Stockholms universitet

Recovering Latent Age Distributions from Misclassified Observations: A Simulation Study Using EM and Method of Moments

Emilio Otero Johansson

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2026:7
Matematisk statistik
Maj 2026

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Matematiska institutionen

Stockholm University

Mathematical Statistics
Stockholm University
Bachelor Thesis **2026:7**
http://www.math.su.se

Recovering Latent Age Distributions from Misclassified Observations: A Simulation Study Using EM and Method of Moments

Emilio Otero Johansson*

May 2026

Abstract

Accurate estimation of population age structure is important in wildlife ecology, but age is often measured with error. This thesis studies how to recover a latent age distribution from misclassified observations using statistical deconvolution in a simulation setting. The true population structure is generated from a Leslie matrix model, while measurement error is introduced through a misclassification mechanism.

Two estimators are evaluated: a method-of-moments (MoM) estimator and a maximum likelihood estimator implemented via the Expectation–Maximization (EM) algorithm. Their performance is assessed through Monte Carlo simulations under different levels of measurement error and sample size, using total variation distance and population forecasts. Results are compared to an oracle baseline and to the naive estimator based on misclassified observations.

The results show that under ill-posed conditions the EM estimator reduces bias but has higher variability, while the naive estimator and MoM are more stable but systematically biased. The study shows that even small bias in age structure can lead to large errors in population forecasts, highlighting that improvements in statistical performance do not necessarily translate into improved ecological performance.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: emiliooterojohansson@gmail.com. Supervisor: Martin Sköld.

Acknowledgements

I would like to thank my supervisor, Martin Sköld, for proposing an interesting thesis topic and for his availability, guidance, and valuable discussions throughout the project. I would also like to thank my parents for their unwavering support during my studies. Finally, I am grateful to my friends, Nils and Fredrika, for their constant encouragement.

AI usage ChatGPT has been used as a tool for language refinement, structuring suggestions, and clarification of statistical concepts. All scientific content, modeling choices are my own.

Contents

1 Introduction **4**

2 Methodology **5**
- 2.1 Deconvolution . . . 5
  - 2.1.1 Latent distribution . . . 6
  - 2.1.2 Measurement error . . . 6
  - 2.1.3 Observed data . . . 7
- 2.2 Estimation methods . . . 7
  - 2.2.1 Oracle and naive estimator . . . 7
  - 2.2.2 Method of Moments estimator . . . 8
  - 2.2.3 Maximum likelihood estimation via the EM algorithm 9
- 2.3 Matrix population models . . . 11
  - 2.3.1 Leslie Matrix . . . 11
  - 2.3.2 Computing age distribution from a Leslie matrix . . . 12
  - 2.3.3 Population forecast . . . 12
- 2.4 Evaluation metric: total variation distance (TV) . . . 13

3 Simulation Study **13**
- 3.1 Simulated Age Distributions . . . 13
- 3.2 Error term . . . 15
- 3.3 Simulation design . . . 16
- 3.4 Simulation procedure . . . 16

4 Results **17**
- 4.1 Recovery of latent age distribution . . . 17
- 4.2 Sample size effect on TV . . . 19
- 4.3 Population forecast . . . 21

5 Discussion **22**
- 5.1 Properties of estimators . . . 23
- 5.2 Practical Implications . . . 24
- 5.3 Viability of assumptions & improvements . . . 24

6 Appendix **26**

1 Introduction

Accurate estimates of population size are of central importance in ecology and wildlife management, as they provide the basis for assessing population resilience, conservation status, and long-term viability. For large carnivores in particular, reliable population estimates are essential for informing management decisions, including conservation planning and hunting regulation (Kindberg et al. 2011).

Although current monitoring programs provide estimates of population size and sex distribution, reliable estimation of age structure remains more challenging. In Sweden, the brown bear (Ursus Arctos) population has been continuously monitored by the Swedish Museum of Natural History, which has served as the national coordinator for bear inventories on behalf of the Swedish Environmental Protection Agency since 2018. Within this framework, population size and sex distributions are estimated using DNA sampling from faeces combined with capture–recapture methods (Åsbrink, Sköld, and Gyllenstrand 2025; Åsbrink, Sköld, Källman, et al. 2023).

While total population size is fundamental, understanding the age composition of a population is also essential because demographic processes such as survival and reproduction vary strongly with age (Morris 2008). The importance of age-structured population data has recently been emphasized in a synthesis report summarizing 40 years of Scandinavian brown bear research (*Björnprojektet – 40 års forskning* 2025), which identified the need for further research on age structure in the Swedish brown bear population.

In bear populations, age typically cannot be measured directly. A common approach involves analyzing tooth cementum layers, corresponding to annual growth rings in teeth. However, this method is invasive and requires the removal of a molar. As a recent development, DNA-based age estimation methods have been proposed (Nakamura et al. 2023). These indirect measurements are typically associated with measurement error. According to Conn and Diefenbach 2007, researchers often respond to substantial measurement error either by excluding noisy data or by ignoring the measurement error entirely.

This motivates statistical approaches that explicitly model measurement error and its relationship to the true age distribution. Such approaches can be formulated as a deconvolution problem. In this setting, the observed age distribution (with measurement error) can be viewed as a blurred version of the true underlying distribution, where the blurring is caused by uncertainty in the age determination method. Deconvolution consists of recovering the true underlying distribution from such blurred data. The effectiveness of these methods depends both on the severity of the measurement error and on the amount of available data.

The present thesis investigates this deconvolution problem in a controlled simulation setting. The true population age structure is generated using a

Leslie matrix model parameterized with data from Norwegian Institute for Nature Research (n.d.). Measurement error is then introduced through a probabilistic misclassification mechanism, producing an observed age distribution that deviates from the latent structure.

The primary objective of this study is to evaluate how accurately the latent age distribution can be recovered using deconvolution approaches based on the Expectation–Maximization (EM) algorithm and the method-of-moments (MoM) estimator. The misclassification mechanism was parameterized to resemble the error structure reported in (Nakamura et al. 2023), while an additional high-measurement-error regime was included to evaluate performance under more severely ill-posed conditions. Using Monte Carlo simulations, the estimators were evaluated across five different sample sizes. Recovery performance was assessed using total variation distance, while the ecological implications of estimation error were examined through population forecasting based on the Leslie matrix model.

2 Methodology

This section focuses on the general theoretical framework and its adaptation to the present setting, while detailed parameter choices and simulation specifications are deferred to Section 3 unless needed to motivate specific modeling choices. We begin by presenting the general deconvolution framework and then specialize it to the present setting of age misclassification in discrete populations. In the subsequent section on estimation methods, we present the two estimation approaches considered in this thesis: a method-of-moments estimator and a likelihood-based estimator based on the Expectation–Maximization algorithm, including a full derivation of the EM procedure. We then introduce matrix population models, which are used both to construct a biologically grounded population age distribution and to generate population forecast error under estimated and true age structures. Finally, we define the evaluation metric used throughout the study, the total variation distance, which quantifies discrepancies between true and estimated age distributions.

2.1 Deconvolution

This study considers a deconvolution problem in which the objective is to recover the distribution of a latent variable from observations of an error-contaminated variable. This can be formulated through a convolution structure in which the observed variable is a noisy transformation of the true signal $Y = Z + \epsilon$ (Carroll et al. 2006, Chapter 12). Here, Z denotes the latent random variable, ϵ denotes measurement error, and Y denotes the observed random variable. The distribution of Y is then the convolution of the distributions of Z and ϵ. This formulation extends naturally to multivariate

settings, where both the latent and observed quantities are represented as random vectors, written as $\mathbf{Y} = \mathbf{Z} + \boldsymbol{\epsilon}$.

