THE EVOLUTION OF TIME PREFERENCE WITH AGGREGATE UNCERTAINTY

Arthur Robson
Department of Economics
Simon Fraser University
Burnaby, B. C. V5A 1S6 Canada
Robson@sfu.ca

Larry Samuelson
Department of Economics
Yale University
New Haven, CT 06520-8281 USA
Larry.Samuelson@yale.edu

February 1, 2008

Abstract: We examine the evolutionary foundations of intertemporal preferences. When all the risk affecting survival and reproduction is idiosyncratic, evolution selects for agents who maximize the the discounted sum of expected utility, discounting at the sum of the population growth rate and the mortality rate. Aggregate uncertainty concerning survival rates leads to discount rates that exceed the sum of population growth rate and death rate, and can push agents away from exponential discounting.

Contents

---

*We thank Georg Nöldeke and Jeroen Swinkels for helpful discussions and thank the Canada Research Chair Program and the National Science Foundation (SES-0241506 and SES-0549946) for financial support.

THE EVOLUTION OF TIME PREFERENCE WITH AGGREGATE UNCERTAINTY

1 Introduction

How much do people discount the future? How does their discounting change as they consider events further in the future?[1] Perhaps more fundamentally, why do people discount at all?

Fisher's [4, pp. 84–85] pioneering study of intertemporal tradeoffs called attention to one reason future rewards are discounted—an intervening death may prevent us from realizing such rewards. This reason has played a recurring role in discussions of discounting (e.g., Yaari [24]). Hansson and Stuart [9] and Rogers [18] argue that evolution should select in favor of people whose discounting reflects the growth rate of the population with whom they are competing (see also Robson and Szentes [17]). Putting these ideas together leads to models in which people discount at the sum of the population growth rate and mortality rate.

One difficulty with this argument is that the numbers don't obviously match. Studies of contemporary subject rates of time preference yield estimates as high as twelve to twenty percent per year (Lawrance [11]). Using data from contemporary hunter-gatherers, Gurven and Kaplan [8] estimate that annual mortality rates during our evolutionary history ranged from one percent for ten-year-olds to four percent for sixty-year-olds, while the average population growth rate over this two-million year period was approximately zero. A second cause for concern is the growing evidence that intertemporal preferences exhibit a present bias not captured by the exponential discounting of standard models.

This paper re-examines the foundations of intertemporal preferences. Like Hansson and Stuart [9] and Rogers [18], we view peoples' preferences as having been shaped by biological evolution. We consider the evolution of intertemporal preferences in age-structured populations, i.e., populations in which each individual can reproduce at different ages, focussing on the simplest question of how people discount future reproduction. When all the risk affecting survival and reproduction is idiosyncratic, we find the standard result that there is a close connection between the evolutionary criterion for success and the simplest criterion for intertemporal choice—the discounted

---

[1] Recent policy discussions, especially those regarding global warming, have focussed attention on the first question (e.g., Nordhaus [15]), while recent work in behavioral economics has directed attention to the latter (Frederick, Loewenstein and O'Donoghue [5]).

sum of expected utility. This result in turn generates the anticipated rate of discount, namely the sum of the population growth rate and the mortality rate.

Our contribution derives from the observation that the risks in our evolutionary environment are unlikely to have been purely idiosyncratic. Fluctuations in the weather, abundant predators, epidemics, and failures of food sources are all bound to have a common effect on death rates. Such aggregate uncertainty breaks the connection between discounting and the sum of the growth and death rates. We first show that aggregate uncertainty "generically" lowers the growth rate below that arising with comparable idiosyncratic uncertainty.[2] In addition, if the environmental fluctuations have a uniform effect on people of different ages, then future reproduction is discounted at a rate exceeding the population growth rate plus the expected mortality rate—aggregate risk may lie behind the gap between discount rates and growth and mortality rates.[3]

What if the effects of aggregate uncertainty differ across ages? We find that discount rates need no longer be constant, and we present natural (but by no means universal) conditions under which the rate of discount falls as a function of age. This "present bias" in discounting is reminiscent of the present bias that has played a central role in behavioral economics. However, the discount rates that emerge from our model are tied to age rather than time, precluding preference reversals.[4]

Section 2 introduces the mechanics of age-structured populations for the simpler case of an environment with only idiosyncratic uncertainty. Section 3 examines aggregate uncertainty. Section 4 discusses some of the many features that are left out of our analysis. Proofs not contained in the body of the paper are collected in Section 5.

---

[2] See Robson [16] for an analogous result for populations without an age structure.

[3] Section 3.4 explains how this model formalizes and generalizes the "sawtooth" explanation sometimes advanced to reconcile an average growth rate near zero in our evolutionary past with the higher growth rates often seen in contemporary hunter-gatherers. This sawtooth model couples long periods of sustained growth with rare, rapid and evolutionarily-neutral population collapses.

[4] Dasgupta and Maskin [3] and Sozou [20] also present evolutionary foundations for presently-biased preferences, including in Dasgupta and Maskin's case the possibility of preference reversals. We discuss these papers in Section 4.

2 Idiosyncratic Uncertainty

It is helpful to first consider the more straightforward case of idiosyncratic uncertainty, drawing on Charlesworth [1] and Houston and McNamara [10].

2.1 The model

Time is discrete, given by $t = 0, 1, \ldots$. We take a census of a population at the start of each period t, letting $N_\tau(t)$ be the number of agents then of age $\tau \in \{1, 2, \ldots, T\}$.

The first event in period t is that each agent of age $\tau \in \{1, 2, \ldots, T\}$ has offspring, with x_τ denoting the expected number of offspring born to an age-τ parent. Each agent of each age $\tau \in \{0, \ldots, T-1\}$ then either dies or survives, with S the probability of survival. Agents of age T disappear from our system. This may reflect either death or a continuing life without reproduction, essentially equivalent fates from a biological point of view.[5] All surviving agents younger than T enter the next period one year older. This brings us to the beginning of period $t+1$, where we take the next census, finding $N_\tau(t+1)$ agents of age $\tau \in \{1, 2, \ldots, T\}$, and begin the process anew with the next round of births.

Section 4 briefly discusses generalizations to the more realistic case of age-dependent mean death rates. In the meantime, taking death rates to be constant allows us to isolate other factors that may lie behind varying discount rates. Depending on the magnitude of the survival rate S, the population may be exploding or shrinking to oblivion. None of the subsequent analysis would be affected if the survival rate were modeled as reflecting an environmental carrying capacity, as long as our S is then interpreted as the endogenously-determined zero-population-growth steady-state survival rate.

A reproduction profile in this environment is a T-tuple $\{x_\tau, \tau \in \{1, 2, \ldots, T\}\}$. This profile is heritable. Our interest centers on which such profiles will be selected by evolution. In particular, we will ask how evolution induces people to trade off x_τ and $x_{\tau'}$, effectively revealing the discounting that evolution builds into preferences.

We are ultimately interested in people's preferences over the wide variety of things they consume, rather than simply reproduction. We view the

---

[5] Continued life without reproduction scales up the population but does not affect its growth rate. A mutation that increased one of $\{x_1, \ldots, x_T\}$ by even a very small amount, while sacrificing all survival beyond age T, would increase the growth rate and hence would be evolutionarily favored.

fertility x_τ as being a function of consumption of food, shelter, status, and a host of other economic goods, with intertemporal preferences over these goods induced by their implications for reproduction. We do not assert that people explicitly consider the reproductive implications of each decision they make. Evolution has instead doubtlessly found it more expedient to simply endow us with preferences over economic goods. But these preferences were originally shaped by the implications of the resulting decisions for reproduction. Identifying intertemporal tradeoffs over the x_τ is thus the first step in understanding discounting.[6]

We also recognize that our modern environment is quite different from that in which we evolved. However, precisely because evolution found it more expedient to simply give us preferences over economic goods rather than make us relentless reproduction calculators, insight into the preferences that shape behavior in our modern world is to be found by examining our evolutionary past.

