Estimating Workforce Attrition Rate Parameters: A Controlled Comparison

Robert Mark Bryce and Jillian Anne Henderson
Defence Research and Development Canada (DRDC)-Centre for Operational Research and Analysis (CORA), Department of National Defence (DND), 60 Moodie Drive, Ottawa, Canada

Keywords: Attrition, Workforce Analytics, Memoryless, Survival Time, Stochastic Process, Estimator.

Abstract: Here we consider workforce attrition rate parameter estimators (three continuous, three discrete), demonstrating that they originate in the Ordinary Differential Equation (ODE) describing a memoryless survival time population. This demonstration reveals how to craft estimators and reason about them. Setting up controlled numerical experiments we test this family of estimators to determine their relative performance, both in steady state and under dynamic change of population level. Additionally, we find the analytic expressions for the residual error of the estimators which provides insight into their bias properties. Our results suggest that, out of common simple estimators, the two-point average population and half-intake estimators are robust choices.

1 INTRODUCTION

Determining the attrition properties of a workforce is informative as a diagnostic (e.g., are more members leaving than typical?) and for planning purposes (e.g., if we want to grow by 10% within three years how many hires per annum will we need to take on?). Comparison to peers can be valuable, providing insight on the relative position ones institution holds, and forecasting is useful for a "heads up" view on institutional health. To this end having a good estimate of the attrition rate parameter describing one's workforce is valuable, where good means, among other things, providing precision and accuracy for modelling and forecasting purposes. In this work we compare estimators of the attrition rate parameter which are in use by various militaries, who have a keen interest in managing occupation levels (sub-populations) that deliver mandated capability, with the aim of providing a controlled comparison. To this end we set up numerical experiments with known attrition rate properties and quantitatively check the performance of the estimators.

We take a simple model of workforce attrition as foundational: workforce attrition is viewed as a stochastic process. A quasi-stationary survival time distribution is taken to describe member lifetime within the institution. This viewpoint allows for stochastic simulations, such as Discrete Event Simulation (DES), to generate population trajectories via random assignment of lifetime from the survival distribution. Notable, the historical record is a particular realization of a stochastic process describing the population. The stochastic process view takes the additional step of positing that other realizations are informative and meaningful, including future population trajectories. The validity of this assumption hinges on the survival distribution capturing the variation expected, and, for forecasts, that the institute and embedding social conditions do not qualitatively change or show "significant" quantitative change over the forecast horizon.

One key advantage of the stochastic process view is contact can be made with theory, for the case of a memoryless population (see Section 2, where the theory is outlined) which maps directly, and uniquely, to an exponential survival distribution (Feller, 1957). Furthermore the theoretical framework speaks to the majority of attrition rate parameter estimators (Section 3), with the possible exception of heuristic estimators where contact with theory is more opaque. As the estimators can be viewed as a family which assumes an identical underlying model (memoryless attrition) it is expected that all estimators will, broadly speaking, be in agreement. Despite this one can expect differences in performance and, for example, one can imagine some estimators showing bias in various contexts. To identify possible bias and speak to relative performance we consider three scenarios: (1) steady, (2) growing, and (3) reduction of populations,

Bryce, R. and Henderson, J.
Estimating Workforce Attrition Rate Parameters: A Controlled Comparison.
DOI: 10.5220/0011748600003396
In *Proceedings of the 12th International Conference on Operations Research and Enterprise Systems (ICORES 2023)*, pages 82-93
ISBN: 978-989-758-627-9; ISSN: 2184-4372

by constructing numerical experiments with known attrition properties (Section 4). These experiments mainly consider the "best case"—a memoryless population—as interpretation is facilitated here, but we also consider a population with a distinctly non-exponential survival distribution (Canadian Armed Forces (CAF) Regular Force (Reg F), Primary Reserves (P Res), and Naval Warfare Officer (NWO)) to extend analysis and speak to practical issues.

2 MODELS OF ATTRITION

It appears that, in practice, the concept of workforce attrition is associated with the number of members of a population leaving in some set period and is measured by a churn, or attrition, rate. The churn rate is defined as the number of members who leave on an interval, *out*, divided by the population size, P, a conceptually viable definition. As, in general, the population level changes when members leave and, typically, new members enter over time there are questions regarding what "the" population is and there are differing specific choices one can make—take the population level at the start of the interval, at the end, the average, etc.

Despite the ambiguity and apparent *ad hoc* choice in how to measure attrition, we will show that there is essentially a single model of attrition—the memoryless model—plus the stochastic process view, which can either match this model (when the survival distribution is exponential) or be general (e.g., non-parametric distributions such as a histogram or cumulative distribution function (CDF), or parametric forms). Viewed from the lens of the memoryless model common means of measuring "attrition rate", which we term the attrition rate *parameter* (for reasons that will become apparent), will be seen as members of the same family and which all naturally follow from the same originating equation.

It should be noted that even a loose definition of churn will approximately measure the attrition and, unless dramatic change in population levels occurs over the interval in question, these will all be similar due to the same originating concept. While this is unarguably true it should be noted that 1) churn rate of customer contracts or employees, say, is used to make financial and hiring decisions and so bias and error in estimates can have real impact on decisions, 2) a careful development will provide a means to select a "good" means of estimating, as even if the choices are similar perhaps some are *better* in some sense, and 3) the development here can speak towards improvements beyond the current estimators in use.

2.1 The Memoryless Model

A simple continuous differential model describing the change of a homogenous population over time due to attrition at a rate αP (outflow) and a constant intake rate in (inflow) is

$$P' = -\alpha P + in \tag{1}$$

where P' is the time derivative of the population. This Ordinary Differential Equation (ODE) has an exact solution

$$P(t) = \frac{in}{\alpha} + \left(P_0 - \frac{in}{\alpha}\right)e^{-\alpha t}, \tag{2}$$

where P_0 is the initial population, as can be verified by plugging this solution back into the originating differential equation. Evaluating the solution (2) at discrete multiples of the unit time $t = 0, 1, 2\ldots$ we recast (2) as a recurrence relation and find

$$\begin{aligned} P_{t+1} &= \frac{in}{\alpha} + \left(P_t - \frac{in}{\alpha}\right)e^{-\alpha} = P_t e^{-\alpha} + \frac{in}{\alpha}(1 - e^{-\alpha}) \\ &= (1-\gamma)P_t + in\left(\frac{\gamma}{\alpha}\right), \end{aligned} \tag{3}$$

where the last expression foreshadows the corrected discrete approximation we will consider in the next section, and which makes use of the mapping between the continuous time attrition rate parameter and the discrete time probability $\gamma = 1 - e^{-\alpha T}$, where T is the unit time.