In many applications, including the present one, the variable of interest is discrete. In such cases, the additive error formulation is replaced by a discrete measurement error mechanism, which will be specified in the following sections.

2.1.1 Latent distribution

The true age structure of the bear population is represented by the parameter vector

$$\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_k), \tag{1}$$

where π_j denotes the probability that an individual belongs to age class j. In particular, π_1 corresponds to the age group 0–1 years, π_2 to 1–2 years, and so on. The discretization of age into fixed classes is consistent with the NINA Bear Model (Norwegian Institute for Nature Research n.d.).

Let n denote the number of independent individuals in the sample. For each individual $r = 1, \ldots, n$, let $T_r \in \{1, \ldots, k\}$ denote the true age class. We assume that the latent variables $T_1, \ldots, T_n$ are independent and identically distributed with

$$P(T_r = j) = \pi_j.$$

The vector of observed class counts is then defined as

$$\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_k)\,,$$

where

$$Z_j = \sum_{r=1}^{n} \mathbf{1}(T_r = j)$$

denotes the number of individuals in age class j. It follows that

$$\boldsymbol{Z} \sim \text{Multinomial}(n, \boldsymbol{\pi}).$$

2.1.2 Measurement error

In this study the measurement error is assumed to be fully known. Moreover, we assume a normal distribution for the error term as a simple and tractable way to control the variability in the observations and to introduce systematic measurement error. While this assumption may not fully reflect the error structure described in Nakamura et al. (2023), the aim of this thesis is to study the recovery of the latent distribution under a simplified but comparable error setting.

The measurement error is assumed to be centred at zero conditional on the true age class. Specifically, we model

$$\epsilon_r \mid T_r = j \sim \mathcal{N}(0, \sigma^2).$$

Applying the measurement error directly at the individual level would yield a continuous latent measurement $T_r + \epsilon_r$, which is incompatible with the discrete age-class structure of the model. We therefore define a misclassification induced by this continuous error model by discretizing the conditional distribution of $\epsilon_r \mid T_r = j$ over the k predefined age classes. Let M_{ji} denote the probability of misclassifying a bear of true age j with age i:

$$M_{ji} = P(O_r = i \mid T_r = j),$$

where $O_r \in \{1, \ldots, k\}$ denotes the observed age class corresponding to individual r such that:

$$\sum_{i=1}^{k} M_{ji} = 1.$$

This leads to the construction of a square misclassification matrix $\boldsymbol{M}$ with k rows and columns, where row j corresponds to the true age class and column i to the observed age class.

2.1.3 Observed data

Given the misclassification matrix $\boldsymbol{M}$ described above. The count vector $\mathbf{Y} = (Y_1, \ldots, Y_k)$ of misclassified observations therefore follow a multinomial distribution with probability vector $\boldsymbol{\pi M}$:

$$\mathbf{Y} \sim \text{Multinomial}(n, \boldsymbol{\pi M}), \tag{2}$$

where

$$Y_i = \sum_{r=1}^{n} \mathbf{1}(O_r = i).$$

2.2 Estimation methods

In this thesis, we consider two methods to recover the latent age structure $\boldsymbol{\pi}$: a method-of-moments (MoM) estimator and a maximum likelihood estimator based on the EM algorithm. We define an oracle baseline based on $\boldsymbol{Z}$ and a naive estimator based on $\boldsymbol{Y}$ as reference points.

2.2.1 Oracle and naive estimator

In the absence of measurement error and with access to the true age counts, a natural benchmark estimator is given by the empirical proportions of the

true age classes:

$$\hat{\boldsymbol{\pi}}_{\text{oracle}} = \frac{\mathbf{Z}}{n}.$$

We refer to this as the oracle estimator, since it assumes access to the unobserved true counts $\mathbf{Z}$. In contrast, when only misclassified observations are available and measurement error is ignored, a natural baseline estimator is given by the empirical proportions of the observed data:

$$\hat{\boldsymbol{\pi}}_{\text{naive}} = \frac{\mathbf{Y}}{n}. \tag{3}$$

It therefore serves as a baseline for assessing whether the proposed deconvolution methods improve upon the naive estimator in estimating $\boldsymbol{\pi}$.

2.2.2 Method of Moments estimator

The method-of-moments assumes that moments, such as $\mathbb{E}[\mathbf{Z}]$, can be expressed as functions of the unknown parameter $\boldsymbol{\pi}$. Parameter estimates are obtained by equating these theoretical moments with their empirical counterparts, see Alm and Britton 2008, p. 275.

From equation (2), the first moment of the observed data is given by

$$\mathbb{E}[\boldsymbol{Y}] = n\boldsymbol{\pi}\boldsymbol{M}.$$

The corresponding empirical moment is given by the observed proportions $\boldsymbol{Y}/n$ see equation (3). Equating theoretical and empirical moments yields

$$\boldsymbol{\pi}\boldsymbol{M} = \frac{\boldsymbol{Y}}{n}.$$

Solving for $\boldsymbol{\pi}$ gives the MoM estimator

$$\hat{\boldsymbol{\pi}}_{MoM} = \frac{\boldsymbol{Y}}{n}\boldsymbol{M}^{-\mathbf{1}},$$

provided that $\boldsymbol{M}$ is invertible.

Matrix regularization In practice, the misclassification matrix $\boldsymbol{M}$ may be nearly singular, which makes direct inversion numerically unstable. To address this issue, we apply Tikhonov regularization (ridge regularization), which stabilizes the inversion by penalizing large fluctuations in the estimated parameter vector.

The regularized estimator is defined as

$$\hat{\boldsymbol{\pi}} = \left(\boldsymbol{M}^\top \boldsymbol{M} + \lambda \boldsymbol{I}\right)^{-1} \boldsymbol{M}^\top \frac{\mathbf{Y}}{n},$$

where $\lambda > 0$ is a tuning parameter controlling the strength of regularization and I denotes the identity matrix. The resulting estimator can be written as a linear operator applied to the observed data,

$$\hat{\boldsymbol{\pi}} = A_\lambda \frac{\mathbf{Y}}{n},$$

where

$$A_\lambda = \left(M^\top M + \lambda I\right)^{-1} M^\top.$$

2.2.3 Maximum likelihood estimation via the EM algorithm

The EM algorithm is a general-purpose method for computing maximum likelihood estimates in the presence of incomplete or latent data (McLachlan and Krishnan 2008). Let $\boldsymbol{Y}$ denote a discrete random vector corresponding to the observed (incomplete) data, and $\boldsymbol{Z}$ a discrete random vector representing the latent data. Then $\boldsymbol{X} = (\boldsymbol{Y}, \boldsymbol{Z})$ is the random vector of the complete data. If $\mathbf{X}$ were fully observable, we could directly compute the complete-data log-likelihood function $\ell_c(\boldsymbol{\pi})$ based on a realization x and estimate $\boldsymbol{\pi}$ by maximum likelihood.

