF KEY OUTPUT QUANTITIES

TABLE B1. TOTAL ACCESSIONS

Case 1

UNCERTAINTY TYPES: Fixed P's and π's, Proportional Accessions

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	27.21007	2.551	0.094	136.051	5.703	0.042	272.102	8.066	0.030
2	29.33047	2.722	0.093	146.652	6.087	0.042	293.304	8.608	0.029
3	28.43757	2.970	0.104	142.188	6.642	0.047	284.376	9.393	0.033
4	30.36673	3.306	0.105	151.834	7.392	0.049	303.667	10.455	0.034
5	27.76970	3.694	0.133	138.849	8.261	0.059	277.698	11.682	0.042
6	29.18818	3.771	0.129	145.941	8.432	0.058	291.881	11.925	0.041
7	28.56967	3.849	0.135	142.848	8.607	0.060	285.697	12.172	0.043
8	29.94012	4.009	0.134	149.701	8.964	0.060	299.402	12.678	0.042
9	28.14908	4.136	0.147	140.745	9.249	0.066	281.491	13.081	0.046
10	29.09697	4.182	0.144	145.485	9.351	0.064	290.969	13.224	0.045

Case 2

UNCERTAINTY TYPES: Fixed P's and π's, Multinomial Accessions

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	27.210	2.551	0.094	136.051	5.703	0.042	272.102	8.066	0.030
2	29.330	2.723	0.093	146.652	6.090	0.042	293.304	8.612	0.029
3	28.438	2.973	0.105	142.188	6.647	0.047	284.376	9.400	0.033
4	30.367	3.308	0.109	151.834	7.398	0.049	303.667	10.462	0.034
5	27.770	3.700	0.133	138.849	8.274	0.060	277.698	11.701	0.042
6	29.188	3.777	0.129	145.941	8.446	0.058	291.881	11.945	0.041
7	28.570	3.855	0.135	142.848	8.621	0.060	285.697	12.192	0.043
8	29.940	4.016	0.134	149.701	8.979	0.060	299.402	12.608	0.042
9	28.149	4.144	0.147	140.745	9.266	0.066	281.491	13.105	0.047
10	29.097	4.190	0.144	145.485	9.369	0.064	290.969	13.249	0.046

Case 3

UNCERTAINTY TYPES: Random P's and π's, Multinomial Accessions

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	28.276	3.357	0.119	136.664	7.892	0.058	272.452	11.254	0.041
2	30.028	3.451	0.115	147.104	7.798	0.053	293.647	11.057	0.038
3	28.740	3.523	0.123	142.248	7.765	0.054	284.355	10.874	0.038
4	30.308	3.482	0.114	151.986	7.507	0.049	303.836	10.580	0.035
5	28.583	4.237	0.148	139.229	9.685	0.070	277.850	13.752	0.049
6	29.784	4.232	0.142	146.289	9.496	0.065	292.119	13.448	0.046
7	28.974	4.260	0.14	143.008	9.416	0.066	285.765	13.308	0.047
8	30.243	4.158	0.137	150.001	9.099	0.061	299.699	12.848	0.043
9	28.809	4.558	0.158	140.993	10.376	0.074	281.539	14.717	0.052
10	29.643	4.522	0.153	145.779	10.177	0.070	291.152	14.415	0.050

Case 4

UNCERTAINTY TYPE: Random P's and π's, Random Multinomial Accessions ($\alpha = 0.9$, $\beta = 0.9$)

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	25.448	5.880	0.231	122.998	[illegible]	0.107	245.207	18.649	0.076
2	29.261	7.395	0.253	143.296	[illegible]	0.115	286.929	23.265	0.081
3	28.477	7.546	0.265	140.991	[illegible]	0.119	281.840	23.686	0.084
4	30.050	7.602	0.253	149.830	[illegible]	0.113	299.568	23.879	0.080
5	26.774	8.290	0.310	130.528	18.609	0.143	260.497	26.336	0.101
6	28.733	8.611	0.300	141.023	[illegible]	0.137	281.581	27.298	[illegible]
7	28.489	8.708	0.306	140.694	[illegible]	0.139	281.143	27.581	0.098
8	29.521	8.754	0.297	146.570	19.618	0.134	292.906	27.742	0.095
9	27.530	8.944	0.325	134.895	[illegible]	0.149	269.361	28.490	0.106
10	28.517	9.020	0.316	140.131	[illegible]	0.145	279.854	28.699	0.103

TABLE B2. ACCESSIONS (LARGE CELL)

Case 1

UNCERTAINTY TYPES: Fixed p's and π's, Proportional Accessions

PLANNING YEAR	FORCE SIZE N = 100 μ	σ	c.v.	N = 500 μ	σ	c.v.	N = 1000 μ	σ	c.v.
1	12.245	1.1476	.1186	61.223	2.5665	.0419	122.446	3.6295	.0296
2	13.199	1.2247	.0928	65.993	2.7390	.0415	131.987	3.8735	.0293
3	12.797	1.3368	.1045	63.984	2.9890	.0467	127.969	4.2271	.0330
4	13.665	1.4876	.1089	68.325	3.3266	.0487	136.650	4.7046	.0344
5	12.496	1.6625	.1330	62.482	3.7174	.0595	124.964	5.2571	.0421
6	13.135	1.6971	.1292	65.673	3.7945	.0578	131.346	5.3662	.0409
7	12.856	1.7321	.1347	64.282	3.8730	.0603	128.564	5.4772	.0426
8	13.473	1.8042	.1339	67.365	4.0340	.0599	134.731	5.7050	.0423
9	12.667	1.8615	.1470	63.335	4.1622	.0657	126.671	5.8863	.0465
10	13.094	1.8818	.1437	65.468	4.2078	.0643	130.936	5.9508	.0454

Case 2

UNCERTAINTY TYPES: Fixed p's and π's, Multinomial Accessions

PLANNING YEAR	FORCE SIZE N = 100 μ	σ	c.v.	N = 500 μ	σ	c.v.	N = 1000 μ	σ	c.v.
1	12.245	2.8376	.2317	61.223	6.3450	.1036	122.446	8.9732	.0733
2	13.199	2.9599	.2243	65.993	6.6186	.1003	131.987	9.3601	.0709
3	12.797	2.9712	.2322	63.984	6.6436	.1038	127.969	9.3956	.0734
4	13.665	3.1196	.2283	68.325	6.9757	.1021	136.650	9.8652	.0722
5	12.496	3.1056	.2485	62.482	6.9446	.1111	124.964	9.8211	.0786
6	13.135	3.1801	.2421	65.673	7.1110	.1083	131.346	10.0564	.0766
7	12.856	3.1751	.2470	64.282	7.0996	.1104	128.564	10.0404	.0781
8	13.473	3.2673	.2425	67.365	7.3060	.1085	134.731	10.3322	.0767
9	12.667	3.2319	.2551	63.335	7.2265	.1141	126.671	10.2198	.0807
10	13.094	3.2796	.2505	65.468	7.3336	.1120	130.936	10.3712	.0792

