



The optimal momentum of population growth and decline[☆]

Gustav Feichtinger^{a,b}, Stefan Wrzaczek^{c,a,*}

^a Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/OeAW, University of Vienna), Austria

^b Institute of Statistics and Mathematical Methods in Economics (Research unit VADOR), Vienna University of Technology, Vienna, Austria

^c International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, Laxenburg, Austria

ARTICLE INFO

Keywords:

Population momentum
Age-structured optimal control
Extended Maximum Principle
Adaptation of net reproduction rate

ABSTRACT

About 50 years ago, Keyfitz (1971) asked how much further a growing human population would increase if its fertility rate were immediately to be reduced to replacement level and remain there forever. The reason for demographic momentum is an age-structure inertia due to relatively many potential parents because of past high fertility. Although nobody expects such a miraculous reduction in reproductive behavior, a gradual decline in fertility in rapidly growing populations seems inevitable. As any delay in fertility decline to a stationary level leads to an increase in the momentum, it makes sense to think about the timing and the quantum of the reduction in reproduction. More specifically, we consider an intertemporal trade-off between costly pro- and anti-natalistic measures and the demographic momentum at the end of the planning period. This paper uses the McKendrick–von Foerster partial differential equation of age-structured population dynamics to study a sketched problem in a distributed parameter control framework. Among the results obtained by applying an appropriate extension of Pontryagin's Maximum Principle are the following: (i) monotony of adaptation efforts to net reproduction rate and convex decrease/concave increase (if initial net reproduction rate exceeds 1/is below 1); and (ii) oscillating efforts and reproduction rate if, additionally, the size of the total population does not deviate from a fixed level.

1. Introduction

In the decades after World War II, politicians in some developing countries refused to accept birth control, fearing that it would stop their populations from increasing. At the 1974 Bucharest World Population Conference, the official delegate from Brazil argued strictly against zero population growth given the ample, still uninhabited spaces in that country. When confronted by demographers in the audience with the huge momentum of population growth – namely that Brazil's population would further increase by about two-thirds even if by some miracle fertility decreased to bare replacement level – the politician refused to listen.

Revisiting a discussion of French demographers around the middle of 20th century (P. Vincent, J. Bourgeois-Pichat), Nathan Keyfitz, a path-breaking population mathematician, was able to derive one of the most discussed results in the whole of population dynamics: the momentum formula. According to Keyfitz, the demographic momentum is the amount of further population growth that will occur if the net reproduction rate (NRR) instantaneously shrinks from a value greater than 1 to bare replacement level and stays there forever. The essential feature is that although stationarity conditions are fulfilled, zero population

growth is reached only after a while, as the population will continue to have a young age composition for the not too distant future. The Keyfitz formula came in the 1970's just in time to prove that the pro-natalistic population policy outlined above is simply wrong. Although Keyfitz's result is valid for arbitrary initial populations, something of which he was well aware, he assumed in his seminal paper (Keyfitz, 1971; see also Keyfitz and Caswell, 2005, section 6.6) a stable age-structure with growth rate r and crude birth rate b .

Denoting by e_0 the life expectancy (mortality assumed to be constant), by μ the mean age of childbearing in the stationary population, and by R_0 the net reproduction rate (NRR) of the initial population, the population momentum for an immediate drop in fertility to replacement level, that is, to an NRR of 1 is given by:

$$M = \frac{e_0 b R_0 - 1}{\mu r R_0}. \quad (1)$$

Note that the momentum expresses the quotient of the total stationary population divided by the stock of the initial population. Keyfitz's calculations typically deliver values of M between 1.6 and 1.7 for several developing countries (among them Brazil).

[☆] The authors thank Tomas Sobotka, Andreas Novak and two anonymous reviewers for valuable comments and suggestions.

* Corresponding author at: International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, Laxenburg, Austria.

E-mail addresses: gustav@eos.tuwien.ac.at (G. Feichtinger), wrzaczek@iiasa.ac.at (S. Wrzaczek).

Note further that (1) also remains valid for negative population growth. In that case the stable growth rate is below zero, corresponding to an $NRR = e^{\mu}$ smaller than 1. A decreasing population shrinks further for a while until the intended stationary level is reached. The proof of (1) uses the reproductive value of a woman at a certain age and measures the expected number of girl-children that woman will have for the rest of her reproductive life.

As already noted, the result of the momentum also remains valid for general age distributions. In Preston (1997) (p. 162) the general formula for the population momentum is given as:

$$M = M(t) = \int_0^{\beta} \frac{c(t, a)}{c_s(a)} w(a) da, \tag{2}$$

where $c(t, a)$ denotes the proportionate age distribution of the population at time t at replacement fertility level. The age distribution in the denominator of (2) is the stationary one, which emerges after the replacement fertility level has been in place for many years, namely, $c_s(a) = l(a)$ ($l(a)$ denotes the survival probability until age a). The third quantity in (2), $w(a)$, is given as

$$w(a) = \frac{l(a)v(a)}{\mu}, \tag{3}$$

where

$$v(a) = \frac{\int_a^{\beta} l(x)m^*(x) dx}{l(a)} \tag{4}$$

denotes the reproductive value (see Fisher, 1930) of an a -year old woman as explained above, with replacement fertility rate $m^*(x)$. α and β denote the minimal and maximal age of childbearing, respectively.

To simplify the following analysis, we follow the approach of Coale (1972) assuming a unique single age at birth denoted by μ . Starting with the time at which replacement-level fertility is imposed, each female below age μ expects one birth (none above μ). Thus, we have

$$\begin{aligned} v(a) &= 1 && \text{for } a \leq \mu \\ &= 0 && \text{for } a > \mu. \end{aligned} \tag{5}$$

Hence, from (2) we obtain

$$M(t) = \frac{e_0}{\mu} \int_0^{\mu} \frac{c(t, a)}{l(a)} da. \tag{6}$$

After Keyfitz’s path-breaking paper on the population momentum (Keyfitz, 1971), several authors relaxed the assumption of an immediate drop in fertility to replacement level. Schoen and Kim (1998) assumed that the growth rate of births linearly declines to zero over a specified time interval, and found an expression for the resulting population momentum. Later, Goldstein (2002) also studied the momentum for a gradual demographic transition to replacement fertility. By specifying a linear frontier in the Lexis surface, he derived a simple formula for the ultimate size of the population (for other relevant work, we refer to Li and Tuljapurkar, 1999 and Li and Tuljapurkar, 2000). Feichtinger et al. (2023) studied the development of the momentum in a pseudostable context.

Let us mention an inherent dichotomy in the social sciences referring to the *descriptive* and *normative* approach of a discipline. While sociology and demography mainly be seen as descriptive/explanatory, one of the core paradigms of economics is an agent maximizing his/her utility (or minimizing costs) faced with scarce resources.

A major purpose of the present paper is to illustrate the issue of optimization in population sciences. This is relevant not only in population policy, where normative aspects play a central role, but also in other areas of demography (see e.g., Baudisch, 2008; Leon, 1976; Goodman, 1982; Schaffer, 1983; Robson and Kaplan, 2003; Chu and Lee, 2006; Robson and Kaplan, 2007; Chu et al., 2008; Dawid et al., 2009, and Kaplan and Robson, 2009). Consider a government faced with a rapidly growing population, that is, one with an NRR substantially greater than one. If a central planner intends to reach

zero growth of a *closed* population, then fertility reduction is the single instrument to reach this goal.

Fertility reduction happens not instantaneously, but gradually over time. A core question, therefore, posed (and answered at least partially) by the following analysis is how the central planner should distribute their braking efforts over a given planning period. Is it better to start with a slow decline and to increase it successively, or should it be the other way round? And in what sense can *better* be understood? Let us for the moment avoid the important question as to whether the politician responsible for population has possible instruments for reducing fertility and whether those instruments are *efficient* enough to fulfill the politician’s targets.¹ The question of ‘better’ leads us to the normative approach, that is, to an optimization, where the planner/decision maker has to strike a balance between birth control, which is clearly costly, and the target of reaching zero population growth (and also avoiding fluctuations in the size of the population). In the following section we thus propose a one-sex age-structured intertemporal decision model to derive qualitative results on the shape of the birth control path. In a nutshell, the goal of our research is to gain insights into the solution structure in an *analytic* way. One of our results, for instance, is that it is optimal to apply anti- and pronatalistic planning most intensively at the beginning of the planning period, followed by a gradual relaxation.

Let us finally stress the fact that we are primarily interested in the qualitative shape of the optimal solution and that the analysis is therefore *highly stylized* in the sense that it is not related to any special data sets, that is, with a non-validated model.

The paper is organized as follows. Section 2 introduces a model, in which a decision maker tries to drive the population momentum close to 1 in a cost-efficient way over a given planning period. The analysis and analytic properties are presented in Section 3, followed by numerical results and a discussion in Section 4. An extension of the model can be found in Section 5, and the paper concludes with Section 6.

2. The model

In the following we consider development of the population as a dynamical system. The reproduction and mortality processes are responsible for the population structure (i.e., what the population pyramid looks like) and their increase or decrease (in absolute numbers). Mathematically, this is defined by the well-known McKendrick–von Foerster equation (see Keyfitz and Keyfitz, 1997) which reads

$$\begin{aligned} P_t(t, a) + P_a(t, a) &= -p(t, a)P(t, a), & P(0, a) &= P_0(a), \\ B(t) &:= P(t, 0) = \int_0^{\omega} v(t, a)P(t, a) da, \end{aligned} \tag{7}$$

where $P(t, a)$ denotes the number of a -year old individuals living at time t (i.e., in mathematical terms the population density at t) which for the sake of simplicity is considered in a one-sex (i.e., female) population.² $p(t, a)$ denotes the age-specific mortality rate at t . The condition $\int_0^{\omega} p(t, a) da = +\infty$ (for all t) guarantees ω to be the maximal attainable age (see Anița, 2000). $P(0, a)$ denotes the distribution of the initial population, exogenously given by $P_0(a)$. $B(t)$ (as already used in (7)) denotes the number of newborn girls at t with $v(t, a)$ as the fertility rate of an a -year old woman at t .

At this stage we make the following two crucial simplifications to the model that ease the mathematical considerations of the intended derivation of the optimal path of the population momentum:

- Mortality: realized only at the maximal age, i.e., $p(t, a) = \delta(a - \omega)$ for $t \geq 0$

¹ For a brief discussion on that we refer to Section 2.

² Note, that $P_t(t, a) + P_a(t, a)$ denotes the directional derivative of $P(t, a)$ in direction (1, 1) (since time and age evolves at the same pace).

- Reproduction: concentrated at a unique age μ , i.e., $v(t, a) = R(t)\delta(a - \mu)$ for $t \geq 0$

where $\delta(x)$ denotes the Dirac delta function and $R(t)$ the female fertility rate at t .

The decision maker is able to influence the NRR by effort $k(t)$ denoting the relative change of the NRR (i.e., offering education to women, making efforts toward gender equality³ for a dynamic modeling approach to the transition of the traditional family structure to an egalitarian one). For a discussion on the dependence of fertility on policy measures, see Bergsvik et al. (2021), Luci-Greulich and Thévenon (2013), Sobotka et al. (2019), Gauthier (2007), and Björklund (2006).

Remark. As Tomas Sobotka stresses in a personal communication, the issue of influencing fertility is somewhat controversial and is being misused by authoritarian governments from Iran to Russia. But this controversy aside, the problem also lies in the fact that policies have an impact that is context-specific, and dependent on other institutions: the labor market, welfare system, economic development, education, gender equality, and also culture and norms. The impact also depends on the initial fertility level, can be short-term or long-term, mostly due to the tempo effect, or can also impact cohort fertility.

Negative $k(t)$ decreases reproduction $R(t)$. Positive $k(t)$, on the other hand, (i.e., supporting women, establishing childcare facilities) increases reproduction. For this reason we model $R(t)$ by means of an ordinary differential equation⁴

$$\dot{R}(t) = k(t)R(t), \quad R(0) = R_0, \tag{8}$$

where R_0 denotes the initial NRR. Both, negative and positive efforts are equally costly with decreasing marginal efficiency reflected by convex (i.e., in our case quadratic) costs $Ck^2(t)$. C is a model parameter, that measures the cost of decreasing or increasing $R(t)$ (for simplicity) by the same extent. The decision maker has two objectives over a fixed time horizon T . The first intention is to end up with a NRR of 1. From the literature it is known (as discussed in Section 1) that a change in reproduction affects the population structure with delay. This is expressed by the population momentum, which equals 1 for a stationary population. Therefore, the second intention for the decision maker is to opt for a population momentum of 1 at T .

In the following we assume the first objective ($R(T) = 1$) to be a *hard* constraint, that is, the decision maker has to drive reproduction to 1 at T , no matter what the costs may be. The second constraint ($M(T) = 1$) will be considered *soft*, which means that the decision maker is interested in meeting the constraint as closely as possible (relative to an evaluation of the mismatch). Hence, the decision maker tries minimizing

$$\mathcal{V}(R_0) := \int_0^T Ck^2(t) dt + G(M(T) - 1)^2, \tag{9}$$

where the parameter G values the mismatch of meeting a population momentum of 1 at the end of the time horizon. In mathematical terms $\mathcal{V}(R_0)$ denotes the objective function, $Ck^2(t)$ the objective functional, and $G(M(T) - 1)^2$ the salvage value function.