We follow the standard approach in assuming the population is large, captured formally by viewing the set of agents as a continuum. This allows us to construct a convenient deterministic model of the population. Each agent faces *idiosyncratic* uncertainty, in the sense that the agent may have more or fewer offspring in a given period and may or may not survive until the next, but the average number of offspring born to agents of age τ can be taken to be precisely x_τ and the proportion of survivors can be taken to be precisely S.[7]

---

[6] To be more precise, if fertility x_τ were a function f_τ of consumption at date τ, then attitudes to intertemporal inequality in consumption would be affected by the properties of f_τ (its concavity, for example) as well as the way in which the x_τ combined to yield a growth rate. Since the first effect is relatively familiar, we concentrate here on the derivation of the growth rate criterion from the x_τ.

[7] Intuitively, each agent of age τ takes an independent (across agents and across periods) draw from an offspring lottery with mean x_τ, determining the agent's number of offspring, and from a survival lottery that yields survival with probability S and death with probability $1-S$. The law of large numbers then ensures that realized and expected numbers of offspring, as well as realized and expected numbers of total surviving agents, coincide. More precisely, it is well known that one cannot appeal to such a law-of-large-numbers result with a continuum of random variables (cf. Al-Najjar [14]). In our case, as in many applications, independence is not necessary, allowing one to construct explicit probability spaces yielding random variables with the properties that are important for our results.

The population evolves according to

$$[N_1(t+1),\dots,N_T(t+1)] \quad = \quad [N_1(t),\dots,N_T(t)] \begin{bmatrix} Sx_1 & S & 0 & \dots & 0 \\ Sx_2 & 0 & S & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ Sx_{T-1} & 0 & 0 & \dots & S \\ Sx_T & 0 & 0 & \dots & 0 \end{bmatrix} \tag{1}$$

$$\equiv \quad N'(t)X, \tag{2}$$

where $'$ denotes transpose. The transition matrix X is the *Leslie matrix* (Leslie [12, 13]). The population at time t is given by

$$N'(t) = N'(0)X^t,$$

The fate of the population thus hinges on the properties of X^t. We assume that the Leslie matrix X is primitive, in that there exists some $k > 0$ for which X^k is strictly positive.[8] This allows us to bring standard results in matrix theory to bear in examining X^t.

2.2 Discounted expected utility

The Perron-Frobenius theorem (Seneta [19, Theorem 1.1]) implies that the Leslie matrix has a "dominant" eigenvalue ϕ that is real, positive, of multiplicity one, and that strictly exceeds the modulus of all other eigenvalues. This eigenvalue is the population growth factor, and its log is the corresponding growth rate, in that (Seneta [19, Theorem 1.2])

$$\lim_{t\to\infty} \frac{X^t}{\phi^t} = vu' \qquad \text{and hence} \qquad \lim_{t\to\infty} \frac{N'(t)}{\phi^t} = N'(0)vu',$$

where the vectors u and v are the strictly positive left ($u'X = \phi u'$) and right ($Xv = \phi v$) eigenvectors associated with ϕ, normalized so that $u'v = 1$ and $\sum_{\tau=1}^{T} u_\tau = 1$. Regardless of the initial condition $N'(0)$, the proportion of the population of each age τ approaches u_τ. The vector u thus describes the limiting age distribution of the population. The vector v then gives the

---

[8] A sufficient condition for this is that there exist two relatively prime ages τ and τ' for which x_τ and $x_{\tau'}$ are both nonzero. It suffices, for example, that τ and τ' are adjacent. Note that $x_T > 0$ by assumption, since otherwise agents of age T would be past reproductive age and removed from our consideration.

"reproductive value" of an individual of each age, or the relative contribution that each such individual makes to the long run population.

The long-run growth rate of the total reproductive value of the population, can be obtained by premultiplying the first equation by the vector u', postmultiplying by v and then taking logs so that

$$\lim_{t\to\infty} \frac{1}{t} \ln(u' X^t v) = \ln \phi. \tag{3}$$

Evolution must select for behavior that maximizes the eigenvalue ϕ, or equivalently, that maximizes the long-run growth growth rate $\ln \phi$. This eigenvalue solves the characteristic equation[9]

$$\Phi = x_1 + \frac{x_2}{\Phi} + \frac{x_3}{\Phi^2} + \ldots + \frac{x_T}{\Phi^{T-1}}, \tag{4}$$

where

$$\Phi = \frac{\phi}{S}.$$

Evolution would endow an agent with preferences (or more precisely, would endow an agent with behavior consistent with such preferences) whose indifference curves are described by the right side of (4). In particular, choices $(x_1, \ldots, x_T)$ that preserve the right side of (4) (given Φ) give rise to this same value of Φ and hence the same growth rate $\ln \phi$, while choices leading to smaller values of the right side lead to smaller growth rates.

We can readily derive marginal rates of substitution from (4) of the form:

$$\frac{dx_{\tau+1}}{dx_\tau} = \Phi. \tag{5}$$

Marginal rates of substitution between $x_{\tau+1}$ and x_τ are thus independent of τ and independent of the magnitudes of $x_{\tau+1}$ and x_τ. Equivalently, we can capture the preferences represented by (4) in a utility function of the form:

$$U(x_1, \ldots, x_T) = \sum_{\tau=1}^{T} \Phi^{-(\tau-1)} x_\tau = \sum_{\tau=1}^{T} e^{-(\ln \phi - \ln S)(\tau-1)} x_\tau. \tag{6}$$

---

[9] This is a rearrangement of

$$\begin{vmatrix} Sx_1 - \phi & S & 0 & \ldots & 0 \\ Sx_2 & -\phi & S & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ Sx_{T-1} & 0 & 0 & \ldots & S \\ Sx_T & 0 & 0 & \ldots & -\phi \end{vmatrix} = 0.$$

The agent thus discounts exponentially at the rate $\ln \Phi$, that is, at the sum of the population growth rate ($\ln \phi$) and the death rate ($-\ln S$).[10] This exponential discounting has an intuitive interpretation. As one delays a birth, one falls behind the rest of the population at rate $\ln \Phi$, since one's death occurs at rate $-\ln S$ and the rest of the population is growing at rate $\ln \phi$. The delay must then be compensated by an increment in births sufficient to balance this loss. At the same time, notice that changes in the death rate will have no effect on the discount rate, since they will be matched by equivalent changes in the population growth rate that preserve their sum.

In the environment described by this simple model, we would observe only one equilibrium profile $(x_1, \ldots, x_T)$, from which we could infer marginal rates of substitution and hence discount factors (via (4) and (5)) and conclude that behavior is consistent with the utility function given in (6). To generalize this result, suppose that newborn agents are independently (across time and agents) assigned a feasible set $X^T \subset \Re^T_+$ of possible reproductive profiles. Some agents may find themselves in more favorable circumstances than others, with more favorable prospects for reproduction. We would then observe a potentially vast variety of choices, all consistent with preferences that are again described by (4), with Φ now being the average population growth rate. We could again infer that marginal rates of substitution are given by (5) and the utility function given by (6).