Case 3

UNCERTAINTY TYPES: Random p's and π's, Multinomial Accessions

PLANNING YEAR	FORCE SIZE N = 100 μ	σ	c.v.	N = 500 μ	σ	c.v.	N = 1000 μ	σ	c.v.
1	12.752	3.8902	.3051	61.247	8.5099	.1389	121.999	12.2910	.1007
2	13.544	4.0660	.3002	65.916	8.8080	.1336	131.501	12.9623	.0986
3	12.959	3.9552	.3052	63.747	8.6674	.1360	127.333	12.6046	.0990
4	13.753	4.0962	.2978	68.111	8.9838	.1319	136.057	13.1072	.0963
5	12.891	4.0943	.3176	62.396	8.9633	.1437	124.418	12.9551	.1041
6	13.432	4.1927	.3121	65.552	9.1185	.1391	130.815	13.3525	.1021
7	13.066	4.1286	.3160	64.088	9.0292	.1409	127.964	13.1038	.1024
8	13.634	4.2048	.3084	67.220	9.2008	.1369	134.204	13.3953	.0998
9	12.992	4.1849	.3221	63.186	9.1762	.1453	126.071	13.2754	.1053
10	13.367	4.2410	.3173	65.325	9.2552	.1417	130.381	13.5177	.1037

Case 4

UNCERTAINTY TYPES: Random p's and π's, Random Multinomial Accessions
($\alpha = 0.9$, $\beta = 0.9$)

PLANNING YEAR	FORCE SIZE N = 100 μ	σ	c.v.	N = 500 μ	σ	c.v.	N = 1000 μ	σ	c.v.
1	11.477	4.2700	.3720	55.122	9.3037	.1688	109.799	13.3330	.1214
2	13.198	5.0008	.3789	64.210	10.8468	.1689	128.088	15.7019	.1226
3	12.840	4.9852	.3882	63.184	10.9144	.1727	126.206	15.6906	.1243
4	13.547	5.1064	.3769	67.144	11.2066	.1669	134.146	5.9983	.044
5	12.074	5.1252	.4245	58.496	11.2214	.1918	116.648	16.0592	.1377
6	12.958	5.3573	.4134	63.193	11.7031	.1852	126.095	4.0435	.0321
7	12.847	5.3720	.4189	63.050	11.7904	.1870	125.895	16.908	.1343
8	13.310	5.4478	.4093	65.682	11.9786	.1824	131.162	17.2086	.1312
9	12.415	5.3968	.4347	60.452	11.8520	.1961	120.619	16.9681	.1407
10	12.860	5.4812	.4262	62.795	12.0188	.1914	125.321	17.2719	.1378

TABLE B3. ACCESSIONS (SMALL CELL)

Case 1

UNCERTAINTY TYPES: Fixed p's and π's, Proportional Accessions

PLANNING YEAR	FORCE SIZE N = 100 μ	σ	c.v.	N = 500 μ	σ	c.v.	N = 1000 μ	σ	c.v.
1	1.361	.1265	.0929	6.803	.2846	.0418	13.605	.4037	.0297
2	1.467	.1378	.0939	7.333	.3050	.0416	14.665	.4301	.0293
3	1.422	.1483	.1043	7.109	.3317	.0467	14.219	.4701	.0331
4	1.518	.1643	.1082	7.592	.3701	.0487	15.183	.5225	.0344
5	1.388	.1844	.1329	6.942	.4135	.0596	13.885	.5840	.0421
6	1.459	.1897	.1300	7.297	.4219	.0578	14.594	.5967	.0409
7	1.428	.1924	.1347	7.142	.4301	.0602	14.285	.6083	.0426
8	1.497	.2000	.1336	7.485	.4483	.0599	14.970	.6340	.0424
9	1.407	.2074	.1474	7.037	.4626	.0657	14.075	.6542	.0465
10	1.455	.2098	.1442	7.274	.4680	.0643	14.548	.6611	.0454

Case 2

UNCERTAINTY TYPES: Fixed p's and π's, Multinomial Accessions

PLANNING YEAR	FORCE SIZE N = 100 μ	σ	c.v.	N = 500 μ	σ	c.v.	N = 1000 μ	σ	c.v.
1	1.361	1.1441	.8406	6.803	2.5581	.3760	13.605	3.6178	.2659
2	1.467	1.1883	.8100	7.333	2.6569	.3623	14.665	3.7573	.2562
3	1.422	1.1718	.8241	7.109	2.6199	.3685	14.219	3.7053	.2606
4	1.518	1.2124	.7987	7.592	2.7109	.3571	15.183	3.8338	.2525
5	1.388	1.1632	.8380	6.942	2.6012	.3747	13.885	3.6787	.2649
6	1.459	1.1925	.8173	7.297	2.6666	.3654	14.594	3.7711	.2584
7	1.428	1.1807	.8268	7.142	2.6403	.3697	14.285	3.7334	.2614
8	1.497	1.2091	.8077	7.485	2.7041	.3613	14.970	3.8243	.2555
9	1.407	1.1747	.8349	7.037	2.6268	.3733	14.075	3.7148	.2639
10	1.455	1.1942	.8208	7.274	2.6702	.3671	14.548	3.7762	.2596

Case 3

UNCERTAINTY TYPES: Random p's and π's, Multinomial Accessions

PLANNING YEAR	FORCE SIZE N = 100 μ	σ	c.v.	N = 500 μ	σ	c.v.	N = 1000 μ	σ	c.v.
1	1.388	1.5284	1.1012	6.747	3.4312	.5086	13.715	4.8963	.3570
2	1.474	1.5931	1.0808	7.266	3.6212	.4984	14.783	5.1718	.3498
3	1.412	1.5476	1.0960	7.021	3.5195	.5013	14.313	5.0463	.3526
4	1.503	1.6260	1.0818	7.505	3.6951	.4924	15.301	5.3159	.3474
5	1.404	1.5485	1.1029	6.874	3.4875	.5073	13.986	4.9816	.3562
6	1.462	1.5919	1.0889	7.225	3.6166	.5006	14.707	5.1690	.3515
7	1.424	1.5649	1.0989	7.059	[illegible]	.5023	14.387	5.0806	.3532
8	1.489	1.6193	1.0875	7.407	3.6707	.4956	15.092	5.2746	.3495
9	1.415	1.5611	1.1033	6.963	3.5232	.5061	14.172	5.0364	.3554
10	1.456	1.5903	1.0922	7.209	3.6115	.5016	14.659	5.1665	.3523