Here it is clear that the attrition rate αP is the attrition rate parameter α times the population level, which matches simple dimensional analysis where α has units of inverse (unit) time and P units of members as so multiplied together they have units of a rate (members per (unit) time leaving).

It should be noted that in (1) a constant intake rate is assumed. This facilitates solution (2), but does not lead to meaningful loss of generality as functions with a finite number of jump discontinuities can be approximated by piecewise constant functions with arbitrary precision—one simply solves (1) on each piecewise constant region.

2.2 Discrete Approximation

Discrete Time Markov Models (DTMMs) have long been a popular approach to workforce analysis (Seal, 1945; Young and Almond, 1961; Merck and Hall, 1971; Vajda, 1978; Bartholomew and Forbes, 1979), most likely due to the ease of implementation and widespread knowledge of Markov chains. However, they are typically misspecified in that entry events can happen between the time steps and therefore as time

steps increase in size error will increasingly accrue. We will first outline the standard discrete time approach, in the single component workforce case, provide an approximate correction given in (Okazawa, 2007), and demonstrate the exact correction which recovers equation (3).

The basic idea behind the DTMM for workforces is that entry processes occur at discrete points in time. For a homogeneous population only two processes occur: entry into the population, and exit from the population (for heterogeneous populations flow between subcomponents also occurs). Entry is treated deterministically with a set number of entrants entering at the end of a time period, while exit is considered probabilistically with a given probability of $\gamma = 1 - e^{-\alpha T}$ of leaving (Vajda, 1978; Bartholomew and Forbes, 1979); note the explicit relationship with the continuous time attrition rate parameter. The result is the following equation describing the population over time

$$P_{t+1} = P_t - out + in = (1-\gamma)P_t + in, \qquad (4)$$

where $out = \gamma P_t$ is the number of personnel who leave, in expectation, over a time step and in is the intake occurring at the end of the time period.

Notice that (3) differs from (4) in that in has a modifying factor—this is due to the discrete time steps neglecting intake events happening between the time steps. The main limitation of (4), for modelling continuous entry into a population, is that entry can occur between the time steps leading to error. In (Okazawa, 2007) an approximate correction that accounts for continuous entry between time steps is developed leading to

$$P_{t+1} = (1-\gamma)P_t + (1-\gamma/2)in. \qquad (5)$$

This correction is denoted the "half-intake" formulation for obvious reasons. With regards to this approximate correction we note that by assuming continuous intake over the time period one models situations where entry happens uniformly over time, where in is considered as an intake density (number who enter, divided by the time period; for simplicity we will work with unit time). The density entering at time t will experience exponential decay for a time of $(1-t)$, i.e., to the end of the time step. This modifies (4) in the following manner

$$\begin{aligned} P_{t+1} &= (1-\gamma)P_t + \int_0^1 in \cdot e^{-\alpha(1-t)}dt \\ &= (1-\gamma)P_t + in\Big(\frac{1-e^{-\alpha}}{\alpha}\Big) \\ &= (1-\gamma)P_t + in\Big(\frac{\gamma}{\alpha}\Big), \end{aligned} \qquad (6)$$

and (3) is recovered exactly.

Finally, we note that the multiple component DTMM model is a simple (matrix) extension of the single component model outlined here (Bartholomew and Forbes, 1979). While not widely recognized, one can also directly solve the matrix version of (1) either exactly (Higham, 2008; Henderson and Bryce, 2019) or approximately by finding the Linear Dynamical System (LDS) associated with the matrix version of (1) (i.e., use Euler's method). The LDS is identical in form with the DTMM, with the discrete transition probabilities replaced with continuous attrition rate parameters. Note that as the time step diminishes to zero the LDS and DTMM converge to the same (exact) limit. If a continuous entry process describes a population the continuous model is a better fit than a discrete time model, and as the LDS is identical in form to the DTMM the advantages of DTMM (mainly simple implementation) can be had in the continuous domain while reducing modelling error due to taking finite time steps. It should also be noted that simply reducing the time step towards zero will lead to improved results, for either approximate approach (DTMM, LDS).

2.3 The General Formula (GF)

The GF for the attrition rate parameter is an heuristically derived approach (Vincent et al., 2021; Vincent, 2019; Vincent et al., 2018). The GF approach offers an alternative perspective on attrition to that taken here; Ref. (Vincent et al., 2021) is a readable and careful overview of the GF approach which includes empirical results and reviews prior work.

Briefly, the approach taken is to consider subintervals defined by intake events and imposing the constraint that the attrition formula should have some desired properties (namely, the attrition rate parameter is the proportion that departs and the overall rate is related to the sub-period rates) which leads to

$$1 - \gamma = \prod_{i=1}^{n} \frac{p_i}{p_i + out_i} \qquad \text{(GF)}$$

where γ is the discrete attrition rate parameter, i indexes the n sub-intervals, p_i is the population at the end of the *i-th* sub-interval, and out_i is the attrition on the *i-th* sub-interval.

Note that there is no explicit time scale in the GF and that a change in sub-interval size does not affect the measured attrition rate parameter; as the attrition rate parameter describes the expected time to attrition there must be a time scale, and so the GF relies on an implicit unit time (e.g., the measurement period). Further note that as intake events define the sub-intervals

these are free fall (no intake) regions, on which discrete time approaches are expected to perform well, with the important distinction that the time intervals are not fixed.

2.4 Stochastic Processes

We first remark on the relationship $\gamma = 1 - e^{-\alpha T}$ that maps the continuous attrition rate α to the discrete one γ. This relationship describes decay between impulse intakes at discrete times and can be obtained directly from Equation 2 by setting *in* to zero to find the exponential decay on free fall periods $P_0 e^{-\alpha T}$ and noting that the fraction of the initial population remaining will be $P_0(1-\gamma)$ where we introduce γ as the fraction of the population that will attrite on the interval. We can generalize our view by considering the stochastic process perspective where for some initial population members have lifetimes that will dictate their eventual exit time. Here, for an exponential survival curve the decay is, by definition, exponential and the continuous-discrete relationship holds. A survival curve is defined as $1 - CDF(t)$, where $CDF(t)$ is the cumulative distribution function, and for an exponential survival curve the distribution is the exponential distribution. We see then that a stochastic process with an exponential survival distribution will match, in expectation, the ODE describing a memoryless population—at least when intake occurs at fixed discrete times.

We strengthen the connection between a (exponential survival distribution) stochastic processes and the memoryless population ODE view by noting the following. First, we do not need intake to occur at fixed points in time. Rather we can allow variable free fall periods with boundaries defined by intake events (e.g., as for the GF), and again the survival curve will match the ODE, in expectation, on the intervals and the continuous-discrete relationship holds. Second, if we have *uniform* intake of *in* over a unit time interval than the solution to the ODE with intake rate *in* (Equation (2)) is recovered, in expectation, as can be seen by considering Equation (6) and associated discussion. We see that a stochastic process, defined by an exponential survival distribution and (piecewise constant) uniform intake regions, will have average behaviour described by the memoryless ODE describing the population.

