Fitting Markovian binary trees using global and individual population data

Sophie Hautphenne*

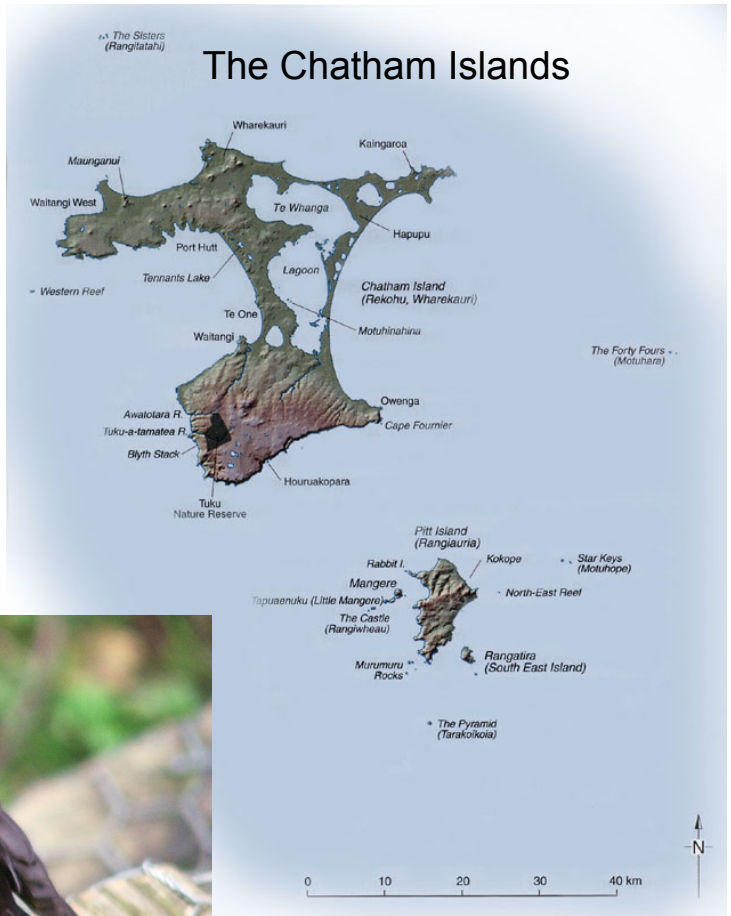
joint work with Melanie Massaro and Kate Turner

*The University of Melbourne & EPFL

MAM9

Budapest, 28–30 June 2016

Motivation : Modelling the black robin population



Saving the world's most endangered bird

By 1980, the population of black robins had declined to five birds, including only a single successful breeding pair.

In 1980-1989, intensive conservation efforts helped the population recover to 93 birds by 1990 [1].

In 1990-1998, the population was closely monitored, but without human intervention, and continued to grow rapidly to 197 adults by 1998.

From 1999, the population growth slowed considerably and it only reached about 250 adults in 2014.

1. D. Butler and D. Merton. *The Black Robin : Saving the World's Most Endangered Bird*. Oxford University Press, Auckland (1992)

Motivation : Modelling the black robin population

We work in collaboration with field biologists who have been collecting raw data on the population for more than 30 years

species_name	year	nest____ID	status__	island___	mID	mage	fID	fage	laid	rim	hatched	fled	ind	o1ID	o2ID	o3ID	cf
Black_Robin	1980	80MAN-DM01	breeding	Mangere__	A2	3	A1	9	2	0	1	0	0				yes
Black_Robin	1980	80MAN-DM02	breeding	Mangere__	A2	3	A1	9	2	0	2	2	2	B1	B2		yes
Black_Robin	1980	80MAN-DM03	breeding	Mangere__	A2	3	A1	9	2	0	1	1	1	B3			no
Black_Robin	1981	81MAN-DM01	breeding	Mangere__	A2	4	A1	10	2	0	2	2	2	C1	C2		yes
Black_Robin	1981	81MAN-DM02	breeding	Mangere__	A2	4	A1	10	2	0	1	1	1	C3			yes
Black_Robin	1981	81MAN-DM03	breeding	Mangere__	A2	4	A1	10	2	0	2	2	2	C4	C5		no
Black_Robin	1982	82MAN-DM01	breeding	Mangere__	A2	5	A1	11	1	0	0	0	0				na
Black_Robin	1982	82MAN-DM02	breeding	Mangere__	A2	5	A1	11	2	0	1	1	1	D1			no
Black_Robin	1983	83MAN-DM01	breeding	Mangere__	A2	6	D1	1	2	0	1	1	0	E1			yes
Black_Robin	1983	83MAN-DM02	breeding	Mangere__	A2	6	D1	1	2	0	2	1	1	E2			no
Black_Robin	1983	83MAN-DM06	breeding	Mangere__	A2	6	A1	12	2	0	0	0	0				yes
Black_Robin	1983	83MAN-DM07	breeding	Mangere__	A2	6	C5	2	2	0	0	0	0				no
Black_Robin	1984	84MAN-DM01	breeding	Mangere__	A2	7	C5	3	3	0	3	3	3	F11	F12	F13	yes
Black_Robin	1984	84MAN-DM02	breeding	Mangere__	A2	7	C5	3	3	0	1	1	1	F14			yes
Black_Robin	1984	84MAN-DM03	breeding	Mangere__	A2	7	C5	3	3	0	3	2	2	F15	F16		no
Black_Robin	1984	84MAN-DM04	breeding	Mangere__	A2	7	C5	3	2	0	1	1	0	F17			no
Black_Robin	1985	85MAN-DM01	breeding	Mangere__	A2	8	E8	2	1	0	1	0	0				yes
Black_Robin	1985	85MAN-DM02	breeding	Mangere__	A2	8	E8	2	3	0	2	2	1	G14	G15		yes
Black_Robin	1985	85MAN-DM03	breeding	Mangere__	A2	8	E8	2	3	1	0	0	0				no
Black_Robin	1985	85MAN-DM04	breeding	Mangere__	A2	8	E8	2	2	1	1	1	1	G16			no
Black_Robin	1980	80MAN-DM04	breeding	Mangere__	A4	6	A3	2	2	0	1	1	1	B4			yes
Black_Robin	1980	80MAN-DM05	breeding	Mangere__	A4	6	A3	2	2	0	1	0	0				yes
Black_Robin	1981	81MAN-DM04	breeding	Mangere__	A4	7	A3	3	2	0	1	0	0				yes
Black_Robin	1981	81MAN-DM05	breeding	Mangere__	A4	7	A3	3	2	0	2	0	0				yes
Black_Robin	1981	81MAN-DM06	breeding	Mangere__	A4	7	A3	3	2	0	1	1	0	C6			no
Black_Robin	1982	82MAN-DM03	breeding	Mangere__	A4	8	A3	4	2	0	0	0	0				no
Black_Robin	1982	82MAN-DM04	breeding	Mangere__	A4	8	A3	4	1	0	0	0	0				no
Black_Robin	1982	82MAN-DM05	breeding	Mangere__	A4	8	A3	4	1	0	0	0	0				no
Black_Robin	1981	81MAN-PAIR07	paired__	Mangere__	B1	1	B3	1	0	0	0	0	0				na

We want to fit branching processes to these data to make an age-specific demographic analysis of the population

Markovian binary trees : Trade-off between realism and tractability

Linear birth-and-death process :

- Lifetimes follow an exponential distribution
- Reproduction occurs according to a Poisson process

Not realistic enough !

Bellman-Harris branching processes :

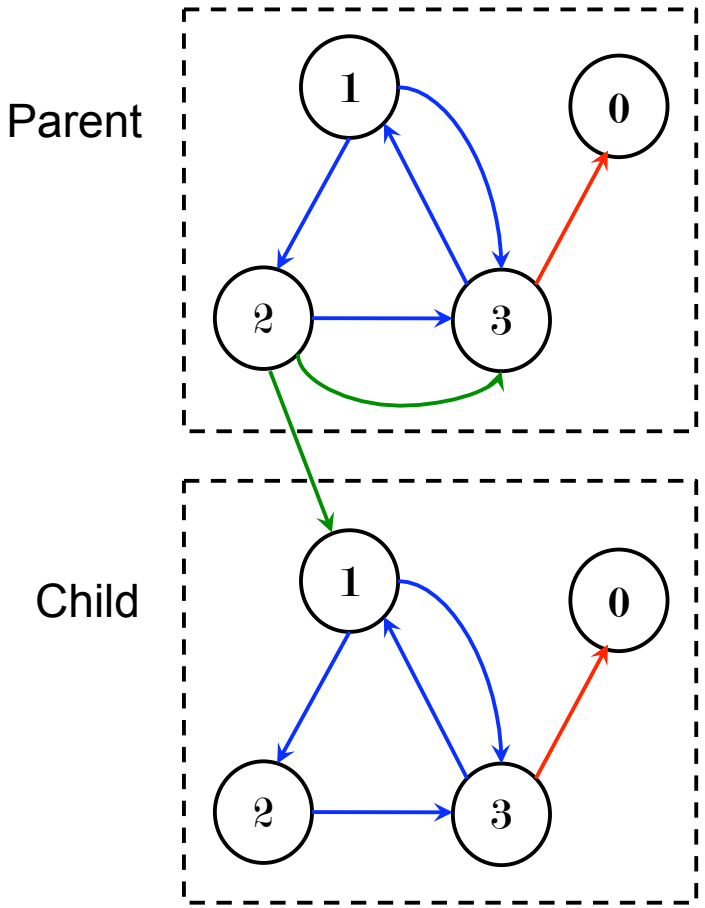
- Lifetimes follow an arbitrary distribution
- Reproduction occurs according to a more general process

Not tractable enough !

We consider an intermediate type of branching process, called the *Markovian binary tree* (MBT), which is at the same time very general and tractable.

In an MBT, individuals' lifetime is structured into phases.