Case 4

UNCERTAINTY TYPES: Random p's and π's, Random Multinomial Accessions (α = 0.9, β = 0.9)

PLANNING YEAR	FORCE SIZE N = 100 μ	σ	c.v.	N = 500 μ	σ	c.v.	N = 1000 μ	σ	c.v.
1	1.249	1.4498	1.1608	6.073	3.2319	.5322	12.343	4.6092	.3734
2	1.436	1.6140	1.1240	7.078	3.6329	.5133	14.400	5.1856	.3601
3	1.400	1.5906	1.1361	6.959	3.5833	.5149	14.187	5.1333	.3618
4	1.480	1.6592	1.1211	7.398	3.7407	.5056	15.086	5.3764	.3564
5	1.315	1.5427	1.1732	6.434	3.4329	.5327	13.112	4.8001	.3663
6	1.411	1.6183	1.1469	6.965	3.6300	.5212	14.176	5.1830	.3657
7	1.400	1.6143	1.1531	6.946	3.6180	.5209	14.153	[illegible]	[illegible]
8	1.453	1.6583	1.1413	7.237	3.7214	.5142	14.750	5.3411	.3621
9	1.353	1.5843	1.1710	6.660	3.5306	.5301	13.559	[illegible]	.3719
10	1.431	1.6199	1.1362	6.921	3.6278	.5242	14.080	[illegible]	.3674

TABLE B4. TOTAL FOURTH-YEAR-GROUP SIZE

Case 1

UNCERTAINTY TYPES: Fixed p's and π's, Proportional Accessions

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	22.42	1.21	.0540	112.10	2.71	.0242	224.20	3.83	.0171
2	21.42	1.75	.0817	107.10	3.90	.0364	214.21	5.52	.0258
3	23.51	2.37	.1008	117.52	5.31	.0452	235.05	7.51	.0320
4	20.65	2.95	.1429	103.25	6.59	.0638	206.50	9.32	.0451
5	22.26	3.10	.1393	111.29	6.93	.0623	222.59	9.80	.0440
6	21.58	3.20	.1483	107.91	7.16	.0664	215.81	10.13	.0470
7	23.05	3.44	.1492	115.23	7.69	.0667	230.45	10.87	.0472
8	21.07	3.59	.1704	105.37	8.03	.0762	210.75	11.36	.0539
9	22.15	3.67	.1657	110.76	8.22	.0742	221.51	11.62	.0525
10	21.68	3.71	.1711	108.41	8.28	.0764	216.82	11.72	.0541

Case 2

UNCERTAINTY TYPES: Fixed p's and π's, Multinomial Accessions

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	22.420	1.213	.0541	112.100	2.711	.0242	224.200	3.834	.0171
2	21.421	1.746	.0815	107.103	3.903	.0364	214.207	5.520	.0258
3	23.505	2.374	.1010	117.523	5.310	.0452	235.046	7.509	.0319
4	20.650	2.954	.1431	103.249	6.605	.0640	206.498	9.341	.0452
5	22.259	3.105	.1395	111.294	6.942	.0624	222.589	9.817	.0441
6	21.581	3.208	.1486	107.906	7.174	.0665	215.813	10.145	.0470
7	23.045	3.444	.1494	115.227	7.701	.0668	230.454	10.890	.0473
8	21.074	3.601	.1709	105.372	8.052	.0764	210.745	11.387	.0540
9	22.151	3.682	.1662	110.755	8.233	.0743	221.509	11.644	.0526
10	21.682	3.713	.1712	108.408	8.303	.0766	216.816	11.742	.0542

Case 3

UNCERTAINTY TYPES: **Random p's and π's, Multinomial Accessions**

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	22.007	1.637	.0744	111.811	3.803	.0340	224.016	5.470	.0244
2	20.632	2.303	.1116	106.474	5.432	.0510	213.712	7.791	.0365
3	22.545	3.111	.1380	116.971	7.454	.0637	234.707	10.638	.0453
4	20.534	3.083	.1501	103.106	6.715	.0651	206.314	9.491	.0460
5	21.836	3.343	.1531	111.032	7.389	.0665	222.418	10.492	.0472
6	20.920	3.473	.1660	107.406	7.868	.0733	215.421	11.182	.0519
7	22.269	3.880	.1742	114.844	9.045	.0788	230.266	12.873	.0559
8	20.769	3.677	.1770	105.057	8.143	.0775	210.418	11.526	.0548
9	21.669	3.845	.1774	110.428	8.593	.0778	221.272	12.194	.0551
10	21.086	3.872	.1836	107.971	8.785	.0814	216.481	12.476	.0576

Case 4

UNCERTAINTY TYPES: Random p's and π's, Random Multinomial Accessions
(α = 0.9, β = 0.9)

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	22.007	1.637	.0744	111.811	3.803	.0340	224.016	5.470	.0244
2	20.632	2.303	.1116	106.474	5.432	.0510	213.712	7.791	.0365
3	22.545	3.111	.1380	116.971	7.454	.0637	234.707	10.638	.0453
4	18.481	4.662	.2523	92.795	10.437	.1125	185.682	14.772	.0796
5	21.282	5.844	.2746	108.159	13.218	.1222	216.649	18.737	.0865
6	20.733	6.014	.2901	106.461	13.754	.1292	213.520	19.510	.0914
7	21.935	6.290	.2868	113.214	14.554	.1286	227.031	20.663	.0910
8	19.455	6.396	.3288	98.497	14.558	.1478	197.282	20.624	.1045
9	20.907	6.725	.3217	106.453	15.329	.1440	213.289	21.727	.1019
10	20.739	6.827	.3292	106.231	15.662	.1474	212.986	22.207	.1043

TABLE B5. FOURTH-YEAR-GROUP SIZE (LARGE CELL)