However, in practice we only observe $y \in \mathcal{Y}$. Thus, the EM algorithm considers expectation with respect to the conditional distribution of $\boldsymbol{Z} \mid \boldsymbol{Y}$ and a current parameter estimate $\boldsymbol{\pi}^{(t)}$ McLachlan and Krishnan 2008:

$$Q(\boldsymbol{\pi} \mid \boldsymbol{\pi}^{(t)}) = \mathbb{E}_{\boldsymbol{\pi}^{(t)}}\left[\ell_c(\boldsymbol{\pi}) \mid \boldsymbol{Y} = \boldsymbol{y}\right], \tag{4}$$

this corresponds to the Expectation step (E-step) of the EM algorithm. In the Maximization step (M-step), the parameter is updated by maximizing this function:

$$\boldsymbol{\pi}^{(t+1)} = \arg\max_{\boldsymbol{\pi}} Q(\boldsymbol{\pi} \mid \boldsymbol{\pi}^{(t)}). \tag{5}$$

The process is then iterated upon until $\ell(\boldsymbol{\pi}^{(t+1)}) - \ell(\boldsymbol{\pi}^{(t)}) \leq \epsilon$ for some arbitrary small ϵ value, which is then interpreted as a local/global maxima.

Monotonicity A key property for the algorithm to work is for the likelihood to be monotonically increasing for each iteration. $L(\boldsymbol{\pi}^{(t+1)}) \geq L(\boldsymbol{\pi}^{(t)}))$ which is discussed in Section 3.2 of McLachlan and Krishnan 2008.

Applying EM-algorithm Let $\boldsymbol{X} = X_1, X_2, \ldots, X_n$ denote the complete data which can be represented by individual-level pairs $X_r = (O_r, T_r)$. Since, in the present multinomial setting, it is convenient to use the sufficient statistic N_{ji}, where N_{ji} denotes the number of individuals with true age class j that are observed in age class i.

The joint distribution for a single individual of true age class T and observed age O classes is:

$$P(T = j, O = i) = (\pi_j M_{ji}),$$

where $\boldsymbol{M}$ is the misclassification matrix defined in Section 2.1.2 and for n independent observations $\boldsymbol{X}$ we obtain:

$$p_X(\boldsymbol{X}; \boldsymbol{\pi}) = \prod_i \prod_j (\pi_j M_{ji})^{n_{ji}}.$$

The complete data log-likelihood is then:

$$\ell_c(\boldsymbol{\pi}) = \sum_j \sum_i n_{ji} \log(\pi_j M_{ji}) = \sum_j \sum_i n_{ji} \log M_{ji} + \sum_j \sum_i n_{ji} \log \pi_j. \quad (6)$$

From equation (4) we take the conditional expectation on n_{ji} with an initial parameter $\boldsymbol{\pi^{(0)}}$ conditioned on $\boldsymbol{Y}$ and obtain:

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\pi}^{(0)}}\left[\ell_c(\boldsymbol{\pi}) \mid \mathbf{Y} = \mathbf{y}\right] &= \mathbf{E}_{\boldsymbol{\pi}^{(0)}}\left[\sum_j \sum_i n_{ji} \log(\pi_j M_{ji}) \,\middle|\, \mathbf{Y} = \mathbf{y}\right] \\ &= \sum_i \sum_j \mathbf{E}_{\boldsymbol{\pi}^{(0)}}\left[n_{ji} \mid \mathbf{Y} = \mathbf{y}\right] \log(\pi_j M_{ji}). \end{aligned}$$

The conditional expectation $E_{\boldsymbol{\pi^{(0)}}}[n_{ji} \mid \mathbf{Y} = \mathbf{y}]$ is the expected number of bears with true age class j that are observed in age class i, given the observed data $\mathbf{Y} = \mathbf{y}$. This depends on the assumed latent age distribution $\boldsymbol{\pi^{(0)}}$.

The probability that an individual belongs to true age class $T = j$ given that it is observed in age class $O = i$, under the current parameter estimate $\boldsymbol{\pi}^{(t)}$, is given by

$$P(T = j \mid O = i, \boldsymbol{\pi}^{(t)}) = \gamma_{ji}^{(t)} = \frac{\pi_j^{(t)} M_{ji}}{\sum_{a=1}^{k} \pi_a^{(t)} M_{ai}}. \quad (7)$$

Thus, the expected number of individuals with true age class j among the y_i individuals observed in age class i is

$$\mathbb{E}[n_{ji} \mid Y_i = y_i] = y_i \gamma_{ji}^{(t)} \quad (8)$$

Maximization-step We now choose $\boldsymbol{\pi}^{(t+1)}$ such that, in accordance with equation 5, the conditional expected log-likelihood is maximized. From

equation 6, we note that only the terms involving $\boldsymbol{\pi}$ depend on the parameter. The problem is reduced to maximizing:

$$\sum_j \mathbb{E}[n_{ji} \mid \mathbf{Y} = \mathbf{y}] \log \pi_j,$$

subject to the constraints

$$\sum_j \pi_j = 1, \quad 0 \leq \pi_j \leq 1.$$

From Theorem 14.5 (p. 235) in Wasserman 2004, the maximum likelihood estimator for a multinomial random vector $\mathbf{Z} \sim \text{Multinomial}(n, \boldsymbol{\pi})$ is obtained by maximizing

$$\ell(\boldsymbol{\pi}) = \sum_{j=1}^{k} Z_j \log \pi_j,$$

under the constraint $\sum_j \pi_j = 1$. Using Lagrange multipliers yields the well-known result

$$\hat{\pi}_j = \frac{Z_j}{n}, \quad j = 1, \ldots, k.$$

We now replace the unobserved counts Z_j with their conditional expectations

$$\mathbb{E}[n_{ji} \mid \mathbf{Y} = \mathbf{y}].$$

Using equation (8). The maximization step is then given by

$$\pi_j^{(t+1)} = \frac{\sum_{i=1}^{k} Y_i \gamma_{ji}^{(t)}}{n}. \tag{9}$$

2.3 Matrix population models

Matrix population models are a standard framework in population ecology (Caswell 2001). In this section, we describe the theoretical framework used to construct a hypothetical true age distribution $\boldsymbol{\pi}$, which is used as the parameter of a multinomial model for simulating population data.

2.3.1 Leslie Matrix

To describe the population age structure, we use a matrix population model. Let $j = 1, 2, \ldots, k$ denote discrete age classes. The population at time t is represented by the vector

$$\mathbf{v}_t = (n_1, n_2, \ldots, n_k),$$

where n_j denotes the number of individuals in age class j.

Let f_j denote the per-capita fertility rate of individuals in age class j, defined as the expected number of offspring produced per time step per individual. The total number of offspring produced by age class j is then $n_j f_j$. Survival is modeled by age-specific survival probabilities s_j, representing the probability of surviving from age class j to $j+1$. The Leslie matrix is then given by

$$\mathbf{L} = \begin{pmatrix} f_1 & f_2 & f_3 & \cdots & f_k \\ s_1 & 0 & 0 & \cdots & 0 \\ 0 & s_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & s_{k-1} & 0 \end{pmatrix}.$$

2.3.2 Computing age distribution from a Leslie matrix

A stable age distribution describes the long-term relative proportions of individuals in each age class of a population. It arises under the assumption of constant demographic parameters, meaning that both the survival probabilities $\mathbf{s}$ and fertility rates $\mathbf{f}$ do not change over time. Under these conditions, the total population size may vary, but the relative proportions between age classes converge to a fixed structure. In this simulation study, the stable age distribution is used as a probability distribution over age classes and is taken to define the parameter vector $\boldsymbol{\pi}$ of the latent distribution introduced in Section 1. This stable structure can be obtained from the Leslie matrix $\mathbf{L}$.