One can consider a more general stochastic process, where instead of drawing from an exponential survival distribution we draw from a general distribution and again have uniform piecewise constant intake. While the ODE (1) describing a memoryless population will no longer fully describe the dynamics we can simulate under historical and planned intake plans using, for example, DES (Devroye, 1998). For this work we will be using stochastic simulations where we uniformly create entry events at a given annual intake rate, and for each entry draw a survival time from a distribution (either exponential or an empirical histogram) to find the corresponding exit event. From a given realization we will measure the population at an annual frequency, and measure the outflow *out* on the annual time intervals and use these measurements to assess attrition rate parameter estimators (see next section). Details on the numerical experiments are provided in Section 4.

This direct match between a stochastic process view of the population and the ODE (1) describing the mean population trajectory, for the appropriate *in* profile, is advantageous as one can verify stochastic simulations against theory in this special (memoryless population) case (Henderson and Bryce, 2019) and then extend to more complex survival distributions (Bryce and Henderson, 2020) if warranted by modelling needs. In particular, modelling error due to assuming the memoryless case can be probed; depending on the modelling context this error can be quite limited (Bryce and Henderson, 2020), and for mild changes in intake the memoryless assumption is a reasonable simplification.

3 MEASURING α FROM DATA

3.1 Estimators

Provided a model one would like to determine the parameters, given data. To do this an estimator has to be constructed. In Equation 1 the instantaneous attrition rate (outflow volume per unit time) is $out' = \alpha P(t)$, where the prime is used here to emphasize this is a rate, and we have

$$out = \int_{t_1}^{t_2} out' dt = \alpha \int_{t_1}^{t_2} P(t) dt \tag{7}$$

where we take α as constant and pull it out of the integral. We can define an interval $(t_2 - t_1)$, measure the number *out* who attrite over this interval, and solve the integral of the population to obtain an estimator of the attrition rate parameter

$$\widehat{\alpha} = \frac{\int_{t_1}^{t_2} out' dt}{\int_{t_1}^{t_2} P(t) dt} \tag{8}$$

where we use a hat to emphasize this is an estimator.

Here we focus on simple numerical estimates of the population integral, as these relate to commonly

used estimators, but we note that the population integral over an interval can be determined from Human Resources records and also found for simulated and forecast populations. We construct three estimators using simple numerical integration schemes on a single interval, and then introduce three other estimators used in practice. In the following we use an annual interval, as this is a natural unit for organizations and human resources, and so $(t_2 - t_1)$ falls out as it is unit—but is implicitly present.

- Steady State Estimator (Continuous).
  We consider the ideal case of a constant (steady state) population, where $P(t) = P_{ss}$ so $\int_{t_1}^{t_2} P(t)dt = P_{ss}$ exactly. This gives
  $$\widehat{\alpha}_{ss} = \frac{out}{P_{ss}}. \tag{E0}$$
  This estimator is only suitable when the population does not change which, in general, we do not expect to hold even in steady state (due to stochastic fluctuations). The change in population levels is what motivates a need for various estimators. Here we do not empirically test E0, due to it's limited applicability, but instead derive it to make contact with intuitive development of estimators, which often start with E0 based on the conceptual definition of attrition, as well as to support its use in reasoning about steady state conditions.
- Left Estimator (Continuous).
  We consider the left Riemann sum on a single interval, $\int_{t_1}^{t_2} P(t)dt \approx P(t_1)$ which, substituting into Equation 8, leads to
  $$\widehat{\alpha}_1 = \frac{out}{P_0} \tag{E1}$$
  where we use P_0 to denote $P(t_1)$ for convenience. The Korea Institute for Defence Analyses uses this estimator (Vincent, 2019).
- Mean Estimator (Continuous).
  Using the trapezoidal rule on a single interval $\int_{t_1}^{t_2} P(t)dt \approx (P(t_1) + P(t_2))/2$ leads to
  $$\widehat{\alpha}_2 = \frac{out}{P_{mean}} = \frac{2\ out}{P_0 + P_1} \tag{E2}$$
  where we again change notation for the starting and end populations on the interval. We denote this the mean estimator as the two-point average over the interval is used to estimate the population. This estimator is used by the Australian Army, Swedish Defence Research Agency, and U.K.'s Royal Air Force (Vincent, 2019).
- Right Estimator (Continuous).
  Using the right Riemann sum on a single interval leads to
  $$\widehat{\alpha}_3 = \frac{out}{P_1} \tag{E3}$$
  where we use P_1 for $P(t_2)$.
- Half-Intake Estimator (Discrete).
  $$\widehat{\gamma}_4 = \frac{out}{P_0 + in/2} \tag{E4}$$
  This estimator is used by the CAF (Vincent, 2019), and is a correction of a DTMM approximation (Okazawa, 2007). This is a semi-discrete model as it does not correspond to a probability, as in the discrete case, as can be seen by considering the high attrition steady state limit where *in* and $out \gg P_0$ such that the estimate could take on values > 1. At lower attrition and time steps the estimator should become increasingly well behaved.
- Markov Estimator (Discrete).
  Here attrition is measured in free fall conditions (no intake) over a single time step. $\int_{t_1}^{t_2} P(t)dt = \frac{P_0}{\alpha}\gamma$, which is obtained by plugging in the solution to Equation 1 (i.e., Equation 2) with an intake rate of zero, and using the discrete-continuous map definition of γ. Plugging this into Equation 8 allows us to cancel the α from both sides and move the γ to the left-hand side and we find
  $$\widehat{\gamma}_5 = \frac{out}{P_0}. \tag{E5}$$
  Note that free fall conditions is misspecified when there is a non-zero intake rate, e.g. a discrete time step process is misspecified for continuous entry populations. This estimator is recommended in (Bartholomew and Forbes, 1979) as this form is known to be the Maximum Likelihood Estimation (MLE) for DTMM (Anderson and Goodman, 1957).
- General Formula (GF) Estimator (Discrete).
  Here we start as for the Markov estimator. In free fall the following holds: $P_1 = P_0 - out$, we use this constraint to find the GF estimator as follows. Apply the constraint to the numerator *and* denominator of E5, and simplify to get the single time interval GF estimator
  $$\widehat{\gamma}_6 = 1 - \frac{P_1}{P_1 + out}. \tag{E6}$$
  As for the Markov estimator this will be misspecified for non-zero intake over an interval. This estimator is recommended in Ref. (Vincent et al., 2021).