3 Aggregate Uncertainty

We now examine the case of aggregate uncertainty. There are a number of ways such uncertainty might matter, but we focus on the particularly salient possibility that death rates may have a common component across individuals. Perhaps a particularly severe winter or dry summer decreases all survival probabilities, or a good growing season for food or an epidemic among predators increases them. On top of this, we will then also allow these aggregate fluctuations to have varying effects on agents of different ages. An infestation of predators may especially affect younger agents, for example, or an epidemic may disproportionately affect older agents.

---

[10] We can write the survival probability from one period to the next as $S = e^{-\delta}$, where δ is the continuously compounded death rate, and then take logs to express the death rate as $\delta = -\ln S$.

3.1 Why does aggregate uncertainty matter?

Why does it make a difference whether uncertainty is aggregate or idiosyncratic? It is helpful here to consider the model of Robson [16], in which the population has a trivial age structure. Agents survive from age zero to age one with probability S. At age one they have x expected offspring and then die. With purely idiosyncratic uncertainty, the population size $N(t)$ in period t is given by

$$N(t) = (Sx)N(t-1) = (Sx)^t N(0).$$

Hence the growth rate is $\ln Sx$ (and Sx is the leading eigenvector ϕ of the trivial Leslie matrix $[Sx]$).

Now suppose that instead of a fraction S of the agents surviving from age zero to age one, an independent random draw in each period determines whether all agents survive or all perish, with the probability of survival being S. This shift from idiosyncratic to aggregate uncertainty leaves *expected* survival rates untouched but has a profound effect on the population, whose fate is now eventual extinction with probability one.

More generally, suppose that with probability p a fraction $S_1 > 0$ of the agents survive, while with probability $1-p$ fraction $S_2 > 0$ survive, where $pS_1 + (1-p)S_2 = S$. The draws determining the fraction of surviving agents are independent over time. It follows that the population size in period t is given by

$$N(t) = (S_1 x)^{n(t)} (S_2 x)^{t-n(t)} N(0),$$

where $n(t)$ is the random number of times that the environment is such that fraction S_1 survive. The asymptotic growth rate of the population is then, almost surely,

$$\begin{aligned}
\lim_{t\to\infty} \frac{1}{t} \ln N(t) &= \lim_{t\to\infty} \left(\frac{n(t)}{t} \ln(S_1 x) + \frac{t-n(t)}{t} \ln(S_2 x) + \frac{\ln N(0)}{t} \right) \\
&= p \ln(S_1 x) + (1-p) \ln(S_2 x) \\
&< \ln(Sx).
\end{aligned}$$

The growth factor under idiosyncratic uncertainty is the arithmetic mean, Sx, while the growth factor under comparable aggregate uncertainty is the geometric mean, $(S_1 x)^p (S_2 x)^{1-p} < Sx$. This finding that the growth rate in unstructured populations under aggregate uncertainty falls short of the growth rate in the corresponding mean economy is general, depending on none of the specific structure of this example.

3.2 Aggregate uncertainty in age-structured population

Our task now is to extend the model of aggregate uncertainty to age-structured populations. Let $\tilde{S}_\tau$ be a random variable giving the probability that an agent of age $\tau \in \{0, \dots, T-1\}$ survives until the next period, with mean S. Hence, we think of each agent of age $\tau \in \{1, \dots, T-1\}$ as first receiving a common realization $\tilde{S}_\tau$ with support in $(0, 1]$, identifying the probability that this agent will survive until the next period, from a distribution with mean S. The agent then takes an idiosyncratic draw from a Bernoulli random variable that gives survival with probability $\tilde{S}_\tau$ and death otherwise. Draws of $\tilde{S}_\tau$ are independent over time.

The *mean Leslie matrix* is given by

$$X = \begin{bmatrix} Sx_1 & S & 0 & \dots & 0 \\ Sx_2 & 0 & S & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ Sx_{T-1} & 0 & 0 & \dots & S \\ Sx_T & 0 & 0 & \dots & 0 \end{bmatrix}, \tag{7}$$

and we continue to let ϕ denote the leading eigenvalue of this matrix, so that $\ln \phi$ is the population growth rate that would prevail in a population with the same mean behavior but no aggregate uncertainty. The Leslie matrix in period t is a random variable denoted by

$$\tilde{X}(t) = \begin{bmatrix} x_1\tilde{S}_0(t) & \tilde{S}_1(t) & 0 & \dots & 0 \\ x_2\tilde{S}_0(t) & 0 & \tilde{S}_2(t) & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ x_{T-1}\tilde{S}_0(t) & 0 & 0 & \dots & \tilde{S}_{T-1}(t) \\ x_T\tilde{S}_0(t) & 0 & 0 & \dots & 0 \end{bmatrix}. \tag{8}$$

Analogously to (3), we are interested in the growth rate

$$\lim_{t \to \infty} \frac{1}{t} \ln u' \tilde{X}(1) \dots \tilde{X}(t) v,$$

where u and v are the eigenvectors associated with the mean Leslie matrix X. We can interpret this as the long-run growth rate of total reproductive value, evaluated with the population proportions u and reproductive values v from the mean Leslie matrix.[11] This is now a product of random matrices. Not only can we not apply the Perron-Frobenius theorem, but it is

---

[11] No difficulty arises out of using the eigenvectors u and v from the mean Leslie matrix in forming this measure of the growth rate. Proposition 1 below holds for any norm

no longer obvious that the limit exists. Fortunately, we have the following remarkable result (first established by Furstenberg and Kesten [6, Theorem 2] and extended in Tanny [21, Theorem 7.1]):[12]

Proposition 1 *If* $E \ln u'\tilde{X}v < \infty$*, then there exists a finite* $\lambda \in \Re_{++}$ *such that, almost surely,*

$$\lim_{t\to\infty} \frac{1}{t} \ln u'\tilde{X}(1)\dots\tilde{X}(t)v = \ln\lambda.$$

We refer to $\ln\lambda$ as the growth rate under aggregate uncertainty.

3.3 Aggregate uncertainty slows growth

Our first result is a generalization to age-structured populations of the finding that aggregate uncertainty slows the population's growth rate.[13] Section 5.1 proves:

Proposition 2

$$\lambda \le \phi.$$

Hence, aggregate uncertainty can only reduce the growth rate.[14] "Generically," it does so strictly, in the sense that the growth rate is lower under aggregate uncertainty unless all of the possible realized Leslie matrices happen to have the same leading eigenvalue and associated left eigenvector. The following example illustrates this latter possibility.

$||\tilde{X}(1)\dots\tilde{X}(t)||$ (indeed for any positive submultiplicative function of $\tilde{X}(1)\dots\tilde{X}(t)$). We retain our assumption that the mean Leslie matrix X is primitive. Together with the restriction $\tilde{S} > 0$ (the population avoids extinction), this ensures that asymptotically, all elements of $\tilde{X}(1)\dots\tilde{X}(t)$ grow at the same rate.

[12] Taking each $\tilde{S}_\tau = 1$ gives us an upper bound an $u'\tilde{X}v$, ensuring that we satisfy the sufficient condition $E\ln u'\tilde{X}v < \infty$.

[13] This result depends on the assumption that the idiosyncratic uncertainty is independent across periods. For example, a stochastic environment in which the Leslie matrices X_1 and X_2 alternate gives a higher population growth rate than does the mean Leslie matrix X, where

$$X_1 = \begin{bmatrix} 0 & 1 \\ 8 & 0 \end{bmatrix} \qquad X_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad X = \begin{bmatrix} 0 & 1 \\ 4 & 0 \end{bmatrix}.$$

[14] See Curry [2], Gillespie [7], and Houston and McNamara [10, Chapter 10] (as well as Robson [16]) for similar results for the case of $T = 1$.