Phase-structured lifetime

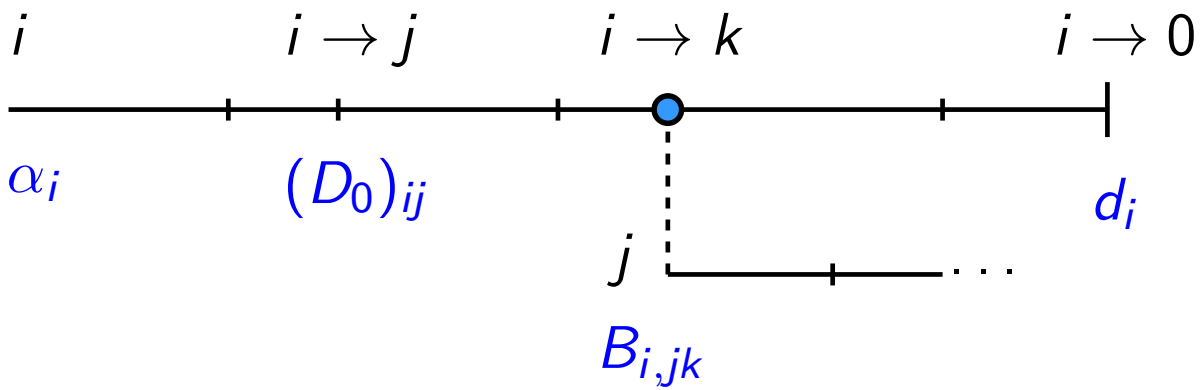


3 types of transitions:

- ``Hidden'' transitions
- Birth
- Death

The individuals' lifetime in an MBT

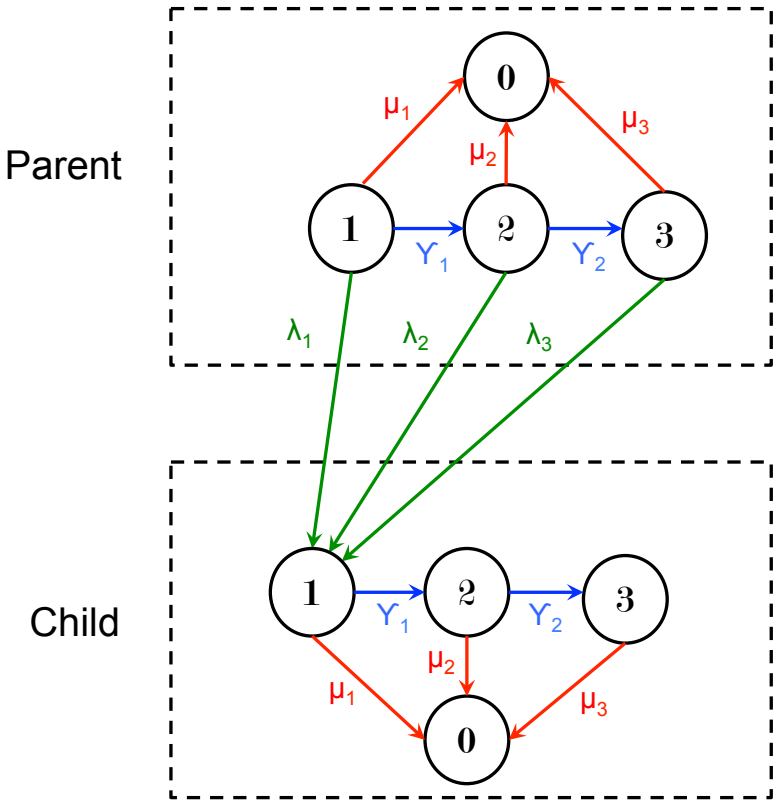
- Lifetime controlled by an underlying Markov process with n transient phases and one absorbing phase ;



- $\boldsymbol{\alpha}$: initial phase distribution ($1 \times n$ vector) ;
- D_0 : hidden phase transition rates ($n \times n$ matrix) ;
- B : transition rates associated with a birth ($n \times n^2$ matrix) ;
- $\mathbf{d}$: transition rates associated with the death ($n \times 1$ vector).

n phases $\longrightarrow$ $n^3 + n^2 + n - 1$ parameters !

Simple structure



Now, $\alpha_1 = 1, \quad (D_0)_{i,i+1} = \gamma_i, \quad B_{i,1i} = \lambda_i, \quad d_i = \mu_i$

n phases $\rightarrow$ $3n-1$ parameters $\gamma_1, \ldots, \gamma_{n-1}, \lambda_1, \ldots, \lambda_n, \mu_1, \ldots, \mu_n$

Reproduction and lifetime in the simple MBT

In this MBT model,

- reproduction events occur according to a transient Markov modulated Poisson process,
- the lifetime of an individual has a Coxian phase-type distribution.

Parameter fitting using population data

Aim : to fit the parameters of the MBT to different types of population data sets, distinguishing between

1. *global population data :* averaged age-specific fertility and mortality rates over the whole population, or
2. *individual population data :* age-specific fertility and mortality counts per individual.

The estimation method depends on the precision of the available data.

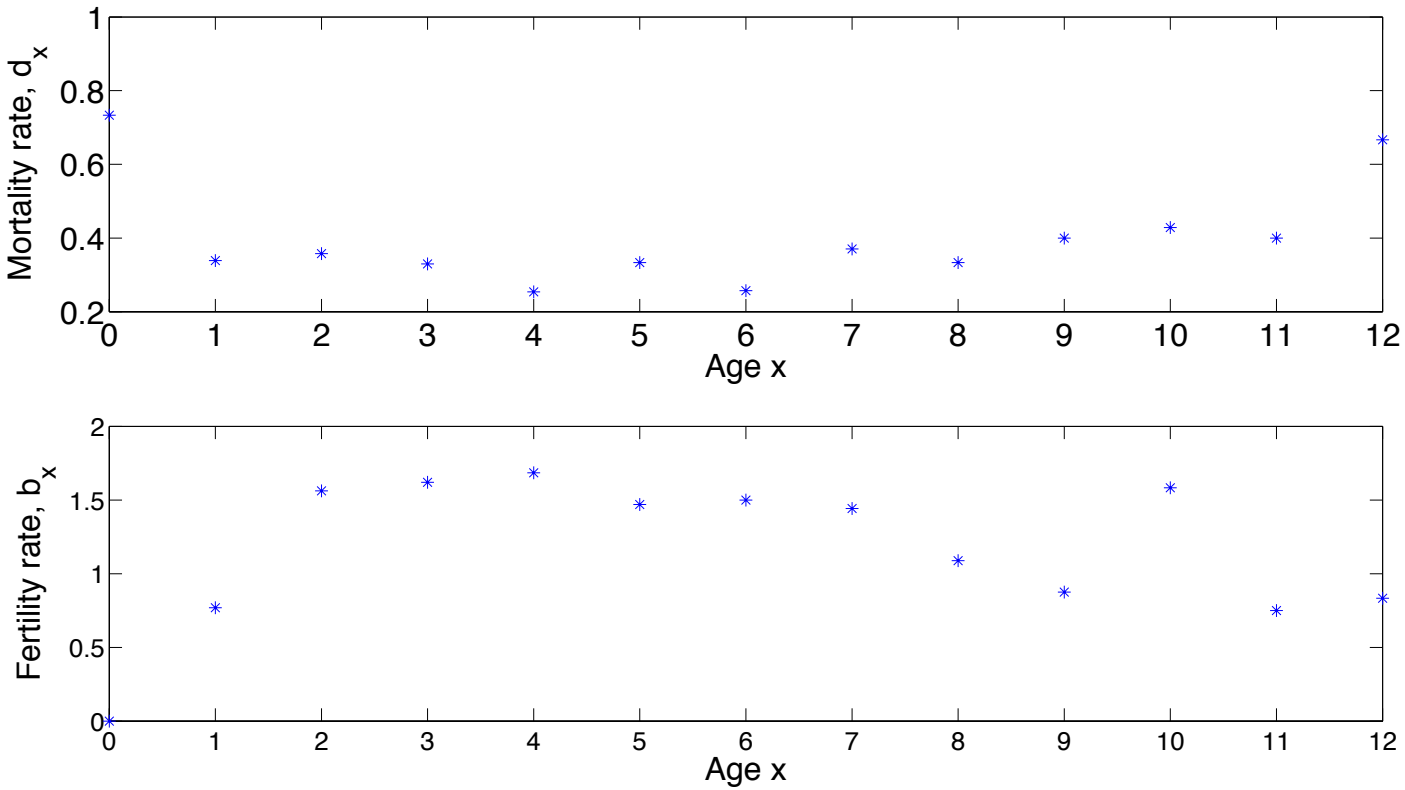
We aim at choosing the optimal number of phases n using some validation methods.

Once an estimator is known for the parameters, we derive confidence intervals for the model outcomes.

Global population data and non-linear regression

Assume that we have estimates of the mean mortality and fertility rates at age x, denoted respectively by d_x and b_x, for ages $x = 0, 1, \ldots, M$. These are our data points.

Example : black robin data, close monitoring period 1990-2001



Mean age-specific fertility and mortality rates

Let L be the lifetime of an individual, and $D = D_0 + \text{diag}(\boldsymbol{\lambda})$.

We can show that the expressions for the mean mortality and fertility rates at age x in the MBT model are respectively given by

$$\bar{d}(x) = \mathrm{P}(x < L \leq x+1 \,|\, L > x) = \frac{\boldsymbol{\alpha} e^{Dx}(I - e^{D})\mathbf{1}}{\boldsymbol{\alpha} e^{Dx}\mathbf{1}}$$

$$\bar{\beta}(x) = \mathrm{E}(N([x, x+1)) \,|\, L > x) = \frac{\boldsymbol{\alpha} e^{Dx}(I - e^{D})(-D)^{-1}\boldsymbol{\lambda}}{\boldsymbol{\alpha} e^{Dx}\mathbf{1}}$$

These functions are non-linear in both the input variable x and in the parameters.

Weighted non-linear least square estimates

We want the same MBT model to fit both the age-specific fertility and mortality data.

The $3n-1$ parameters are estimated by minimizing the sum of weighted squared errors

$$F = \sum_{x=0}^{M} \left[(d_x - \bar{d}(x))^2 + (\beta_x - \bar{\beta}(x))^2 \right] S_x,$$

where

$$S_x = (1-d_0)(1-d_1)\cdots(1-d_{x-1})$$

is the observed probability of survival until age x.