Together with the differential Eqs. (7) and (8) as constraints, the decision maker faces the following finite time optimal control problem

$$\min_{k(t)} \mathcal{V}(R_0) \tag{10a}$$

$$\text{s.t. } P_t(t, a) + P_a(t, a) = 0, \quad P(0, a) = P_0(a) \tag{10b}$$

$$B(t) = P(t, 0) = P(t, \mu)R(t) \tag{10c}$$

³ See e.g., Feichtinger et al. (2017).

⁴ In contrast, the McKendrick–von Foerster Eq. (7) is a partial differential equation.

Table 1

Control and state variables.	
Independent variables:	
time	t
age	a
Control variable:	
efforts to adapt the NRR	$k(t)$
State variables:	
Population density of age a at t	$P(t, a)$
NRR at t	$R(t)$
Total population at t	$N(t)$
State variables (aggregated):	
Births at t	$B(t)$
Deaths at t	$D(t)$
Population momentum at t	$M(t)$

$$\dot{R}(t) = k(t)R(t), \quad R(0) = R_0, R(T) = 1 \tag{10d}$$

$$M(t) = \frac{\omega}{\mu} \int_0^\mu c(t, a) da \tag{10e}$$

$$c(t, a) = \frac{P(t, a)}{\int_0^\omega P(t, a) da}. \tag{10f}$$

Expression (10e) for the population momentum can be obtained from (6) as presented in Preston et al. (2001) (see also Feichtinger et al., 2023). Through a reformulation of the Eqs. (10e) and (10f), this problem can be considered as a non-standard form of an age-structured optimal control model. First, we observe that the denominator in (10f) denotes the total population $N(t)$, which increases by births $B(t)$ and decreases by deaths $D(t)$ and thus can be considered as a simple ordinary differential equation

$$\begin{aligned} \dot{N}(t) &= B(t) - D(t), \quad N(0) = N_0 \\ D(t) &= P(t, \omega). \end{aligned} \tag{11}$$

Here, $N_0 := \int_0^\omega P(0, a) da$. Secondly, the $c(t, a)$ can be inserted into the expression for $M(t)$ to obtain an expression that depends only on state variables and can be used for the salvage value function in (9).

Using this for (10), the transformed model reads

$$\min_{k(t)} \mathcal{V}(R_0) \tag{12a}$$

$$\text{s.t. } P_t + P_a = 0, \quad P(0, a) = P_0(a) \tag{12b}$$

$$P(t, 0) = B(t) \tag{12c}$$

$$\dot{R}(t) = k(t)R(t), \quad R(0) = R_0, R(T) = 1 \tag{12d}$$

$$\dot{N}(t) = B(t) - D(t), \quad N(0) = N_0 \tag{12e}$$

$$B(t) = \int_0^\omega \delta(a - \mu)P(t, a)R(t) da \tag{12f}$$

$$D(t) = \int_0^\omega \delta(a - \omega)P(t, a) da \tag{12g}$$

where we used the Dirac delta function to allow the application of a suitable Maximum Principle. From a control theoretic point of view, $k(t)$ denotes the control variable, $P(t, a)$, $R(t)$, $N(t)$ the state variables, and $B(t)$ and $D(t)$ the (aggregated) state variables. The population momentum, denoted by $M(t)$, which is at the center of the problem, is a function of state variables. Table 1 summarizes the variables and the notation of the model at a glance.

The full problem (12) is a non-standard age-structured optimal control problem. The standard Maximum Principle for these problems allows for age-structured state variables, as well as for aggregated states (see Brokate, 1985), and additionally for aggregated state variables thereby capturing the interactions between age-groups (see Feichtinger et al., 2003). Applications can be found, for example, in demography and population economics (see e.g., Arthur and McNicoll, 1977; Feichtinger and Veliov, 2007), in economics (see e.g., Feichtinger et al., 2006; Prskawetz and Veliov, 2007, or Wrzaczek et al., 2014), in health economics (see e.g., Kuhn et al., 2010, 2011; Frankovic

et al., 2020a, or Frankovic et al., 2020b), and in epidemiology (see e.g., Hethcote, 1988). The above model, however, also includes non-distributed state variables (e.g., evolving over time only). Thus the standard Maximum Principle must be extended by applying needle variations to the Lagrangian formulation of the problem. A sketch of the proof (for a general formulation of the above problem) can be found in Appendix. An analogous Maximum Principle has already been applied for the solution of multi-stage optimal control problems with random switching time. These problems can be transformed to age-structured optimal control problems including age-structured and non-distributed (concentrated) state variables,⁵ which offers a number of advantages compared to the standard backwards approach.

3. Optimality conditions and analytic properties

For the solution of the transformed problem (12) we apply the extended Maximum Principle⁶ (see Appendix) and define the Hamiltonian⁷ as (skipping t and a for notation reasons if they are not of particular importance)

$$H = -\frac{1}{\omega} Ck^2 + \underbrace{\xi \cdot 0}_{=0} + \frac{1}{\omega} \lambda_R kR + \frac{1}{\omega} \lambda_N (B - D) + \eta_B \delta_\mu PR + \eta_D \delta_\omega P, \quad (13)$$

where $\xi(t, a)$ denotes the age-structured adjoint variable of the population for age a at t . $\lambda_R(t)$ and $\lambda_N(t)$ denote the (non-distributed) adjoint variables for reproduction and total population at t respectively, and $\eta_B(t)$ and $\eta_D(t)$ the adjoint variables for (the aggregated state variables) births and deaths at t . As is standard in optimal control theory, adjoint variables can be interpreted as dynamic shadow prices of the corresponding optimal control problem (i.e., denoting the change of the objective function if the corresponding state variable is increased marginally at a and/or t). Moreover, further analysis of the adjoint variable (i.e., backwards solution of the corresponding differential equation, see (15) in this model) shows that adjoint variables are future-oriented (i.e., future effects of a marginal increase in the state variables are aggregated). See, for example, Kuhn et al. (2010) or Wrzaczek et al. (2010) who exploit this effect in the context of population dynamics.

The Hamiltonian directly implies the (necessary) first order condition for $k(t)$,

$$H_k \omega = -2Ck + \lambda_R R = 0 \quad \Rightarrow \quad k = \frac{\lambda_R R}{2C}. \quad (14)$$

This implies that efforts to adapt the NRR depend not only on (i) the current reproduction value (i.e., the state value $R(t)$), but also on (ii) the future effect (i.e., the adjoint value $\lambda_R(t)$). Moreover, it is crucial to mention that the sign of the efforts is only determined by the future effects (i.e., by the sign of the adjoint variable) and does not depend on whether reproduction lies above or below the target value of 1 (which has to be reached at the end of the planning horizon T).

The adjoint equations and transversality conditions complete the necessary conditions of the Maximum Principle. We obtain

$$\begin{aligned} \xi_t + \xi_a &= -\eta_B \delta_\mu R - \eta_D \delta_\omega \\ \dot{\lambda}_R &= -\lambda_R k - \int_0^\omega \eta_B \delta_\mu P da \\ \dot{\lambda}_N &= 0 \\ \eta_B &= \lambda_N + \xi(t, 0) \end{aligned}$$

⁵ See Wrzaczek et al. (2020) for the proof and a discussions and Wrzaczek (2021), Buratto et al. (2022) for applications.

⁶ Note that the existence of an optimal solution as well as sufficiency conditions are not considered within this contribution.

⁷ The $\frac{1}{\omega}$ term enters for concentrated state variables and parts of the objective function, as the Hamiltonian corresponds to both age and time. See also the formulation of the Lagrangian (38) and the Hamiltonian (41) in the proof of the extended Maximum Principle (Appendix).

$$\eta_D = -\lambda_N \quad (15)$$

for the adjoint equations, and

$$\begin{aligned} \xi(t, \omega) &= 0 \\ \xi(T, a) &= -2G \left(\frac{\omega \int_0^\omega \kappa(a) P(T, a) da}{N(T)} - 1 \right) \frac{\omega \kappa(a)}{\mu N(T)} \\ \dot{\lambda}_N(T) &= 2G (M(T) - 1) \frac{M(T)}{N(T)} \end{aligned} \quad (16)$$

for the transversality conditions. κ (1 below μ , 0 above μ) is a continuous function that allows the transformation of the population momentum (10e) into a mathematically suitable form that can be dealt with using the Maximum Principle. As (12) includes initial and end conditions for $R(t)$, there is no transversality condition for the corresponding adjoint variable $\lambda_R(T)$. To each condition of (16) it is possible to give a comprehensive interpretation. $\xi(t, \omega)$ denotes the marginal value of an increase of ω -year old individuals at t . The value is certainly zero, as these people will exit the model (i.e., die) immediately and do not, moreover, contribute to reproduction. In contrast, $\xi(T, a)$ evaluates the reproductive effect of an increase of a -year old individuals at the end of the planning horizon. This effect is active for individuals not older than the reproduction age μ , as these people contribute directly to the population momentum. Older individuals do not matter for the population momentum, which implies that $\xi(T, a)$ is nil for $a > \mu$ (mathematically expressed by the function $\kappa(a)$, which is 1 for $a \leq \mu$ and zero for $a > \mu$). $\lambda_N(t)$ covers the (marginal) effect of the (current) total population to the deviation of the population momentum at T and is therefore constant over time (see (15)). This in turn enters the adjoint of the (age-structured) population before T via the other adjoints for births and deaths.

Exploiting (16) we arrive at a simpler representation for $\xi(t, a)$ and $\lambda_R(t)$.

$$\begin{aligned} \xi_t + \xi_a &= -(\bar{\lambda}_N + \xi(t, 0)) \delta_\mu R + \bar{\lambda}_N \delta_\omega \\ &= \begin{cases} -(\bar{\lambda}_N + \xi(t, 0)) \delta_\mu R & \text{for } a = \mu \\ \bar{\lambda}_N \delta_\omega & \text{for } a = \omega \\ 0 & \text{else} \end{cases} \\ \dot{\lambda}_R &= -\lambda_R k - (\bar{\lambda}_N + \xi(t, 0)) P(t, \mu), \end{aligned} \quad (17)$$

where

$$\bar{\lambda}_N := 2G (M(T) - 1) \frac{M(T)}{N(T)}, \quad (18)$$

is a constant. $\lambda_N(t)$, $\eta_B(t)$, and $\eta_D(t)$ can be solved easily for all t ,

$$\begin{aligned} \lambda_N(t) &= \bar{\lambda}_N \\ \eta_B(t) &= \bar{\lambda}_N + \xi(t, 0) \\ \eta_D(t) &= -\bar{\lambda}_N. \end{aligned} \quad (19)$$

Interesting in that respect is the equality $\lambda_N(t) = \bar{\lambda}_N$, which means that an additional individual of the total population increases the objective function (i.e., decreases costs) if the momentum is too high at T , i.e.,

$$\begin{aligned} M(T) > 1 &\Leftrightarrow \text{increase of } N(t) \text{ reduces costs} \\ &\quad \text{(momentum too large): } \lambda_N(t) > 0 \\ M(T) < 1 &\Leftrightarrow \text{increase of } N(t) \text{ increases costs} \\ &\quad \text{(momentum too low): } \lambda_N(t) < 0 \end{aligned}$$

It is important to understand that $\lambda_N(t)$ measures the effect assuming that $P(t, a)$ hypothetically remains at the same level. This principle is well understood in optimal control theory and is the reason why $\xi(t, 0)$ additionally enters $\eta_B(t)$, which is discussed in the course of Lemma 1.

Using the adjoint variables and the interpretation of the transversality conditions gives an important insight into the optimal value of the reproduction adaptation efforts from a demographic point of view. This is summarized in the following Lemma and discussed after the proof.

Lemma 1. Along the optimal transitional path the optimal efforts for adapting the NRR changes depend negatively on the marginal deviation of the population momentum at T and on the generalized reproductive value at t of one expected progeny (normalized by marginal costs of adaptation efforts).

Proof. Differentiating the (necessary) first order condition (14) with respect to time implies the dynamics of $k(t)$, also known as Euler equation, that is:

$$\dot{k}(t) = -\frac{R(t)P(t, \mu)}{2C} (\bar{\lambda}_N + \xi(t, 0)). \tag{20}$$

As $R(t)P(t, \mu)$ is strictly positive, the sign of the change of k depends on the two terms $\bar{\lambda}_N$ and $\xi(t, 0)$ (positive sign means a decrease, and vice versa). ■

To understand this interesting result we first recall that $R(t)P(t, \mu)$ denotes the expected number of progeny at t (as fertility is concentrated at age μ) and that $2C$ are marginal costs of the adaptation efforts. With this in mind it is clear that the sign of \dot{k} is determined by the two terms in brackets, which interestingly focus on different points in time. First, $\bar{\lambda}_N$ captures that the expected progeny influences the population momentum at the end of the time horizon. The effect is not evaluated at t , as the corresponding dynamics is nil, which means that the effect of an additional individual at t shifts forwards to the end of the time horizon. Secondly, $\xi(t, 0)$ denotes a generalization of Fisher’s reproductive value (see Fisher, 1930). As discussed and proven in Wrzaczek et al. (2010) and Feichtinger et al. (2011) (see also the discussion in Kuhn et al., 2010), the generalization appears as part of the shadow price of the population within an age-structured optimal control model if (and only if) births are endogenously determined in the model. The value is a generalization of Fisher’s value, as the pure reproductive value (well-known in demography) is weighted by the effects that one additional progeny has on the objective function (of the corresponding problem). Before we give further insights into a decomposition of this generalized value in our model, it is important and interesting to observe that the evolution of adaptation efforts depends on the direct effect on population momentum at the end of the planning horizon and on the indirect effect of expected progeny.⁸

A complete decomposition of the generalized reproductive value reveals the different forces of the model that are active for different ages of different cohorts.