Case 1

UNCERTAINTY TYPES: Fixed p's and π's, Proportional Accessions

PLANNING YEAR	FORCE SIZE N = 100 μ	σ	c.v.	N = 500 μ	σ	c.v.	N = 1000 μ	σ	c.v.
1	7.520	.6716	.0893	37.600	1.5020	.0399	75.200	2.1241	.0282
2	8.554	1.1122	.1300	42.770	2.4870	.0581	85.540	3.5170	.0411
3	9.033	1.4943	.1654	45.165	3.3417	.0740	90.330	4.7259	.0523
4	9.217	1.7393	.1887	46.086	3.8893	.0844	92.171	5.5003	.0597
5	9.935	1.8185	.1830	49.677	4.0661	.0819	99.353	5.7504	.0579
6	9.633	1.8423	.1912	48.164	4.1196	.0855	96.329	5.8260	.0605
7	10.286	1.9486	.1894	51.432	4.3574	.0847	102.864	6.1623	.0599
8	9.407	1.9728	.2097	47.033	4.4112	.0938	94.067	6.2384	.0663
9	9.887	2.0189	.2042	49.436	4.5145	.0913	98.871	6.3845	.0646
10	9.678	2.023	.2090	48.388	4.5236	.0935	96.777	6.3974	.0661

Case 2

UNCERTAINTY TYPES: Fixed p's and π's, Multinomial Accessions

PLANNING YEAR	FORCE SIZE N = 100 μ	σ	c.v.	N = 500 μ	σ	c.v.	N = 1000 μ	σ	c.v.
1	7.520	.6716	.0893	37.600	1.5020	.0399	75.200	2.1241	.0282
2	8.554	1.1122	.1300	42.770	2.4870	.0581	85.540	3.5170	.0411
3	9.033	1.4943	.1654	45.165	3.3417	.0740	90.330	4.7259	.0523
4	9.217	2.6155	.2838	46.086	5.8487	.1269	92.171	8.2713	.0897
5	9.935	2.7242	.2742	49.677	6.0914	.1226	99.353	8.6145	.0867
6	9.633	2.7174	.2821	48.164	6.0761	.1262	96.329	8.5929	.0892
7	10.286	2.8387	.2760	51.432	6.3474	.1234	102.864	8.9766	.0873
8	9.407	2.7912	.2967	47.033	6.2415	.1327	94.067	8.8268	.0938
9	9.887	2.8592	.2892	49.436	6.3934	.1293	98.871	9.0416	.0914
10	9.678	2.8469	.2942	48.388	6.3659	.1316	96.777	9.0028	.0930

Case 3

UNCERTAINTY TYPES: Random p's and π's, Multinomial Accessions

PLANNING YEAR	FORCE SIZE N = 100 μ	σ	c.v.	N = 500 μ	σ	c.v.	N = 1000 μ	σ	c.v.
1	7.399	.8871	.1199	37.442	2.0169	.0534	74.986	2.8567	.0381
2	8.361	1.4601	.1746	42.578	3.4116	.0801	85.279	4.8375	.0567
3	8.852	1.9204	.2169	45.114	4.5231	.1003	90.272	6.4169	.0711
4	9.326	3.2812	.3518	45.911	7.4987	.1633	91.703	10.6781	.1164
5	9.913	3.4376	.3468	49.512	7.7473	.1565	98.861	11.3809	.1151
6	9.493	3.3697	.3550	47.898	7.8267	.1634	95.742	11.2145	.1171
7	10.101	3.5683	.3533	51.209	8.3443	.1629	102.336	11.9813	.1171
8	9.430	3.4224	.3629	46.857	7.8356	.1672	93.525	11.1784	.1195
9	9.836	3.5239	.3583	49.244	8.0978	.1644	98.350	11.6685	.1186
10	9.569	3.4731	.3630	48.151	8.0454	.1671	96.214	11.5191	.1197

Case 4

UNCERTAINTY TYPES: Random p's and π's, Random Multinomial Accessions (α = 0.9, β = 0.9)

PLANNING YEAR	FORCE SIZE N = 100 μ	σ	c.v.	N = 500 μ	σ	c.v.	N = 1000 μ	σ	c.v.
1	7.399	.8871	.1199	37.442	2.0169	.0539	74.986	2.8567	.0381
2	8.361	1.4601	.1746	42.568	3.4116	.0801	85.279	4.8375	.0567
3	8.852	1.9204	.2169	45.114	4.5291	.1004	90.272	6.4169	.0711
4	8.391	3.4900	.4158	41.392	7.8976	.1908	82.533	11.2147	.1359
5	9.662	4.0531	.4195	48.232	9.1848	.1904	96.297	13.1635	.1367
6	9.409	4.0540	.4309	47.477	9.2863	.1956	94.898	13.2390	.1395
7	9.950	4.2189	.4240	50.483	9.7305	.1927	100.898	13.8992	.1504
8	8.833	4.0970	.4638	43.930	9.2830	.2113	87.686	13.1909	.1504
9	9.490	4.3073	.4539	47.473	9.7699	.2058	94.802	13.9532	.1472
10	9.413	4.3252	.4595	47.374	9.8693	.2083	94.661	14.0506	.1484

TABLE B6. FOURTH-YEAR-GROUP SIZE (SMALL CELL)

Case 1

UNCERTAINTY TYPES: Fixed p's and π's, Proportional Accessions

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	1.900	0.308	0.162	9.500	0.689	0.073	19.00	0.975	0.051
2	2.707	0.514	0.190	13.537	1.149	0.085	27.075	1.625	0.060
3	1.661	0.531	0.320	8.303	1.187	0.143	16.606	1.679	0.101
4	1.130	0.450	0.399	5.648	1.007	0.178	11.296	1.424	0.126
5	1.218	0.468	0.385	6.088	1.047	0.172	12.177	1.481	0.122
6	1.181	0.464	0.393	5.903	1.038	0.176	11.806	1.468	0.124
7	1.261	0.482	0.383	6.303	1.079	0.171	12.607	1.524	0.121
8	1.153	0.468	0.406	5.764	1.047	0.182	11.529	1.480	0.128
9	1.212	0.480	0.396	6.059	1.073	0.177	12.117	1.517	0.125
10	1.186	0.476	0.402	5.930	1.065	0.180	11.861	1.506	0.127

Case 2

UNCERTAINTY TYPES: Fixed p's and π's, Multinomial Accessions

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	1.900	.3082	.1622	9.500	.6892	.0725	19.000	.9747	.0513
2	2.707	.5138	.1898	13.537	1.489	.1100	27.075	1.6248	.0600
3	1.661	.5310	.3197	8.303	1.1870	.1430	16.606	1.6787	.1011
4	1.130	1.0459	.9256	5.648	2.3388	.4141	11.296	3.3074	.2928
5	1.218	1.0863	.8919	6.088	2.4288	.3989	12.177	3.4350	.2821
6	1.181	1.010	.8552	5.903	2.3946	.4057	11.806	3.3864	.2868
7	1.261	1.1077	.8784	6.303	2.4771	.3930	12.607	3.5031	.2779
8	1.153	1.0625	.9215	5.764	2.3755	.4121	11.529	3.3595	.2914
9	1.212	1.0890	.8985	6.059	2.4352	.4019	12.117	3.4438	.2842
10	1.186	1.0780	.9089	5.930	2.4108	.4065	11.861	3.4094	.2874