Individual population data and maximum likelihood

Instead of using the mean age-specific mortality and fertility rates, we can do much better by directly exploiting the individual data counts

species_name	year	nest____ID	status__	island___	mID	mage	fID	fage	laid	rim	hatched	fled	ind	o1ID	o2ID	o3ID	cf
Black_Robin	1980	80MAN-DM01	breeding	Mangere__	A2	3	A1	9	2	0	1	0	0				yes
Black_Robin	1980	80MAN-DM02	breeding	Mangere__	A2	3	A1	9	2	0	2	2	2	B1	B2		yes
Black_Robin	1980	80MAN-DM03	breeding	Mangere__	A2	3	A1	9	2	0	1	1	1	B3			no
Black_Robin	1981	81MAN-DM01	breeding	Mangere__	A2	4	A1	10	2	0	2	2	2	C1	C2		yes
Black_Robin	1981	81MAN-DM02	breeding	Mangere__	A2	4	A1	10	2	0	1	1	1	C3			yes
Black_Robin	1981	81MAN-DM03	breeding	Mangere__	A2	4	A1	10	2	0	2	2	2	C4	C5		no
Black_Robin	1982	82MAN-DM01	breeding	Mangere__	A2	5	A1	11	1	0	0	0	0				na
Black_Robin	1982	82MAN-DM02	breeding	Mangere__	A2	5	A1	11	2	0	1	1	1	D1			no
Black_Robin	1983	83MAN-DM01	breeding	Mangere__	A2	6	D1	1	2	0	1	1	0	E1			yes
Black_Robin	1983	83MAN-DM02	breeding	Mangere__	A2	6	D1	1	2	0	2	1	1	E2			no
Black_Robin	1983	83MAN-DM06	breeding	Mangere__	A2	6	A1	12	2	0	0	0	0				yes
Black_Robin	1983	83MAN-DM07	breeding	Mangere__	A2	6	C5	2	2	0	0	0	0				no
Black_Robin	1984	84MAN-DM01	breeding	Mangere__	A2	7	C5	3	3	0	3	3	3	F11	F12	F13	yes
Black_Robin	1984	84MAN-DM02	breeding	Mangere__	A2	7	C5	3	3	0	1	1	1	F14			yes
Black_Robin	1984	84MAN-DM03	breeding	Mangere__	A2	7	C5	3	3	0	3	2	2	F15	F16		no
Black_Robin	1984	84MAN-DM04	breeding	Mangere__	A2	7	C5	3	2	0	1	1	0	F17			no
Black_Robin	1985	85MAN-DM01	breeding	Mangere__	A2	8	E8	2	1	0	1	0	0				yes
Black_Robin	1985	85MAN-DM02	breeding	Mangere__	A2	8	E8	2	3	0	2	2	1	G14	G15		yes
Black_Robin	1985	85MAN-DM03	breeding	Mangere__	A2	8	E8	2	3	1	0	0	0				no
Black_Robin	1985	85MAN-DM04	breeding	Mangere__	A2	8	E8	2	2	1	1	1	1	G16			no
Black_Robin	1980	80MAN-DM04	breeding	Mangere__	A4	6	A3	2	2	0	1	1	1	B4			yes
Black_Robin	1980	80MAN-DM05	breeding	Mangere__	A4	6	A3	2	2	0	1	0	0				yes
Black_Robin	1981	81MAN-DM04	breeding	Mangere__	A4	7	A3	3	2	0	1	0	0				yes
Black_Robin	1981	81MAN-DM05	breeding	Mangere__	A4	7	A3	3	2	0	2	0	0				yes
Black_Robin	1981	81MAN-DM06	breeding	Mangere__	A4	7	A3	3	2	0	1	1	0	C6			no
Black_Robin	1982	82MAN-DM03	breeding	Mangere__	A4	8	A3	4	2	0	0	0	0				no
Black_Robin	1982	82MAN-DM04	breeding	Mangere__	A4	8	A3	4	1	0	0	0	0				no
Black_Robin	1982	82MAN-DM05	breeding	Mangere__	A4	8	A3	4	1	0	0	0	0				no
Black_Robin	1981	81MAN-PAIR07	paired__	Mangere__	B1	1	B3	1	0	0	0	0	0				na

We organise these data into samples of i.i.d. individual life vectors

$$\mathbf{v}^{(i)} = [0, -2, 0, 2, 6, 2, 2, 5, 0, -1, -1, -1, -1], \quad 1 \leq i \leq N,$$

recording life and reproduction data for each age class (allowing for missing information)

Maximum likelihood estimation

The log-likelihood function is

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(N)}) = \sum_{j=1}^{N} \log p(\mathbf{v}^{(j)}|\boldsymbol{\theta}), \tag{1}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, D_0, \boldsymbol{\lambda}, \mathbf{d}\}$, and $p(\mathbf{v}^{(j)}|\boldsymbol{\theta})$ is the probability of observing $\mathbf{v}^{(j)}$, under the model parameter $\boldsymbol{\theta}$.

The probabilities $p(\mathbf{v}^{(j)}|\boldsymbol{\theta})$ can be decomposed into products involving matrices $P(k)$ for $1 \leq k \leq K$, where

$$P_{ij}(k) = P[N(1) = k, \varphi(1) = j \,|\, N(0) = 0, \varphi(0) = i],$$

and $K = \max_{i,j}\{v_i^{(j)} : 1 \leq i, 1 \leq j \leq N\}$.

Computing $P(k)$ (I)

- $P_{ij}(k,t) := P[N(t) = k, \varphi(t) = j | N(0) = 0, \varphi(0) = i]$
- $P(k,t) = (P_{ij}(k,t))_{1 \leq i,j \leq n}$
- $P^*(z,t) := \sum_{k \geq 0} P(k,t) z^k, \quad P(k,t) = \left. \frac{\partial^k P^*(z,t)}{k! (\partial z)^k} \right|_{z=0}$
- Kolmogorov equations :

$$\begin{aligned} \partial P^*(z,t)/\partial t &= (D_0 + zD_1) P^*(z,t) \\ \rightarrow P^*(z,t) &= \exp[(D_0 + zD_1)t]. \end{aligned}$$

Computing $P(k)$ (II)

- Successive differentiations w.r. to z :

$$\begin{aligned}
\partial P^*(z,t)/\partial t &= (D_0 + zD_1)P^*(z,t) \\
\partial^2 P^*(z,t)/(\partial t)(\partial z) &= D_1 P^*(z,t) + (D_0 + zD_1)\partial P^*(z,t)/\partial z \\
&\vdots \\
\partial^{(K+1)} P^*(z,t)/(\partial t)(\partial z)^K &= KD_1 \partial^{(K-1)} P^*(z,t)/(\partial z)^{(K-1)} \\
&\quad +(D_0 + zD_1)\partial^K P^*(z,t)/(\partial z)^K .
\end{aligned}$$

- Equivalent to $\partial_t Y(z,t) = A(z)Y(z,t)$ with

$$Y_i(z,t) = \partial^{i-1} P^*(z,t)/(\partial z)^{i-1} \quad 1 \le i \le K+1,$$

$$A(z) = \begin{bmatrix}
(D_0 + zD_1) & & & & \\
D_1 & (D_0 + zD_1) & & & \\
 & 2D_1 & (D_0 + zD_1) & & \\
 & & \ddots & & \\
 & & & KD_1 & (D_0 + zD_1)
\end{bmatrix}$$

Computing $P(k)$ (III)

- We obtain

$$P(k) = P(k,1) = \frac{1}{k!}(\mathbf{e}_k \otimes I)\, e^{\mathcal{M}}\, (\mathbf{e}_1^\top \otimes I)$$

with

$$\mathcal{M} = \begin{bmatrix} D_0 & & & & \\ D_1 & D_0 & & & \\ & 2D_1 & D_0 & & \\ & & \ddots & & \\ & & & KD_1 & D_0 \end{bmatrix}.$$

Optimal number of phases (I) : Akaike Information Criterion

We compute

$$\mathrm{AIC} = 2p - 2\mathcal{L}(\hat{\boldsymbol{\theta}}; \mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(N)}),$$

where $p = 3n - 1$ is the number of parameters, for different values of n, and choose the value of n which minimizes the AIC.

Optimal number of phases (II) : Cross-validation method

We perform a K-fold cross validation over the data sample of individual life vectors (with $K = 5$).

The idea is to randomly divide the data into K equal-sized parts. We leave out part k, fit the model to the other $K - 1$ parts (combined), and then evaluate the likelihood of the left-out kth part under the estimated parameters.

This is done in turn for each part $k = 1, 2, ...K$, and then the results are averaged. We choose the model which maximizes this averaged value.

Confidence intervals

When the real model is unknown, we estimate point-wise confidence intervals for the model outcomes

- in an empirical way, using bootstrapped datasets, or
- in a theoretical way, using the fact that for any function $g(t, \boldsymbol{\theta})$,

$$g(t, \hat{\boldsymbol{\theta}}) \sim \mathcal{N}(g(t, \boldsymbol{\theta}), \nabla g(t, \boldsymbol{\theta})\, J^{-1}\, \nabla g(t, \boldsymbol{\theta})^{\top}),$$