Example 1 *Suppose there are two equally likely Leslie matrices, X' and X'', with mean matrix X, given by*

$$X' = \begin{bmatrix} x & 0 \\ 0 & 0 \end{bmatrix} \qquad X'' = \begin{bmatrix} 0 & 0 \\ x^2 & 0 \end{bmatrix} \qquad X = \begin{bmatrix} \frac{x}{2} & 1 \\ \frac{x^2}{2} & 0 \end{bmatrix}.$$

In each period, the realized Leslie matrix is independently drawn to be either X' or X''. The mean matrix X has leading eigenvalue x (and hence growth rate $\ln x$), left eigenvector $u' = \left[\frac{x}{1+x}, \frac{1}{1+x}\right]$, and right eigenvector $v = \left[\frac{2(1+x)}{3x}, \frac{1+x}{3}\right]$. The matrices X' and X'' each have the same leading eigenvalue and left eigenvector. For any t, any product of the form $u'X(1)X(2)\ldots X(t)v$, where each $X(t')$ is either X' or X'', gives the same outcome. As a result, the growth rate without aggregate uncertainty (i.e., with $X(t') = X$ for all t') matches that with aggregate uncertainty.

3.4 Common survival rates

Perhaps the most natural case to consider is that in which the aggregate shocks affect the survival rates of all ages equally.

Proposition 3 *Let the random variables $\tilde{S}_0, \ldots, \tilde{S}_{T-1}$ be identical. Then evolution selects for preferences under which*

$$\frac{dx_{\tau+1}}{dx_\tau} = \Phi = \frac{\phi}{S}. \tag{9}$$

As before, ϕ and S are the leading eigenvalue and mean death rate associated with the mean Leslie matrix (7). Comparing with (5), we thus see that aggregate uncertainty has no effect on marginal rates of substitution, and hence discounting. At the same time, it decreases the growth rate if the random variables $\tilde{S}_\tau$ are nondegenerate (to $\ln \lambda < \ln \phi$; cf. Proposition 2). Under aggregate uncertainty, the discount rate will thus exceed the sum of the actual growth rate and the mean death rate.

Proof. Let $\tilde{S}(t)$ denote the common realization in period t of the random variables $\tilde{S}_0, \ldots, \tilde{S}_{T-1}$. Then, almost surely

$$\begin{aligned}
\ln\lambda &= \lim_{t\to\infty}\frac{1}{t}\ln\left(u'\tilde{X}(1)\ldots\tilde{X}(t)v\right) = \lim_{t\to\infty}\frac{1}{t}\ln\left(\frac{\tilde{S}(1)}{S}\ldots\frac{\tilde{S}(t)}{S}u'X^t v\right) \\
&= \ln\phi + \lim_{t\to\infty}\frac{1}{t}\ln\left(\frac{\tilde{S}(1)}{S}\ldots\frac{\tilde{S}(t)}{S}\right) \\
&= \ln\phi + E\ln\tilde{S} - \ln S \\
&< \ln\phi,
\end{aligned}$$

where the inequality follows from the fact that $E\tilde{S} = S$ and hence $E\ln\tilde{S} < \ln S$. Since the fertilities $(x_1, \ldots, x_T)$ appear only in $\ln\phi$, the arguments of Section 2.2 ensure that evolution will select for marginal rates of substitution given by (9). ■

Intuitively, shocks that are common across ages distort none of the intertemporal trade-offs captured by the marginal rate of substitution. The marginal rate of substitution and hence the discount rate is then fixed at the specification appropriate for the mean Leslie matrix. If the aggregate uncertainty is severe, the growth rate λ may fall well short of ϕ, giving us discounting at a rate significantly exceeding the the sum of the growth rate and the mean death rate. Hence, as long as our ancestral environment featured aggregate uncertainty, there is no puzzle in our having evolved to have discount rates higher than can be justified on the basis of the long-run average population growth rate and mean death rates.

Gurven and Kaplan [8, pp. 346–348] note that contemporary hunter-gatherer groups often exhibit annual growth rates in excess of two percent, considerably higher than the approximately zero growth rate that prevailed over the vast bulk of our evolutionary history. They suggest two explanations. First, contemporary hunter-gatherers may not reflect our evolutionary past. Second, population dynamics may exhibit a saw-tooth pattern, with intermixed periods of relatively strong growth and occasional and perhaps quite rapid population crashes, and with the former bound to be disproportionately represented among contemporary data. As long as the population crashes are evolutionarily neutral, in that they do not change the population age structure, this argument is formalized and generalized by the model presented in this section. The rare and rapid population crashes could keep long-term growth rates hovering near zero, while the marginal rate of substitution would be adapted to the (possibly much higher) growth rate of the mean Leslie matrix.

3.5 Imperfectly correlated survival rates

We now turn to the case in which fluctuations in the aggregate environment have potentially different effects on the survival of different ages. In doing so, our attention turns from the level to the pattern of discounting. Our general finding is that imperfectly correlated survival rates push marginal rates of substitution away from exponential discounting. The nature of the departure from exponential discounting depends on the precise nature of the aggregate uncertainty. We explore here a plausible case that gives rise to a present bias.

We model survival rates as being affected by relatively small age-specific perturbations around a age-independent common shock. This allows environmental fluctuations to have different effects on agents of different ages, while still capturing our intuition that there will be considerable commonality across ages in survival rates. Our focus on small age-specific perturbations also allows us to use a convenient approximation method for the analysis.

As before, a random variable $\tilde{S}(t)$ is drawn in each period t, identically and independently distributed over time, with mean S. In the proportion $1-\epsilon$ of the population, each individual then receives an idiosyncratic draw giving a survival with probability $\tilde{S}$ and death otherwise. In addition, random variables $(\hat{S}_0, \ldots, \hat{S}_{T-1})$ are also drawn each period, again identically and independently distributed over time, each with mean 0. For the remaining ϵ proportion of the population, each agent of age τ then obtains an idiosyncratic draw giving survival with probability $\tilde{S}+\hat{S}_\tau$ and death otherwise (with $\tilde{S}+\hat{S}_\tau$ having support in $(0,1]$). The larger is ϵ, the greater is the variation across ages in the aggregate death rate. We consider the case of small ϵ.