It should be noted that, in general, there can be confusion between whether the discrete or continuous attrition rate parameter is being estimated from the data, for example (E1) and (E5) are identical in form but measure the continuous and discrete rate parameters, respectively. For comparison and modelling purposes it is important to be clear on this point. For clarity here we map all discrete estimates back to the continuous domain by inverting $\gamma = 1 - e^{-\alpha T}$, this allows all estimators to be directly compared to each other and to the set attrition rate. For example, for (E5) we report as $\widehat{\alpha}_5 = -ln(1-\widehat{\gamma}_5) = -ln(1-\frac{out}{P_0})$.

The discrete approach follows the same underlying attrition model and is a discrete time approximation to the ODE describing a memoryless population, and so the discrete estimators (E4–E6) fall into the same fundamental family of estimators. For finite time steps these estimators will have discretization error, but for small α this error will be small and in general all estimators are expected to converge and perform well in the vanishing α limit.

We can relate the three continuous estimators we constructed from single step numerical integration, as

$$\frac{P_0}{out} + \frac{P_1}{out} = \frac{2(P_0+P_1)}{2out} \tag{9}$$

and so

$$\frac{1}{\widehat{\alpha}_1} + \frac{1}{\widehat{\alpha}_3} = \frac{2}{\widehat{\alpha}_2}. \tag{10}$$

We note the following mnemonic to help identify which index refers to which estimator: the left estimator takes the first point in the interval ("one"), the mean estimator uses the midpoint approximation ("two"), and the right estimator takes the last point ("three").

3.2 Residual Error of Estimators

As we are considering the performance of the estimators we derive analytic forms for the expected residual error. We first consider correlation effects in combination with steady state considerations to find residuals for the continuous estimators, and then use steady state arguments to find residuals for the discrete estimators. We note that the residual is equivalent to the bias of the estimator, and emphasize that the residuals apply in steady state.

3.2.1 Residual Error for Continuous Estimators

Observe the following. In steady state if there is an atypically large loss in an interval (*out*) then the population level at the end (P_1) will be smaller than normal; in this case E3 be strongly effected as the numerator increases while the denominator decreases. Likewise for unusually small outflow. This makes E3 strongly (anti) correlated with P_1. Note that both E1 and E2 do not suffer from such a strong effect, and in particular we expect less correlation for E1 and moderate correlation for E3. We will make use of these observations below.

To assess the performance of the estimators we will consider the residual error between the average over simulations (expectation) and the set α. We derive an analytic expression for the residual, making use of covariance between estimators and population level, where we use $\widehat{\alpha}_*$ and P_* as placeholders

$$COV[\widehat{\alpha}_*, P_*] \equiv \mathbb{E}[\widehat{\alpha}_* P_*] - \mathbb{E}[\widehat{\alpha}_*]\mathbb{E}[P_*] \tag{11}$$

where $\mathbb{E}$ is the expectation. We expand the definition of $\widehat{\alpha}_*$ and use the steady state result $\mathbb{E}[P_*] = in/\alpha$ (see E0)

$$COV[\widehat{\alpha}_*, P_*] = \mathbb{E}[\frac{out}{P_*} P_*] - \mathbb{E}[\widehat{\alpha}_*]\frac{in}{\alpha} \tag{12}$$

in steady state $\mathbb{E}[out] = in$ and so

$$COV[\widehat{\alpha}_*, P_*] = in - \mathbb{E}[\widehat{\alpha}_*]\frac{in}{\alpha} \tag{13}$$

multiplying both sides by $-\alpha/in$ we obtain the following for the residual

$$\mathbb{E}[\widehat{\alpha}_*] - \alpha = -\frac{\alpha}{in} COV[\widehat{\alpha}_*, P_*]. \tag{R1–R3}$$

Note that the observations above suggests that the residual (error) of E3 will be the largest, E1 the smallest, and E2 intermediate. As we expect *anti* correlation the form of the residual indicates a positive bias.

It is worth noting that α can be considered as a time dilation factor (see Equation 2) and so as α approaches zero we expect 'time to slow' relative to larger α. The result is P_* will be quasi-static in a time region and *out* (and hence $\widehat{\alpha}_*$) will be ≈ 0, and so the covariance in Equation R1–R3 will go to zero likewise. Applying the time dilation argument in the other direction, we expect covariance to grow with α.

3.2.2 Residual Error for Discrete Estimators

To find the residual of the discrete estimators we consider steady state arguments, where we use E0 to assert that $\mathbb{E}[P_*] = in/\alpha$ and the fact that $\mathbb{E}[out] = in$ to eliminate all occurrences of variables, other than α, in the estimator equations (where we take the expectation to allow elimination). Applying algebraic simplification leads to the following:

- Half intake estimator residual

$$\mathbb{E}[\widehat{\alpha}_4] - \alpha = -\ln\left(\frac{1-\frac{\alpha}{2}}{1+\frac{\alpha}{2}}\right) - \alpha. \tag{R4}$$

- Markov estimator residual

$$\mathbb{E}[\hat{\alpha}_5] - \alpha = -\ln(1-\alpha) - \alpha. \quad \text{(R5)}$$

- General Formula (GF) estimator residual

$$\mathbb{E}[\hat{\alpha}_6] - \alpha = \ln(1+\alpha) - \alpha. \quad \text{(R6)}$$

By inspection, we see that the Markov and half-intake estimator residuals will have a positive bias (i.e., the estimates are expected to be too large), which disappears as α goes to zero. In contrast the GF estimator residual is negative in expectation, and again goes to zero as α does.

We note two simplifying aspects about our derivation of discrete residual error here. First, we do not consider correlation effects as we expect (and will see in Section 4) these will be small relative to the misspecification error inherent in the discrete estimators. Second, we take $\mathbb{E}[1/X] \approx 1/\mathbb{E}[X]$ which is strictly untrue (Jensen's inequality) as we, again, expect the introduced error to be relatively small.

4 NUMERICAL EXPERIMENTS

To test the performance of the estimators we will measure them on a workforce population for which we control the parameters. We chose a single component (i.e., homogenous) population described by the ODE (1) for simplicity allowing us to investigate bias due to estimator choice without complicating factors.

As previewed in Section 2.4, we use stochastic simulations to create realizations of our population which is defined by a characteristic survival distribution and grown from zero to steady state under a constant intake rate. After considering this steady state regime we introduce an intake rate step growth or decline to examine the dynamic regime. The estimators from Section 3.1 were measured over each simulation run (1000 replications) and on an annual interval (chosen as it is a common frequency used in institutional planning and reporting). Estimator bias was then determined by examining the residual error between the final mean estimator, $\hat{\alpha}$, and the set attrition rate parameter, α.