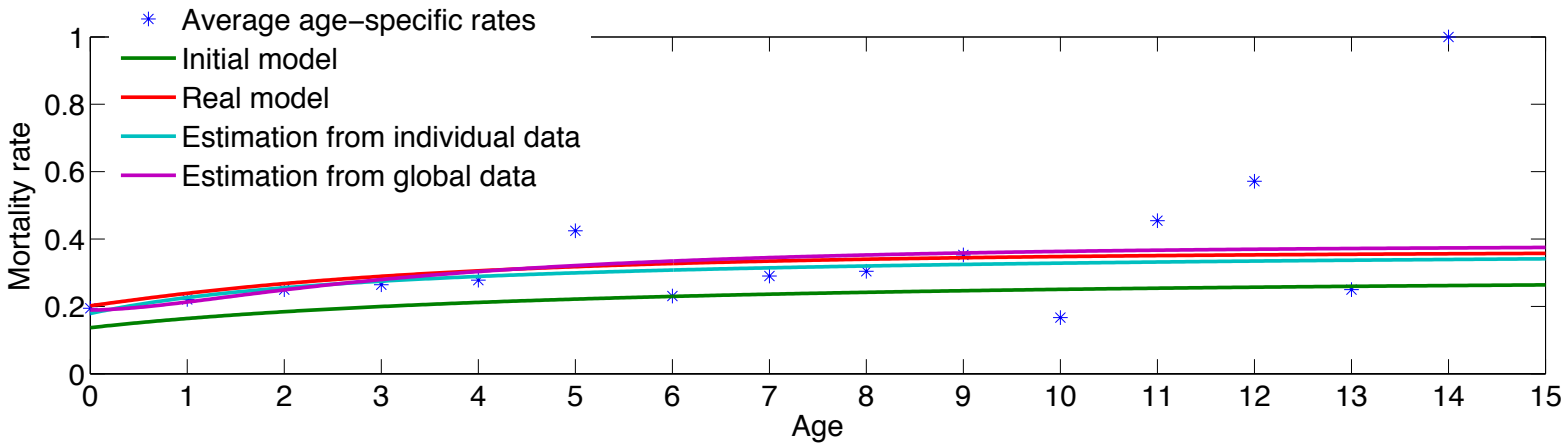
where

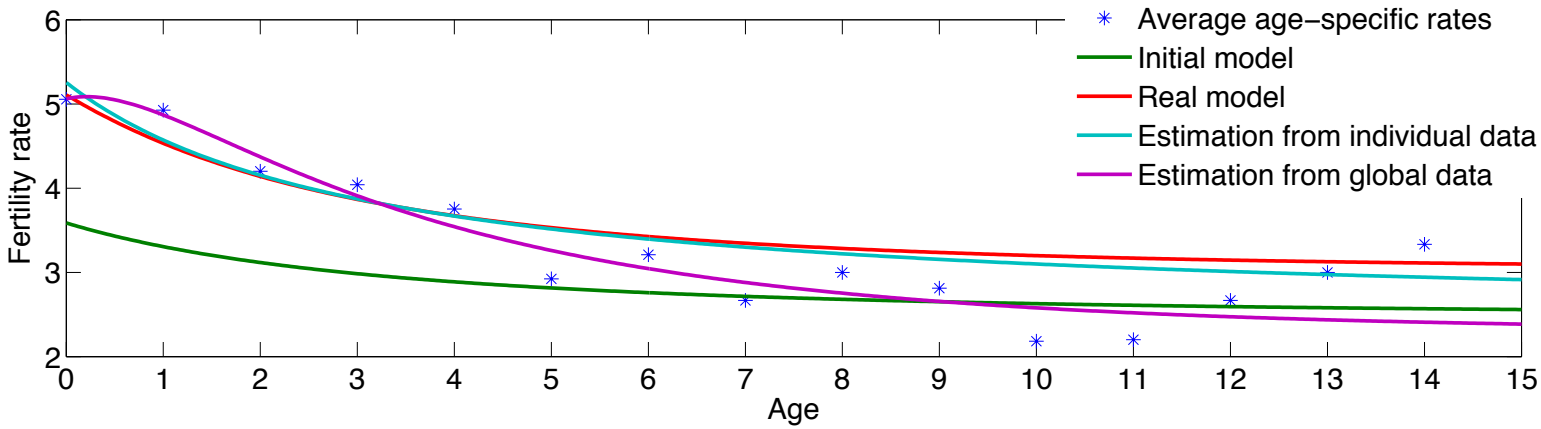
$$J = -\left.\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\top}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

is the observed information matrix.

Toy Example 1 : Comparison of model fits

Example with $n = 3$ phases. We simulated a dataset of $N = 250$ life vectors and recorded results for the first 15 age classes.

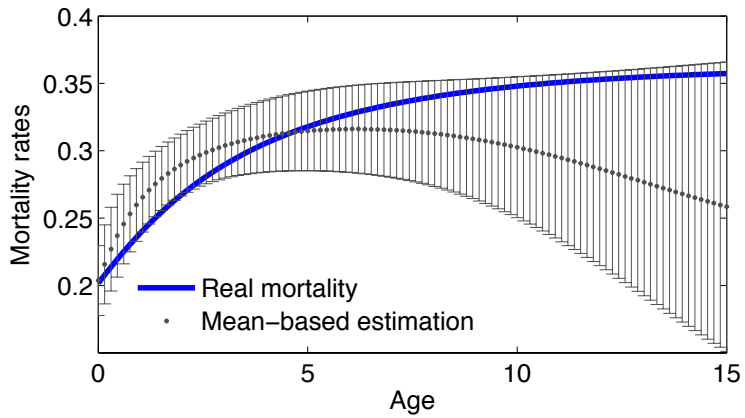


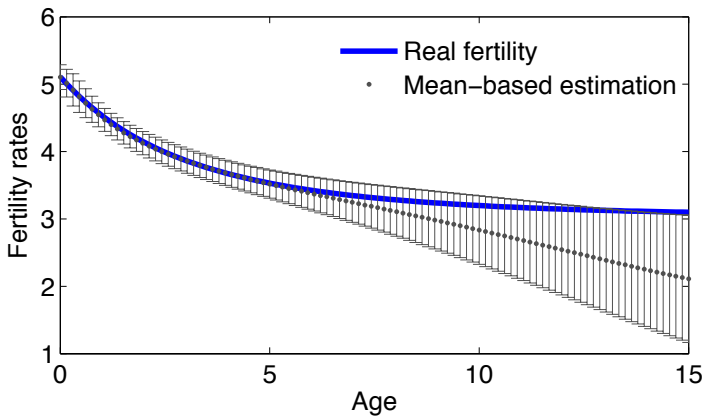


Toy Example 1 : Confidence intervals (I)

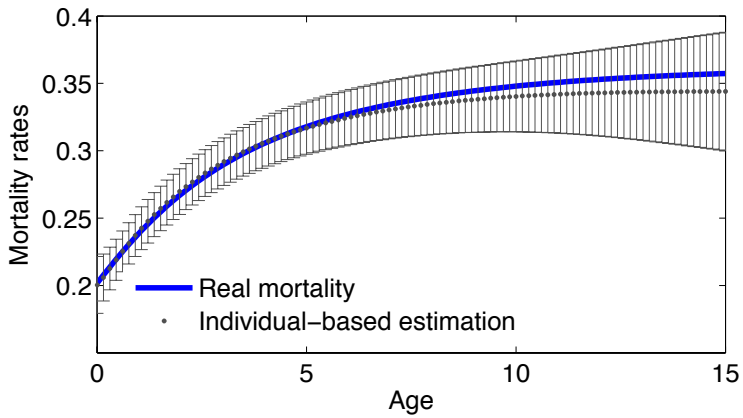
Mean and 95% pointwise CIs of the model fits corresponding to 50 simulations from the real model

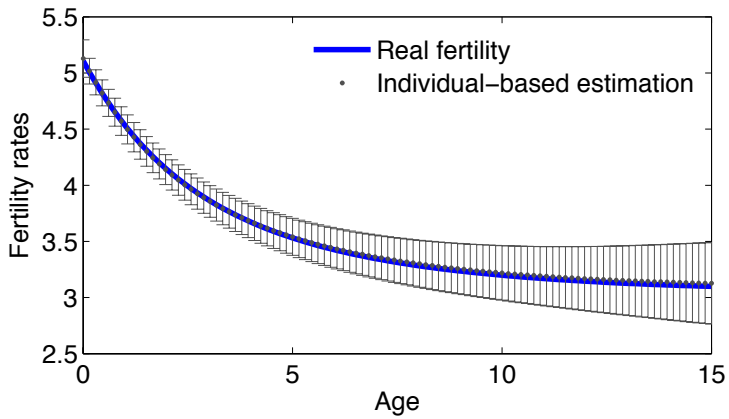
Global population data





Individual population data

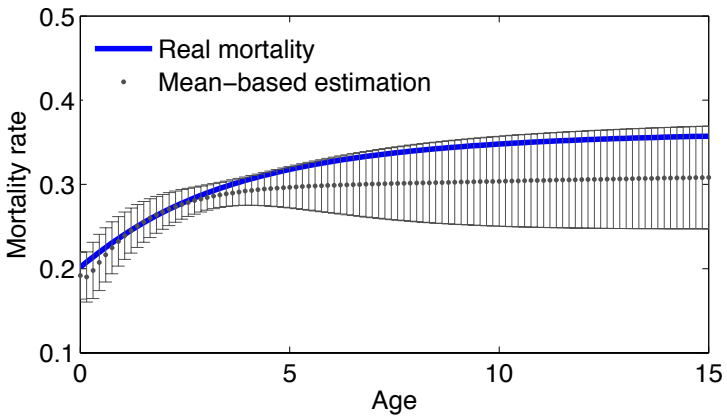


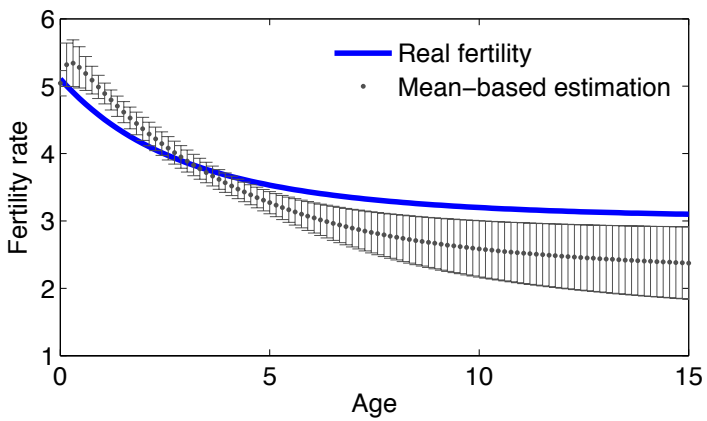


Toy Example 1 : Confidence intervals (II)

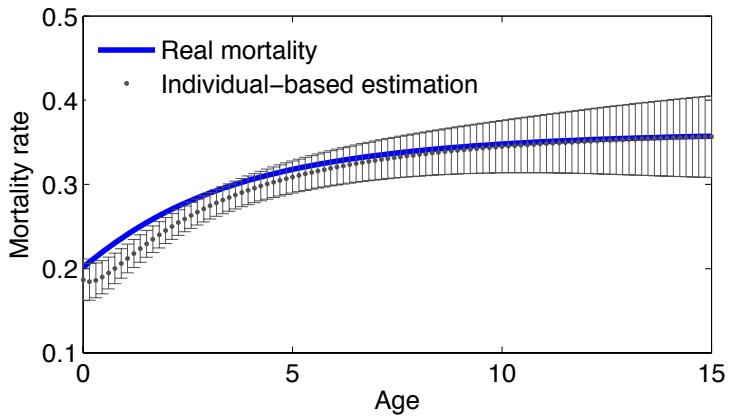
Mean and 95% pointwise CIs of the model fits corresponding to 50 bootstrapped datasets generated from the first dataset