Proposition 1. The generalized reproductive value can be decomposed into three regions illustrated in Fig. 1: cohorts that have already died before T (region 1), cohorts that are alive at T but are older than the reproductive age (region 2), and cohorts that are alive and younger than, or exactly at, the reproductive age (region 3). For these cohorts $\xi(t, 0)$ can be reduced to

$$t + \omega < T \text{ (region 1)} : \xi_1(t, 0) = \bar{\lambda}_N R(t + \mu) - \bar{\lambda}_N + \xi(t + \mu, 0)R(t + \mu) \tag{21a}$$

$$t + \mu < T \leq t + \omega \text{ (region 2)} : \xi_2(t, 0) = \bar{\lambda}_N R(t + \mu) + \xi(t + \mu, 0)R(t + \mu) \tag{21b}$$

$$t + \mu = T \text{ (region 2/3)} : \xi_{2/3}(t, 0) = \bar{\lambda}_N R(T) + \xi(T, 0)R(T) - 2G(M(T) - 1) \frac{\omega}{\mu} \frac{1}{N(T)} \tag{21c}$$

$$t + \mu > T \text{ (region 3)} : \xi_3(t, 0) = -2G(M(T) - 1) \frac{\omega}{\mu} \frac{1}{N(T)}. \tag{21d}$$

Proof. From solving the adjoint equation for $\xi(t, a)$ backwards we obtain the following general expression

$$\xi(t, a) = \int_a^\omega ((\bar{\lambda}_N + \xi(t - a + s, 0)) \delta_\mu R(t - a + s) - \bar{\lambda}_N \delta_\omega) ds + \xi(t - a + \omega, \omega), \tag{22}$$

⁸ The reason why (20) gets rid of $\lambda_R(t)$ is a backward solution of the adjoint equation for $\lambda_R(t)$.

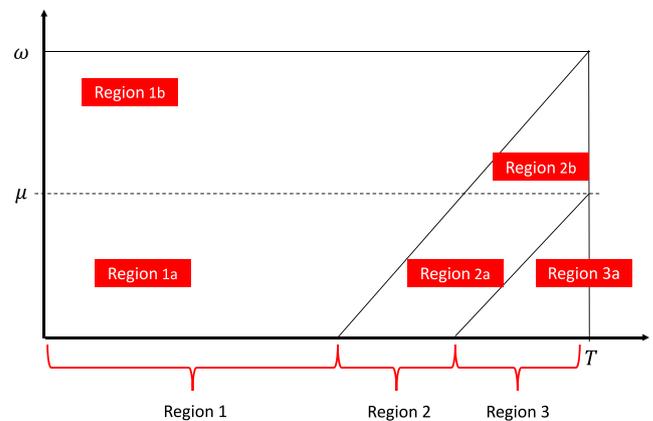


Fig. 1. Lexis diagram with time along the horizontal axis and age along the vertical axis. Individuals born within region 1 die before T . Individuals born within region 2 reproduce but do not die before T . Individuals born within region 3 neither reproduce nor die before T .

for a cohort that dies before T . For cohorts that die after T the expression is slightly adapted by changing the upper limit of the integral.

In the following we evaluate this expression for $a = 0$ explicitly for different cohorts. As the decision makers’ objective is to drive the population momentum as close to 1 as possible at t , and as the population momentum is driven by the population size and the relative size of the population that has not already reproduced, we consider cohorts (i) that have already died before T (region 1 in Fig. 1), (ii) cohorts that are alive at T but have already reproduced (region 2 in Fig. 1), and (iii) cohorts that are alive at T and have not yet reproduced (region 3 in Fig. 1). The cohort born at $T - \mu$ is special in the sense that it reproduces exactly at the end of the planning horizon (lying on the transition from region 2 to 3). As this cohort is isolated in the sense that it is only a single cohort (marginally deviating from $T - \mu$ means belonging to region 2 or 3) with zero measure, it is denoted by region 2/3. Moreover, Fig. 1 distinguishes a part ‘a’ and ‘b’ within each region, where ‘a’ and ‘b’ correspond to individuals who have not reproduced or have already reproduced at (t, a) . The value for newborns $\xi(t, 0)$, however, always lies at the lower boundary of the diagram (at the horizontal axis).

- For $\xi_1(t, 0)$: set $a = 0$, $t + \omega < T$ implies $\xi(t + \omega, \omega) = 0$. Using the definition of the Dirac delta function implies (21a).
- For $\xi_2(t, 0)$: set $a = 0$, since $a > \mu$ again $\xi(t + \omega, \omega) = 0$ (recall the definition of the function κ). Using the definition of the Dirac delta function implies (21b), where in contrast to $\xi_1(t, 0)$ the age of ω will not be reached within the planning horizon.
- For $\xi_3(t, 0)$: set $a = 0$, $a \leq \mu$ implies the above expression (21d). As the cohort dies after T the Dirac delta function for death ($\delta(a - \omega)$) is nil. The integral over $\delta(a - \mu)$ equals 1 for cohorts born at $T - \mu$ (region 2/3) and zero otherwise. ■

The result just derived allows us to gain valuable insights into the structure of optimal adaptation rates of fertility. In a first step we are able to show that any policy with an ‘up and down’ of the NRR is suboptimal. Although such a behavior might be expected, we are able to prove it analytically (i.e., in a logical consistent way). As the proof of the following proposition will show, the validity of the following inequality of the demographic momentum is required:

$$M(T) < \frac{\omega}{\mu}, \tag{23}$$

which follows directly from (6). A formal proof can alternatively be found in Feichtinger et al. (2023) (Proposition 1). Although (23) was

shown for rectangular mortality and a unique birth age of mothers, we suspect that it holds true for general mortality and fertility schedules.

Proposition 2. *Assume that an optimal solution for (12) exists and let $k^*(t)$ denote the optimal efforts to adapt the NRR. Then $k^*(t)$ does not change the sign.*

Proof. In an optimal solution the first order condition (14) has to hold and $R(t)$ is trivially strictly positive. Hence, the sign of $k^*(t)$ is determined by $\lambda_R(t)$. Backwards solution of the corresponding adjoint equation (solving (17) backwards reverses the sign of k in the exponential function) yields

$$\lambda_R(t) = \lambda_R(T)e^{\int_t^T k(s') ds'} + \int_t^T e^{\int_t^{s'} k(s'') ds''} (\bar{\lambda}_N + \xi(s', 0)) P(s', \mu) ds', \tag{24}$$

where $\bar{\lambda}_N$ is given by (18) and $\xi(s', 0)$ by Proposition 1.

Due to (21d) the sign of $\xi(t, 0)$ does not change the sign within region 3. For region 2 we consider a cohort that is sufficiently near to region 3, that is, such that the corresponding progeny already belong to region 3 ($(t + \mu) + \mu > T$). Using (21b) and the definition of \bar{N} we obtain

$$\begin{aligned} \xi_2(t, 0) &= (\bar{\lambda}_N + \xi(t + \mu, 0)) R(t + \mu) \\ &= 2G(M(T) - 1) \frac{1}{N(T)} \left(M(T) - \frac{\omega}{\mu} \right) R(t + \mu). \end{aligned} \tag{25}$$

A comparison with $\xi_3(t, 0)$ shows that the sign does not change if $M(T) < \frac{\omega}{\mu}$ (see (23)). Reproducing this argument analogously implies that $\xi(t, 0)$ has the same sign for all t .

Since further $\bar{\lambda}_N - \xi_3(t, 0)$ vanishes in the definition of $\lambda_R(t)$, the sign of $\lambda_R(t)$ does not change for all t due to (23). ■

The result, that the optimal adaptation efforts do not change the sign in any planning interval, can be sharpened such that it always behaves monotonic. Note that such a shape of efficient birth control is both intuitively plausible as well as generally true.

Corollary 1. *Assume that an optimal solution for (12) exists and let $k^*(t)$ denote the optimal efforts to adapt the NRR. Moreover, assume that the mortality at the begin of the planning horizon is rectangular, i.e., $M(0) < \frac{\omega}{\mu}$. Then $k^*(t)$ is strictly monotonic over the entire time horizon according to*

$$\begin{aligned} M(0) > 1 &\implies k^*(t) < 0, \dot{k}^*(t) > 0, \ddot{k}^*(t) < 0 \\ &\text{for } t \in [0, T] : \text{ decreasing momentum} \\ M(0) = 1 &\implies k^*(t) = 0, \dot{k}^*(t) = 0, \ddot{k}^*(t) = 0 \\ &\text{for } t \in [0, T] : \text{ constant momentum} \\ M(0) < 1 &\implies k^*(t) > 0, \dot{k}^*(t) < 0, \ddot{k}^*(t) > 0 \\ &\text{for } t \in [0, T] : \text{ increasing momentum.} \end{aligned} \tag{26}$$

Proof. Consider the last part of the time horizon $[T - \Delta, T]$ and $M(t) > 1$. Δ is sufficiently small such that $M(t) - 1$ does not change the sign. Then Proposition 2 implies that $M(T - \Delta) > 1$ implies $M(T) > 1$ and negative $k^*(t)$ in $t \in [T - \Delta, T]$. Observe, also that for negative $k^*(t)$ the condition $M(0) < \frac{\omega}{\mu}$ is stronger than $M(T) < \frac{\omega}{\mu}$. Applying Proposition 2 and going back in time implies that $k^*(t)$ (for all t) is negative for $M(0) > 1$. Again, the condition on the rectangular mortality at $t = 0$ is stronger than at $t = T$ and therefore sufficient.

To derive the sign of $\dot{k}(t)$ we plug (18) and (21) into the Euler Eq. (20). Evaluation of the four cases of (21) together with the assumption $M(t) > 1$ for $t \in [0, T]$ proves the assertion. For the curvature of $k(t)$ we take the derivative of the Euler Eq. (20) w.r.t. time. By manipulation we obtain the curvature as the product of $\dot{k}(t)$ (which is positive) and a negative term, which finishes the proof.

For $M(T) < 1$ the proofs work analogously. For $M(T) = 1$ the population starts in a stationary situation and no intervention is necessary. ■

Finally, the paths of optimal adaptation rates turn out to be concave/convex for $R(0) \geq 1$. These results are intuitively plausible. A decision maker who is concerned about the shrinking population should start with intensive pro-natalistic measures followed by a gradual relaxation until stationary conditions occur, that is, $R(T) = 1$ are reached at terminal time T .

Note that in all three properties of the efficient fertility adaptation trajectories – that is, unique signs of $k^*(t)$, the sign’s monotony, and most intensive efforts occur at the beginning, followed by a successive decline – and are lost under more general targets, for example, by including a penalty for fluctuations of the population size as considered in Section 5.

4. Numerical results and discussion

To illustrate the optimal paths of the efforts to adapt the reproduction and related trajectories over time, we solve (12) numerically with a gradient based optimization algorithm (for theoretical details see Veliov, 2003). At this stage we are interested in observing and interpreting the qualitative behavior and use artificial parameters. We are not interested in a realistic calibration, which can only be achieved with a considerable model extension at the cost of losing (and this is another reason for not following this) all the analytical conclusions we have drawn in the previous section.

Hence, we assume life up to the age $\omega = 100$ and a reproduction age of $\mu = 30$ years. For the cost parameters we choose $C = 10$ (control costs) and $G = 10^5$ (cost of mismatch of the population momentum at T). For the planning period we use $T = 200$ years.

Numerous numerical runs with different parameters show robustness with respect to changes in C and G . Additional numerical runs for different net reproduction ages are presented in Section 4.1. The time horizon has also been varied. The results are briefly discussed at the end of this section, but not presented here.

We consider four scenarios corresponding to the initial NRR: very high (very young population, $R_0 = 5$), high (young population, $R_0 = 2.5$), low (old population, $R_0 = 0.75$), and very low (very old population, $R_0 = 0.5$). The population is normalized to an initial size of $N_0 = 100$ for all cases with a density ($P_0(a)$) of a stable population corresponding to the (initial) reproduction scenario.

The results are presented in Figs. 2–6, and are discussed later in this paper.