In particular, it is determined by the dominant eigenvalue λ_1 and the corresponding right eigenvector $\mathbf{w}$. The existence of a unique dominant eigenvalue λ_1, which is real, non-negative, and greater in magnitude than all other eigenvalues λ_i, is guaranteed for non-negative primitive matrices by the Perron Frobenius theorem (Caswell 2001). If $\mathbf{L}$ is primitive, then as $t \to \infty$, the population vector $\mathbf{v}_t$ satisfies

$$\frac{\mathbf{v}_t}{\sum_j n_{j,t}} \to \boldsymbol{w},$$

where $\boldsymbol{w}$ is the normalized right eigenvector associated with λ_1, which defines the stable age distribution used in this study, i.e. $\boldsymbol{\pi} = \boldsymbol{w}$.

2.3.3 Population forecast

In this thesis, we investigate how errors in the estimated age structure affect population forecasts as a way to evaluate the performance of the deconvolution method in an applied setting. In particular, we compare the population forecast obtained from the true stable age distribution $\boldsymbol{\pi}$ with the forecast based on an estimated age distribution $\hat{\boldsymbol{\pi}}$.

The population dynamics n time steps into the future can be expressed as

$$\mathbf{v}_{t+n} = \mathbf{L}^n \mathbf{v}_t.$$

The resulting deviation in population vectors can be written as

$$\mathbf{L}^n \hat{\boldsymbol{\pi}} - \mathbf{L}^n \boldsymbol{\pi} = \mathbf{L}^n (\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}).$$

To evaluate the effect on total population size, we define

$$N_{t+n}(\boldsymbol{x}) = \sum_{j=1}^{k} \left(\mathbf{L}^n \boldsymbol{x}\right)_j,$$

which gives the total population size at time $t + n$ given an initial age distribution $\boldsymbol{x}$. The relative deviation in total population size is then defined as

$$\text{RelErr}_{t+n} = \frac{N_{t+n}(\hat{\boldsymbol{\pi}}) - N_{t+n}(\boldsymbol{\pi})}{N_{t+n}(\boldsymbol{\pi})}, \tag{10}$$

which quantifies how errors in the initial age structure estimates propagate through time under stable population dynamics defined by $\mathbf{L}$.

2.4 Evaluation metric: total variation distance (TV)

As an evaluation metric for the deconvolution methods, we use the Total Variation (TV) distance. Intuitively, the TV distance represents the smallest amount of probability mass that must be redistributed to transform one probability distribution into another.

Formally, for two probability distributions P and Q defined on a finite sample space $\mathcal{X}$, the TV distance is given by

$$\text{TV}(P, Q) = \frac{1}{2} \sum_{x \in \mathcal{X}} |P(x) - Q(x)|.$$

The TV distance takes values in $[0, 1]$, where 0 corresponds to identical distributions and 1 no overlap in probability mass.

In this thesis, P and Q correspond to discrete age distributions, in particular the true age distribution $\boldsymbol{\pi}$ and its estimate $\hat{\boldsymbol{\pi}}$. The TV distance therefore provides a natural measure of discrepancy between the estimated and true population structure.

3 Simulation Study

3.1 Simulated Age Distributions

We consider a population structured into $k = 20$ age classes, assuming a maximum age of 20 years due to the negligible frequency of individuals

older than 20. Consistent with the discretization in the NINA Bear Model (Norwegian Institute for Nature Research n.d.). Based on this age structure, we defined a 20×20 Leslie matrix $\mathbf{L}$ as described in Section 2.3. Each age class represents a one-year interval, where the age-class index $j \in \{1, \ldots, 20\}$ corresponds to the interval $[j-1, j)$ for $j = 1, \ldots, 19$, while the final class $j = 20$ represents individuals aged 19 years and older.

We define the fertility rate of individuals in age class j as

$$f_j = s_j \cdot p_{\text{coy},j} \cdot \text{litter}_j,$$

where:

- s_j is the survival probability from age j to $j+1$,
- $p_{\text{coy},j}$ is the probability of conceiving at age j,
- litter_j is the expected litter size at age j.

The survival and reproduction parameters were approximated by visual inspection of published graphs from the NINA Bear Model. The extracted values used in the construction of $\mathbf{L}$ are provided in Appendix 6.

The stable age distribution, shown in Figure 1, is computed as the normalized dominant right eigenvector of $\mathbf{L}$ using the method described in Section 2.3.2. This distribution is denoted by $\boldsymbol{\pi}$ and serves as the latent age distribution used throughout the simulation study.

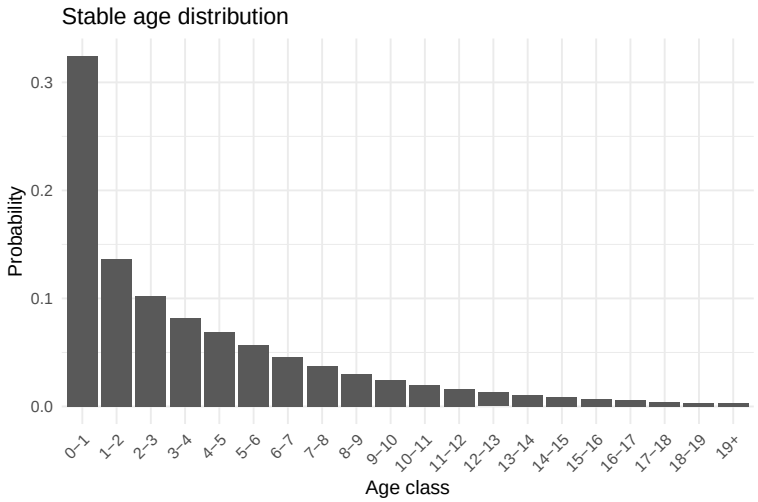

Figure 1: Latent age distribution $\boldsymbol{\pi}$ obtained from the normalized dominant eigenvector of the Leslie matrix $\mathbf{L}$.

3.2 Error term

In this section, we define the discretisation procedure used to construct the misclassification matrix $\mathbf{M}$ described in Section 2.1.2. The chosen error model is designed to approximately reflect the error pattern shown in Figure 5(a) in Nakamura et al. 2023. An illustrative example is provided in the Appendix (Figure 9).

For each individual with true age class $T_r = j$, we assume an additive Gaussian measurement error as specified in 2.1.2 with two distinct values of σ, namely $\sigma \in \{1, 4\}$:

$$\epsilon_r \mid T_r \sim \mathcal{N}(0, \sigma^2).$$

The observed value $j - 0.5 + \epsilon_r$ is discretised by assigning it to the corresponding age interval $[i-1, i)$.

To account for the lower boundary, we condition on the event that the perturbed value lies within the support of the age scale, resulting in a truncated distribution. The probability of observing class $i = 1, \ldots, 19$ is therefore

$$M_{ij} = \frac{\Phi\left(\frac{i-(j-0.5)}{\sigma}\right) - \Phi\left(\frac{i-1-(j-0.5)}{\sigma}\right)}{1 - \Phi\left(\frac{0-(j-0.5)}{\sigma}\right)}.$$

For the upper boundary, the last age class absorbs all remaining probability mass:

$$M_{20,j} = 1 - \sum_{i=1}^{19} M_{ij}.$$

This ensures that $\sum_{i=1}^{20} M_{ij} = 1$.