Case 3

UNCERTAINTY TYPES: Random p's and π's, Multinomial Accessions

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	1.764	.5148	.2918	9.312	1.1296	.1213	18.830	1.6239	.0862
2	2.394	.8307	.3470	13.098	1.9071	.1456	26.667	2.7814	.1043
3	1.395	.7113	.5099	7.963	1.6637	.2089	16.324	2.4308	.1489
4	.962	1.2157	1.2637	5.373	3.0378	.5654	11.169	4.3493	.3894
5	1.025	1.2700	1.2390	5.791	3.2125	.5547	12.044	4.6020	.3821
6	.985	1.2398	1.2587	5.600	3.1335	.5596	11.666	4.5048	.3863
7	1.048	1.2954	1.2361	5.987	3.2903	.5496	12.471	4.7401	.3801
8	.974	1.2313	1.2642	5.476	3.0889	.5641	11.391	4.4261	.3886
9	1.017	1.2685	1.2473	5.759	3.2081	.5571	11.982	4.5988	.3838
10	.992	1.2498	1.2599	5.629	3.1528	.5601	11.722	4.5296	.3864

Case 4

UNCERTAINTY TYPES: Random p's and π's, Random Multinomial Accessions
($\alpha = 0.9$, $\beta = 0.9$)

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	1.764	.51478	.2918	9.312	1.1296	.1213	18.830	1.6239	.0862
2	2.394	.8307	.3470	13.098	1.9071	.1456	26.667	2.7814	.1043
3	1.395	.7113	.5099	7.963	1.6637	.2089	16.324	2.4308	.1489
4	.866	1.1515	1.3297	4.836	2.8555	.5905	10.052	4.0894	.4068
5	.998	1.2795	1.2821	5.641	3.2108	.5692	11.732	4.5995	.3920
6	.976	1.2657	1.2968	5.550	3.1763	.5723	11.563	4.5651	.3948
7	1.032	1.3157	1.2749	5.902	3.3181	.5622	12.296	4.7793	.3887
8	.913	1.2190	1.3352	5.134	3.0269	.5896	10.680	4.3376	.4061
9	.981	1.2806	1.3054	5.551	3.2037	.5771	11.550	4.5923	.3976
10	.975	1.2795	1.3123	5.538	3.1991	.5777	11.533	4.5948	.3984

TABLE B7. TOTAL REQUIRED REENLISTMENT RATE

Case 1

UNCERTAINTY TYPES: Fixed p's and π's, Proportional Accessions

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	0.3937	0.0212	0.0539	0.3927	0.0095	0.0242	0.3926	0.0067	0.0171
2	0.4135	0.0335	0.0810	0.4114	0.0150	0.0364	0.4111	0.0106	0.0258
3	0.3782	0.0378	0.1000	0.3752	0.0169	0.0451	0.3748	0.0120	0.0319
4	0.4348	0.0608	0.1399	0.4279	0.0272	0.0636	0.4270	0.0192	0.0451
5	0.4030	0.0550	0.1366	0.3969	0.0246	0.0620	0.3961	0.0174	0.0439
6	0.4167	0.0605	0.1452	0.4096	0.0271	0.0661	0.4087	0.0191	0.0468
7	0.3904	0.0570	0.1459	0.3836	0.0255	0.0664	0.3827	0.0180	0.0471
8	0.4297	0.0712	0.1657	0.4200	0.0318	0.0758	0.4188	0.0225	0.0538
9	0.4082	0.0659	0.1614	0.3995	0.0295	0.0738	0.3984	0.0208	0.0523
10	0.4177	0.0694	0.1660	0.4082	0.0310	0.0760	0.4071	0.0219	0.0539

Case 2

UNCERTAINTY TYPES: Fixed p's and π's, Multinomial Accessions

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	0.3937	0.0212	0.0539	0.3927	0.0095	0.0242	0.3926	0.0067	0.0171
2	0.4135	0.0335	0.0810	0.4114	0.0150	0.0364	0.4111	0.0106	0.0258
3	0.3782	0.0378	0.1000	0.3752	0.0169	0.0451	0.3748	0.0120	0.0319
4	0.4349	0.0610	0.1402	0.4279	0.0273	0.0637	0.4270	0.0193	0.0451
5	0.4030	0.0551	0.1368	0.3969	0.0247	0.0621	0.3961	0.0174	0.0440
6	0.4168	0.0606	0.1454	0.4096	0.0271	0.0662	0.4087	0.0192	0.0469
7	0.3904	0.0571	0.1462	0.3836	0.0255	0.0665	0.3827	0.0180	0.0472
8	0.4298	0.0713	0.1660	0.4200	0.0319	0.0760	0.4188	0.0226	0.0539
9	0.4083	0.0660	0.1618	0.3995	0.0295	0.0739	0.3984	0.0209	0.0524
10	0.4178	0.0695	0.1664	0.4083	0.0311	0.0761	0.4071	0.0220	0.0540

Case 3

UNCERTAINTY TYPES: **Random p's and π's, Multinomial Accessions**

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	0.4021	0.0298	0.0740	0.3990	0.0134	0.0340	0.3931	0.0096	0.0244
2	0.4318	0.0476	0.1103	0.4143	0.0211	0.0509	0.4123	0.0150	0.0364
3	0.3978	0.0539	0.1354	0.3777	0.0240	0.0635	0.3757	0.0170	0.0452
4	0.4382	0.0643	0.1468	0.4286	0.0278	0.0648	0.4274	0.0196	0.0459
5	0.4124	0.0617	0.1496	0.3980	0.0264	0.0663	0.3965	0.0187	0.0471
6	0.4322	0.0696	0.1616	0.4119	0.0300	0.0729	0.4096	0.0212	0.0518
7	0.4072	0.0688	0.1691	0.3855	0.0302	0.0783	0.3834	0.0214	0.0557
8	0.4370	0.0750	0.1717	0.4213	0.0325	0.0770	0.4195	0.0229	0.0546
9	0.4189	0.0721	0.1720	0.4009	0.0310	0.0773	0.3989	0.0219	0.0549
10	0.4314	0.0766	0.1776	0.4102	0.0332	0.0808	0.4079	0.0234	0.0574