Fig. 2 illustrates the optimal adaptation efforts of reproduction over time and reveals three qualitative properties of the time paths depending on the scenario (i.e., on the initial reproduction value R_0 , which can be greater or smaller than the stationary value 1). For $R_0 > 1$ the optimal efforts $k^*(t)$ are always negative, monotonically increasing and concave, while for $R_0 < 1$ it holds that $k^*(t) > 0$, and gradually decreases in a convex manner. All these properties have also been proven analytically as summarized in Corollary 1. While the first two properties are straightforward, the concavity/convexity is more remarkable. It means that both efficient anti- and pro-natalistic measures should be applied most intensively at the beginning of a planning interval, but gradually relaxed in the period that follows. Mathematically, the reason lies in the effectiveness of the efforts, which becomes higher the larger the reproduction value $R(t)$ deviates from its stationary level 1. If this assumption (multiplicativity of $k(t)$ and $R(t)$ in the (8)) were relaxed, Corollary 1 would hold without the concavity/convexity result.

Fig. 3 shows how $k^*(t)$ affects the NRR over time (as defined by (8)). The convex/concave shape carries over, but is markedly different compared to the efforts: this is due to multiplication. Moreover, it is obvious that the NRR of a (very) young population ($R_0 > 1$) stays above 1 for $t \in [0, T]$ until it reaches 1 at T (forced by the constraint $R(T) = 1$). For a (very) old population the opposite can be observed. This effect also follows from Corollary 1 and the classic Bellman principle.

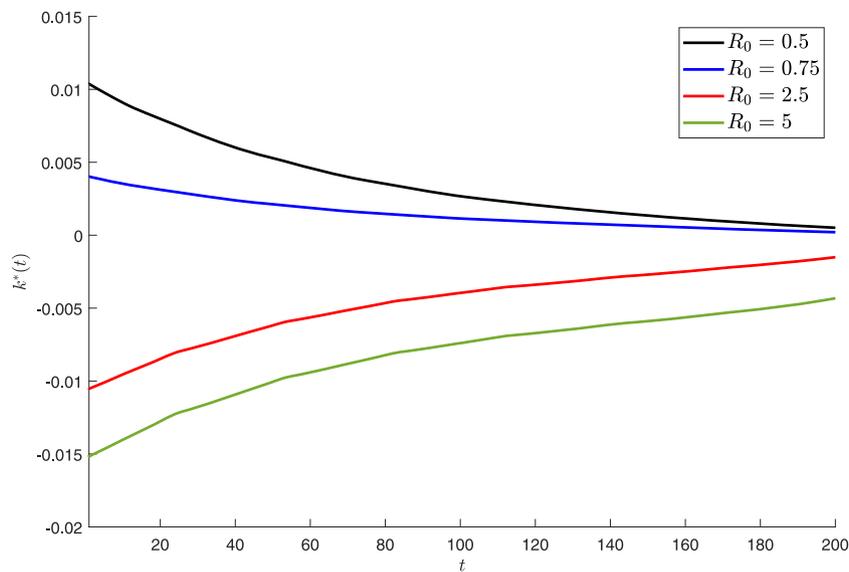


Fig. 2. Optimal adaptation of reproduction $k^*(t)$ over time.

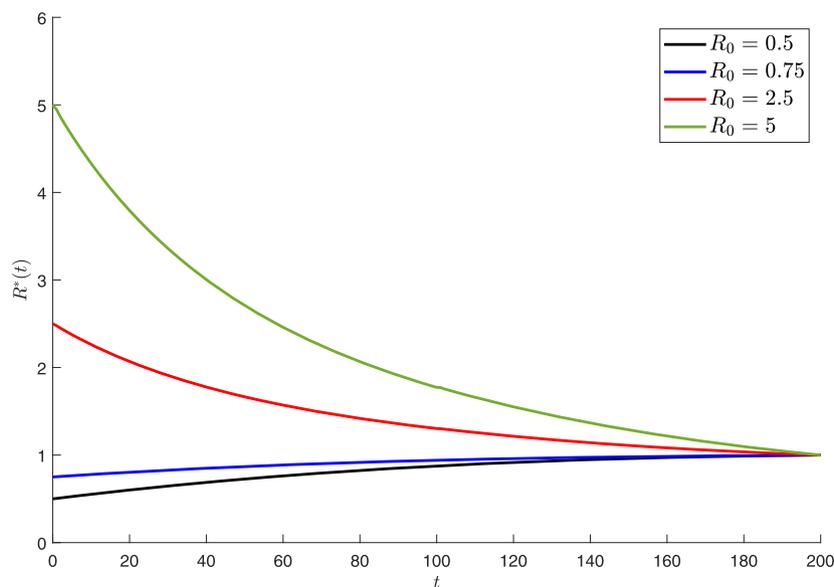


Fig. 3. Reproduction $R^*(t)$ over time.

Although the decision maker aims for a population momentum of 1 at the end of the time horizon, we are interested in the development over time as shown in Fig. 4 (colored lines: $M(t)$ for $t \in [0, T)$, colored bullets: $M(T)$). As for $k^*(t)$ and $R^*(t)$, the population momentum stays above/below 1 for a young/old population throughout the planning period. However, the convex/concave behavior disappears due to the intertwined time-lag effects of the total population and the corresponding age-structure. Moreover, we see that $M(T)$ will match 1 equally. The higher the initial reproduction number deviates from the stable one, the higher the population momentum deviates from 1 at T as well, as the decision maker balances the costs of the adaptation efforts (which are considerably higher for adapting larger deviations) with the mismatching costs of the momentum.

Fig. 5 presents the total population over time. It is remarkable that the shape of these trajectories is homologous to that of the population momentum. In addition, we see that the size of the total population changes considerably over the 200 years. In our scenario of a very old population (i.e., $R_0 = 0.5$) the total population goes down to a little over 10% of the initial value. In the case of a very young population

(i.e., $R_0 = 5$) it increases by a factor of almost 180. This clearly relates to the fairly restrictive assumption that the total size of the population is not considered at all in the objective function. This will be relaxed in Section 5, where the decision maker aims for the total population to remain around a constant level, in addition to the end constraint $R(T) = 1$ and the objective of bringing the momentum as close to 1 as possible at T .

Finally, Fig. 6 shows the progressive aging/rejuvenation of a young/old initial population for $R_0 > 1$ and $R_0 < 1$, respectively, and shows an almost stationary population after 200 years for every scenario at a different population size corresponding to the initial NRR.

Remark 1. It should be stressed that these results are numerical ones. Although the qualitative results are supported by numerous numerical runs with different model parameters, we should note that analytical proof(s) that go beyond those of Section 3 (formulated in Proposition 2 and Corollary 1) are not available. However, analytical proofs for the shape of the optimal solution of optimal control models are inherently difficult to obtain (in particular for the case of the age-structured

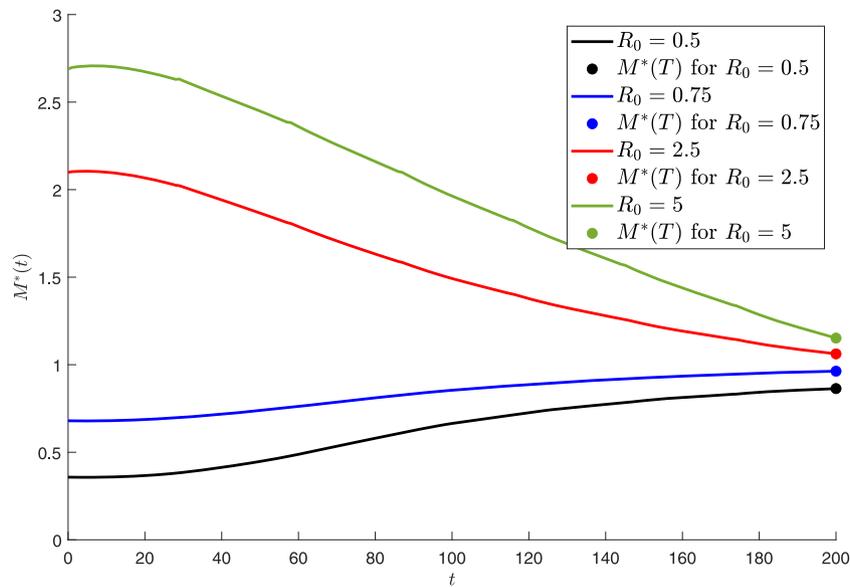


Fig. 4. Optimal population momentum $M^*(t)$ over time.

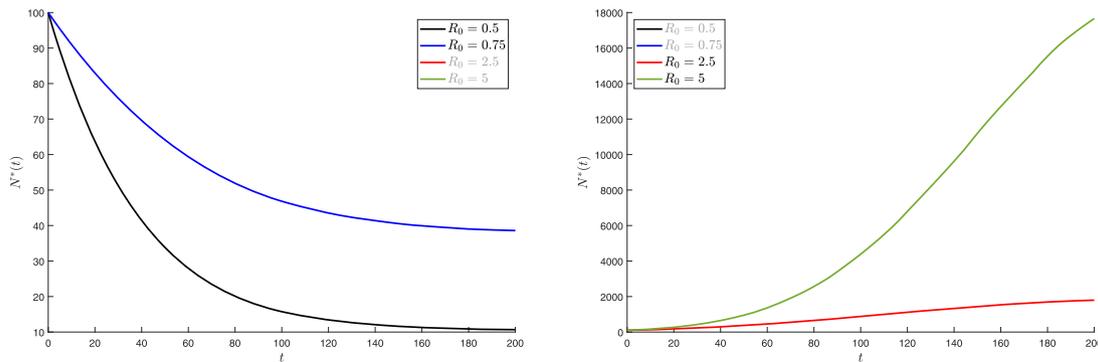


Fig. 5. Total population $N^*(t)$ over time (left-hand panel: $R_0 = 0.5$ and $R_0 = 0.75$, right-hand panel: $R_0 = 2.5$ and $R_0 = 5$).

optimal control model), which should be noted as a special feature of this model.

Remark 2. Further note, that the tiny edges across the age-time space which are visible in Fig. 6 are no numerical mistake, but due to our base assumptions regarding the mortality and reproduction sets at one fixed age (i.e., $\mu = 30$ and $\omega = 100$). Technically the edges emerge at the boundaries between regions 1–2 and 2–3 of Fig. 1. At these lines, the transversality conditions change disruptively (see (16)), which continues iteratively from the back to the front of the planning period.

The results discussed within this section are robust against variations of the time horizon T . The optimal efforts as well as the NRR show the same qualitative shape, but are more/less pronounced for a smaller/larger T , while the match of the population momentum to 1 deteriorates/improves. The population develops accordingly.

4.1. Variable age of fertility μ

The results of Figs. 2–6 are based on a fixed reproduction age of $\mu = 30$ years. Now, as already mentioned before, we are varying μ and comparing the base case with a 5-year lower and higher age (i.e., 25 and 35 years). Figs. 7–8 show the results for the high (young population, $R_0 = 2.5$) and low (old population, $R_0 = 0.75$) scenario. Throughout the figures, the black line shows the result for the younger reproduction age ($\mu = 25$), the red one for the benchmark reproduction age ($\mu = 30$, as

used for Figs. 2–6), and the blue one for the older reproduction age ($\mu = 35$).

The very high (very young population, $R_0 = 5$) and very low (very old population, $R_0 = 0.5$) scenarios give rise to analogous conclusions, but are not shown in order not to overload the figures.

The left-hand panel of Fig. 7 shows the optimal efforts to adapt the NRR for a young population (left-hand panel). The behavior of every line follows the shape discussed in the previous section (monotony, concavity/convexity). However, for a young population with a younger reproduction age (black line) the efforts are (i) stronger initially compared to the benchmark case (red line), but (ii) lower after approximately 70 years. In the case of an older reproduction age it is the other way around. The first effect (i) corresponds to a higher initial population momentum as can be seen from the right-hand panel in Fig. 7, which implies that the decision maker has to put more effort into bringing it down. The second effect (ii) then results from the initial increase, which implies a more rapid reduction of the NRR. To reach the end constraint $R(T) = 1$, however, the efforts have to be decreased afterwards. The optimal behavior of $k^*(t)$ also explains the shape of the NRR (see middle panel of Fig. 7) which shows a higher/lower curvature for a younger/older reproduction age.

For an older initial population ($R_0 < 1$) the effects are just the opposite, as shown in the right-hand panels of Fig. 8.