We start with the case of memoryless (exponential) attrition, where the attrition rate parameter is varied over a wide range in order to observe both typical and more extreme, but plausible, behaviour ($\alpha \in [0.01, 2.0]$). We then extend the experiment with the general stochastic process case by using historical personnel records of several CAF sub-groups.

The CAF considers various sub-populations of interest for strategic planning and here we include the two largest groups: Regular Force (Reg F) and Primary Reserves (P Res) ($\sim 120,000$ and $\sim 80,000$ observations, respectively). Also included is the Reg F sub-population of Naval Warfare Officer (NWO)– a single occupation of particular interest for study to the Royal Canadian Navy (RCN) during ongoing fleet recapitalization ($\sim$ 3600 observations). Figure 1 shows the probability distribution functions (PDFs) constructed from the histograms of empirical service lifetime data (binned annually). We determine the average lifetime $\mathbb{E}[t] = t_{hist}$ from the empirical histogram and use this to determine the effective attrition rate parameter $\alpha_{hist} = 1/t_{hist}$ ($\alpha_{RegF} \approx 0.063$, $\alpha_{NWO} \approx 0.099$, $\alpha_{PRes} \approx 0.164$) (see (Bryce and Henderson, 2020) for further details).

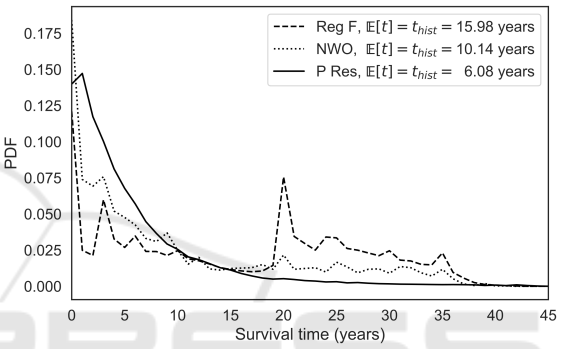

Figure 1: Empirical survival time PDFs for three CAF subpopulations: Reg F, P Res, and NWO [Reg F].

4.1 Steady State Regime

Each population is grown from zero to a steady state population of $P_{ss} = 1000$. The intake rate is constant year over year and set to the steady state value given by $in_{ss} = \alpha P_{ss}$ (E0). Intake times are uniformly distributed over the yearly interval with associated attrition times determined by sampling lifetimes from the population's survival distribution. A total simulation time of 1300 years was chosen in order to give an abundance of time for each population to reach steady state. Recall that the e-folding time for a memoryless population is given by the characteristic mean lifetime according to $e^{-\alpha t} = e^{-t/\mathbb{E}[t]}$ (1 e-folding time ranges from 100 to 0.5 years for $\alpha \in [0.01, 2.0]$). Refer to (Bryce and Henderson, 2020) for a discussion on transient time to steady state for the populations considered here.

Estimator bias was measured over each replication and for every interval during the steady state region (600 simulation years). Relative residual error ($[\hat{\alpha} - \alpha]/\alpha$) for all simulated populations is presented in Figure 2 (solid curves) with precision shown using the standard error of the mean (SEM) (shaded re-

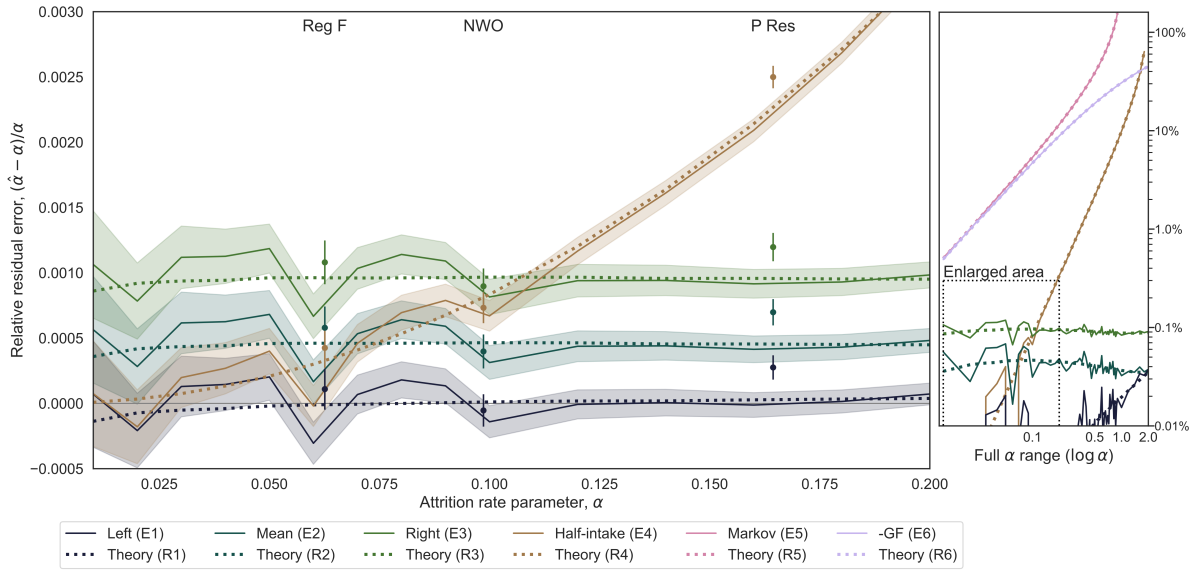

Figure 2: Attrition rate parameter estimator performance in the steady state regime. Mean relative residual error results (solid curves) with corresponding theoretical expectation (dotted curves) are plotted against set α. The main plot (left-hand side) focuses on the region of $\alpha \leq 0.2$. Estimator results from populations simulated using empirical survival lifetimes are shown for comparison (points). Precision on mean results is presented as a shaded region (memoryless populations) or error bars (empirical populations) representing $\pm 1SEM$. Results for the full parameter range ($\alpha \in [0.01, 2.0]$) are shown on the right-hand side plot (log-log scale). Note that the negative of the GF estimator residual error (E6) is plotted here to facilitate comparison.

gion $\pm 1SEM$). The main plot (left-hand side) focuses on $\alpha \leq 0.2$ (i.e., often observed for real workforces) while the right-hand side plot shows the full experimental results over a larger range of relative residual error (up to $\sim 100\%$). To aid in the visualization, at an attrition rate set to $\alpha = 0.1$, the measured mean estimators results in the following order from smallest (lowest curve) to largest (highest curve) residual error: Left (E1: gray), Mean (E2: blue), Right (E3: green), Half-intake (E4: brown), General Formula (E6: violet, note that the negative residual error is plotted here for ease of comparison, see Section 3.2.2), and Markov discrete (E5: pink).