The resulting distribution corresponding to $\sigma = 1$ and $\sigma = 4$ is shown in Figure 2.

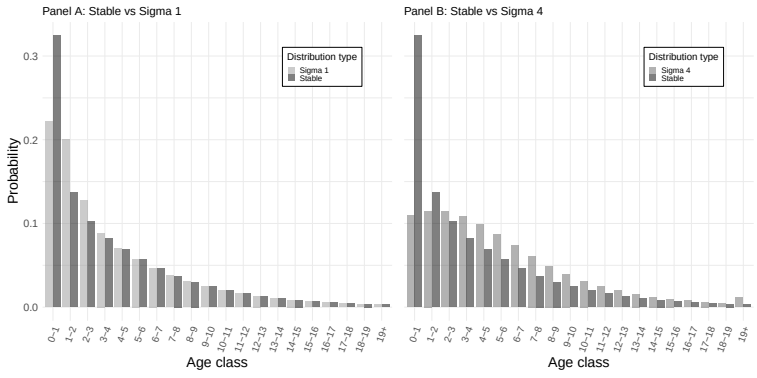

Figure 2: The stable age distribution π and the corresponding observed age distributions $M\pi$ under a measurement error model $\epsilon_r \mid T_r \sim \mathcal{N}(0, \sigma^2)$ with $\sigma \in \{1, 4\}$.

3.3 Simulation design

This section describes the experimental setup. The latent age distribution $\boldsymbol{\pi}$ shown in Figure 1 is fixed across all simulations. Measurement error is introduced through a misclassification matrix $\boldsymbol{M}$, constructed from the Gaussian error model described in Section 3.2 with two choices of σ, $\sigma = 1$ and $\sigma = 4$.

To study the effect of sample size, we consider

$$n \in \{50, 100, 250, 500, 1000\}.$$

We compare four estimators of the latent distribution:

- oracle estimator $\hat{\boldsymbol{\pi}}_{oracle} = \mathbf{Z}/n$
- the naive estimator $\hat{\boldsymbol{\pi}}_{naive} = \mathbf{Y}/n$,
- the method-of-moments estimator $\hat{\boldsymbol{\pi}}_{MoM}$,
- the EM-based maximum likelihood estimator $\hat{\boldsymbol{\pi}}_{ML}$ with naive initialization $\hat{\boldsymbol{\pi}}_{naive}$

For each configuration, the simulation is repeated $R = 1000$ times. The EM algorithm, convergence is declared when

$$\ell(\boldsymbol{\pi}^{(t+1)}) - \ell(\boldsymbol{\pi}^{(t)}) < \epsilon,$$

with $\epsilon = 10^{-4}$, or after a maximum of 1000 iterations.The regularization parameter for the MoM estimator, λ, was chosen by visual inspection of Figure 4, selecting the smallest value that still produced a sufficiently smooth estimate. This resulted in $\lambda = 10^{-1}$ for $\sigma = 1$ and $\lambda = 10^{-2}$ for $\sigma = 4$.

3.4 Simulation procedure

We evaluate the estimators using a Monte Carlo simulation. Each replication proceeds as follows:

Step 1: Generate latent data Sample age classes independently as

$$\boldsymbol{Z} \sim \text{Multinomial}(n, \boldsymbol{\pi})$$

Step 2: Apply measurement error For each individual with true age class $T_r = j$, the observed class is sampled according to the misclassification probabilities:

$$O_r \mid T_r = j \sim \text{Categorical}(M_{j\cdot}).$$

Step 3: Estimate distribution Compute $\hat{\boldsymbol{\pi}}$ using each estimator.

Step 4: Evaluate performance Compute the total variation distance

$$\mathrm{TV}(\hat{\boldsymbol{\pi}}, \boldsymbol{\pi}) = \frac{1}{2} \sum_{j=1}^{k} |\hat{\pi}_j - \pi_j|.$$

Step 5: Forecasting analysis Forecasts based on $\hat{\boldsymbol{\pi}}$ are compared to those obtained from the true distribution $\boldsymbol{\pi}$ using the relative error in total population size defined in Section 2.3.

4 Results

4.1 Recovery of latent age distribution

Overall, the simulation indicates that the EM maximum likelihood (ML) estimator provides the most accurate recovery of the latent age distribution in terms of diminished bias, but at the cost of increased variability. In contrast, the naive and MoM estimators are systematically biased, particularly in lower age classes, but exhibits substantially lower variability. While the ML estimator better captures the expected structure of the latent distribution, its estimates are more sensitive to sampling variation and measurement error, especially for the older age classes.

From Figures 3, it can be observed that the naive estimator produces smoother and more regular age profiles across adjacent bins. In contrast, the ML estimator yields more irregular patterns, with greater local fluctuations between neighboring age classes, while the MoM estimator produces comparatively smooth but biased estimates. This behavior is consistently observed across simulation replications and is not specific to this individual realization of the estimate.

Figure 3: Recovered latent age distributions for $n = 500$ using the ML, method-of-moments (MoM), and naive estimators under noise levels $\sigma = 1$ and $\sigma = 4$. The displayed distribution corresponds to a single reconstruction, selected as the one whose total variation (TV) distance to the true latent distribution is closest to the median across all 1000 Monte Carlo simulations. The dashed line represents the true latent age distribution.

The singular estimate in Figure 3, while informative, do not reflect the overall performance of the recovery methods. In Figure 4, we therefore present the median recovered age distribution across all simulations, together with the corresponding 2.5% and 97.5% Monte Carlo percentile bands based on 1000 replications.

Consistent with the previous figure, the naive estimator systematically underestimates lower age classes (0–3 years), with a more pronounced effect under higher noise ($\sigma = 4$). It can also be observed that the corresponding confidence bands for naive and MoM are considerably narrower than those of the ML estimator across all age classes. The ML estimator exhibits a close fit to the true latent age distribution in terms of the median, but shows substantially higher variability compared to the naive estimator. This variability also becomes less smooth under the higher noise setting ($\sigma = 4$). The MoM estimator displays a low variability high bias with additional mass being shifted toward older age classes under ($\sigma = 4$).

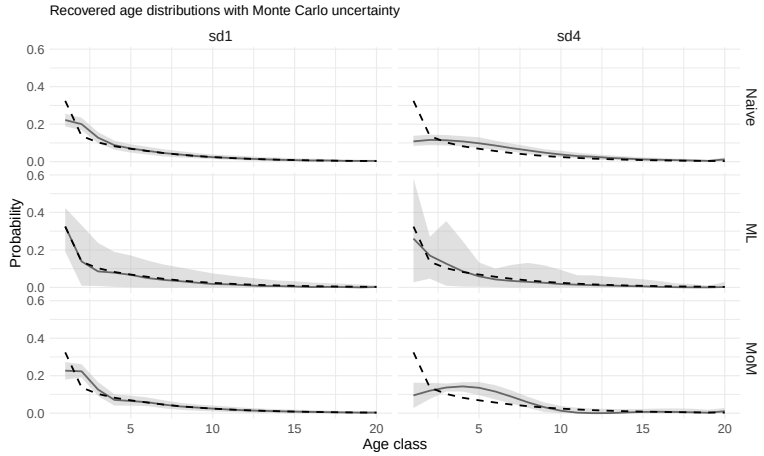

Figure 4: Recovered latent age distributions for $n = 500$ using the ML, method-of-moments (MoM), and naive estimators under noise levels $\sigma = 1$ and $\sigma = 4$. For each setting, the median estimate across 1000 Monte Carlo simulations is shown, with shaded regions indicating the 2.5% and 97.5% Monte Carlo percentiles. The dashed line represents the true latent age distribution.