3.5.1 An approximation

To introduce the approximation method with which we work,[15] it is useful to return to the case of a trivial age structure presented in Section 3.1. Suppose the survival proportion in period t is given by

$$(1-\epsilon)S+\epsilon\hat{S}(t), \tag{10}$$

---

[15] Proposition 2 assures us that the limit of the product of random Leslie matrices describing the population exists, but there are no general methods for calculating this limit and characterizing its dependence on the elements of the Leslie matrix or parameters describing the stochastic process.

where S is the mean survival rate (i.e., $\tilde{S}(t)$ is degenerate) and $\hat{S}(t)$ is a random variable drawn independently across periods and (for the purposes of this example only) equally likely to take the values $S+\delta$ or $S-\delta$. Hence, over the course of many periods, the realized survival proportion will be $S-\epsilon\delta$ about half the time and $S+\epsilon\delta$ about half the time. The counterpart of (3) is then

$$\lim_{t\to\infty}\frac{1}{t}\ln\frac{N(t)}{N(0)}=\frac{1}{2}\ln((S-\epsilon\delta)x)+\frac{1}{2}\ln((S+\epsilon\delta)x). \tag{11}$$

We can take a Taylor expansion of the right side of (11) around $\epsilon=0$ to obtain

$$\begin{aligned}\lim_{t\to\infty}\frac{1}{t}\ln\frac{N(t)}{N(0)} &= \frac{1}{2}\left(\ln(Sx)-\epsilon\frac{\delta x}{Sx}-\frac{\epsilon^2}{2}\frac{(\delta x)^2}{(Sx)^2}\right)+\frac{1}{2}\left(\ln(Sx)+\epsilon\frac{\delta x}{Sx}-\frac{\epsilon^2}{2}\frac{(\delta x)^2}{(Sx)^2}\right)+O(\epsilon^3)\\ &= \ln\phi-\frac{\epsilon^2}{2\phi^2}x^2V+O(\epsilon^3),\end{aligned} \tag{12}$$

where $V=\delta^2$ is the variance of $\hat{S}(t)$ and hence x^2V is the variance of the growth factor. The growth rate of the population under aggregate uncertainty thus falls short of the corresponding "idiosyncratic" growth rate $\ln\phi$ by a term that is proportional to the variance of the aggregate survival factor. Increasing x increases the idiosyncratic growth rate $\ln\phi$, but also increases the relevant variance and hence increases the extent to which the actual growth rate falls short of $\ln\phi$.

Because the age structure in this example is trivial, we could just as well have worked with an explicit solution such as (11) and avoided the approximation altogether. However, a similar approximation is vital to our consideration of nontrivial age structures.

To examine age-structured populations, we can write the realized Leslie matrix for period t as

$$\tilde{Z}(t)=\tilde{X}(t)+\epsilon\tilde{H}(t), \tag{13}$$

where $\tilde{X}(t)$ is the commonly perturbed Leslie matrix as in (8), under the assumption that the $\tilde{S}_\tau$ are identical, and $\tilde{H}(t)$ is the perturbation matrix

$$\tilde{H}(t)=\begin{bmatrix} x_1\hat{S}_0(t) & \hat{S}_1(t) & 0 & \dots & 0\\ x_2\hat{S}_0(t) & 0 & \hat{S}_2(t) & \dots & 0\\ \vdots & \vdots & \vdots & & \vdots\\ x_{T-1}\hat{S}_0(t) & 0 & 0 & \dots & \hat{S}_{T-1}(t)\\ x_T\hat{S}_0(t) & 0 & 0 & \dots & 0\end{bmatrix}.$$

Each of the random variables in the matrix $\tilde{H}(t)$ has a zero mean. We use the following generalization (cf. Tuljapurkar [22, Chapter 12]) of (12) to age-structured populations (with the proof in Section 5):

Proposition 4 *Suppose the matrices $\tilde{H}(t)$ in (13) are independent across periods and have a zero expected value. Then, almost surely,*

$$\begin{aligned} &\lim_{t\to\infty} \frac{1}{t} \ln u' \tilde{Z}(1) \dots \tilde{Z}(t) v \\ = \quad & \ln \phi + E \ln \tilde{S} - \ln S - \frac{\epsilon^2}{2\phi^2} E \left(\frac{S}{\tilde{S}} \right)^2 E\{(u' \tilde{H} v)^2\} + O(\epsilon^3). \end{aligned} \tag{14}$$

The term $E\{(u'\tilde{H}v)^2\}$ is the variance of the growth factor of total reproductive value, evaluated in the long run using the population proportions u and reproductive values v derived from the mean Leslie matrix.

Revisiting some previous points, it is immediate that $E\{(u'\tilde{H}v)^2\} \geq 0$, and hence that introducing variation in the effects of aggregate uncertainty across ages cannot increase the population growth rate. Independence across the various components of $\tilde{H}$ is one formulation ensuring that $E\{(u'\tilde{H}v)^2\} > 0$ and hence that variation in aggregate uncertainty slows growth. In Example 1, we have $\tilde{X}(t) = X$ for all t, $H(t)$ equals either X' or X'', $\epsilon = 1$, $E \ln \tilde{S} - \ln S = 0$ and $E\{(u'Hv)^2\} = 0$.

3.5.2 Marginal rates of substitution

When aggregate effects on survival vary across ages in a symmetric way, marginal rates of substitution fall short of their value without such variance (while still exceeding the sum of the growth rate and the mean mortality rate, if ϵ is small) and now also decline over time:

Proposition 5 *Suppose the random variables $(\hat{S}_0, \dots \hat{S}_{T-1})$ share common variance V and common covariances C. Then for sufficiently small ϵ, the marginal rate of substitution falls short of the value Φ it would take aggregate shocks were identical across ages and is decreasing in τ, i.e.,*

$$\frac{dx_{\tau+1}}{dx_\tau} \leq \frac{dx_{\tau+2}}{dx_{\tau+1}} \leq \Phi$$

both strictly so if $x_{\tau+1} > 0$ and $C < V$.

The random shocks $\hat{S}_\tau$ to the survival probabilities may range from being independent across agents ($C = 0$) to being perfectly correlated ($C = V$) (notice that, necessarily, $C \leq V$). As long as the aggregate shocks are not perfectly correlated across ages, marginal rates of substitution are decreasing in τ, i.e., intertemporal preferences exhibit a present bias.

The common-variance and common-covariance assumptions are sufficient but not necessary for this result. It is clear that this present bias will continue to obtain as long as the distributions of the various aggregate shocks are not too dissimilar. Indeed, the method of proof can be applied to ascertain the implications of any configuration of distributions, though with possibly much more tedious calculations.

3.5.3 Why not exponential?

What lies behind these results? As in the one-dimensional case in (12), the growth rate under aggregate uncertainty is given by the growth rate under certainty minus a term reflecting the variance of the long run growth factor of total reproductive value (cf. (14)). The issue then becomes how the variances of the shocks to the survival rates of various ages combine to yield the variance of this growth factor.

To provide the simplest illustration of these complex variance effects, let $T = 3$ and let $\tilde{S} = S \in (0,1)$ with probability one, so there is no common component to the aggregate shocks to survival and $\Phi = \phi/\tilde{S}$. Assume also that the age-dependent aggregate shocks are contemporaneously independent and share common variances $V_0 = V_1 = V_2 = V$. The variance component $\{(u'\tilde{H}v)\}$ in (14) is then[16]

$$\begin{aligned}
u'\tilde{H}v &= [u_1, u_2, u_3] \begin{bmatrix} \hat{S}_0 x_1 & \hat{S}_1 & 0 \\ \hat{S}_0 x_2 & 0 & \hat{S}_2 \\ \hat{S}_0 x_2 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \\
&= v_1 \sum_{\tau=1}^{3} u_\tau x_\tau \hat{S}_0 + v_2 u_1 \hat{S}_1 + v_3 u_2 \hat{S}_2 \\
&= v_1 u_1 \Phi \hat{S}_0 + v_2 u_1 \hat{S}_1 + v_3 u_2 \hat{S}_2
\end{aligned}$$

Squaring and taking the expectation, using the independence of the aggre-

---

[16] The last equality uses (4) and $u_\tau = \Phi u_{\tau+1}$.

gate shocks across ages, we have[17]

$$\begin{aligned} E(u'\tilde{H}v)^2 &= v_1^2 u_1^2 \Phi^2 V + v_2^2 u_1^2 V + v_3^2 u_2^2 V \\ &= V u_1^2 \Phi^2 \left[v_1^2 + \frac{v_2^2}{\Phi^2} + \frac{v_3^2}{\Phi^4} \right] \qquad (15) \end{aligned}$$