Case 4

UNCERTAINTY TYPES: Random p's and π's, Random Multinomial Accessions
(α = 0.9, β = 0.9)

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	0.4021	0.0298	0.0740	0.3940	0.0134	0.0340	0.3931	0.0096	0.0244
2	0.4318	0.0476	0.1103	0.4143	0.0211	0.0509	0.4123	0.0150	0.0364
3	0.3978	0.0539	0.1354	0.3777	0.0240	0.0635	0.3757	0.0170	0.0452
4	0.5065	0.1201	0.2372	0.4802	0.0533	0.1111	0.4769	0.0377	0.0791
5	0.4447	0.1135	0.2554	0.4129	0.0497	0.1204	0.4092	0.0351	0.0858
6	0.4602	0.1231	0.2676	0.4202	0.0534	0.1271	0.4156	0.0377	0.0906
7	0.4342	0.1150	0.2650	0.3951	0.0500	0.1265	0.3908	0.0353	0.0903
8	0.5012	0.1487	0.2967	0.4565	0.0660	0.1446	0.4509	0.0466	0.1034
9	0.4645	0.1354	0.2915	0.4219	0.0595	0.1411	0.4169	0.0420	0.1008
10	0.4703	0.1397	0.2970	0.4232	0.0611	0.1443	0.4177	0.0431	0.1031

TABLE B8. REQUIRED REENLISTMENT RATE (LARGE CELL)

Case 1

UNCERTAINTY TYPES: Fixed p's and π's, Proportional Accessions

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	.5309	.0470	.0886	.5275	.0210	.0399	.5271	.0149	.0282
2	.4709	.0602	.1279	.4646	.0269	.0579	.4638	.0190	.0410
3	.4505	.0725	.1610	.4409	.0324	.0736	.4397	.0229	.0522
4	.4450	.0811	.1822	.4328	.0363	.0838	.4313	.0256	.0595
5	.4120	.0730	.1771	.4013	.0326	.0813	.4000	.0231	.0577
6	.4262	.0786	.1845	.4142	.0352	.0849	.4127	.0249	.0603
7	.3989	.0729	.1829	.3878	.0326	.0841	.3864	.0231	.0597
8	.4396	.0883	.2009	.4248	.0395	.0930	.4229	.0279	.0660
9	.4173	.0818	.1960	.4039	.0366	.0906	.4023	.0259	.0643
10	.4272	.0856	.2003	.4129	.0383	.0927	.4111	.0271	.0658

Case 2

UNCERTAINTY TYPES: Fixed p's and π's, Multinomial Accessions

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	.5309	.0470	.0886	.5275	.0210	.0399	.5271	.0149	.0282
2	.4709	.0602	.1279	.4646	.0269	.0579	.4638	.0190	.0410
3	.4505	.0725	.1610	.4409	.0324	.0736	.4397	.0229	.0522
4	.4643	.1219	.2626	.4366	.0545	.1249	.4332	.0386	.0890
5	.4286	.1093	.2550	.4047	.0489	.1208	.4017	.0346	.0861
6	.4439	.1160	.2613	.4177	.0519	.1242	.4144	.0367	.0885
7	.4144	.1063	.2564	.3909	.0475	.1216	.3880	.0336	.0866
8	.4581	.1249	.2727	.4285	.0559	.1304	.4248	.0395	.0930
9	.4341	.1158	.2669	.4073	.0518	.1272	.4040	.0366	.0907
10	.4447	.1204	.2707	.4164	.0538	.1293	.4128	.0381	.0922

Case 3

UNCERTAINTY TYPES: Random p's and π's, Multinomial Accessions

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	.5430	.0642	.1182	.5305	.0285	.0537	.5290	.0201	.0380
2	.4882	.0827	.1695	.4681	.0373	.0796	.4660	.0263	.0565
3	.4685	.0971	.2072	.4434	.0441	.0994	.4410	.0312	.0707
4	.4773	.1494	.3131	.4421	.0702	.1588	.4378	.0503	.1149
5	.4476	.1386	.3095	.4101	.0636	.1551	.4060	.0461	.1136
6	.4698	.1481	.3152	.4245	.0676	.1592	.4194	.0485	.1155
7	.4411	.1385	.3141	.3970	.0630	.1587	.3923	.0453	.1155
8	.4753	.1524	.3207	.4345	.0707	.1627	.4296	.0506	.1178
9	.4545	.1445	.3179	.4130	.0661	.1601	.4084	.0478	.1170
10	.4686	.1505	.3211	.4228	.0687	.1626	.4176	.0493	.1180

Case 4

UNCERTAINTY TYPES: Random p's and π's, Random Multinomial Accessions ($\alpha = 0.9$, $\beta = 0.9$)

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	.5430	.0642	.1182	.5305	.0285	.0537	.5290	.0201	.0380
2	.4882	.0827	.1695	.4681	.0373	.0796	.4660	.0263	.0585
3	.4685	.0971	.2072	.4434	.0441	.0994	.4410	.0312	.0707
4	.5535	.1962	.3545	.4959	.0913	.1841	.4898	.0652	.1334
5	.4821	.1720	.3567	.4255	.0782	.1838	.4190	.0562	.1342
6	.4991	.1814	.3634	.4331	.0816	.1884	.4255	.0582	.1368
7	.4697	.1688	.3594	.4069	.0756	.1858	.4000	.0541	.1352
8	.5449	.2080	.3817	.4709	.0953	.2023	.4619	.0680	.1471
9	.5033	.1894	.3763	.4348	.0859	.1974	.4268	.0615	.1441
10	.5096	.1934	.3794	.4362	.0871	.1997	.4276	.0621	.1452

TABLE B9. REQUIRED REENLISTMENT RATE (SMALL CELL)

Case 1

UNCERTAINTY TYPES: Fixed p's and π's, Proportional Accessions

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	.2372	.0375	.1581	.2324	.0168	.0722	.2318	.0119	.0512
2	.1681	.0308	.1832	.1634	.0138	.0843	.1628	.0097	.0598
3	.2915	.0845	.2900	.2699	.0378	.1401	.2672	.0267	.1001
4	.4506	.1550	.3441	.4011	.0693	.1728	.3950	.0490	.1241
5	.4141	.1387	.3351	.3714	.0620	.1671	.3660	.0439	.1199
6	.4295	.1463	.3406	.3835	.0654	.1706	.3778	.0463	.1225
7	.3994	.1333	.3338	.3586	.0596	.1663	.3535	.0422	.1193
8	.4438	.1547	.3486	.3935	.0692	.1758	.3872	.0489	.1263
9	.4192	.1435	.3423	.3738	.0642	.1717	.3681	.0454	.1233
10	.4300	.1487	.3458	.3822	.0665	.1740	.3763	.0470	.1250