4.2 Sample size effect on TV

Having characterized the recovered distributions and their variability, we next quantify how estimator performance changes with sample size using total variation (TV) distance.

In Figure 5, we observe how sample size affects the total variation (TV) distance across estimators. The method of moments (MoM) estimator exhibits a slightly higher TV distance across both noise settings relative to the naive estimator, with lower sensitivity to sample size. The naive estimator tracks the oracle baseline more closely than the ML and MoM estimators in both noise settings except for $\sigma = 4$ and $n > 500$.

The maximum likelihood (ML) estimator shows consistent improvement with increasing sample size in both error settings, with a higher baseline TV distance under $\sigma = 4$ than with $\sigma = 1$. Among all estimators, ML exhibits the strongest negative dependence with increased sample size within the considered range, and this trend does not appear to taper off.

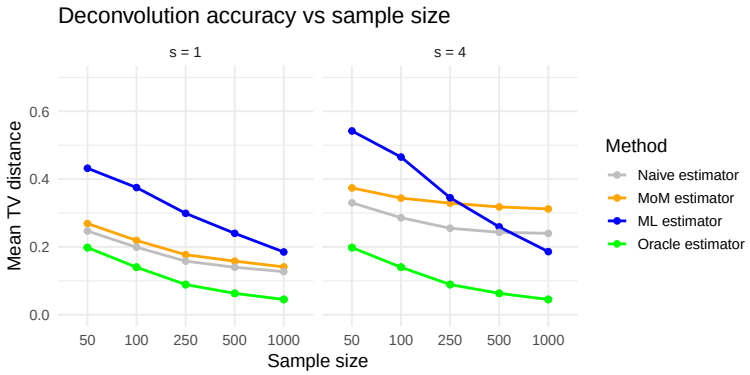

Figure 5: Mean total variation (TV) distance between estimated $\hat{\boldsymbol{\pi}}$ and true age distribution $\boldsymbol{\pi}$ across five sample sizes, based on 1000 Monte Carlo replications. Results are shown for four estimators: the naive estimator, method of moments (MoM), EM-based maximum likelihood (EM), and an oracle baseline obtained by sampling directly from the true latent distribution $\boldsymbol{\pi}$. Panels correspond to different measurement error levels ($\sigma = 1$ and $\sigma = 4$).

To isolate the effect of sampling variability on the total variation (TV) distance, we adjust the observed TV distances by subtracting the corresponding (TV) distance obtained from the oracle baseline. Consistent with Figure 5, the maximum likelihood (ML) estimator exhibits a strong dependence on sample size, with the TV distance decreasing as sample size increases. In contrast, both the MoM and naive estimators show increasing TV distance with sample size.

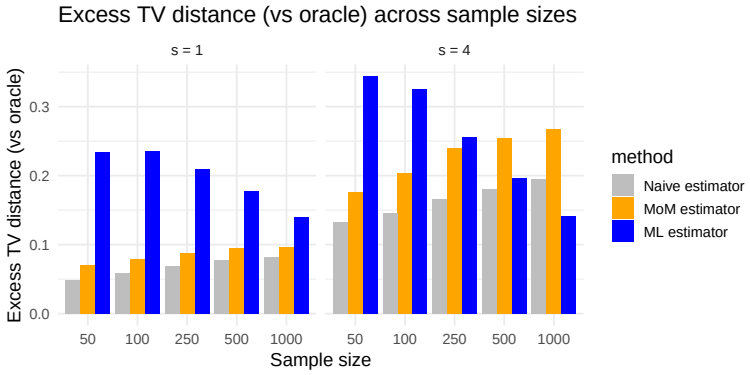

Figure 6: Mean excess total variation (TV) distance relative to the oracle baseline across sample sizes and estimation methods, based on 1000 Monte Carlo simulations. The excess TV is defined as $\mathrm{TV}(\hat{\boldsymbol{\pi}}, \boldsymbol{\pi}) - \mathrm{TV}(\hat{\boldsymbol{\pi}}_{\mathrm{oracle}}, \boldsymbol{\pi})$.

We now proceed to examine the distribution and variability in total

variation (TV) distance across estimators. The previous plots have focused on mean performance, whereas the violin plot in Figure 7 illustrates the full distribution of TV distances across Monte Carlo replications.

The results show that the TV distance for the naive and MoM estimators is more tightly concentrated around their respective medians, with the MoM estimator exhibiting slightly higher variability. In contrast, the ML estimator displays a substantially wider distribution of TV distances. This spread becomes more pronounced for both the MoM and ML estimators under the higher noise level $\sigma = 4$.

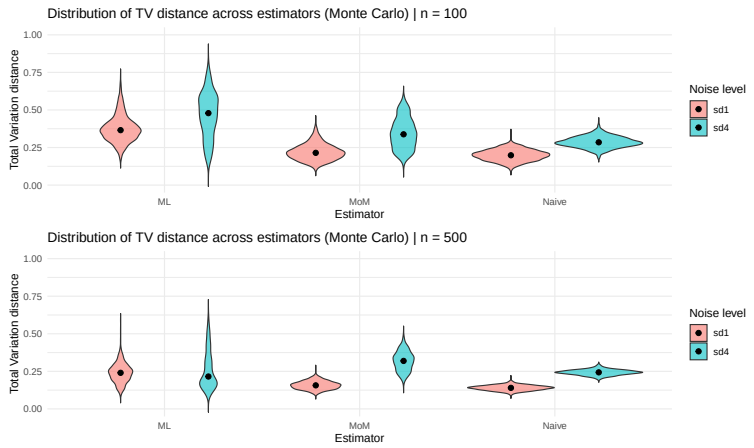

Figure 7: Violin plot of total variation (TV) distance across 1000 Monte Carlo replications for three estimators under noise levels $\sigma = 1$ and $\sigma = 4$. The upper panel corresponds to sample size $n = 100$, while the lower panel corresponds to $n = 500$.

4.3 Population forecast

As mentioned in Section 2.3.1, population dynamics are governed by the Leslie matrix framework. We evaluate how estimation errors in the initial age distribution propagate into five-year total population forecasts. Figure 8 shows the deviation of estimated total population sizes from the true latent population. A mean value close to zero indicates an unbiased population forecast.

The naive estimator exhibits a smaller bias when $\sigma = 1$ compared to $\sigma = 4$, corresponding to approximately 16% and 76% higher total population at year five, respectively. Overall, the naive estimator shows relatively low variability compared to the other estimators. The MoM estimator exhibits slightly lower variability than the ML projection but shows a positive bias, consistently overestimating the total population by approximately 10% for $\sigma = 1$ and 59% for $\sigma = 4$. The EM-based ML estimator appears ap-

proximately unbiased in terms of its expected total population forecast, but exhibits substantially higher uncertainty compared to the naive estimator.