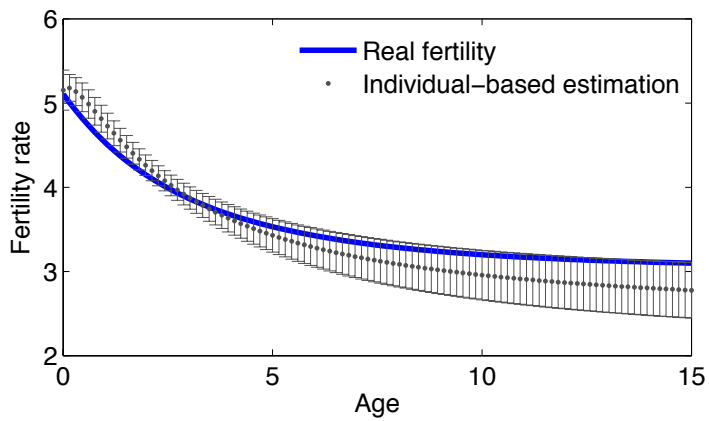
Global population data





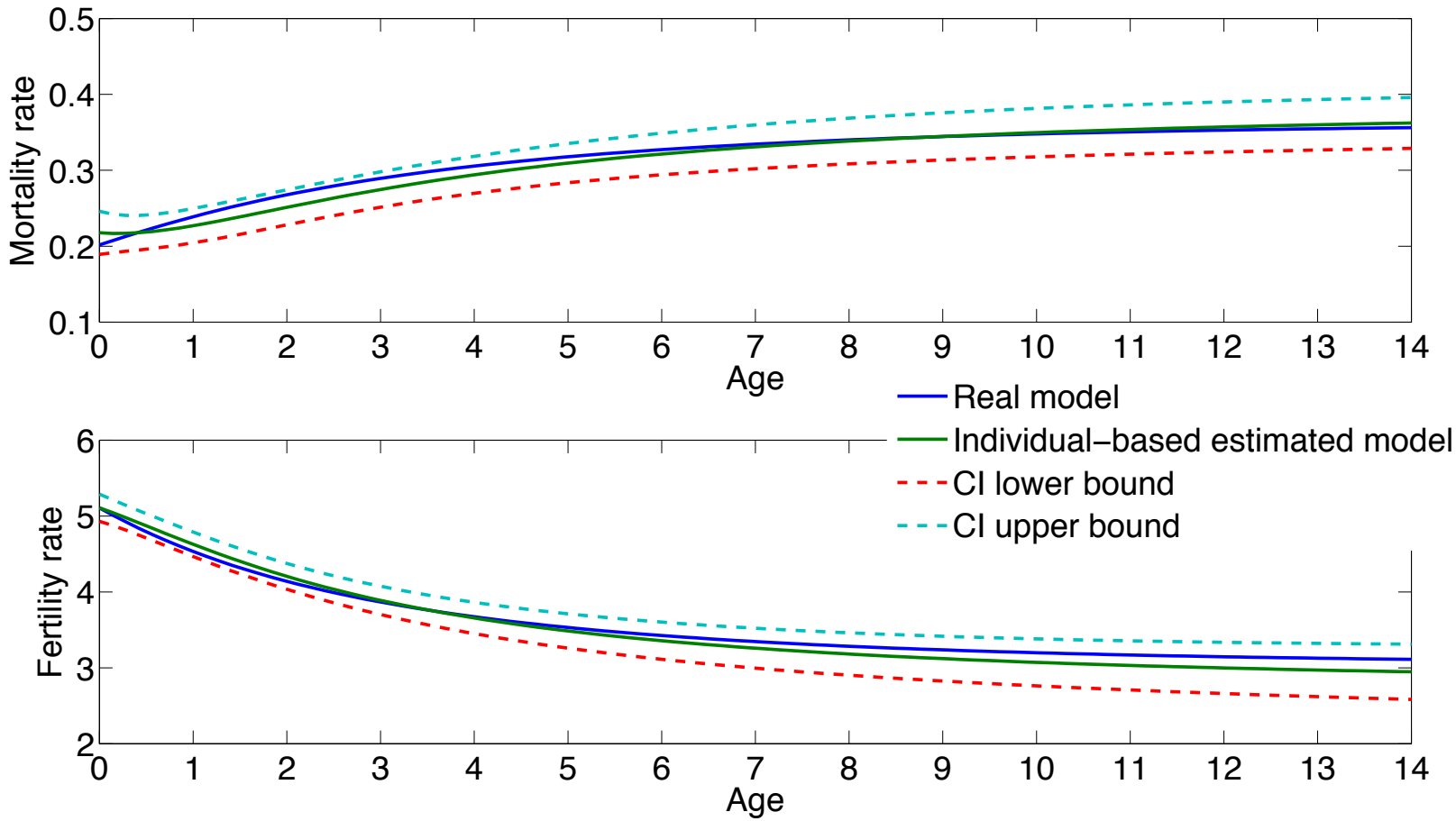
Individual population data





Toy Example 1 : Confidence intervals (III)

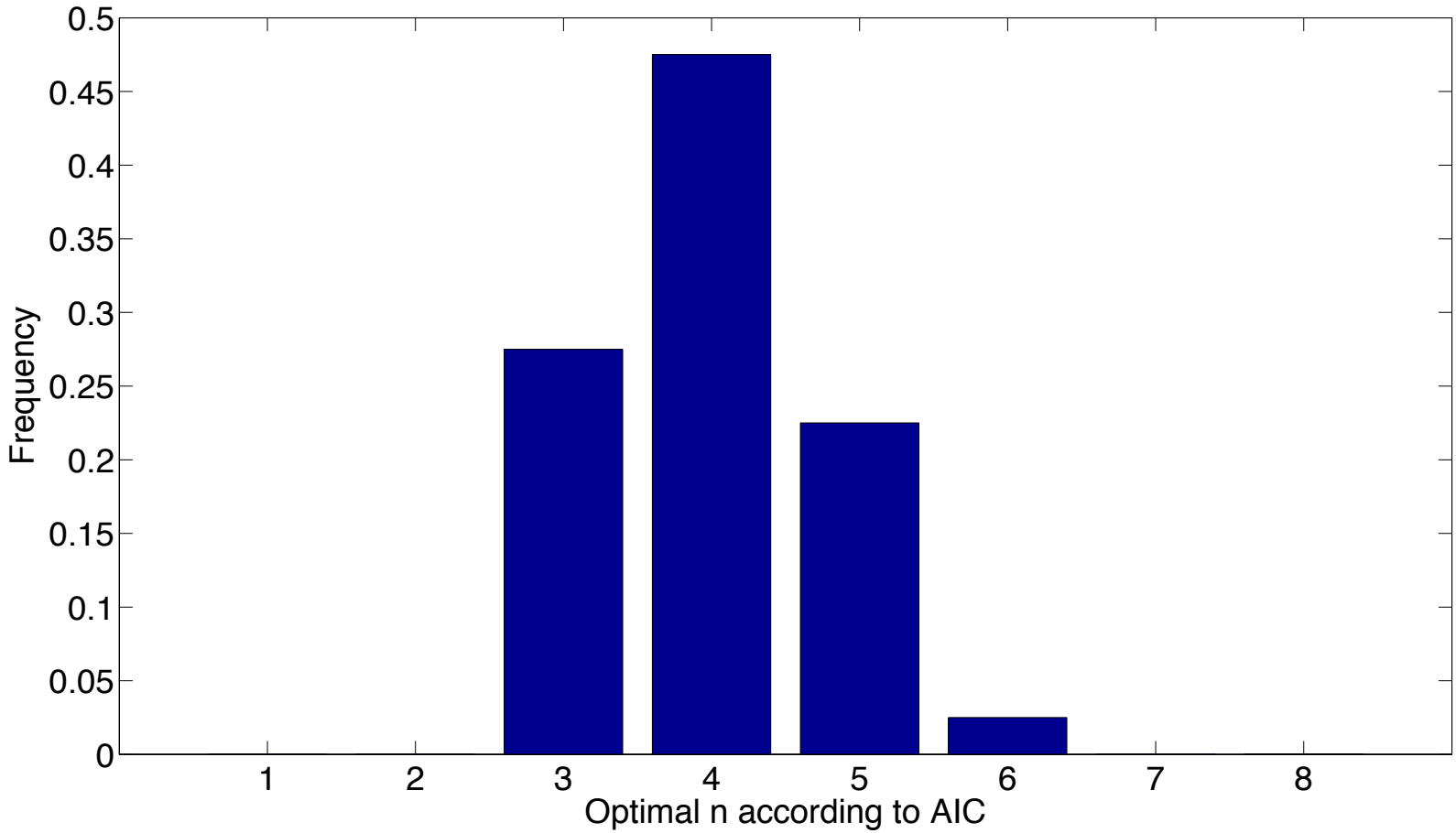
Theoretical CI in the individual population data case