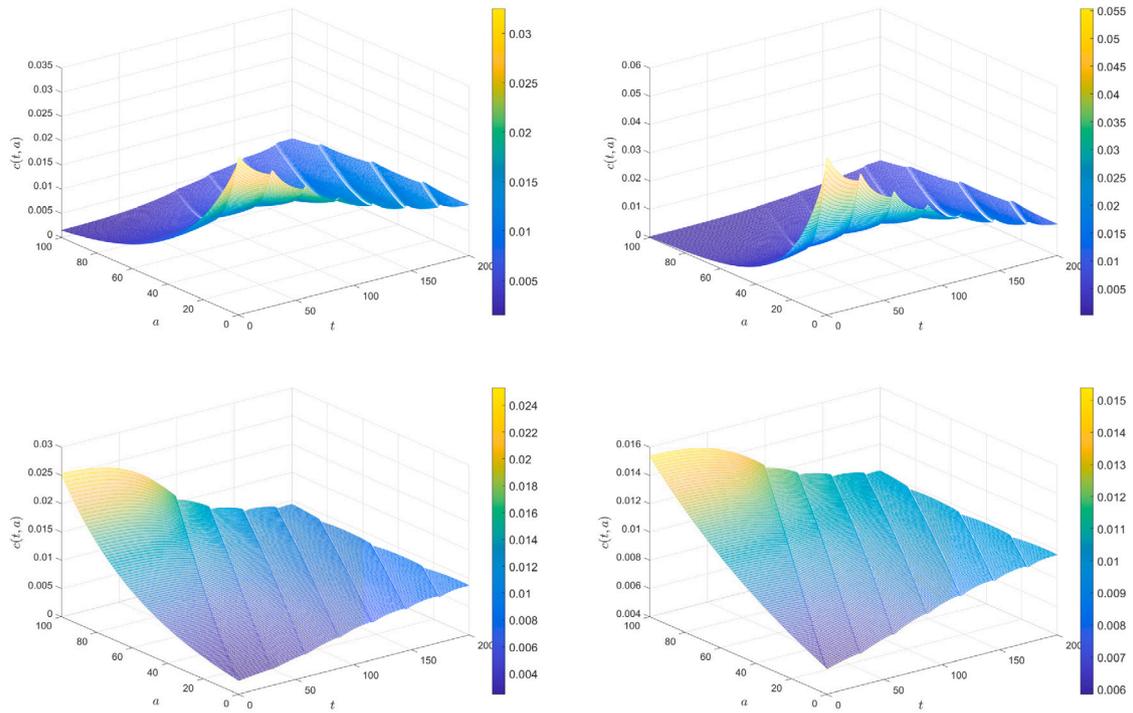


Fig. 6. Population density $P(t, a)$ over time (top-left-hand panel: $R_0 = 2.5$, top-right-hand panel: $R_0 = 5$, lower-left-panel: $R_0 = 0.75$, lower-right-panel: $R_0 = 0.5$).

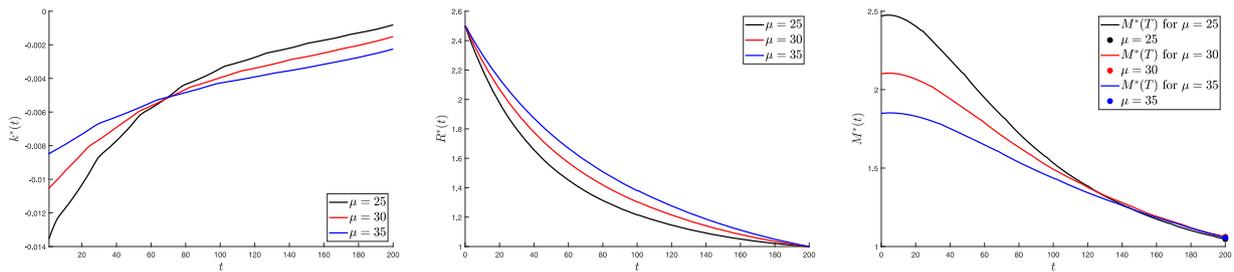


Fig. 7. Optimal adaptation of reproduction $k^*(t)$ (left-hand panel), NRR $R^*(t)$ (middle panel) and population momentum $M^*(t)$ (right-hand panel) for different μ for a young population $R_0 = 2.5$.

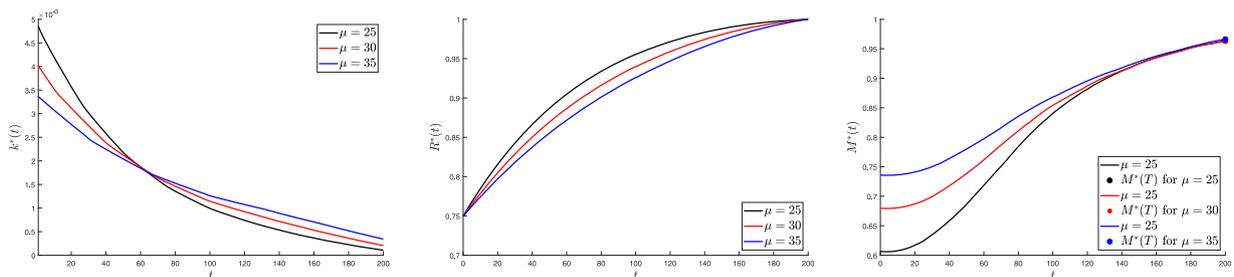


Fig. 8. Optimal adaptation of reproduction $k^*(t)$ (left-hand panel), NRR $R^*(t)$ (middle panel) and population momentum $M^*(t)$ (right-hand panel) for different μ for an old population $R_0 = 0.75$.

4.2. Robustness check: Effect of model simplifications

The theoretical results of Section 3 are possible due to the simplifying assumption that mortality and fertility are concentrated at ages ω and ν , respectively. While the concentrated fertility rate does not appear in the proofs and therefore seems to be unproblematic, the

concentrated mortality rate is explicitly used in the proof of Proposition 2. To study the implications of the simplification, we thus rely on numerical runs in two steps. First, we allow the fertility to be realistically distributed between the ages of 15 and 40 years keeping the mortality rate concentrated at ω . Secondly, we additionally assume a continuous mortality rate between all ages.

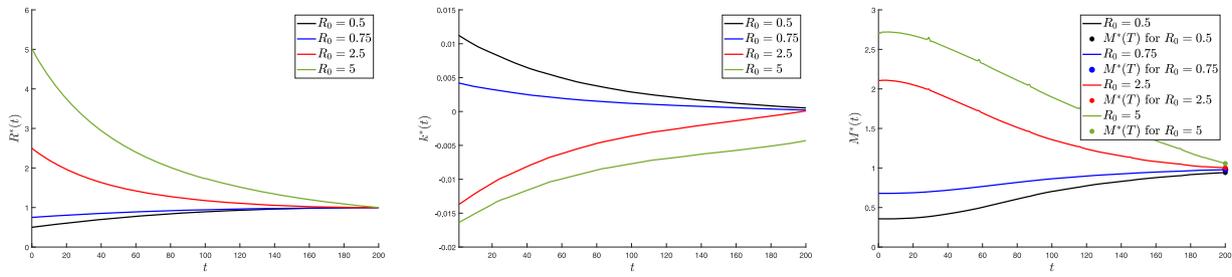


Fig. 9. Case distribution fertility: Optimal adaptation of reproduction $k^*(t)$ (left panel), $R^*(t)$ (middle panel) and population momentum $M^*(t)$ (right panel) for $R_0 = 0.75$ (old population) and $R_0 = 2.5$ (young population).

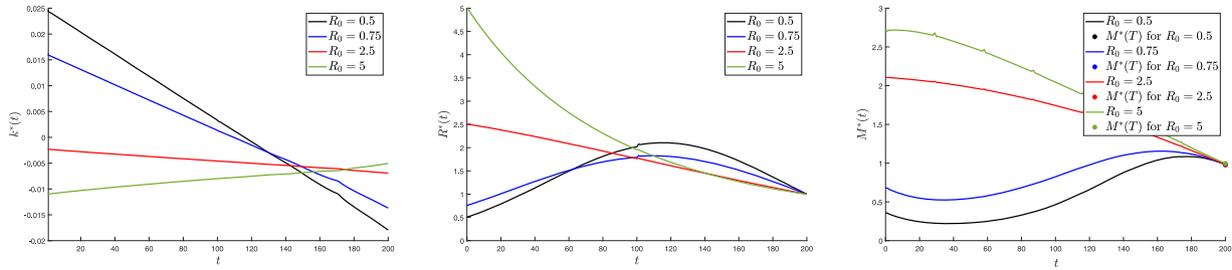


Fig. 10. Case distribution fertility and mortality: Optimal adaptation of reproduction $k^*(t)$ (left-hand panel), $R^*(t)$ (middle panel) and population momentum $M^*(t)$ (right-hand panel) for $R_0 = 0.75$ (old population) and $R_0 = 2.5$ (young population).

Let us quickly revisit the general model and discuss the implications of a distribution of the fertility and mortality rate. The McKendrick–von Foerster equation and the equations for births and deaths now read

$$P_t + P_a = -p(t, a)P(t, a), \quad P(0, a) = P_0(a), P(t, 0) = B(t) \quad (27a)$$

$$B(t) = \int_0^\omega v(a)P(t, a)R(t) da \quad (27b)$$

$$D(t) = \int_0^\omega p(t, a)P(t, a) da. \quad (27c)$$

The rest of the general problem (12) remains the same, except that the general expression for the optimal momentum (2) has to be used in the objective function. The Hamiltonian of the general problem reads

$$H = -\frac{1}{\omega}Ck^2 - \xi pP + \frac{1}{\omega}\lambda_R kR + \frac{1}{\omega}\lambda_N(B - D) + \nu PR + pP, \quad (28)$$

which implies that the first order condition (14) does not change. The mortality and fertility rate, however, change the adjoint equations for $P(t, a)$ and $R(t)$ which become

$$\xi_t + \xi_a = p\xi - \nu R - \eta_D P \quad (29a)$$

$$\dot{\lambda}_R = -\lambda_R k - \int_0^\omega \nu P da. \quad (29b)$$

with the corresponding transversality conditions. Thus, the relaxation of the simplification of the concentrated fertility rate enters (29b) into the second term on the right-hand side, where fertility is distributed among all cohorts within the fertile age at t . The relaxation of the concentrated mortality rate enters twice in (29a). The first term is added to the adjoint equation and accounts for the decrease in the population (along the life-cycle of a cohort) due to mortality. Secondly, the third term on the right-hand side is no longer concentrated at the maximal age ω , but distributed according to the mortality rate.

Fig. 9 shows the optimal adaptation of reproduction (left-hand panel), the NRR (middle panel) and the population momentum over time (right-hand panel) for a fertility distributed between 15 and 40 years. The plots show the results for (very) young ($R_0 > 1$) and (very) old populations ($R_0 < 1$), which are only slightly different from the ones in Figs. 2–6. We therefore conclude (running a series of examples with various parametric constellations) that for this relaxation the main analytical properties (Proposition 2 and Corollary 1) still hold. The analytical reason for this is that the fertility distribution only slightly

adapts the adjoint Eq. (29b) via the second term and this term is not critical in the proofs of Section 3.

Fig. 10 illustrates the optimal adaptation of reproduction (left-hand panel), the NRR (middle panel), and the population momentum over time (right-hand panel) if, additionally, the mortality rate is distributed (realistically) over the entire life course (i.e., from birth to the maximal age ω). In this case, the numerical runs show robust behavior for (very) young ($R_0 > 0$), but different behavior for (very) old populations ($R_0 < 1$). In the latter case, the monotony of the optimal adaptation efforts remains, but can change the sign during the time horizon. This in turn implies a wave in the NRR and the optimal momentum over time. Analytically this lies in the fact that rectangular mortality is crucial for the proof of Proposition 2 (states that $k(t)$ does not change the sign), which does not hold in general. The intuition for this effect is hidden in the last term of the adjoint equation (and the transversality condition). If mortality is concentrated at ω , this term is zero for the cohorts that are born within $(T, T - \omega]$ (region 2 and 3 in Fig. 1) but equal to $-\eta_D p = \bar{\lambda}_N p$ for the continuous mortality rate. So now, these cohorts also change the population momentum (negatively) by contributing to deaths. This reversing effect dominates the sign of $\lambda_R(t)$ at the end of the time horizon. This in turn implies that this wave is an anticipation effect of the predefined finite time horizon. The result of Corollary 1 seems to be partially true. While the monotony has already been discussed, the slope (decreasing or increasing) seems to be robust.

5. Extension: Penalizing fluctuations of the total population

Within the basic model (12) of Section 2, the decision maker aims to bring the population momentum as close to 1 as possible within a given planning period T . Adaptations to the NRR are costly, which is why they should also be minimized. The theoretical analysis (i.e., unique sign of $k(t)$ as proven in Corollary 1) as well as the numerical runs (see Fig. 5) show a drastic change in the size of the total population, although the population momentum at T is close to 1.

This section enriches the basic model (12) by penalizing deviations from a predefined size of the population over the entire time horizon. Hence, we add the quadratic distance from an exogenously given \bar{N} (aggregated over time) to the objective function (9). This yields

$$\bar{V}(R_0) := \mathcal{V}(R_0) + \int_0^T (N(t) - \bar{N})^2 dt, \quad (30)$$

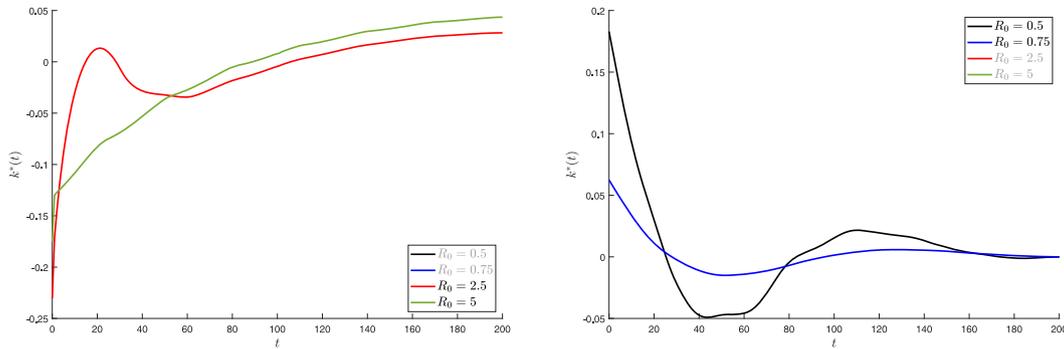


Fig. 11. Optimal adaptation of reproduction $k^*(t)$ over time (left-hand panel: $R_0 = 2.5$ and $R_0 = 5$, right-hand panel: $R_0 = 0.5$ and $R_0 = 0.75$).

which is minimized with respect to the same dynamical system.