If we convert the reproductive values v_2 and v_3 to their age-one equivalents, we find[18]

$$v_2 = \left(\frac{x_2}{\Phi} + \frac{x_3}{\Phi^2} \right) v_1, \qquad v_3 = \frac{x_3}{\Phi} v_1, \qquad (16)$$

so that

$$E(u'\tilde{H}v)^2 = V u_1^2 v_1^2 \Phi^2 \left[1 + \frac{1}{\Phi^2} \left(\frac{x_2}{\Phi} + \frac{x_3}{\Phi^2} \right)^2 + \frac{1}{\Phi^4} \left(\frac{x_3}{\Phi} \right)^2 \right].$$

This expression immediately suggests that we should expect to be pushed away from exponential discounting. We have values of x_2 and x_3 discounted by a variety of powers of ϕ in addition to the exponential discounting appearing in (4). To pursue the details, note that the long run growth rate is given by the expression

$$\Lambda = \ln \lambda = \ln \phi - \frac{\epsilon^2}{2} E\left(u'\tilde{H}v \right)^2 = \ln \phi - \frac{\epsilon^2 V u_1^2 v_1^2}{2S^2} \left[1 + \frac{1}{\Phi^2} \left(\frac{x_2}{\Phi} + \frac{x_3}{\Phi^2} \right)^2 + \frac{1}{\Phi^4} \left(\frac{x_3}{\Phi} \right)^2 \right].$$

We have a present bias if and only if (letting $d\Lambda/dx_\tau = \Lambda_\tau$)

$$\frac{\Lambda_1}{\Lambda_2} > \frac{\Lambda_2}{\Lambda_3} \quad \text{or} \quad \Lambda_2^2 < \Lambda_1 \Lambda_3.$$

In order to find complete expressions for these derivatives, we must account for the dependence of the endogenous variables Φ, u_1, and v_1 on x_1 x_2 and x_3. However, it is only the explicit dependence of Λ on x_2 and x_3 that introduces a distortion away from exponential discounting. In addition, this explicit dependence affects only the term of order ϵ^2. Letting $\phi = 1$ for simplicity, this allows us to calculate that we have a present bias if and only if

$$\frac{d}{dx_2} \left[1 + (x_2 + x_3)^2 + (x_3)^2 \right] > \frac{1}{2} \frac{d}{dx_3} \left[1 + (x_2 + x_3)^2 + (x_3)^2 \right]$$

From (16), this condition is equivalent to $v_2 > v_3$, which holds as long as $x_2 > 0$ (and thus the condition $x_{\tau+1} > 0$ in the statement of Proposition

---

[17] The second equality uses $u_\tau = \Phi u_{\tau+1}$.

[18] For example, 1-period-olds produce a total of x_1 1-period-olds one period later (worth $v_1 \Phi$), x_2 one-period-olds two periods later (worth v_1/Φ^2), and x_3 1-period-olds two periods later (worth $v_1 \Phi^3$).

5). Hence, when $\Phi = 1$, the presence of a present bias is equivalent to the condition that reproductive values decline with age. This decline in turn reflects the fertility x_2 available to a two-period-old agent that is lost to a three-period-old agent. An analogous but slightly more complex argument yields a similarly unambiguous result when $\Phi \neq 1$.

3.6 Infant mortality varying by parental age

Our first message was that aggregate uncertainty drives a wedge between discount rates and the sum of the population growth and mortality rates. On top of this, we have now seen that aggregate uncertainty can push discounting away from the exponential pattern of discounted expected utility. We view the conditions under which we have obtained a present bias (Proposition 5) as being quite plausible, but we offer here a demonstration that there are clearly limitations to the generality of this result.

Suppose that newborns whose parents are of different ages have different infant mortality rates. For example, older parents may be larger and better able to nourish themselves, in turn allowing them to produce larger or better-nourished offspring (cf. Charlesworth [1, Chapter 5]). If these infant mortality rates were idiosyncratic, there would be no difficulty in simply folding them into the values x_τ, with no other change in the analysis. However, the case that these newborn survival rates are subject to aggregate uncertainty requires a new analysis.

To isolate the effects of this uncertainty, we assume that $\tilde{S}(t)$ is degenerate, that there is no aggregate randomness in other survival rates, and that parent age has no impact lasting beyond infant mortality. These are clearly restrictive assumptions. Our purpose is to illustrate that the structure of discounting depends on the underlying stochastic specification, rather than to advance the model being developed here (and its attendant future-biased discounting) as being realistic.

We can again write the realized Leslie matrix for period t as in (13), with $\tilde{X}(t)$ given by X from (7) for each t and with the perturbation matrix $\tilde{H}(t)$ now given by

$$\begin{bmatrix} x_1\hat{\mathbb{S}}_1(t) & 0 & 0 & \dots & 0 \\ x_2\hat{\mathbb{S}}_2(t) & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ x_{T-1}\hat{\mathbb{S}}_{T-1}(t) & 0 & 0 & \dots & 0 \\ x_T\hat{\mathbb{S}}_T(t) & 0 & 0 & \dots & 0 \end{bmatrix}, \tag{17}$$

where each of the random variables $\mathbb{S}_\tau$ in the matrix $\tilde{H}(t)$ again has a zero mean. We have:[19]

Proposition 6 *Let* $x_1 = x_2 = \ldots = x_T \equiv x$. *Let* $(\hat{\mathbb{S}}_1(t), \ldots, \hat{\mathbb{S}}_T(t))$ *share common variance* V *and common covariances* C. *Then for small* ϵ

$$\frac{dx_{\tau+1}}{dx_\tau}$$

is increasing in τ, *strictly so if* $\Phi \neq 1$ *(i.e. the mean Leslie matrix gives rise to nontrivial discounting) and* $C < V$.

4 Discussion

We have framed our discussion in terms of marginal rates of substitution, and hence discount rates, over fertility at different ages. In practice, we are interested in the intertemporal tradeoffs involved in all sorts of consumption decisions, only a few of them explicitly involving the timing of births. Our view here is that reproduction is the currency of evolution, and hence the foundation for evolved preferences. The next step is to consider how fertility x_τ is determined by consumption at age τ and previous ages, with these consumption decisions inheriting the discounting induced by their implications for reproduction. Again, there is no reason to believe that people explicitly calculate the implications for reproduction when deciding whether to consume now or later, but there is good reason to believe that their preferences over such trade-offs are shaped by evolutionary forces ultimately rooted in reproduction. Extending the analysis from reproduction to consumption is straightforward if reproduction at age τ is primarily a function of consumption at age τ and becomes more complicated as we move away from this simple case.

We have found that evolutionarily-induced intertemporal preferences may exhibit a present bias. However, this present bias does not lead to preference reversals. The marginal rate of substitution between $x_{\tau+1}$ and x_τ may decline in τ, but this decline is linked to age and not to time relative to the present. A tradeoff between x_9 and x_{10} that confers evolutionary advantages when made at age 1 will still confer such advantages when made

---

[19] This result examines a symmetric setting in which $x_1 = x_2 = \ldots = x_T \equiv x$. When uncertainty is idiosyncratic, the marginal rate of substitution between x_τ and $x_{\tau'}$ is independent of the levels of x_τ and $x_{\tau'}$ (cf. (5)), but this need no longer be the case with aggregate uncertainty. Setting $x_1 = x_2 = \ldots = x_T \equiv x$ is the obvious way to isolate systematic preferences over timing.

at age 5 or at age 9. A 1-period-old will accordingly make intertemporal choices that cannot be rationalized by exponential discounting, but will not reverse those choices later.