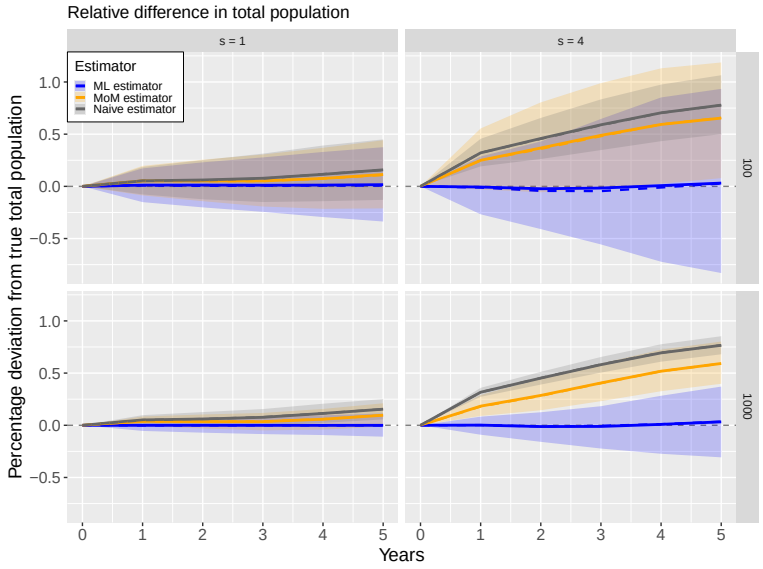

Figure 8: Five-year population forecast error under different estimators and measurement error levels ($\sigma = 1$ and $\sigma = 4$), based on 1000 Monte Carlo simulations. The forecasts are obtained by propagating the estimated initial age distributions $\hat{\boldsymbol{\pi}}$ through the Leslie matrix dynamics defined in Equation (10). Results are shown for the naive estimator, method-of-moments (MoM), and EM-based maximum likelihood (ML) estimator with sample size 100 and 1000.

5 Discussion

The results of this simulation study show that the latent age distribution can be recovered with a less biased estimate using the EM-based maximum likelihood estimator, but at the cost of increased variability under both low ($\sigma = 1$) and high ($\sigma = 4$) measurement error. In contrast, the naive estimator and MoM estimator yielded a biased but smoother age distribution. This systematically underestimates the proportion of younger individuals (ages < 3 years), which results in a corresponding overestimation of older age classes.

Although this bias may appear small in magnitude, its impact is non-negligible in the context of population forecasting. Shifts in probability mass away from demographically critical age classes, especially those contributing most to reproduction, can lead to disproportionately large errors in projected population size. This highlights that even modest bias in age structure estimates can have substantial downstream consequences when

used in population models.

5.1 Properties of estimators

As expected from Figure 2, the naive estimator produces biased estimates due to the misclassification matrix $\boldsymbol{M}$. This bias becomes more pronounced as the sample size increases, as illustrated in Figure 6.

For small sample sizes, the total variation (TV) distance of the naive estimator is largely dominated by sampling variability. As a result, the difference between the naive estimator and the oracle baseline is relatively small. However, as the sample size increases, the effect of sampling variability diminishes, and the remaining error is primarily driven by the systematic bias introduced by the misclassification process.

Formally, by the law of large numbers and the continuous mapping theorem (see p. 356 in Robert and Casella 2020), we have that as $n \to \infty$:

$$\mathrm{TV}(\hat{\boldsymbol{\pi}}_{\text{oracle}}, \boldsymbol{\pi}) \to 0,$$

while the naive estimator converges to the biased distribution $M\boldsymbol{\pi}$, such that

$$\mathrm{TV}(\hat{\boldsymbol{\pi}}_{naive}, \boldsymbol{\pi}) \to \mathrm{TV}(\boldsymbol{\pi}\boldsymbol{M}, \boldsymbol{\pi}).$$

This explains the increase in TV with sample size for the naive estimator in Figure 6. Similarly to the naive estimator the MoM would roughly converge with $n \to \infty$ to:

$$\mathrm{TV}(\hat{\boldsymbol{\pi}}_{\text{MoM}}, \boldsymbol{\pi}) \to \mathrm{TV}(A_\lambda \boldsymbol{\pi}, \boldsymbol{\pi}) \tag{11}$$

The method-of-moments (MoM) estimator is limited in its recovery of the latent age distribution by the instability of matrix inversion. Matrix inversion does not guarantee constraints such as non-negative probabilities or summing to one, which is accounted for by truncation to 0 and normalization. This projection violates the assumptions of the continuous mapping theorem; thus, we do not expect exact convergence to Equation (11). The MoM is therefore limited in its recovery of the latent age distribution by this bias.

The EM maximum likelihood (ML) estimator, as shown in the results, exhibits a decreasing TV distance with increasing sample size relative to the oracle baseline see Figure 6. However, it shows a significantly larger spread in its TV distribution, as illustrated in Figure 7. The overall smooth shape of the violin plot, with no clear clustering, suggests that the EM algorithm does not converge to distinct local maxima or plateaus. Instead, it appears to operate on a relatively flat log-likelihood surface, consistent with the convergence diagnostics in Appendix 2, which shows that the maximum number of iterations is rarely reached. This is symptomatic of an ill-posed EM deconvolution problem, where weak identifiability leads

to unstable convergence behavior (McLachlan and Krishnan 2008, p. 28). The results suggest that this instability is partially mitigated as sample size increases.

We identify in this simulation study that deconvolution through EM maximum likelihood and method-of-moments (MoM) exhibit different failure modes. While the MoM estimator is affected by numerical instability due to inversion of near-singular matrices, as well as the lack of enforced probability constraints (non-negativity and unit sum), the EM algorithm instead suffers from poor identifiability, reflected in a relatively flat log-likelihood surface. In the MoM case, this leads to unstable and often non-admissible estimates unless regularized, whereas for EM the main issue is not feasibility but sensitivity to the observed data. This distinction highlights that both methods are affected by the ill-posed nature of the deconvolution problem, but in fundamentally different ways: MoM through numerical instability, and EM through weak identifiability and flat likelihood.

5.2 Practical Implications

The results above have focused on statistical properties and estimator behavior. We now turn to the practical motivation for estimating age distributions and evaluate how the different estimators affect total population forecasts in an ecological context.

As shown in Figure 8, both MoM and ML outperform the naive estimator in terms of mean and median under both high and low noise, as well as for small and large sample sizes. In particular, we observe that MoM exhibits a slightly wider Monte Carlo percentile band in Figure 4, closer to that of the naive estimator, but with a differently skewed bias.

This suggests that for ecological indicators such as these, the direction and structure of the bias may be more important than its magnitude alone. Specifically we know from the table in Appendix 6 that misclassifying individuals of fertile ≥ 4 years to none fertile ages < 4 or vice-versa will significantly alter the forecast. This is consistent with our expected measurement error in Figure 2 thus leading to the naive estimator overestimating individuals of fertile age.

Whether the MoM or ML estimator is preferable over the naive estimator therefore depends on the decision-makers tolerance for variability, the forecast horizon, and constraints related to sampling cost and sample size.

5.3 Viability of assumptions & improvements

In order to study the deconvolution methods in a controlled setting, several simplifying assumptions were made. The misclassification matrix $\boldsymbol{M}$ was assumed to be fully known. In practical applications, however, $\boldsymbol{M}$ must typically be estimated from auxiliary data such as the one observed in panel

a shown in Figure 5 of Nakamura et al. 2023, introducing additional uncertainty. The results presented in this study should therefore be interpreted with caution, as they represent a best-case scenario for deconvolution performance.

Regarding the method-of-moments estimator, the choice of regularization parameter λ was based on visual inspection and is therefore unlikely to be optimal. A more objective approach, such as cross-validation or other data-driven selection criteria, would provide a more appropriate level of regularization and yield a fairer comparison between estimators.