Toy Example 2 : Akaike Information Criterion

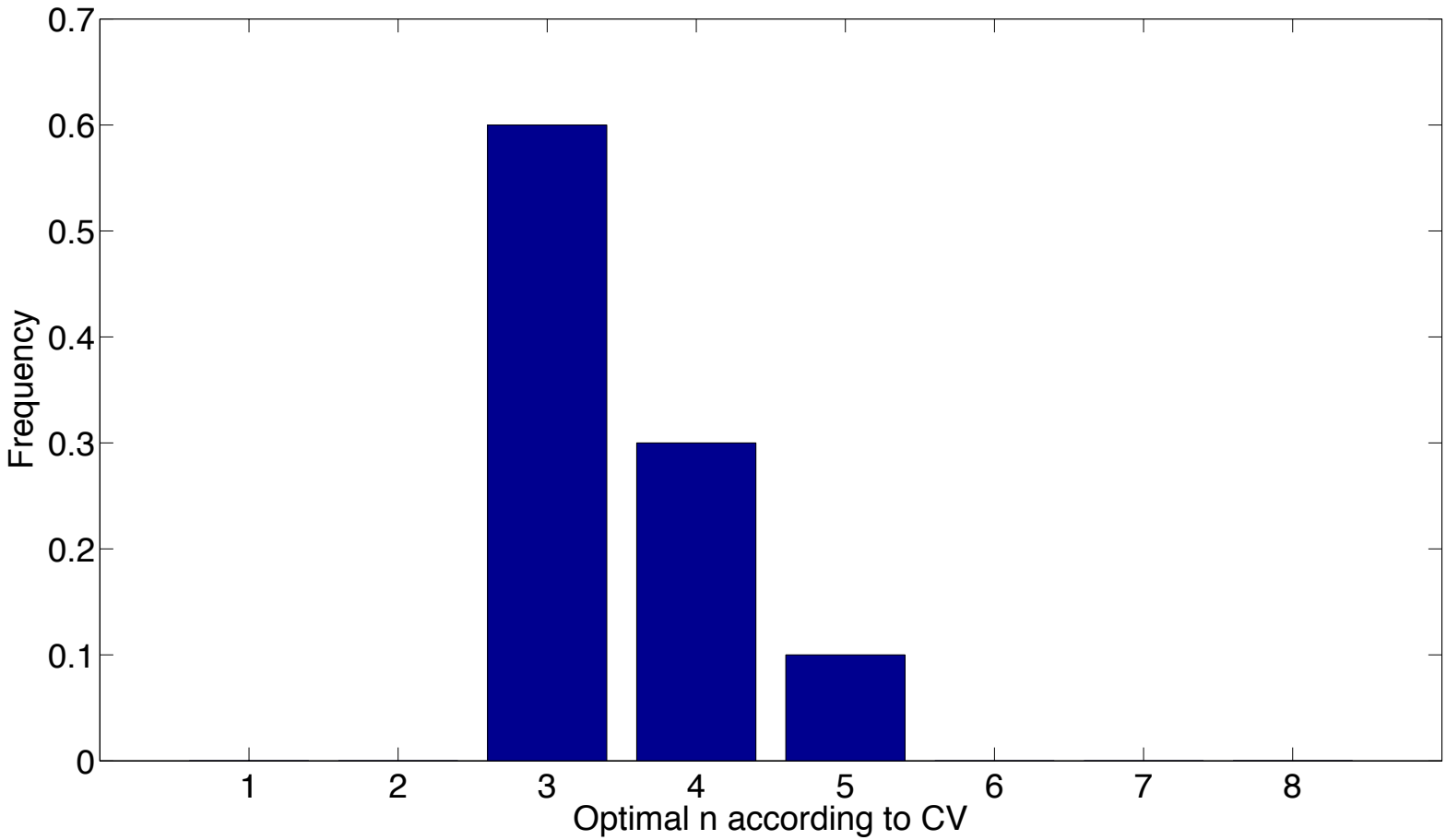
Example with $n = 4$ phases. We simulated a dataset of $N = 500$ life vectors and recorded results for the first 25 age classes.

Frequency of optimal n according to AIC based on 40 simulations



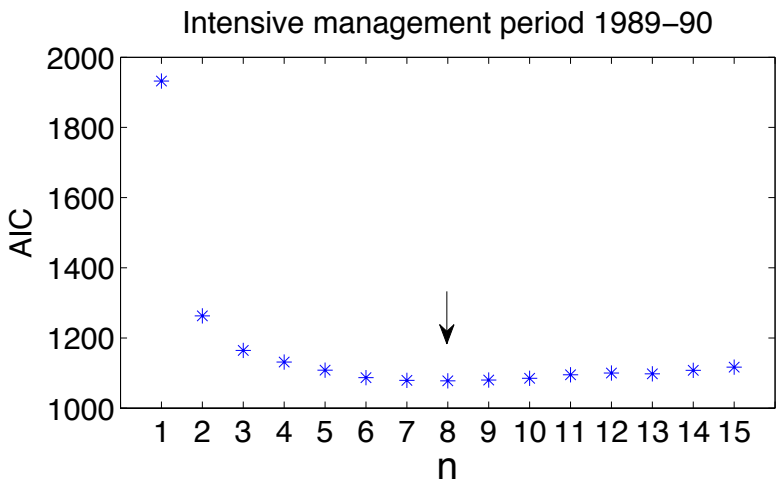
Toy Example 2 : Cross-validation

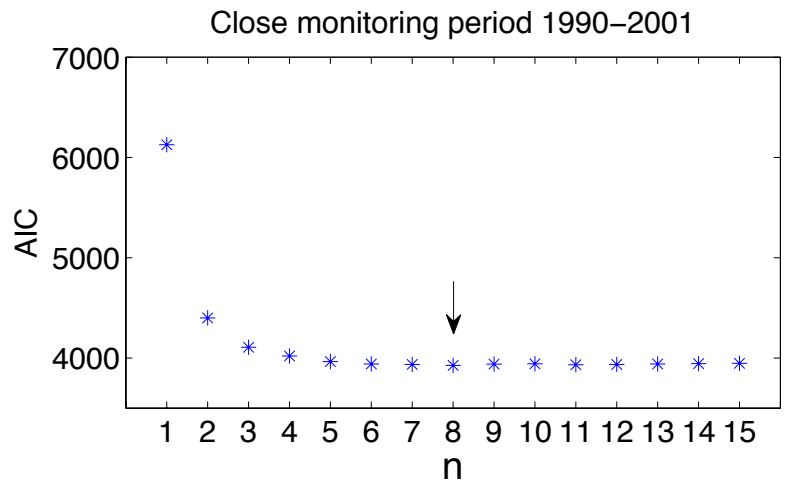
Frequency of optimal n according to CV based on 20 simulations

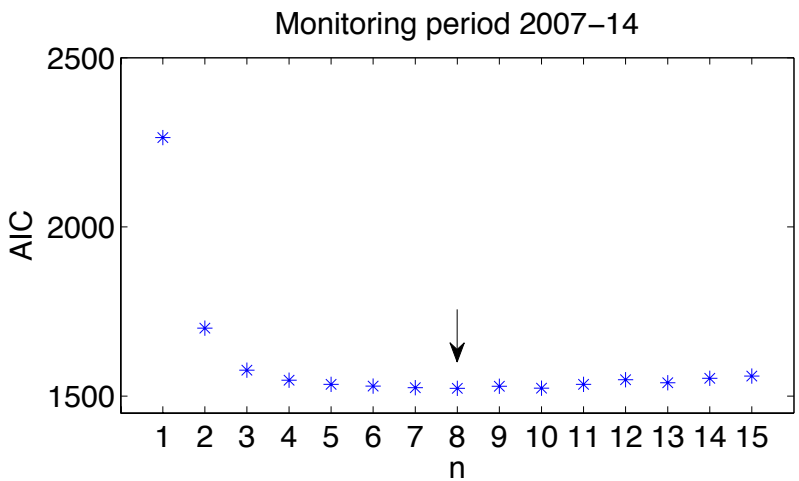


Black robin population : Optimal number of phases (I)

AIC optimal value : $n = 8$

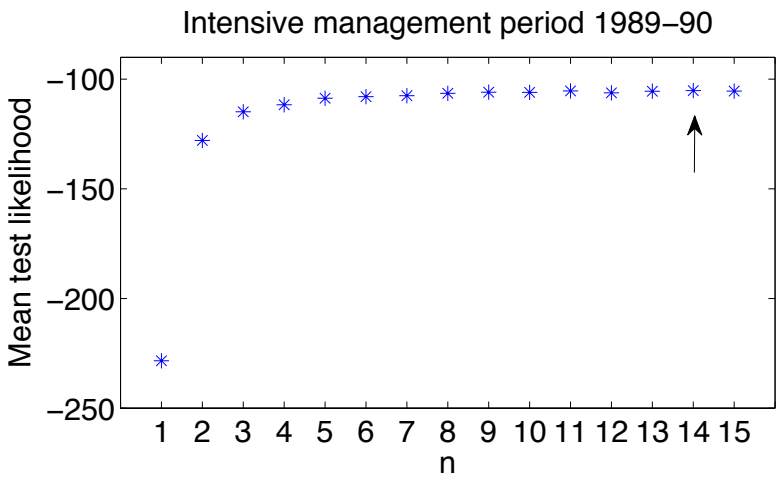


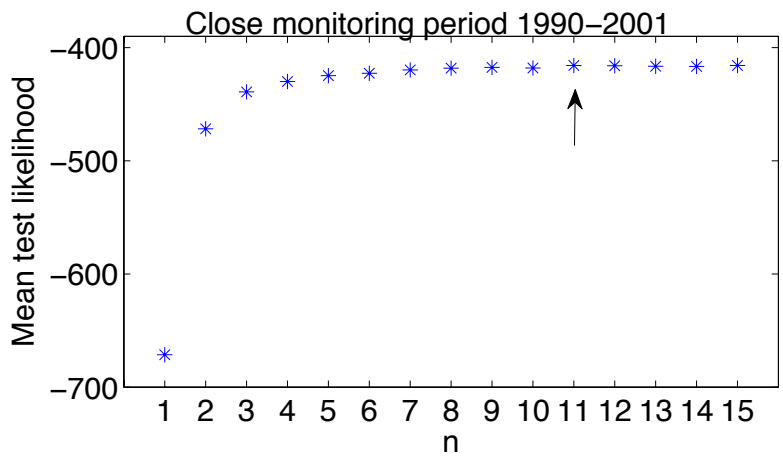


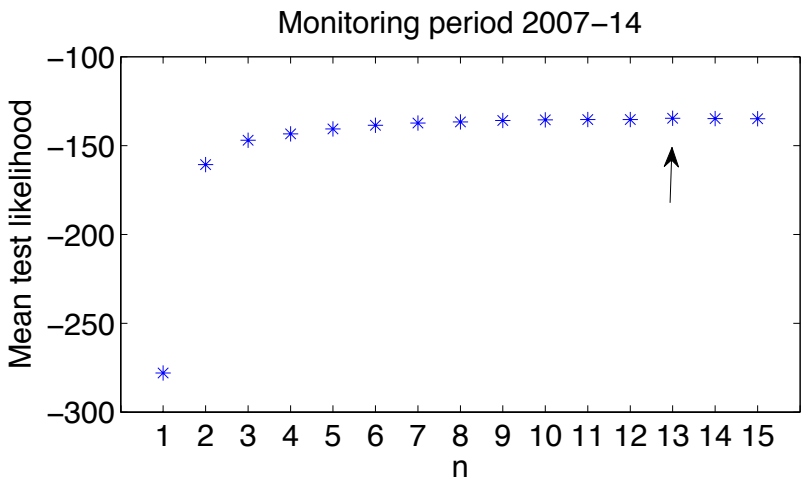


Black robin population : Optimal number of phases (II)