Dasgupta and Maskin [3] and Sozou [20] also present evolutionary models leading to a present bias in discounting, including in Dasgupta and Maskin's case the prospect of preference reversals. The force driving discounting in both models is the prospect that an opportunity for future consumption may disappear before it can be realized. A source of future food may be seized by a hungry rival or access blocked by a predator.[20] We have no doubt that uncertainty is an important element of intertemporal decision making, but have two compelling reasons for not proceeding in a similar fashion. First, these models assume that the basic evolutionary goal is to maximize total undiscounted consumption. In contrast, we derive the appropriate basic goal from a more primitive analysis of population growth rates. Indeed, our analysis implies that future consumption will be discounted even if there is no uncertainty at all. Second, we wish to maintain the conventional dividing line between our preferences and the feasible sets over which these preferences are defined. Dasgupta and Maskin suppose, on the other hand, that evolutionarily important feasibility considerations were built into our preferences, so that contemporary choices between goods are evaluated as if they are choices between their uncertainty-adjusted evolutionary equivalents. Evolution may have have endowed us with such preferences, but it is important to check whether such a hypothesis is necessary in explaining our intertemporal behavior. Our inclination is accordingly to begin by examining discounting over consumption opportunities that are not subject to risk, allowing us to isolate rates of time preference.

Our analysis is based on an age-independent mortality rate, best interpreted as the "background" death risk that is relatively constant throughout our adult lives.[21] However, we would expect discounting to vary systematically over one's life span, especially near the beginning and end. We would then expect discount rates to be relatively high among the young children—who act as if there is no tomorrow—and the elderly, reflecting then the typical human U-shaped mortality pattern. However, intergenerational transfers

---

[20] Discounting is then pushed toward a present bias by the prospect of learning about the hazard rate at which the consumption opportunity disappears (in Sozou [20]) or by the prospect that the consumption opportunity may arrive early (in Dasgupta and Maskin [3]).

[21] Wilson and Daly [23] report that women in Chicago neighborhoods with higher mortality rates tend to reproduce earlier, consistent with the higher discount rates that such mortality rates may induce.

may well blunt the increases in discounting that would otherwise appear once one passes reproductive age, by allowing indirect ways of enhancing effective reproduction by pushing resources into the future. There is thus considerable scope for pushing our model beyond its current narrow focus to capture other considerations.

5 Proofs

5.1 Proof of Proposition 2

Let X be the mean Leslie matrix and let $N(t)$ be the associated population process. Let $\tilde{X}(t)$ be the period-t matrix under aggregate uncertainty, drawn independently across periods according to a measure μ satisfying $E \ln u'\tilde{X}v < \infty$, with

$$\int \tilde{X}(t) d\mu = X.$$

Let $\tilde{N}'(t)$ be a random vector describing the size of each age class in the population at time t under aggregate uncertainty and $N'(t)$ its counterpart under the mean Leslie matrix X.

Our first observation is that

$$E\{\tilde{N}(t)\} = N(t).$$

To see this, notice first that we have

$$E\tilde{N}'(1) = EN'(0)\tilde{X}(1) = N'(0)X = N'(1),$$

with the penultimate inequality following from the fact that each element of X is the expected value of the corresponding element in $\tilde{X}$. Now we construct an argument by induction. Suppose $EN'(0)\tilde{X}(1) \cdots \tilde{X}(t-1) = N'(0)X^{t-1}$. Then

$$E\tilde{N}'(t) = EN'(0)\tilde{X}(1) \cdots \tilde{X}(t) = EN'(0)X^{t-1}\tilde{X}(t) = N'(0)X^t = N'(t),$$

where the second equality follows from the induction hypothesis and the fact that every random variable in the period-t Leslie matrix $\tilde{X}$ is independent of the random variable in the Leslie matrices for periods $1, \ldots, t-1$, and the next equality again follows from the fact that each term in X is the expected value of the corresponding term in $\tilde{X}$.

This gives $E\{\tilde{N}(t)\} = N(t)$ and hence $E\tilde{\mathbb{N}}(t) = \mathbb{N}(t)$, where the latter are the period-t total population sizes under aggregate uncertainty and under the mean matrix, respectively. We can then apply Jensen's inequality to show that this expectation is never higher under aggregate uncertainty than under the corresponding deterministic process:

$$\frac{\ln \mathbb{N}(t)}{t} = \frac{\ln E\{\tilde{\mathbb{N}}(t)\}}{t} \geq \frac{E\{\ln \tilde{\mathbb{N}}(t)\}}{t}.$$

The inequality is strict if the distribution of $\tilde{X}(t)$ is nondegenerate, in the sense that different realizations of $\tilde{X}$ have different left eigenvectors or different eigenvalues. The argument is completed by noting that the long-run average growth rate under under the mean matrix is $\lim_{\lim_{t\to\infty}} \frac{\mathbb{N}(t)}{t}$ and under aggregate uncertainty is $\lim_{t\to\infty}\{\ln \tilde{\mathbb{N}}(t)\}/t$, and that almost surely $\lim_{t\to\infty} \frac{\ln \tilde{\mathbb{N}}(t)}{t} = \lim_{t\to\infty} \frac{E \ln \tilde{\mathbb{N}}(t)}{t}$.

5.2 Proof of Proposition 4

Define,

$$F(t,\varepsilon) = E \ln \left[u'(X(1)+\varepsilon H(1))...(X(t)+\varepsilon H(t))v\right].$$

We then have

$$\frac{dF(t,\varepsilon)}{d\varepsilon} = E\left[\frac{\sum_{i=1}^{t} u'\left[...H(i)...\right]v}{\left[u'(X(1)+\varepsilon H(1))...(X(t)+\varepsilon H(t))v\right]}\right]$$

where $u'\left[...H(i)...\right]v$ is given by

$$u'(X(t)+\varepsilon H(1)) \times \ldots \times H(i) \times \ldots \times (X(t)+\varepsilon H(t))v.$$

Hence, $\sum_{i=1}^{t} u'\left[...H(i)...\right]v$ is the sum of t terms of the form $u'\left[...H(i)...\right]v$, each of which is in turn the product of t matrices, the ith of which is the perturbation matrix $H(i)$, and the remainder of which are realized Leslie matrices of the form $X(j)+\epsilon H(j)$ for $j \neq i$. Similarly,

$$\begin{aligned} &\frac{d^2F(t,\varepsilon)}{d\varepsilon^2} \\ = \quad & 2E\left[\frac{\sum_{j>i} u'\left[...H(i)...H(j)...\right]v}{\left[u'(X(1)+\varepsilon H(1))...(X(t)+\varepsilon H(t))v\right])}\right] - E\left[\frac{\left(\sum_i u'\left[...H(i)...\right]v\right)^2}{(\left[u'(X(1)+\varepsilon H(1))...(X(t)+\varepsilon H(t))v\right])^2}\right]\end{aligned}$$

with analogous notation. Then, using the facts that $u'X = u'\phi$, $Xv = \phi v$ and $EH(t) = 0$, we have

$$\begin{aligned}
\lim_{t\to\infty} \frac{1}{t} F(t,0) &= \ln\phi + E\ln\tilde{S} - \ln S \\
\frac{dF(t,0)}{d\varepsilon} &= 0 \\
\lim_{t\to\infty} \frac{1}{t}\frac{d^2F(t,0)}{d\varepsilon^2} &= -\frac{E\left(u'H(1)v\right)^2}{\phi^2} E\left(\frac{S}{\tilde{S}}\right)^2 .
\end{aligned}$$