Performance results for the empirical populations (points with error bars $\pm 1SEM$) were determined by comparing the measured mean estimator against α_{hist} (recall Section 4 and Figure 1).

4.2 Dynamic Regime

The population set that should capture most workforce situations, $\alpha \in [0.1, 0.5]$, was chosen as a representative sample to examine estimator performance during population growth and reduction scenarios. In this experiment, populations are grown with a constant intake rate of $in = 100$/year from zero to a steady state population of $P_{ss} = in/\alpha$. A growth multiplier, $\delta \in [0, 2]$, is then applied to the intake, causing a step change in the rate. Note that $\delta = 0$ corresponds to a free fall in population (i.e., no intake), $\delta = 1$ corresponds to the steady state, and $\delta = 2$ will lead to a doubling of the intake rate (and eventual doubling of the population). Due to computational time constraints, the total simulation time varies with each population in order to allow enough time to reach steady state.

The first time interval after δ is applied, t_1, corresponds to the greatest population change that occurs during the dynamical regime (see Equation 2). For this reason it is in the first interval that we compare attrition estimator performance. As in the steady state experiment, the residual error between the mean estimator $\hat{\alpha}$ and the set α is used as an indicator of performance. The relative error ($[\hat{\alpha} - \alpha]/\alpha$) for each estimator at t_1 are shown against δ as a heat map in Figure 3 using a restricted scale from -0.2 (dark blue) to 0.2 (dark red). We cut off at a relative error of 0.2 (i.e., 20%), as the Markov (E5) and General Formula (E6) range is large enough ($-0.21, 0.86$) that we lose visual distinction for the other estimators otherwise. The growth multiplier $\delta = 1$ case corresponds to the steady state case (dotted boxes) and results match those in Section 4.1 and Figure 2.

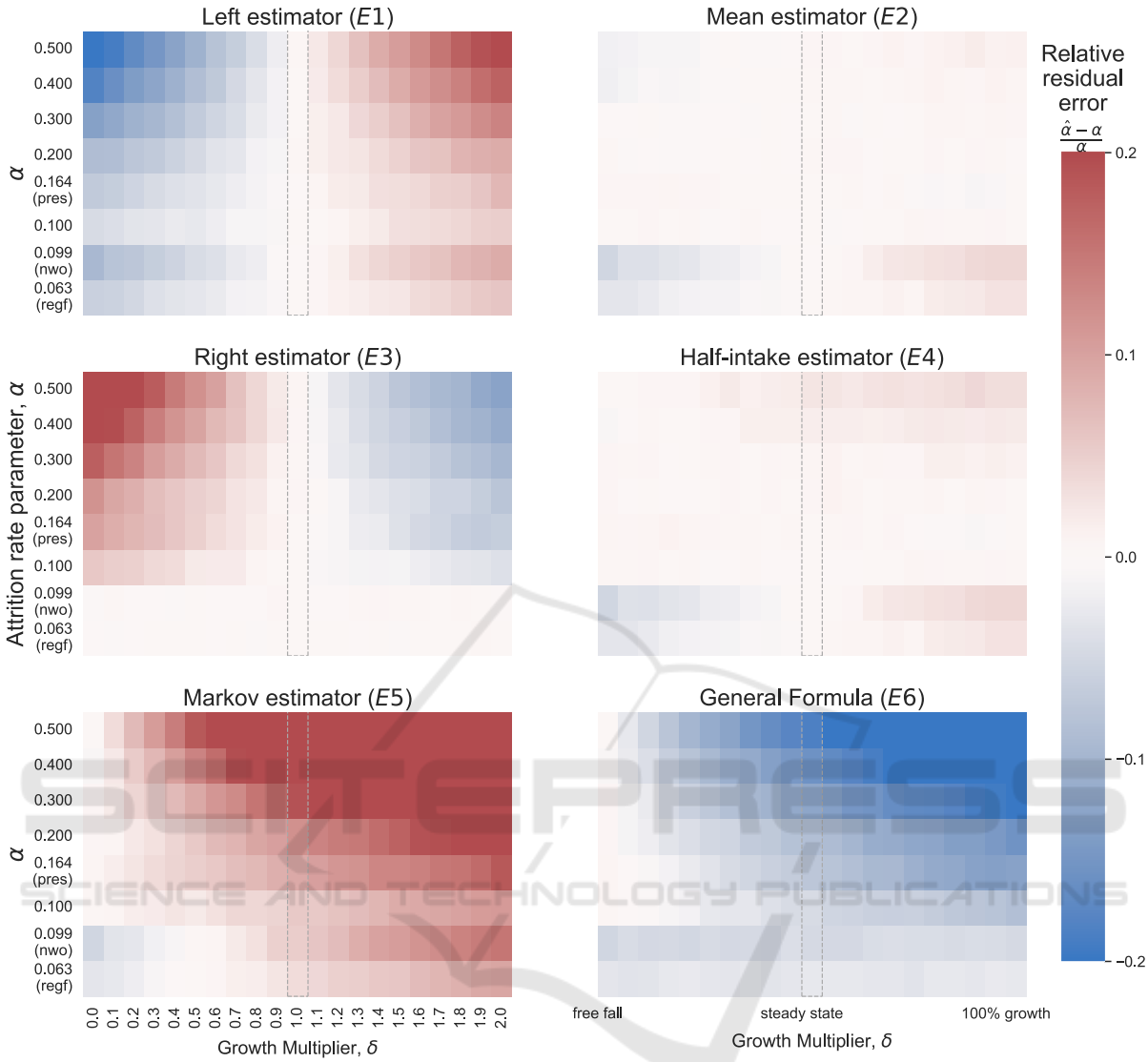

Figure 3: Attrition rate parameter estimator performance during dynamic population change. Relative mean residual error results are shown as heat maps for each estimator using a scale of -0.2 (dark blue) to 0.2 (dark red); we cut off at a relative error of 20% , otherwise we lose visual distinction since the E5 and E6 range is very large. Estimators are measured over the interval immediately after growth ($\delta > 1$) or decay ($\delta < 1$) is initiated in order to examine the maximum effect. The $\delta = 1$ corresponds to the steady state (dashed box) and results here follow those in Figure 2.

5 DISCUSSION

5.1 Estimator Accuracy: Steady State

As expected for memoryless populations in steady state (Figure 2), for small α, we see all estimators converging (mean relative residual error going to zero). The log scale on the right-hand side of Figure 2 shows relative error in % α, a gauge of the error an analyst might expect. For common values of α (≤ 0.2), the magnitude of the bias is very small for most estimators ($< 0.4\%$ for E1 - E4) and becoming uncomfortably large yet likely remaining within tolerance levels for the remaining ($\sim 12\%$ for E5 and $\sim 9\%$ for E6).