Case 2

UNCERTAINTY TYPES: Fixed p's and π's, Multinomial Accessions

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	.2372	.0375	.1581	.2324	.0168	.0722	.2318	.0119	.0512
2	.1681	.0308	.1832	.1634	.0138	.0843	.1628	.0097	.0598
3	.2915	.0845	.2900	.2699	.0378	.1401	.2672	.0267	.1001
4	.7221	.3600	.4985	.4554	.1610	.3535	.4221	.1138	.2697
5	.6477	.3217	[illegible]	.4181	.1439	.3442	.3894	.1017	.2613
6	.6781	.3374	.4976	.4332	.1509	.3483	.4026	.1067	.2650
7	.6174	.3061	.4959	.4022	.1369	.3404	.3753	.0968	.2580
8	.7044	.3510	.4983	.4456	.1570	.3523	.4133	.1110	.2686
9	.6552	.3257	.4972	.4210	.145[illegible]	.3460	.3917	.1030	.2630
10	.6763	.3366	.4977	.4315	.1505	.3489	.4009	.1064	.2655

Case 3

UNCERTAINTY TYPES: Random p's and π's, Multinomial Accessions

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	.2702	.0727	.2689	.2393	.0286	.1195	.2350	.0201	.0856
2	.2055	.0636	.3096	.1712	.0244	.1426	.1665	.0172	.1032
3	.3967	.1606	.4048	.2878	.0576	.2002	.2750	.0401	.1457
4	1.1848	.5766	.4867	.5393	.2310	.4284	.4529	.1531	.3381
5	1.0870	.5312	.4887	.4959	.2104	.4242	.4179	.1393	.3334
6	1.1521	.5612	.4871	.5150	.2195	.4262	.4326	.1454	.3361
7	1.0600	.5183	.4889	.4776	.2016	.4221	.4030	.1338	.3321
8	1.1713	.5700	.4866	.5287	.2262	.4279	.4437	.1498	.3376
9	1.1035	.5386	.4880	.4996	.2124	.4251	.4205	.1407	.3345
10	1.1451	.5576	.4870	.5125	.2185	.4264	.4306	.1448	.3362

Case 4

UNCERTAINTY TYPES: Random p's and π's, Random Multinomial Accessions
($\alpha = 0.9$, $\beta = 0.9$)

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	.2702	.0727	.2689	.2393	.0286	.1195	.2350	.0201	.0856
2	.2055	.0636	.3096	.1712	.0244	.1426	.1665	.0172	.1032
3	.3967	.1606	.4048	.2878	.0576	.2002	.2750	.0401	.1457
4	1.4039	.6744	.4804	.6124	.2681	.4378	.5092	.1777	.3491
5	1.1623	.5637	.4850	.5154	.2216	.4299	.4319	.1468	.3398
6	1.2072	.5837	.4836	.5252	.2264	.4311	.4390	.1500	.3416
7	1.1170	.5425	.4856	.4897	.2092	.4272	.4111	.1388	[illegible]
8	1.3395	.6427	.4798	.5764	.2522	[illegible]	.4790	.1670	[illegible]
9	1.2102	.5842	.4828	.5273	.2283	.4329	.4404	.1512	.3433
10	1.2250	.5906	.4821	.5288	.2290	.4331	[illegible]	[illegible]	.3438

TABLE B10. TOTAL COSTS

Case 1

UNCERTAINTY TYPES: Fixed p's and r's, Proportional Accessions

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	846963.23	60427.20	.0713	4234818.94	301623.69	.0712	8469638.33	603119.05	.0712
2	846371.77	60679.93	.0717	4231859.81	302646.60	.0715	8463718.02	605104.55	.0715
3	847577.12	60746.15	.0717	4237885.64	302759.63	.0714	8475775.15	605276.35	.0714
4	843656.57	60516.73	.0717	4218285.36	301191.11	.0714	8436570.74	602035.57	.0714
5	847335.61	60710.15	.0716	4236678.33	301957.17	.0713	8473360.41	603515.07	.0712
6	846662.52	60879.40	.0719	4233312.91	302668.69	.0715	8466627.69	604904.03	.0714
7	847329.12	60920.64	.0719	4236647.10	302775.49	.0715	8473292.84	605092.39	.0714
8	844491.30	60738.57	.0719	4222456.73	301658.12	.0714	8444913.84	602805.63	.0714
9	846979.26	60848.77	.0718	4234896.24	302130.16	.0713	8469791.36	603729.75	.0713
10	846565.07	60963.84	.0720	[illegible]	302629.60	.0715	8465655.19	604709.54	.0714

Case 2

UNCERTAINTY TYPES: Fixed p's and r's, Multinomial Accessions

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	846963.2270	60427.8752	.0713	4234818.9436	301624.27	.0712	8469638.3316	603119.63	.0712
2	846371.7682	60681.6020	.0717	4231859.8146	302648.28	.0715	8463718.0174	605106.22	.0715
3	847577.1203	60749.1151	.0717	4237885.6396	302762.60	.0714	8475775.1470	605279.32	.0714
4	843656.5654	60518.6526	.0717	4218285.3592	301195.04	.0714	8436570.7629	602039.51	.0714
5	847335.6090	60714.7261	.0717	4236678.3309	301961.76	.0713	8473360.4085	603519.67	.0712
6	846662.5164	60884.0528	.0719	4233312.9137	302673.36	.0715	8466627.6905	604908.71	.0716
7	847329.1187	60925.3756	.0719	4236647.1044	302780.24	.0715	8473292.8359	605097.15	.0716
8	844691.0995	60743.6070	.0719	4222456.7256	301663.18	.0714	8444913.8412	602810.70	.0716
9	846979.2584	60854.0750	.0718	4234896.2376	302135.49	.0713	8469791.3585	603735.09	.0713
10	846565.0718	60969.2107	.0720	4232827.4729	302635.00	.0715	8465655.1899	604714.94	.0714