Cross validation optimal value : $n = 14, 11, 13$

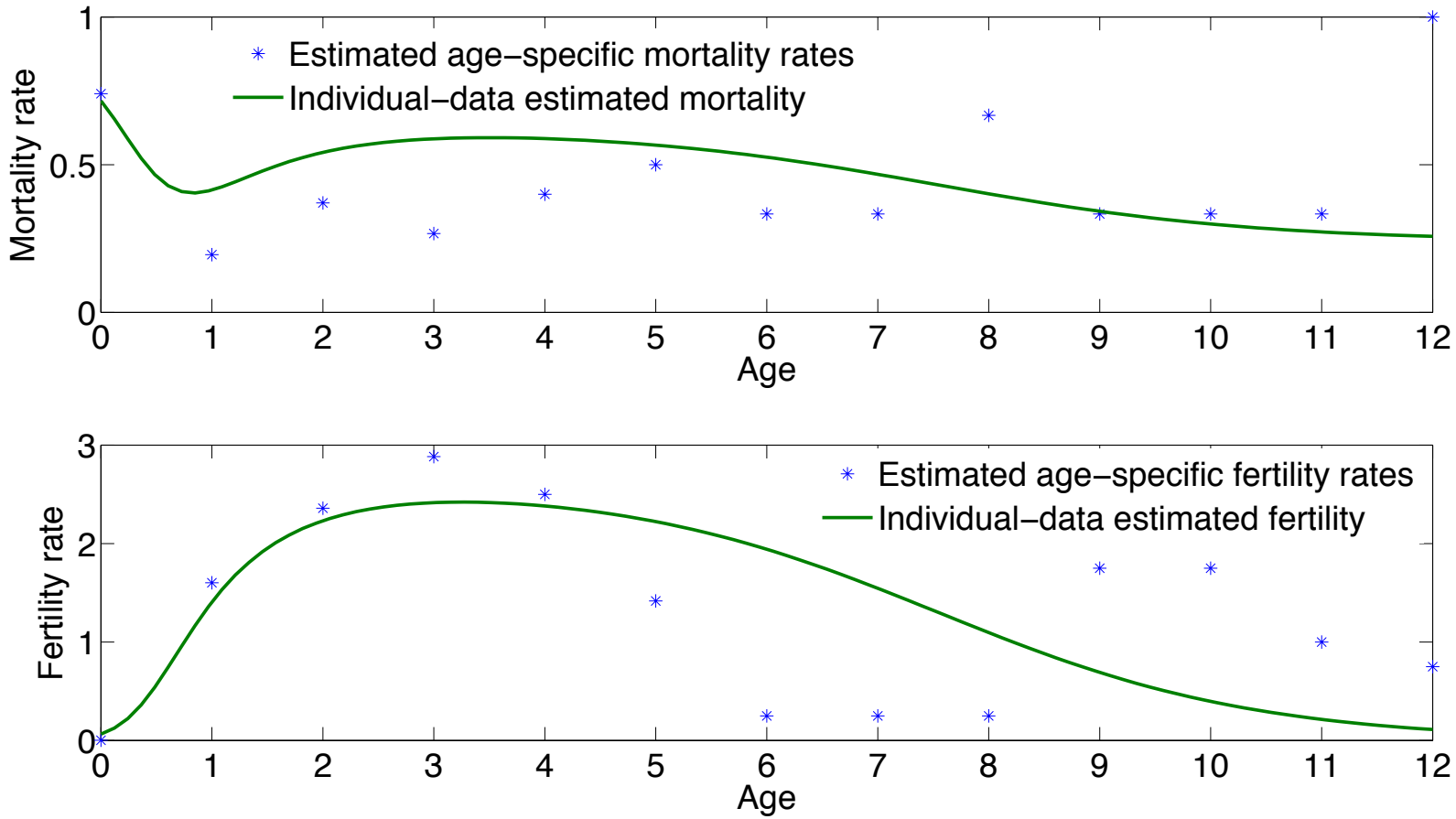






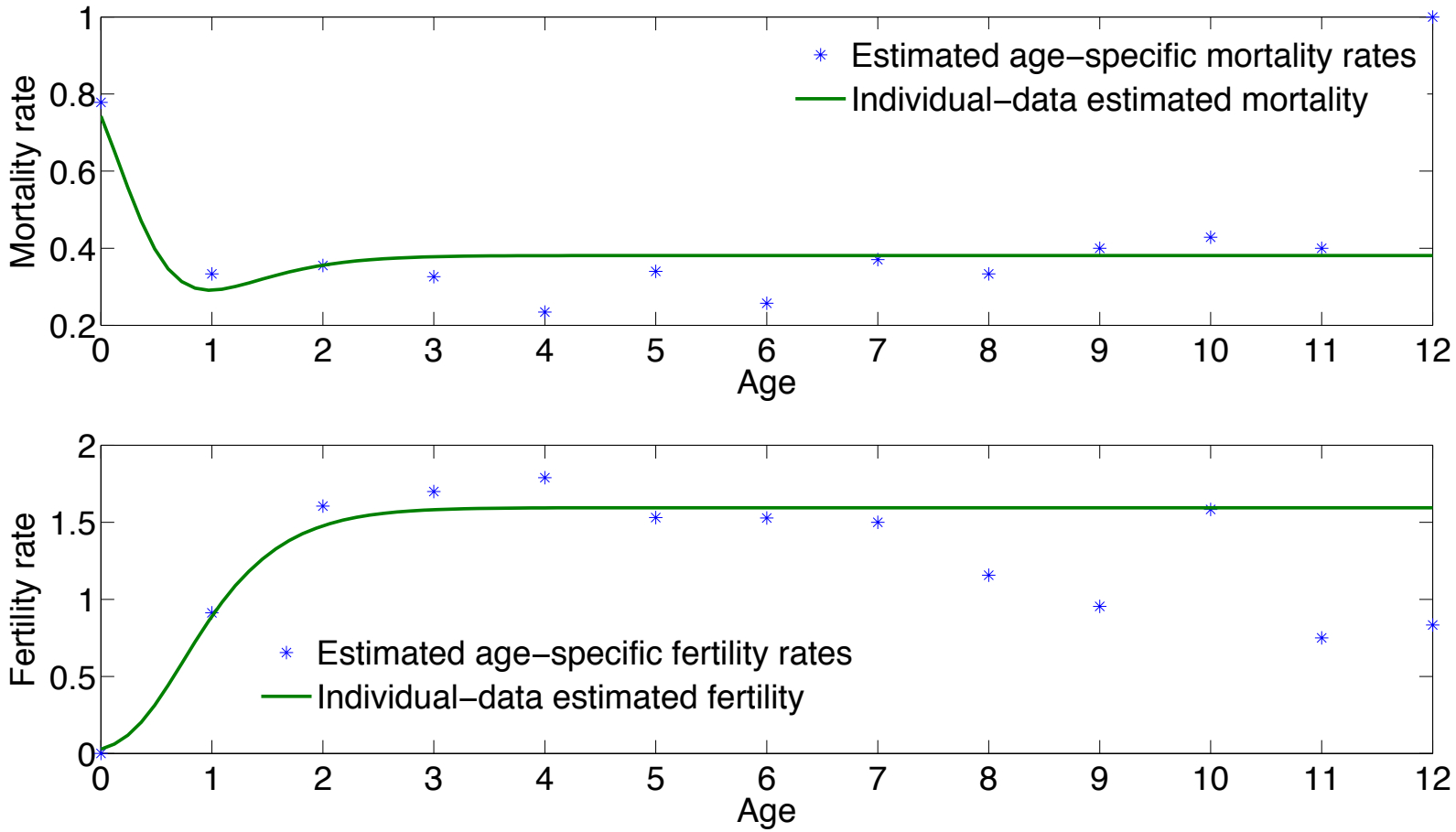
Black robin population : Parameter estimation results (I)