In all cases, results follow analytical expressions for residual error of the estimators (Equations R1–R3, R4–R6, dotted curves in Figure 2), with a few interesting features. At lower α (< 0.15) there is an apparent cyclic pattern that diverges from the analytically derived residual, with a magnitude of order 0.0001. We attribute this to finite sample error and happenstance (spurious pattern); which we con-

firmed by running additional trials and finding larger error than implied by the SEM. This can be understood as, for correlated data, the effective sample size is smaller than the sample size of 1000 replications x 600 years for our experiments. As the correlation length is related to the e-folding time we have an effective $SEM_{eff} \sim SEM/\sqrt{\alpha}$ and we see a rapid increase in variation as α decreases.

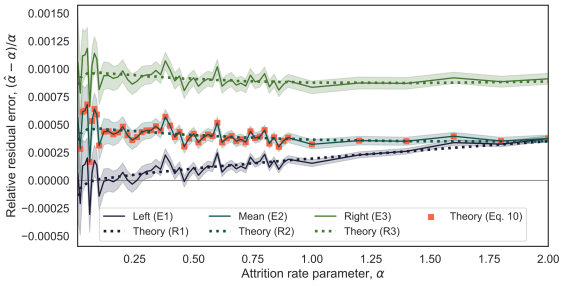

Figure 4: Mean relative residual error (E1–E3, steady state).

Additionally, as α gets larger the error on E1 increases and approaches E2. This is highlighted in Figure 4 where we confirm that the relationship between the three continuous estimators defined by Equation 10 holds in expectation for all α considered (i.e., rearrangement of (10) gives $\mathbb{E}[\alpha_2] = 2\mathbb{E}[(\alpha_1\alpha_3)/(\alpha_1 + \alpha_3)]$, red dotted curve).

As for the populations with empirical distributions (Figure 1) the mean relative residual error shows a small positive bias with respect to the memoryless model (see points with error bars in Figure 2, which lie above the theoretical relative residual). Part of this difference can be explained by considering that, unlike the memoryless populations, the empirical survival distributions do not have a tail beyond a defined cut-toff lifetime, T. Take, for example, the NWO survival PDF reproduced in Figure 5.

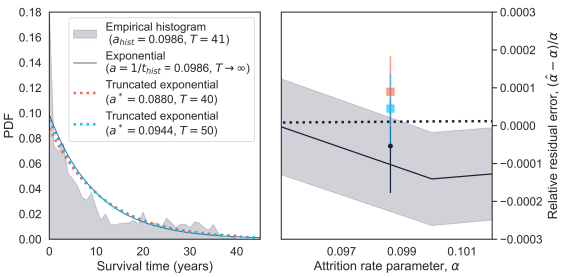

Figure 5: Empirical survival PDF and mean relative residual error of E1 for NWO in steady state.

Visually, the survival time PDF of NWO's shape lies between the other two sub-populations, with an initial spike at small lifetimes and the tail cut off at $T = 41$ (Figure 1). The left-hand side of Figure 5 shows the survival PDF (filled area) along with several exponential fits: an exponential with $\alpha = 1/t_{hist}$ (black curve), and two truncated exponentials cut off at $T = 40$ (red dotted curve) and $T = 50$ (blue dotted curve) with attrition rate parameter α^* such that $\mathbb{E}[t] = t_{hist}$. The right-hand side of Figure 5 shows the corresponding steady state results for the empirical PDF (black point), and the $T = 40$ and $T = 50$ truncated exponentials (red and blue points, respectively). These results illustrate the sensitivity to the tail of the distribution. By implication there is sensitivity to the specifics of the distributions shape, as might be expected. However, we see the absolute error is small and would be well below stochastic measurement error (e.g., in a population of 10,000 or less a difference of a few individuals attriting introduces as large or larger error to the measured α) indicating that, at least near steady state, the memoryless estimators and model can be expected to perform well and the memoryless assumption is well specified. Similar results were found for both Reg F and P Res.

5.2 Estimator Accuracy: Dynamics

Often for institutional planning, a population in steady state is an aspiration rather than a reality. A good method for measuring the attrition rate parameter will likely differ when said population is in equilibrium (Figure 2) and when undergoing changes in intake (Figure 3).

In periods of growth (right-hand side of Figure 3 panels), P_0 is less than P_1 with the difference between them increasing with inflow (δ) and decreasing with outflow (α). Conversely, in periods of decline (left-hand side of Figure 3 panels) P_0 is greater than P_1 with the difference between them increasing with decreasing α and δ.

While E1 performed the best in steady state (Figure 2, dashed box at $\delta = 1$ in Figure 3), it's dependence on P_0 in the denominator guarantees positive bias during growth (red shade) and negative bias during decline (blue shade). The behaviour of E3 mirrors that of E1, though with opposite bias considering it's denominator is P_1. Conversely, both E2 and E4 perform very well during periods of flux since their formulations consider changes in P (mean P and $P_0 + in/2$, respectively). The discrete estimators (E4-E6) all perform well in free fall ($\delta = 0$), as expected as they are correctly specified in this case, but error accrues as intake and outflow increase. E6 retains it's negative bias.

As with the steady state results, the empirical populations in growth and decline vary from the behaviour set by the memoryless populations, with NWO and Reg F standing out in particular in Figure

3. This illustrates the sensitivity of the estimators to the shape of the survival distributions (Figure 1). Similarly to the steady state, the truncated exponential fits to the empirical histograms more closely follow the memoryless results (see Figure 5 for the NWO case).

5.3 General Formula

We have empirically measured the single interval GF here, but note that the multi-interval GF (Equation GF) can be derived from the ODE describing a memoryless population. To do so we consider regions defined by intake events, where E6 gives the estimate on the *i-th* interval

$$\gamma_i = 1 - \frac{P_i}{P_i + out_i}, \tag{14}$$

where we drop the hat for convenience and use i as the index. Using $\gamma_i = 1 - e^{-\alpha_i t_i}$, with t_i being the length of the *i-th* interval we have

$$e^{-\alpha_i t_i} = \frac{P_i}{P_i + out_i}, \tag{15}$$

and we take the natural logarithm of both sides

$$-\alpha_i = \frac{1}{t_i} \ln\left(\frac{P_i}{P_i + out_i}\right). \tag{16}$$

We then find the average α over the N sub-intervals

$$-\overline{\alpha} = \frac{1}{N}\sum_{i=1}^{N} -\alpha_i = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{t_i} \ln\left(\frac{P_i}{P_i + out_i}\right). \tag{17}$$