Case 3

UNCERTAINTY TYPES: Random p's and r's, Multinomial Accessions

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	845279.0177	60422.5320	.0715	4233768.5895	301671.5820	.0713	8468937.9845	603195.2311	.0712
2	844277.6737	60636.0721	.0718	4230466.9937	302725.4915	.0716	8462587.4519	605236.0144	.0715
3	846363.7220	60737.0305	.0718	4237286.8864	302826.9026	.0715	8475282.0825	605372.9893	.0714
4	843654.2339	60556.4656	.0718	4218060.2826	301205.6864	.0714	8436097.9955	602041.8531	.0714
5	845926.8231	60726.4207	.0718	4235849.4996	302002.8750	.0713	8472700.6238	603580.9313	.0712
6	844995.0135	60857.5568	.0720	4232295.7336	302761.5119	.0715	8465289.4885	605037.5209	.0715
7	846103.1219	60918.3871	.0720	4236025.4056	302852.0405	.0715	8472273.0370	605198.6069	.0714
8	843836.7693	60729.9146	.0720	4221790.9655	301633.3565	.0714	8444076.2905	602778.9524	.0714
9	845676.2746	60843.4426	.0719	4234152.1687	302138.5603	.0714	8469187.9733	603777.4387	.0713
10	845139.3202	60942.3883	.0721	[illegible]	302718.1035	.0715	8465023.4399	604833.9876	.0715

Case 4

UNCERTAINTY TYPES: Random p's and r's, Random Multinomial Accessions [illegible]

PLANNING YEAR	FORCE SIZE								
	N = 100			N = 500			N = 1000		
	μ	σ	c.v.	μ	σ	c.v.	μ	σ	c.v.
1	824411.0480	69928.5884	.0848	4132910.0332	305644.7936	.0740	8267872.6424	599754.66	.0725
2	818315.7965	71880.5140	.0878	4103367.8293	308314.2754	.0751	8208947.8038	602819.79	.0736
3	818704.8935	71658.8536	.0875	4100344.6705	307165.0492	.0749	8201560.5043	600404.42	.0732
4	813085.8796	71883.9974	.0884	4066366.9027	304703.5007	.0749	8132774.5735	594391.37	.0731
5	822079.1872	71959.9546	.0875	4119253.6885	307345.4753	.0746	8239962.4103	600537.93	.0729
6	819142.3163	72962.4323	.0891	4105520.8689	309158.5559	.0753	8212690.3557	603026.98	.0734
7	818860.4259	72853.3379	.0890	4101494.5917	308560.1587	.0752	8203980.7423	601764.82	.0734
8	814999.4373	72966.5973	.0895	4078332.3190	306879.3367	.0752	8157277.5202	597723.20	.0733
9	820272.8550	[illegible]	.0887	4109206.0342	307996.6545	.0750	8219624.0214	600751.81	.0731
10	819128.9917	73298.0464	.0895	4104320.4698	309246.9890	.0753	8209996.7560	602660.08	.0734

Appendix C

DERIVATION FOR NUMBER OF REQUIRED RECRUITS

In this Appendix we derive the value a required such that $P(R = c/Y \leq r \mid A = a) \geq b$, where A = a is the number of people that should be recruited to assure a probability of at least b that a reenlistment rate R no higher than r will be required to reenlist c of those people for the career force, and where p is the probability that an individual recruit makes it through his initial four-year obligation.

If R is the (random) required reenlistment rate and Y is the number of people remaining after four years of service of those a who are recruited initially, then R = c/Y, and Y is binomial with parameters a and p. Thus Y has mean pa and variance p(1 - p)a. Then $Z = (Y - pa)/\sqrt{p(1-p)a}$ is approximately a standard normal (mean 0, variance 1) random variable, and we let z_b be the upper quantile point such that $P(Z \geq z_b) = b$.

Since

$$\begin{aligned} P(R \leq r \mid A = a) &= P(c/Y \leq r \mid A = a) \\ &= P(c/r \leq Y \mid A = a) \\ &= P\left(Z \geq \frac{\frac{c}{r} - ap}{\sqrt{p(1-p)a}} \,\middle|\, A = a\right) , \end{aligned}$$

we obtain

$$z_b = \frac{\frac{c}{r} - ap}{\sqrt{p(1-p)a}} \ .$$

Solving for a, we obtain a quadratic equation in a, and its solution

$$a = \frac{2pc/r + z_b^2\ p(1-p) + \sqrt{[2pc/r + z_b^2\ p(1-p)]^2 - 4p^2c^2/r^2}}{2p^2} \ .$$

REFERENCES

[1] Bartholomew, D. J. (1973). Stochastic Models for Social Processes, John Wiley and Sons, New York.

[2] ----- (1975). "Errors of Prediction in Markov Chain Models," Journal of the Royal Statistical Society, Series B, Vol. 37, pp. 444-456.

[3] ----- (1976). "Statistical Problems of Prediction and Control in Manpower Planning," Mathematical Scientist, Vol. 1, pp. 133-144.

[4] ----- (1970). "Some Statistical Techniques in Manpower Planning," A. R. Smith (ed.), Her Majesty's Stationery Office.

[5] Bartholomew, D. J., and A. F. Forbes (1979). Statistical Techniques for Manpower Planning, John Wiley and Sons, New York.

[6] Bartholomew, D. J., R. F. A. Hopes, and A. R. Smith (1976). "Manpower Planning in the Face of Uncertainty," Personnel Review, Vol. 5, pp. 5-17.

[7] Bickel, P. J., and K. A. Doksum (1977). Mathematical Statistics: Basic Ideas and Selected Topics, Holden-Day, San Francisco.

[8] DeGroot, M. H. (1970). Optimal Statistical Decisions, McGraw-Hill, New York.

[9] Doyle, P., and I. Fenwick (1975). "The Pitfalls of AID Analysis," Journal of Marketing Research, Vol. 12, pp. 408-413.

[10] Fishman, G. S. (1979). Principles of Discrete Event Simulation, John Wiley and Sons, New York.

[11] Gotz, G. A., and J. J. McCall (1980). Estimating Military Personnel Retention Rates: Theory and Statistical Method, The Rand Corporation, R-2541-AF.

[12] Grinold, R. C., and K. T. Marshall (1977). Manpower Planning Models, North-Holland Publishing Company, New York.

[13] Parzen, E. (1962). Stochastic Processes, Holden-Day, San Francisco.

[14] Rueter, F. H., et al. (1981). Integrated Simulation Evaluation Model Prototype (ISEM-P) of the Air Force Manpower and Personnel System: Overview and Sensitivity Analysis, TR-81-15, Air Force Human Resources Laboratory, Brooks Air Force Base, Texas.

[15] Vajda, S. (1978). Mathematics of Manpower Planning, John Wiley and Sons, New York.

ATE
LMED

-8