Note that this is a change in the objective function and different from the robustness check in Section 4.2. Here we change the performance measure of the problem (which is very specific to the decision maker), whereas the relaxation corresponds to the level of abstraction of the mathematical model.

5.1. Analysis

As in Section 3 we rely on the Maximum Principle and add the term $-(N - \bar{N})^2$ to the Hamiltonian. From the Hamiltonian maximization we obtain the necessary first order condition

$$k = \frac{\lambda_R R}{2C}. \tag{31}$$

Note that, formally, the condition reads equivalent to (14), but it is different, as the adjoint equation for λ_N differs corresponding to the penalization term in the objective function, i.e.,

$$\dot{\lambda}_N = 2(N - \bar{N}). \tag{32}$$

The rest of the adjoint equations and the transversality conditions are analogous to (15) and (16), respectively. The Euler equation of $k(t)$ also remains formally the same as in the base model (see Eq. (20)), but due to the penalizing term $\lambda_N(t)$ is not constant and depends on t , that is:

$$\dot{k} = -\frac{R(t)P(t, \mu)}{2C} (\lambda_N(t) + \xi(t, 0)), \tag{33}$$

which basically means that the result of Lemma 1 holds in more general form.

By simplification of the adjoint equations we obtain

$$\xi_t + \xi_a = \begin{cases} -(\bar{\lambda}_N - \Delta(t) + \xi(t, 0)) \delta_\mu R & \text{for } a = \mu \\ (\bar{\lambda}_N - \Delta(t)) \delta_\omega & \text{for } a = \omega \\ 0 & \text{else} \end{cases}$$

$$\begin{aligned} \dot{\lambda}_R(t) &= -\lambda_R k - (\bar{\lambda}_N - \Delta(t) + \xi(t, 0)) P(t, \mu) \\ \dot{\lambda}_N(t) &= \bar{\lambda}_N - \Delta(t) \\ \eta_B(t) &= \bar{\lambda}_N - \Delta(t) + \xi(t, 0) \\ \eta_D(t) &= -\bar{\lambda}_N + \Delta(t) \end{aligned} \tag{34}$$

where the new term $\Delta(t) := \int_t^T 2(N(t) - \bar{N}) dt$ denotes the (aggregated) marginal deviation of $N(t)$.

By following the same steps as in Section 3 (without going into detail) we get

$$\lambda_R(t) = \lambda_R(T) e^{\int_t^T k(s') ds'} + \int_t^T e^{\int_t^{s'} k(s'') ds''} (\bar{\lambda}_N - \Delta + \xi(s', 0)) P(s', \mu) ds'. \tag{35}$$

The complete decomposition of the generalized reproductive value (as presented in Proposition 1 for the case of the basic model) can be obtained as follows.

Proposition 3. *The generalized reproductive value can be decomposed into three regions, as illustrated in Fig. 1: cohorts that have already died before T (region 1), cohorts that are alive at T but are older than the reproductive age (region 2), and cohorts that are alive and younger than, or exactly at, the reproductive age (region 3). For these cohorts $\xi(t, 0)$ can be reduced to*

$$\begin{aligned} t + \omega < T \text{ (region 1)} : \xi_1(t, 0) &= (\bar{\lambda}_N - \Delta) R(t + \mu) - (\bar{\lambda}_N - \Delta) \\ &\quad + \xi(t + \mu, 0) R(t + \mu) \tag{36a} \\ t + \mu < T \leq t + \omega \text{ (region 2)} : \xi_2(t, 0) &= (\bar{\lambda}_N - \Delta) R(t + \mu) \\ &\quad + \xi(t + \mu, 0) R(t + \mu) \tag{36b} \\ t + \mu = T \text{ (region 2/3)} : \xi_{2/3}(t, 0) &= (\bar{\lambda}_N - \Delta) R(T) + \xi(T, 0) R(T) \\ &\quad - 2G(M(T) - 1) \frac{\omega}{\mu} \frac{1}{N(T)} + \Delta(T) \tag{36c} \\ t + \mu > T \text{ (region 3)} : \xi_3(t, 0) &= -2G(M(T) - 1) \frac{\omega}{\mu} \frac{1}{N(T)} \\ &\quad + \Delta(t + \mu). \tag{36d} \end{aligned}$$

Proof. The Proposition can be proven by following the same steps as in the proof of Proposition 1. ■

The decomposition of the generalized reproductive value has been used in the basic model to show how the optimal efforts adapt to the NRR if the mortality is rectangular at T . This is formulated in Proposition 2 for the basic model. Following the steps of the proof reveals that $\xi_2(t, 0)$ and $\xi_3(t, 0)$ does not exclude an opposite sign even under the assumption of rectangular mortality rate due to the additional fluctuation factor $\int_{t+\mu}^{t+2\mu} 2(N(t) - \bar{N}) dt$.

As a result, an extension of Proposition 2 cannot be shown, which means that adaptation waves are possible within the planning horizon ($k(t) \leq 0$ for $t \in [0, T]$).

5.2. Numerical example and discussion

For an illustration of the optimal paths we use the same four scenarios as for the benchmark case (very young, young, old, very old population) together with the same reproduction age $\mu = 30$. The cost parameters and the time horizon remain the same. For the intended population size \bar{N} (which is a new parameter), we choose the initial size of 100.

Figs. 11–14 are similarly structured as Figs. 2–5 and differ only in the number of scenarios plotted in one figure. While it was straightforward to distinguish all four scenarios in the benchmark case, now the lines are crossing multiple times, which makes it difficult to observe and explain the main intuition. Therefore, the left-hand panels of Figs. 11–14 plot the case of a very young and young population, while the very old and old can be found in the right-hand one.

Fig. 11 shows the optimal efforts to adapt the NRR for young populations in the left-hand panel and for old ones in the right-hand panel. Analogous to the results of the base model (12) (see Fig. 2) efforts start

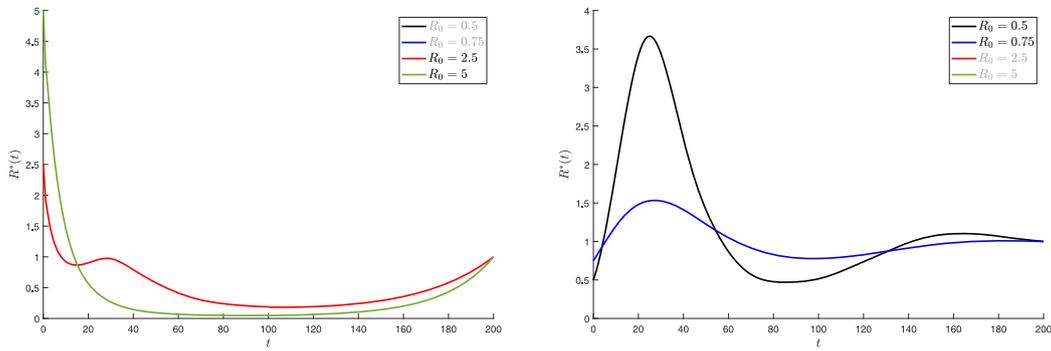


Fig. 12. Reproduction $R^*(t)$ over time (left-hand panel: $R_0 = 2.5$ and $R_0 = 5$, right-hand panel: $R_0 = 0.5$ and $R_0 = 0.75$).

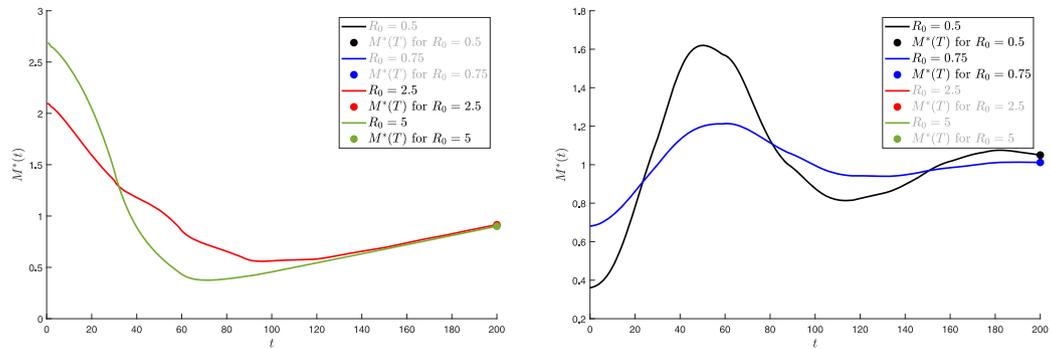


Fig. 13. Optimal population momentum $M^*(t)$ over time (left-hand panel: $R_0 = 2.5$ and $R_0 = 5$, right-hand panel: $R_0 = 0.5$ and $R_0 = 0.75$).

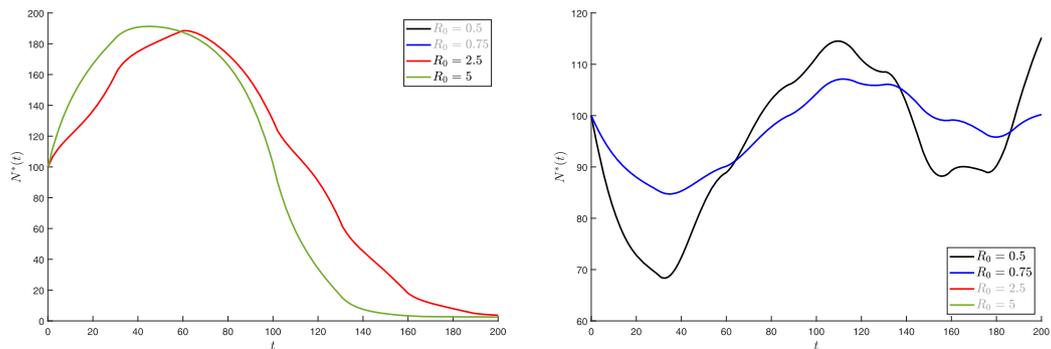


Fig. 14. Total population $N^*(t)$ over time (left-hand panel: $R_0 = 2.5$ and $R_0 = 5$, right-hand panel: $R_0 = 0.5$ and $R_0 = 0.75$).

as negative/positive for the young/old populations. However, now the penalization of fluctuations implies that the efforts can change the sign (potentially) multiple times before they eventually approach zero at T . This behavior, which could be excluded analytically (Proposition 2 and Corollary 1 within the base model), clearly corresponds to the aim of keeping the total size of the population at \bar{N} (i.e., the penalization of the deviation). This means that the decision maker has to counteract a high/low initial NRR immediately and vigorously (i.e., more vigorously than without the penalization) right after the beginning of the planning period. These efforts, however, oppose the end-state constraint $R(T) = 1$ and the aim of ending at a population momentum close to 1. The decision maker must thus turn around after the initial efforts. This interplay may take place more than once, but is more pronounced if the initial deviation of the NRR is higher, as can be seen in the right-hand panel of Fig. 11 (old populations). In the left-hand panel this can only be observed for $R_0 = 2.5$. For $R_0 = 5$ the planning period is just too short for the decision maker to bounce back.

The optimal path of the NRR is plotted in Fig. 12 follows the adaptation efforts and shows the oscillating behavior before the state constraint at the end is fulfilled $R(T) = 1$. Also in case of the NRR

the waves are damped down over time. Due to the multiplicative interaction in the dynamic of $R(t)$ (see (8)) the trajectories in Fig. 12 are more smooth than in Fig. 11, in which the small dents are due to the same reason as discussed in Remark 1.

Fig. 13 plots the optimal trajectory of the population momentum over time. The shape broadly follows that of Figs. 11 and 12. However, in contrast to the results of the base model a higher/lesser initial momentum does not imply a higher/fewer deviation from 1 at T (see discussion of Fig. 4), as the population momentum in this case already crosses 1 (depending on T , R_0 and μ maybe several times) before T .

Fig. 14 finally shows the size of the total population over time. The left-hand panel (young population) supports the statement (made in the discussion of Fig. 12) that the time horizon is too short for the decision maker to sufficiently satisfy all three opposing objectives (which are: end-state constraint, driving the population momentum close to 1, and keeping the population size near \bar{N}). This is obvious from the fact that the trajectories in the left-hand panel show a small population at T . Due to the very high initial NRR, the decision maker had to go in very vigorously to break the rapid population explosion, implying also a sharp decrease of $R(t)$. The rest of the further planning period

Table 2
Descriptive vs normative approach and analytical vs numerical solutions.