We will be generalizing the GF by keeping the t_i's in, but note for equal (constant) t_i and when measuring on a unit interval we can pull t_i out of the sum and $Nt_i = 1$, making the time units implicit. To facilitate generalizing the GF we recast the equation as

$$-\overline{\alpha} = \sum_{i=1}^{N} \ln\left(\frac{P_i}{P_i + out_i}\right)^{\frac{1}{Nt_i}}, \tag{18}$$

and we apply the exponential function to both sides

$$e^{-\overline{\alpha}} = e^{\sum_{i=1}^{N} \ln\left(\frac{P_i}{P_i+out_i}\right)^{\frac{1}{Nt_i}}} = \prod_{i=1}^{N} e^{\ln\left(\frac{P_i}{P_i+out_i}\right)^{\frac{1}{Nt_i}}}. \tag{19}$$

The exponential and natural logarithm cancel each other out, and if we measure over the unit interval we will have $\overline{\gamma} = 1 - e^{-\overline{\alpha}}$ leading to

$$1 - \overline{\gamma} = \prod_{i=1}^{N} \left(\frac{P_i}{P_i + out_i}\right)^{\frac{1}{Nt_i}} \tag{20}$$

which we call the "Generalized General Formula", as it explicitly takes into account differing sub-interval sizes. Note that for equal t_i on the unit interval $Nt_i = 1$ and we recover the GF.

5.4 Estimator Choice

When confronted with annual data we have to consider both the accuracy (bias) and precision (variation) of the estimators. As a simple performance measure we consider the error envelope of the bias plus standard deviation, as an estimate of measurement error. In Figure 6 we see that the continuous estimators, and the half-intake estimator, perform best with about 10% expected measurement error (dotted line). We attribute the half-intake's good performance to the stabilizing effect of including intake (fixed value). Notably, we can find an equivalent half-intake formula from Equation 6 using Taylor expansions, with the identical form but which directly estimates α. The performance of this direct half-intake estimator is *worse* than the classic half-intake. This shows sensitivity to the specific uncontrolled approximations made, and difficulty in reasoning about performance.

Overall, the Mean estimator (E2) is the best choice for attrition rate parameter considering both it's performance (closely following the Left estimator [E1] in steady state [Figure 2] and having the best performance [i.e., least bias] during dynamic population change [Figure 3]) and it's simplicity. The half-intake estimator, which bridges the continuous-discrete divide as it is a partial correction, is the best performing estimator at low to moderate attrition.

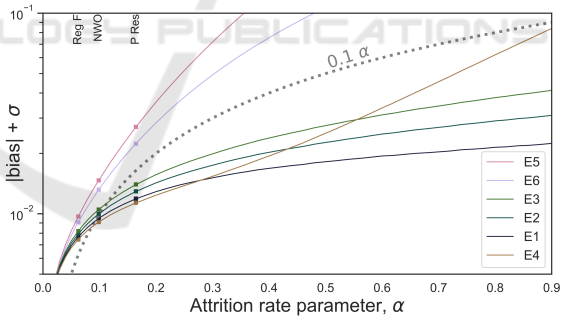

Figure 6: Performance measure of estimator bias + σ in steady state.

It should be noted that use of the mean population estimator is standard practice for measuring employee turnover (i.e., attrition). For example, the U.S. Department of Labor (U.S. Bureau of Labor Statistics, 2022) defines employee turnover as the number of employees who leave divided by the average number of employees. Additionally note that, constrained to measuring at a constant frequency over a unit time interval such that $dt = (1/N)$, that $\int_t^{t+1} P dt = dt \sum P = (1/N)\sum P$ while $\bar{P} = (1/N)\sum P$ and taking the average $\bar{P}$ is identical to integrating the population over

the interval and so the mean average estimator (E2) recovers the exact estimator (Equation 8). While all estimators converge in the limit of diminishing measurement intervals the fact that the average is robust on large (e.g., annual) intervals, has limited bias, matches common labour practice, is used by multiple militaries, and has a particularly simple connection to the exact estimator recommends its use.

6 CONCLUSION

We show that the continuous model of attrition leads to marginal, but real, improvements over the discrete Markov chain view that has dominated for over three quarters of a century (Seal, 1945). We provide insight into three important aspects: 1) we conclusively demonstrate which common estimators work best in practical situations (the two-point average population and half-intake estimators), 2) we find the theoretically exact estimator and demonstrate all common estimators are approximate solutions to this exact estimator, and 3) we derive the analytical residuals for common estimators which provides an understanding of their bias properties.

REFERENCES

Anderson, T. W. and Goodman, L. A. (1957). Statistical inference about Markov chains. *The annals of mathematical statistics*, pages 89–110.

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

Bryce, R. M. and Henderson, J. A. (2020). Workforce populations: Empirical versus Markovian dynamics. In *2020 Winter Simulation Conference (WSC)*, pages 1983–1993. IEEE.

Devroye, L. (1998). *Non-Uniform Random Variate Generation*. Springer, London, 2nd edition.

Feller, W. (1957). *An Introduction to Probability Theory and Its Applications*, volume 1 & 2. John Wiley & Sons, Inc.

Henderson, J. A. and Bryce, R. M. (2019). Verification methodology for Discrete Event Simulation models of personnel in the Canadian Armed Forces. In *2019 Winter Simulation Conference (WSC)*, pages 2479–2490. IEEE.

Higham, N. J. (2008). *Functions of Matrices: Theory and Computation*. SIAM.

Merck, J. W. and Hall, K. (1971). A Markovian flow model: The analysis of movement in large-scale (military) personnel systems. Technical report, RAND CORP SANTA MONICA CALIF.

Okazawa, S. (2007). Measuring attrition rates and forecasting attrition volume. Technical Report DRDC CORA TM 2007-02, Defence Research and Development Canada.

Seal, H. (1945). The mathematics of a population composed of k strata each recruited from the stratum below and supported at the lowest level by a uniform annual number of entrants. *Biometrika*, 33(3):226–230.

U.S. Bureau of Labor Statistics (2022). Job openings and labor turnover - September 2022. https://www.bls.gov/news.release/pdf/jolts.pdf. Accessed: 2022-11-14.

Vajda, S. (1978). *Mathematics of Manpower Planning*. Wiley.

Vincent, E. (2019). The mathematical formalism of growth and discount rate reporting. http://www.corsottawa.org/CORS2019.pdf. Accessed: 2022-11-14.

Vincent, E., Calitoiu, D., and Ueno, R. (2018). Personnel attrition rate reporting. Technical Report DRDC-RDDC-2018-R238, Defence Research and Development Canada.

Vincent, E., Okazawa, S., and Calitoiu, D. (2021). Attrition, promotion, transfer: Reporting rates in personnel operations research. In *ICORES*, pages 115–122.

Young, A. and Almond, G. (1961). Predicting distributions of staff. *The Computer Journal*, 3(4):246–250.