Intensive management period 1980-89



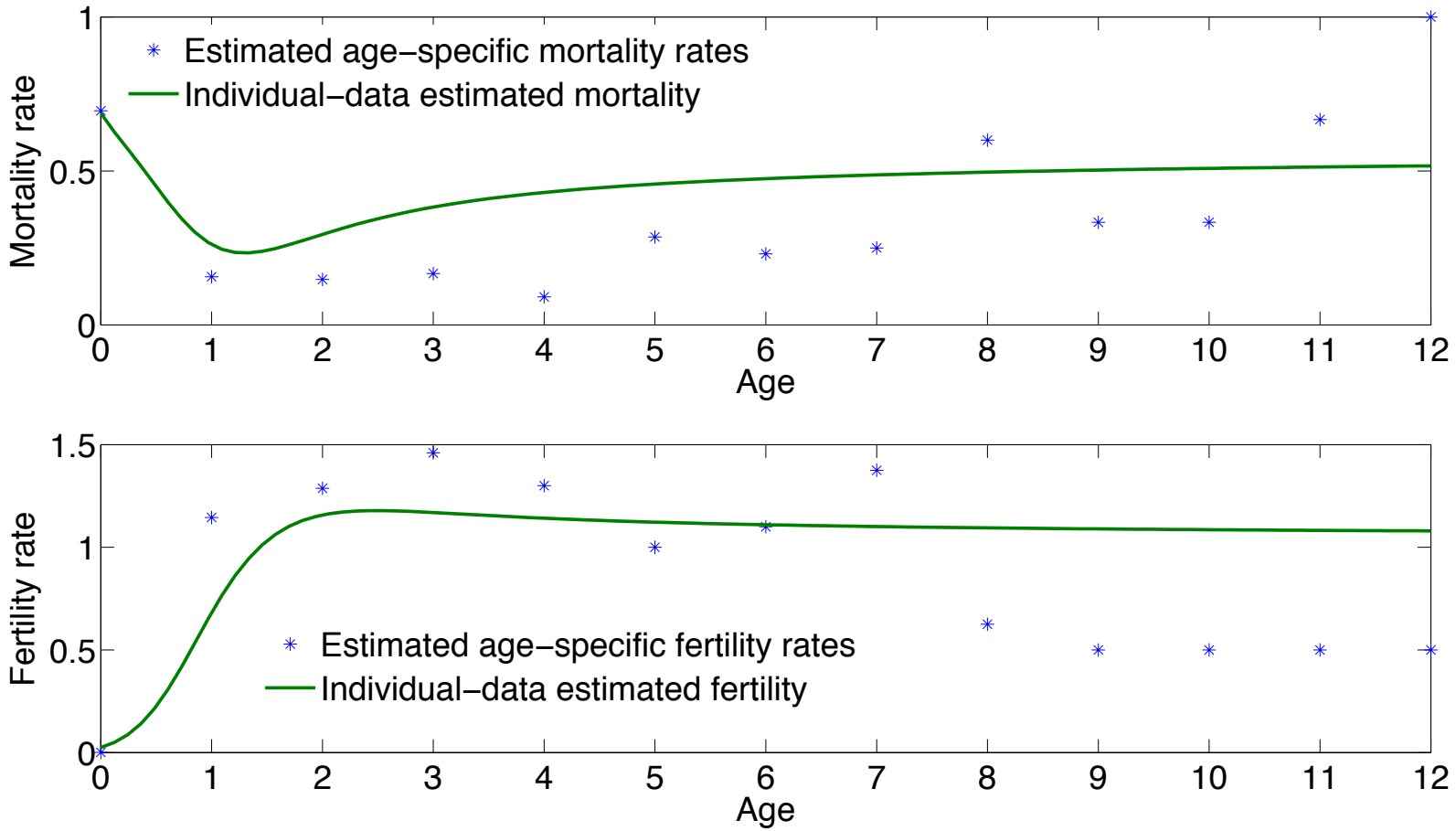
Black robin population : Parameter estimation results (II)

Close monitoring period 1990-2001

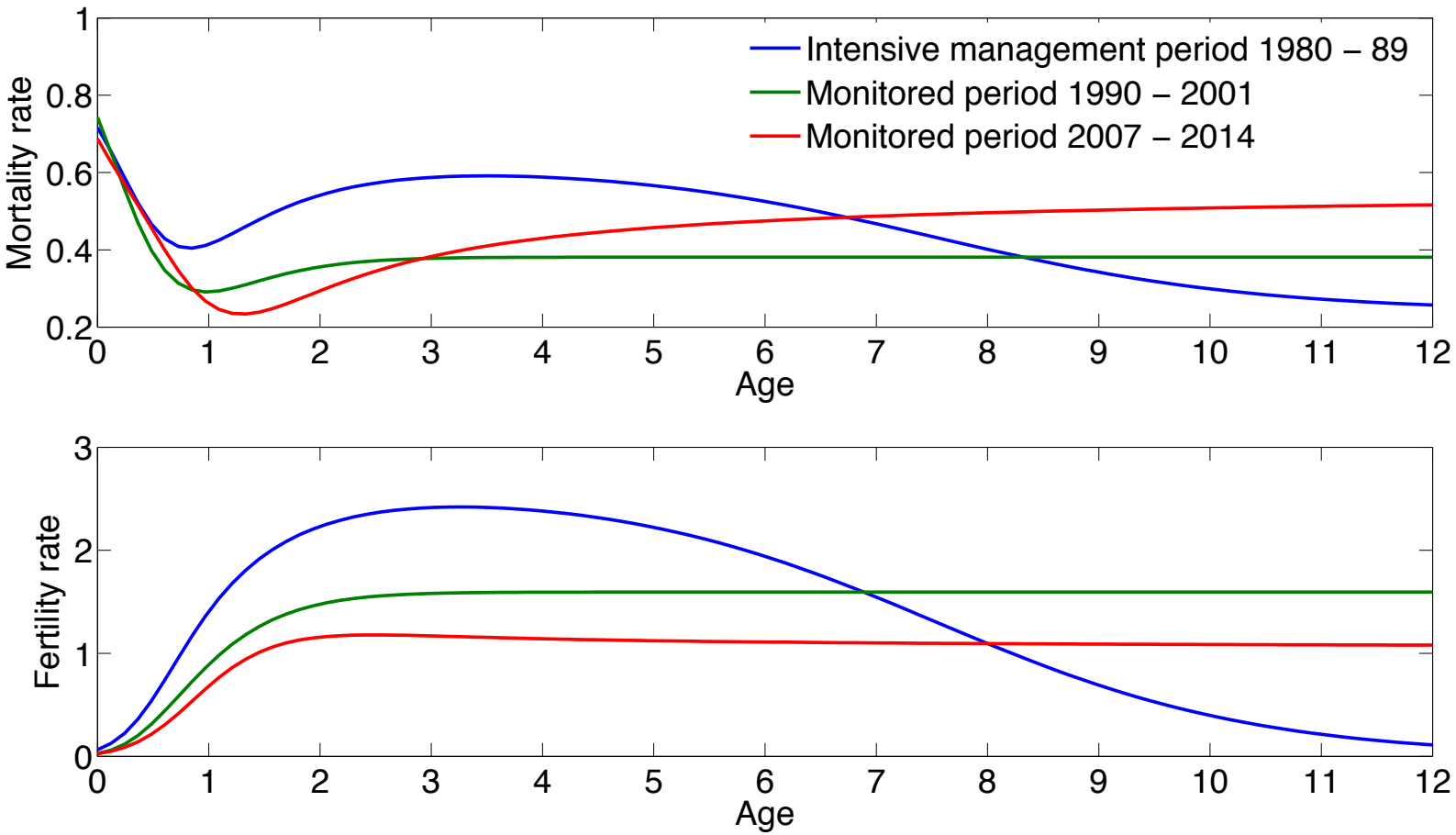


Black robin population : Parameter estimation results (III)

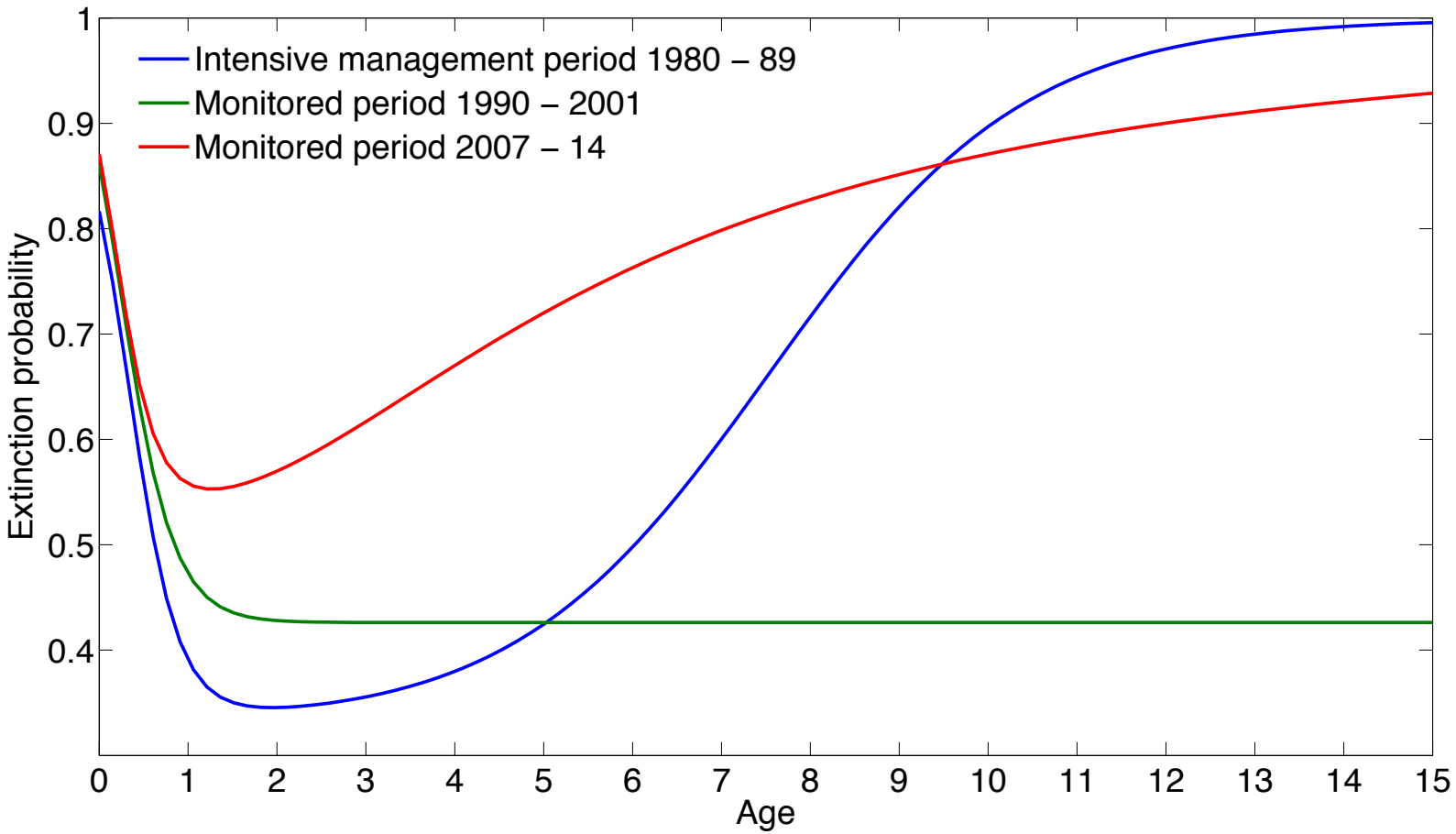
Monitoring period 2007-2014



Black robin population : Age-specific rates comparison



Black robin population : Extinction probability



Thank you for your attention !