Mathematical approach	Results	
	Analytical	Numerical
Descriptive:	stable population theory, pseudostable populations	component method of population projections
Normative:	$k^*(t)$, $N^*(t)$	$P^*(t, a)$

was used to meet $R(T) = 1$ and to approach 1 with the population momentum. Due to the time lag in the population increase $N(t)$ there was not enough time to catch up again with \bar{N} . The right-hand panel (old population) shows that the decision maker is more flexible in the sense that the direction of the adaptation of the NRR can be changed more than once. As a result the total population at the end is closer to \bar{N} than in the left-hand panel.

The numerical solution of the extended model illustrates that many of the conclusions of the basic model do not carry over. In particular, the core results of Proposition 2 and Corollary 1, which cannot be proven here, are not only analytically intractable, but indeed wrong if fluctuations in the population are punished. It turns out to be typical for waves in the adaptation efforts, NRR, the population momentum, and the total population to be optimal if the time horizon is large enough to catch up with the end constraint reasonably before T .

6. Conclusions and extensions

Demography is a descriptive/explanatory science. Moreover, normative aspects are central in population policy, and mathematical optimization models play hardly any role in population dynamics. The purpose of the present research is to illustrate how intertemporal optimization methods can be applied to deal with an important population problem, namely that of the *demographic momentum*.

Before we summarize the results that we obtained in our analysis, let us refer to Table 2 where the dichotomies between a descriptive/normative approach on the one hand and an analytical/numerical approach on the other are exemplified.

As *time* and *age* are the core variables of population dynamics, *distributed parameter control* is an appropriate tool for analyzing problems pertinent to these. For a short introduction to the optimization of age-structured models compare Grass et al. (2008) (chapter 8.3).

Let us stress the following trade-off that arises in that context. The results we obtain by a suitable extension of Pontryagin's Maximum Principle are generally true (i.e., for all pertinent data sets), as they are derived in an analytical way. But one has to pay a price for the generality of these results, namely the simplification of the models considered. To mention just two of them: in our analysis it is assumed that everybody survives to a fixed age ω and every girl in the one-sex model is born at a unique age of her mother, μ .

The main insights from our analysis are as follows. By putting costly efforts into adapting the NRR, the decision maker aims to approach the population momentum as close to 1 as possible under the constraint that the NRR equals 1 at T (see Sections 2 and 4). For this model the analytical analysis proves a monotonic negative concave/positive convex optimal path of the adaptation efforts for negative/positive for an old/young initial population (starting from a stable population in the numerical examples). The curvature cannot be shown for the optimal path of the population momentum.

If fluctuations in the total number of the population are additionally penalized (as analyzed in Section 5), the monotony and the curvature cannot be shown analytically nor obtained numerically. The obtained solutions show oscillating behavior for the optimal adaptation effort paths, which carries over to the NRR, total population, and the population momentum. The reasons for this can be found in the increase/decrease of the population for an initial NRR higher/lower

than 1. To counteract this development (avoiding penalization costs), the decision maker chooses higher adaptation efforts (compared to the base scenario of Sections 2 and 4), which implies a too strong effect followed by opposite efforts, and so forth.

Both models (base model and extension) propose the derivation of an efficient adaptation rate of fertility in an optimal way for arbitrary initial NRRs as the value added of a population dynamic optimization problem presented in this paper.

There are several possible extensions. First, besides including the fluctuations of the total population $N(t)$, penalization of deviations of the age-structure (see (10f)) from a desired one (e.g., the stationary population), should also be considered. In another extension one could relax the symmetry of the (quadratic) penalization of the deviations both in the salvage value as well as in the fluctuations term $(N(t) - \bar{N})^2$. A further task would be to create a model without rectangular mortality and/or a fixed birth age.

A possible extension of the framework proposed in the paper is the application to a pest or endangered species. In an ecological context in both cases it makes sense to consider the control of the fertility schedule, which (because of the time-lagged effect of fertility) affects the population dynamics differently than conventional methods (e.g., use of poisons, releasing animals from breeding programs). See, for instance, Seal (1991) for a general discussion. Dell'Omo and Palmery (2002) who discuss the concept of fertility control in vertebrate pest species in comparison to conventional use of poisons, and Ramsey (2005) who formulates a mathematical model to study the population dynamics of brushtail possums subject to fertility control by using immunocontraceptive vaccines, are just two further examples of the dense literature in this context. Interestingly Bartholow (2007) projects the costs for several contraceptive treatments to manage wild horse populations, which is an important measure in the objective function (i.e., the cost of the control variable) of our proposed model. Certainly the dynamics for non-human species is often modeled in a simplified way. However, the application of our model requires an age-structured population with an endogenous fertility process. Moreover, for several species an 'isolated' model that does not consider other species and to which a predator-prey relation is natural, is not realistic. In this case, our modeling framework has to be extended by an age-structured predator-prey dynamics (see e.g., Gazis et al., 1973 for an early contribution or Wrzaczek et al., 2010 for a continuous-time formulation) which is an interesting task for future research.

CRedit authorship contribution statement

Gustav Feichtinger: Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Stefan Wrzaczek:** Conceptualization, Formal analysis, Methodology, Software, Supervision, Writing – original draft, Writing – review & editing.

Data availability

No data was used for the research described in the article.

Appendix. Extension of the maximum principle

In this section we extend the Maximum Principle for age-structured optimal control problems (see Feichtinger et al., 2003) to 'standard' state variables that only depend on time (and not on age). In the following general model $u(t, a) \in U$ and $v(t) \in V$ are control variables; $y(t, a)$ and $x(t)$ state variables (with corresponding initial condition for x , initial density, and boundary condition for $y(t, a)$); and $Q(t)$ aggregated state variables. Time and age are denoted by t and a , respectively.

In the following we present the sketch of a proof for the following optimal control model

$$\max_{u(t,a) \in U, v(t) \in V} \int_0^T \int_0^\omega L(y(t, a), x(t), Q(t), u(t, a), v(t), t, a) da dt$$

$$+ \int_0^\omega S(x(T), y(T, a), T, a) da \tag{37a}$$

s.t. $y_t(t, a) + y_a(t, a) = f(y(t, a), x(t), Q(t), u(t, a), v(t), t, a)$ (37b)

$$\dot{x}(t) = g(x(t), Q(t), v(t), t) \tag{37c}$$

$$Q(t) = \int_0^\omega h(y(t, a), x(t), Q(t), u(t, a), v(t), t, a) da \tag{37d}$$

$$y(0, a) = y^0(a) \tag{37e}$$

$$y(t, 0) = y^b(x(t), Q(t), v(t), t) \tag{37f}$$

$$x(0) = x_0, \tag{37g}$$

with objective functional $L(\cdot)$, an (age-structured) salvage value function $S(\cdot)$, system dynamics $f(\cdot)$ and $g(\cdot)$, and aggregation density $h(\cdot)$. For details on the regularity conditions of the involved functions we refer to Feichtinger et al. (2003).

For a sketch of the proof of an appropriate Maximum Principle we apply the needle variations principle to the Lagrangian of the above problem (37). The Lagrangian reads (for notational reasons we suppress the functional dependencies of L, f, g, h on state and control variables)

$$\begin{aligned} \mathcal{L}(x, y, Q, u, v, \lambda, \xi, \eta) = & \int_0^T \left\{ \int_0^\omega [L + \xi(t, a) (f - y_t - y_a)] da \right. \\ & + \lambda(t) (g - \dot{x}) + \eta(t) \left(\int_0^\omega h da - Q \right) \Big\} dt \\ & + \int_0^\omega S(x(T), y(T, a), T, a) da. \end{aligned} \tag{38}$$

Integration by parts $\lambda(t)\dot{x}(t)$ and $\xi(t, a) (y_t(t, a) + y_a(t, a))$ implies

$$\int_0^T \lambda(t)\dot{x}(t) dt = \lambda(T)x(T) - \lambda(0)x(0) - \int_0^T \dot{\lambda}(t)x(t) dt, \tag{39a}$$

$$\begin{aligned} \int_0^\omega \xi(t + a, a) (y_t(t + a, a) + y_a(t + a, a)) da \\ = \xi(t + \omega, \omega)y(t + \omega, \omega) - \xi(t, 0)y(t, 0) \\ - \int_0^\omega (\xi_t(t + a, a) + \xi_a(t + a, a)) y(t + a, a) da, \end{aligned} \tag{39b}$$

$$\begin{aligned} \int_{a_0}^\omega \xi(a - a_0, a) (y_t(a - a_0, a) + y_a(a - a_0, a)) da \\ = \xi(\omega - a_0, \omega)y(\omega - a_0, \omega) - \xi(0, a_0)y(0, a_0) \\ - \int_{a_0}^\omega (\xi_t(a - a_0, a) + \xi_a(a - a_0, a)) y(a - a_0, a) da, \end{aligned} \tag{39c}$$

$$\begin{aligned} \int_0^{T-t} \xi(t + a, a) (y_t(t + a, a) + y_a(t + a, a)) da \\ = \xi(T, T - t)y(T, T - t) - \xi(t, 0)y(t, 0) \\ - \int_0^{T-t} (\xi_t(t + a, a) + \xi_a(t + a, a)) y(t + a, a) da, \end{aligned} \tag{39d}$$

where the integration of the population related expression is distinguished into three cases. (39b) corresponds to cohorts that are born and die within the planning period (i.e., for all $t \in [0, T - \omega]$), (39c) corresponds to cohorts that have already been born before $t = 0$ (initial population density, that is, for all $a_0 \in [0, \omega]$ and therefore $t = -a_0 + a$), and (39d) to cohorts that are born within the planning period but have not died at T (i.e., for all $t \in [T - \omega, T]$). Plugging into the Lagrangian gives the following expression,

$$\begin{aligned} \mathcal{L}(\cdot) = & \int_0^T \left\{ \int_0^\omega [L + \xi(t, a)f] da + \lambda(t)g + \eta(t) \left(\int_0^\omega h da - Q \right) \right\} dt \\ & - \left(\lambda(T)x(T) - \lambda(0)x(0) - \int_0^T \dot{\lambda}(t)x(t) dt \right) \\ & - \int_0^{T-\omega} \left(\xi(t + \omega, \omega)y(t + \omega, \omega) - \xi(t, 0)y(t, 0) \right. \\ & \left. - \int_0^\omega (\xi_t(t + a, a) + \xi_a(t + a, a)) y(t + a, a) da \right) dt \\ & - \int_0^\omega \left(\xi(\omega - a_0, \omega)y(\omega - a_0, \omega) - \xi(0, a_0)y(0, a_0) \right. \end{aligned}$$

$$\begin{aligned} & \left. - \int_{a_0}^\omega (\xi_t(a - a_0, a) + \xi_a(a - a_0, a)) y(a - a_0, a) da \right) da_0 \\ & - \int_{T-\omega}^T \left(\xi(T, T - t)y(T, T - t) - \xi(t, 0)y(t, 0) \right. \\ & \left. - \int_0^{T-t} (\xi_t(t + a, a) + \xi_a(t + a, a)) y(t + a, a) da \right) dt \\ & + \int_0^\omega S(x(T), y(T, a), T, a) da. \end{aligned} \tag{40}$$

Defining the Hamiltonian as

$$\mathcal{H}(\cdot) = L + \xi(t, a)f + \frac{1}{\omega} \lambda(t)g + \eta(t)h \tag{41}$$

and plugging in (40) we arrive at

$$\begin{aligned} \mathcal{L}(\cdot) = & \int_0^T \int_0^\omega \mathcal{H} da dt - \int_0^T \eta(t)Q dt \\ & - \left(\lambda(T)x(T) - \lambda(0)x(0) - \int_0^T \dot{\lambda}(t)x(t) dt \right) \\ & - \int_0^{T-\omega} \left(\xi(t + \omega, \omega)y(t + \omega, \omega) - \xi(t, 0)y(t, 0) \right. \\ & \left. - \int_0^\omega (\xi_t(t + a, a) + \xi_a(t + a, a)) y(t, a) da \right) dt \\ & - \int_0^\omega \left(\xi(\omega - a_0, \omega)y(\omega - a_0, \omega) - \xi(0, a_0)y(0, a_0) \right. \\ & \left. - \int_{a_0}^\omega (\xi_t(a - a_0, a) + \xi_a(a - a_0, a)) y(a - a_0, a) da \right) da_0 \\ & - \int_{T-\omega}^T \left(\xi(T, T - t)y(T, T - t) - \xi(t, 0)y(t, 0) \right. \\ & \left. - \int_0^{T-t} (\xi_t(t + a, a) + \xi_a(t + a, a)) y(t + a, a) da \right) dt \\ & + \int_0^\omega S(x(T), y(T, a), T, a) da. \end{aligned} \tag{42}$$

For deriving the first variation we consider needle variations⁹ (see Alekseev et al., 1987) of control and state variables (taking the limit of ϵ from the right), that is

$$x(t) \rightarrow x(t) + \epsilon \delta x(t), \tag{43a}$$

$$v(t) \rightarrow v(t) + \epsilon \delta v(t), \tag{43b}$$

$$y(t, a) \rightarrow y(t, a) + \epsilon \delta y(t, a), \tag{43c}$$

$$u(t, a) \rightarrow u(t, a) + \epsilon \delta u(t, a), \tag{43d}$$

$$Q(t) \rightarrow Q(t) + \epsilon \delta Q(t), \tag{43e}$$

which gives

$$\begin{aligned} \frac{d}{d\epsilon} \mathcal{L}(\cdot) \Big|_{\epsilon=0+} = & \int_0^T \int_0^\omega [H_u \delta u + H_v \delta v + H_x \delta x + H_y \delta y + H_Q \delta Q] da dt \\ & - \int_0^T \eta(t) \delta Q dt \\ & - \lambda(T) \delta x(T) + \int_0^T \dot{\lambda}(t) \delta x(t) dt \\ & - \int_0^{T-\omega} \left(\xi(t + \omega, \omega) \delta y(t + \omega, \omega) \right. \\ & \left. - \xi(t, 0) \left(y_x^b(t, 0) \delta x + y_Q^b(t, 0) \delta Q + y_v^b(t, 0) \delta v \right) \right. \\ & \left. - \int_0^\omega (\xi_t(t + a, a) + \xi_a(t + a, a)) \delta y(t, a) da \right) dt \\ & - \int_0^\omega \left(\xi(\omega - a_0, \omega) \delta y(\omega - a_0, \omega) \right. \end{aligned}$$

⁹ In this context, δ denotes a variation of a curve such that the varied curve is of the same type.