For the EM algorithm, the use of multiple initialization points per sample or the inclusion of additional constraints in the maximization step could improve stability and robustness. In particular, initialization based on the naive estimator may result in zero-valued components, which prevents recovery of those classes in subsequent iterations see Equation (9). This highlights the sensitivity of the EM algorithm to initialization in this setting.

Finally, the population forecasting analysis assumes a fixed Leslie matrix $\mathbf{L}$ and a stable age distribution. In realistic ecological settings, demographic parameters such as fertility and survival vary over time and must themselves be estimated from data, introducing additional sources of uncertainty. It is therefore plausible that, in practice, the combined effect of these uncertainties may dominate the forecasting error, potentially reducing the impact of improvements obtained through deconvolution.

Overall, this thesis provides insight into the behavior of different estimators under simplified and controlled conditions in an ill-posed deconvolution setting, as well as demonstrating a practical use case and its implications in an applied ecological context.

Table 1: Parameters used to construct the Leslie matrix.

Age	Survival	p_coy	Litter_size	F_i
1	0.50	0.00	0.00	0.00
2	0.89	0.00	0.00	0.00
3	0.95	0.00	0.00	0.00
4	0.99	0.00	0.00	0.00
5	0.98	0.15	2.12	0.31
6	0.96	0.35	2.25	0.76
7	0.96	0.55	2.40	1.27
8	0.96	0.55	2.40	1.27
9	0.96	0.55	2.40	1.27
10	0.96	0.55	2.40	1.27
11	0.96	0.55	2.40	1.27
12	0.96	0.70	2.60	1.75
13	0.96	0.70	2.60	1.75
14	0.96	0.70	2.60	1.75
15	0.96	0.70	2.60	1.75
16	0.96	0.70	2.60	1.75
17	0.96	0.58	2.75	1.53
18	0.96	0.58	2.75	1.53
19	0.96	0.58	2.75	1.53
20	0.96	0.58	2.75	1.53

6 Appendix

SD	SampleSize	FailureRate
sd1	50	0.000
sd1	100	0.000
sd1	250	0.000
sd1	500	0.000
sd1	1000	0.000
sd4	50	0.456
sd4	100	0.349
sd4	250	0.174
sd4	500	0.077
sd4	1000	0.022

Table 2: Failure rate of the EM algorithm, defined as the proportion of Monte Carlo replications in which the algorithm did not converge within the maximum number of iterations.

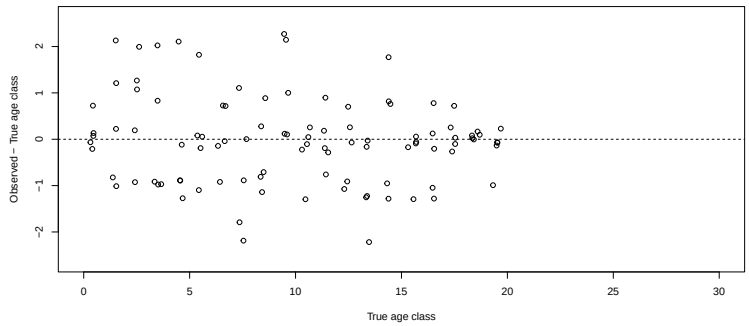

Figure 9: Illustrative simulation of five individuals per age class j under the measurement error model with $\sigma = 1$. The resulting misclassification pattern is designed to replicate that in Figure 5(a) of Nakamura et al. 2023.

References

Alm, Sven Erick and Tom Britton (2008). *Stokastik: sannolikhetsteori och statistikteori med tillämpningar*. 1st ed. Stockholm: Liber. ISBN: 978-91-47-05351-3.

Åsbrink, Jessica, Martin Sköld, and Niclas Gyllenstrand (2025). *Resultat från inventering av brunbjörn i Västerbottens län 2024*. Rapport från Naturhistoriska riksmuseet 2025:3. Naturhistoriska riksmuseets småskriftserie. Stockholm: Naturhistoriska riksmuseet. URL: https://www.nrm.se/download/18.fa4343b199372946183c02c/1758261826003/2025-3-rapport_bjorninventering_2024.pdf.

Åsbrink, Jessica, Martin Sköld, Thomas Källman, et al. (June 30, 2023). *Resultat från inventering av brunbjörn i Dalarnas, Gävleborgs och Värmlands län 2022*. Rapport från Naturhistoriska riksmuseet 2023:2. Naturhistoriska riksmuseets småskriftserie. Rapporten bör citeras: Åsbrink et al. 2023. Stockholm: Naturhistoriska riksmuseet, Enheten för miljöforskning och övervakning. URL: https://www.nrm.se/download/18.fa4343b199372946183c02c/1758261826003/2025-3-rapport_bjorninventering_2024.pdf (visited on 05/08/2026).

Björnprojektet – 40 års forskning (June 19, 2025). Rapport 8942. Stockholm, Sweden: Naturvårdsverket. URL: https://www.naturvardsverket.se/49f29c/globalassets/media/publikationer-pdf/8900/978-91-620-8942-9a.pdf.

Carroll, Raymond J. et al. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. 2nd ed. Second edition. Boca Raton, FL, USA: Chapman and Hall/CRC. ISBN: 9781584886334.

Caswell, Hal (2001). *Matrix Population Models: Construction, Analysis, and Interpretation*. 2nd ed. Second edition. Sunderland, MA, USA: Sinauer Associates. ISBN: 9780878931217.

Conn, Paul B. and Duane R. Diefenbach (2007). "ADJUSTING AGE AND STAGE DISTRIBUTIONS FOR MISCLASSIFICATION ERRORS". In: *Ecology* 88.8, pp. 1977–1983. DOI: https://doi.org/10.1890/07-0369.1. eprint: https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1890/07-0369.1. URL: https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/07-0369.1.

Kindberg, Jonas et al. (2011). "Estimating population size and trends of the Swedish brown bear Ursus arctos population". In: *Wildlife Biology* 17.2, pp. 114–123. DOI: https://doi.org/10.2981/10-100. eprint: https://nsojournals.onlinelibrary.wiley.com/doi/pdf/10.2981/10-100. URL: https://nsojournals.onlinelibrary.wiley.com/doi/abs/10.2981/10-100.

McLachlan, Geoffrey J. and Thriyambakam Krishnan (2008). *The EM Algorithm and Extensions*. 2nd ed. Second edition. Hoboken, NJ, USA: Wiley-Interscience. ISBN: 9780471201700.

Morris, PA (Apr. 2008). "A review of mammalian age determination methods". In: *Mammal Review* 2, pp. 69–104. DOI: 10.1111/j.1365-2907.1972.tb00160.x.

Nakamura, S. et al. (2023). "Age estimation based on blood DNA methylation levels in brown bears". In: *Molecular Ecology Resources* 23, pp. 1211–1225. DOI: 10.1111/1755-0998.13788.

Norwegian Institute for Nature Research (n.d.). *Bear model (age-structured population model)*. URL: https://www.nina.no/bearmodel.

Robert, Christian P. and George Casella (2020). *Likelihood and Bayesian Inference: With Applications in Biology and Medicine*. Latest edition. Boca Raton, FL, USA: CRC Press.

Wasserman, Larry A. (2004). *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics. New York, NY, USA: Springer Science+Business Media. ISBN: 978-0-387-21736-9. DOI: 10.1007/978-0-387-21736-9.