Table 3
Correspondence of the optimal momentum model (12) with the general model (37).

	Optimal momentum model	General model
Control variables	$k(t)$ –	$v(t)$ $u(t, a)$
State variables	$P(t, a)$ $R(t), N(t)$	$y(t, a)$ $x(t)$
Aggregated variables	$B(t), N(t)$	$Q(t)$
Salvage value	$G(M(T) - 1)^2$	$\int_0^\omega S(x(T), y(T, a), a, T) da$

$$\begin{aligned}
 & - \int_{a_0}^\omega (\xi_t(a - a_0, a) + \xi_a(a - a_0, a)) \delta y(a - a_0, a) da \Big) da_0 \\
 & - \int_{T-\omega}^T (\xi(T, T - t) \delta y(T, T - t) \\
 & - \xi(t, 0) (y_x^b(t, 0) \delta x + y_Q^b(t, 0) \delta Q + y_v^b(t, 0) \delta v) \\
 & - \int_0^{T-t} (\xi_t(t + a, a) + \xi_a(t + a, a)) \delta y(t + a, a) da \Big) dt \\
 & + \int_0^\omega [S_x(\cdot) \delta x(T) + S_y(\cdot) \delta y(T, a)] da. \tag{44}
 \end{aligned}$$

The condition $\frac{d}{d\epsilon} \mathcal{L}(\cdot) \Big|_{\epsilon=0+} = 0$ is necessary for optimality of the problem. This implies

$$0 = [H_u] \delta u(t), \tag{45a}$$

$$0 = \left[\int_0^\omega H_v da + \xi(t, 0) y_v^b(t, 0) \right] \delta v(t), \tag{45b}$$

$$0 = \left[\int_0^\omega H_x da + \lambda(t) + \xi(t, 0) y_x^b(t, 0) \right] \delta x(t), \tag{45c}$$

$$0 = [H_y + (\xi_t(t, a) + \xi_a(t, a))] \delta y(t, a), \tag{45d}$$

$$0 = \left[\int_0^\omega H_Q da - \eta(t) + \xi(t, 0) y_Q^b(t, 0) \right] \delta Q(t), \tag{45e}$$

$$0 = \left[-\lambda(T) + \int_0^\omega S_x da \right] \delta x(T), \tag{45f}$$

$$0 = [-\xi(t, \omega)] \delta y(t, a), \tag{45g}$$

$$0 = [-\xi(T, a) + S_y] \delta y(T, a), \tag{45h}$$

which together form a suitable Maximum Principle for problem (37). In particular, (45a)–(45b) corresponds to the Hamiltonian maximizing condition, (45c)–(45e) to the adjoint equations, and (45f)–(45h) to the transversality conditions.

Table 3 summarizes how the optimal momentum model (12) fits in the general model (37).

References

Alekseev, V.M., Tikhomirov, V.M., Fomin, S.V., 1987. *Optimal Control*. Springer US, Boston, MA.

Anița, S., 2000. In: Lowen, R. (Ed.), *Analysis and Control of Age-Dependent Population Dynamics*. In: *Mathematical Modelling: Theory and Applications*, vol. 11, Springer Netherlands, Dordrecht.

Arthur, W.B., McNicoll, G., 1977. Optimal time paths with age-dependence: A theory of population policy. *Rev. Econom. Stud.* 44 (1), 111.

Bartholow, J., 2007. Economic benefit of fertility control in wild horse populations. *J. Wildl. Manage.* 71 (8), 2811–2819.

Baudisch, A., 2008. *Inevitable Aging?: Contributions to Evolutionary-Demographic Theory* | SpringerLink. Springer.

Bergsvik, J., Fauske, A., Hart, R.K., 2021. Can policies stall the fertility fall? A systematic review of the (quasi-) experimental literature. *Popul. Dev. Rev.* 47 (4), 913–964.

Björklund, A., 2006. Does family policy affect fertility?: Lessons from Sweden. *J. Popul. Econ.* 19 (1), 3–24.

Brokate, M., 1985. Pontryagin’s principle for control problems in age-dependent population dynamics. *J. Math. Biol.* 23 (1), 75–101.

Buratto, A., Muttoni, M., Wrzaczek, S., Freiberger, M., 2022. Should the COVID-19 lockdown be relaxed or intensified in case a vaccine becomes available? *PLoS One* 17 (9), e0273557.

Chu, C.C., Chien, H.-K., Lee, R.D., 2008. Explaining the optimality of U-shaped age-specific mortality. *Theor. Popul. Biol.* 73 (2), 171–180.

Chu, C.C., Lee, R.D., 2006. The co-evolution of intergenerational transfers and longevity: An optimal life history approach. *Theor. Popul. Biol.* 69 (2), 193–201.

Coale, A.J., 1972. *Growth and Structure of Human Populations*. A Mathematical Investigation. Princeton University Press.

Dawid, H., Feichtinger, G., Goldstein, J.R., Veliov, V.M., 2009. Keeping a learned society young. *Demogr. Res.* 20, 541–558.

Dell’Omo, G., Palmery, M., 2002. Fertility control in vertebrate pest species. *Contraception* 65 (4), 273–275.

Feichtinger, G., Hartl, R.F., Kort, P.M., Veliov, V.M., 2006. Anticipation effects of technological progress on capital accumulation: a vintage capital approach. *J. Econom. Theory* 126 (1), 143–164.

Feichtinger, G., Kuhn, D.R., Prskawetz, A., Wrzaczek, S., 2011. The reproductive value as part of the shadow price of population. *Demogr. Res.* 24, 709–718.

Feichtinger, G., Prskawetz, A., Seidl, A., Simon, C., Wrzaczek, S., 2017. A bifurcation analysis of gender equality and fertility. *J. Evol. Econ.* 27 (5), 1221–1243.

Feichtinger, G., Rau, R., Novak, A.J., 2023. On the momentum of pseudostable populations. *VID Work. Pap.* 1, 1–25.

Feichtinger, G., Tragler, G., Veliov, V.M., 2003. Optimality conditions for age-structured control systems. *J. Math. Anal. Appl.* 288 (1), 47–68.

Feichtinger, G., Veliov, V.M., 2007. On a distributed control problem arising in dynamic optimization of a fixed-size population. *SIAM J. Optim.* 18 (3), 980–1003.

Fisher, R.A., 1930. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford, Open Library ID: OL7084333M.

Frankovic, I., Kuhn, M., Wrzaczek, S., 2020a. Medical innovation and its diffusion: Implications for economic performance and welfare. *J. Macroecon.* 66, 103262.

Frankovic, I., Kuhn, M., Wrzaczek, S., 2020b. On the anatomy of medical progress within an overlapping generations economy. *De Economist* 168 (2), 215–257.

Gauthier, A.H., 2007. The impact of family policies on fertility in industrialized countries: a review of the literature. *Popul. Res. Policy Rev.* 26 (3), 323–346.

Gazis, D.C., Montroll, E.W., Ryniker, J.E., 1973. Age-specific, deterministic model of predator-prey populations: Application to isle royale. *IBM J. Res. Dev.* 17 (1), 47–53.

Goldstein, J.R., 2002. Population momentum for gradual demographic transitions: an alternative approach. *Demography* 39 (1), 65–73.

Goodman, D., 1982. Optimal life histories, optimal notation, and the value of reproductive value. *Amer. Nat.* 119 (6), 803–823, Publisher: [University of Chicago Press, American Society of Naturalists].

Grass, D., Caulkins, J.P., Feichtinger, G., Tragler, G., Behrens, D.A., 2008. *Optimal Control of Nonlinear Processes: With Applications in Drugs, Corruption and Terror*. Springer.

Hethcote, H.W., 1988. Optimal ages of vaccination for measles. *Math. Biosci.* 89 (1), 29–52.

Kaplan, H.S., Robson, A.J., 2009. We age because we grow. *Proc. R. Soc. B Lond. Biol.* 276 (1663), 1837–1844.

Keyfitz, N., 1971. On the momentum of population growth. *Demography* 8 (1), 71–80, Publisher: Springer.

Keyfitz, N., Caswell, H., 2005. *Applied Mathematical Demography*, third ed. In: *Statistics for Biology and Health*, Springer, New York, NY.

Keyfitz, B., Keyfitz, N., 1997. The mckendrick partial differential equation and its uses in epidemiology and population study. *Math. Comput. Modelling* 26 (6), 1–9.

Kuhn, M., Wrzaczek, S., Oeppen, J., 2010. Recognizing progeny in the value of life. *Econom. Lett.* 107 (1), 17–21.

Kuhn, M., Wrzaczek, S., Prskawetz, A., Feichtinger, G., 2011. Externalities in a life cycle model with endogenous survival. *J. Math. Econom.* 47 (4–5), 627–641.

Leon, J.A., 1976. Life histories as adaptive strategies. *J. Theoret. Biol.* 60, 301–335.

Li, N., Tuljapurkar, S., 1999. Population momentum for gradual demographic transitions. *Popul. Stud.* 53 (2), 255–262, Publisher: [Population Investigation Committee, Taylor & Francis, Ltd.].

Li, N., Tuljapurkar, S., 2000. The solution of time-dependent population models. *Math. Popul. Stud.* 7 (4), 311–329.

Luci-Greulich, A., Thévenon, O., 2013. The impact of family policies on fertility trends in developed countries: l’influence des politiques familiales sur les tendances de la fécondité des pays développés. *Rev. Eur. Démog. (Eur. J. Popul.)* 29 (4), 387–416.

Preston, S.H., 1997. Population dynamics in an age of declining fertility. *Genus* 53 (3), 15–31.

Preston, S.H., Heuveline, P., Guillot, M., 2001. *Demography: Measuring and modeling population processes*, | Wiley. Wiley.com.

Prskawetz, A., Veliov, V.M., 2007. Age-specific dynamic labor demand and human capital investment. *J. Econom. Dynam. Control* 31 (12), 3741–3777.

Ramsey, D., 2005. Population dynamics of brushtail possums subject to fertility control: Fertility control of brushtail possums. *J. Appl. Ecol.* 42 (2), 348–360.

Robson, A.J., Kaplan, H.S., 2003. The evolution of human life expectancy and intelligence in hunter-gatherer economies. *Amer. Econ. Rev.* 93 (1), 150–169.

Robson, A.J., Kaplan, H.S., 2007. Why do we die? Economics, biology, and aging. *Amer. Econ. Rev.* 97 (2), 492–495.

- Schaffer, W.M., 1983. The application of optimal control theory to the general life history problem. *Amer. Nat.* 121 (3), 418–431.
- Schoen, R., Kim, Y.J., 1998. Momentum under a gradual approach to zero growth. *Popul. Stud.* 52 (3), 295–299.
- Seal, U.S., 1991. Fertility control as a tool for regulating captive and free-ranging wildlife populations. *J. Zoo Wildl. Med.* 22 (1), 1–5.
- Sobotka, T., Matysiak, A., Brzozowska, Z., 2019. Policy responses to low fertility: How effective are they? United Nations Population Fund, Working paper No. 1.
- Veliov, V.M., 2003. Newton's method for problems of optimal control of heterogeneous systems. *Optim. Methods Softw.* 18 (6), 689–703, Publisher: Taylor & Francis.
- Wrzaczek, S., 2021. An OLG differential game of pollution control with the risk of a catastrophic climate change. *Int. Game Theory Rev.* 23 (04), 2250002.
- Wrzaczek, S., Kuhn, M., Frankovic, I., 2020. Using age structure for a multi-stage optimal control model with random switching time. *J. Optim. Theory Appl.* 184 (3), 1065–1082.
- Wrzaczek, S., Kuhn, M., Prskawetz, A., Feichtinger, G., 2010. The reproductive value in distributed optimal control models. *Theor. Popul. Biol.* 77 (3), 164–170.
- Wrzaczek, S., Shevkoplyas, E., Kostyunin, S., 2014. A differential game of pollution control with overlapping generations. *Int. Game Theory Rev.* 16 (03), 1450005.