A Taylor expansion then gives

$$\begin{aligned}
\lim_{t\to\infty} \frac{1}{t} F(T,\epsilon) &= \lim_{t\to\infty} \frac{1}{t} E\ln\left[u'X(1)\dots X(t)v\right] \\
&= \ln\phi + E\ln\tilde{S} - \ln S - \frac{\epsilon^2}{2} E\left(\frac{S}{\tilde{S}}\right)^2 \frac{E\left(u'H(1)v\right)^2}{\phi^2} + O(\epsilon^3).
\end{aligned}$$

The proof is now completed by noting that, almost surely (Tuljapurkar [22, Chapter 12]),

$$\lim_{t\to\infty} \frac{1}{t} \ln\left[u'Z(1)\dots Z(t)v\right] = \lim_{t\to\infty} \frac{1}{t} E\ln\left[u'Z(1)\dots Z(t)v\right].$$

■

5.3 Proof of Propositions 5 and 6

We begin with a more general structure that provides the foundation for the proof of Propositions 5 and 6. Let the perturbation matrix $H(t)$ be given by

$$\begin{bmatrix}
x_1\hat{\mathbb{S}}_1(t) & \hat{S}_1 & 0 & \dots & 0 \\
x_2\hat{\mathbb{S}}_2(t) & 0 & \hat{S}_2 & \dots & 0 \\
\vdots & \vdots & \vdots & & \vdots \\
x_{T-1}\hat{\mathbb{S}}_{T-1}(t) & 0 & 0 & \dots & \hat{S}_{T-1} \\
x_T\hat{\mathbb{S}}_T(t) & 0 & 0 & \dots & 0
\end{bmatrix}.$$

Let $\Gamma = \lim_{t\to\infty} \frac{1}{t} \ln u'Z(1)\dots Z(t)v$. Let $C_{\tau\tau'}$ denote the covariance between $\hat{S}_\tau$ and $\hat{S}_{\tau'}$, let $\mathbb{C}_{\tau\tau'}$ denote the covariance between $\hat{\mathbb{S}}_\tau$ and $\hat{\mathbb{S}}_{\tau'}$, and let $\mathfrak{C}_{\tau\tau'}$ denote the covariance between $\hat{\mathbb{S}}_\tau$ and $\hat{S}_{\tau'}$. To make the notation more compact, let

$$\begin{aligned}
x_\tau\hat{\mathbb{S}}_\tau &\equiv \hat{\mathbb{Z}}_\tau \\
\hat{S}_\tau &= \hat{Z}_\tau .
\end{aligned}$$

Expanding (14), define Γ to equal

$$\begin{aligned}
& \ln\phi + E\ln\tilde{S} - \ln S - \frac{\epsilon^2}{2\phi^2} E\left(\frac{S}{\tilde{S}}\right)^2 E\left\{\left(\left[\sum_{i=1}^{T} u_i\hat{\mathbb{Z}}_i, u_1\hat{Z}_1, u_2\hat{Z}_2, \ldots, u_{T-1}\hat{Z}_{T-1}\right] v\right)^2\right\} + O(\epsilon^3) \\
= \quad & \ln\phi + E\ln\tilde{S} - \ln S - \frac{\epsilon^2}{2\phi^2} E\left(\frac{S}{\tilde{S}}\right)^2 E\left\{\left(v_1\sum_{i=1}^{T} u_i\hat{\mathbb{Z}}_i + \sum_{i=1}^{T-1} v_{i+1}u_i\hat{Z}_i\right)^2\right\} + O(\epsilon^3) \\
= \quad & \ln\phi + E\ln\tilde{S} - \ln S \\
& -\frac{\epsilon^2}{2\phi^2} E\left(\frac{S}{\tilde{S}}\right)^2 \left(v_1^2\sum_{i=1}^{T}\sum_{j=1}^{T} u_iu_jx_ix_j\mathbb{C}_{ij} + \sum_{i=1}^{T-1}\sum_{j=1}^{T-1} v_{i+1}v_{j+1}u_iu_jC_{ij} + v_1\sum_{i=1}^{T}\sum_{j=1}^{T-1} x_iu_iu_jv_{j+1}\mathfrak{C}_{ij}\right) \\
& + O(\epsilon^3). \qquad (18)
\end{aligned}$$

5.3.1 Proof of Proposition 6

The proof of Proposition 6 is notationally less involved, and so we present this argument first. From (18), we have

$$\begin{aligned}
\Gamma \quad = \quad & \ln\phi + E\ln\tilde{S} - \ln S\phi - \frac{\epsilon^2}{2\phi^2} E\left(\frac{S}{\tilde{S}}\right)^2 \left(v_1^2\sum_{i=1}^{T}\sum_{j=1}^{T} u_iu_jx_ix_j\mathbb{C}_{ij}\right) + O(\epsilon^3) \\
= \quad & \ln\phi + E\ln\tilde{S} - \ln S - \epsilon^2\frac{u_T^2v_1^2}{2\phi^2} E\left(\frac{S}{\tilde{S}}\right)^2 \left[\sum_{i=1}^{T}\sum_{j=1}^{T} \Phi^{2T-i-j}x_ix_j\mathbb{C}_{ij}\right] + O(\epsilon^3),
\end{aligned}$$

where the final equality uses the fact that

$$\begin{aligned}
u_1 &= \frac{\Phi^{T-1}}{\Phi^{T-1} + \ldots + \Phi + 1} \\
u_2 &= \frac{\Phi^{T-2}}{\Phi^{T-1} + \ldots + \Phi + 1} \\
&\vdots \\
u_T &= \frac{1}{\Phi^{T-1} + \ldots + \Phi + 1}
\end{aligned}$$

Now let

$$\begin{aligned}\frac{D\Gamma}{D\phi} &= \frac{1}{\phi} - \epsilon^2 E\left(\frac{S}{\tilde{S}}\right)^2 \left[\frac{u_T v_1^2}{\phi^2}\frac{du_T}{d\phi} + \frac{u_T^2 v_1}{\phi^2}\frac{dv_1}{d\phi} - \frac{u_T^2 v_1^2}{\phi^3}\right]\left[\sum_{i=1}^{T}\sum_{j=1}^{T}\Phi^{2T-i-j}u_i u_j x_i x_j \mathbb{C}_{ij}\right] \\ &\quad - \epsilon^2 \frac{u_T^2 v_1^2}{2\phi^2 S} E\left(\frac{S}{\tilde{S}}\right)^2 \left[\sum_{i=1}^{T}\sum_{j=1}^{T}(2T-i-j)\Phi^{2T-i-j-1}u_i u_j x_i x_j \mathbb{C}_{ij}\right].\end{aligned}$$

Then we can take the derivatives

$$\frac{d\Gamma}{dx_\tau} = \frac{D\Gamma}{D\phi}\frac{d\phi}{dx_\tau} + \frac{d\Gamma}{dv_1}\frac{dv_1}{dx_\tau} + \frac{d\Gamma}{dx_\tau}. \tag{19}$$

This formulation recognizes that u_T depends only on ϕ, while v_1 is given by,

$$v_1 = \frac{\sum_{\tau=1}^{T}\Phi^\tau}{\Phi^T + \sum_{\tau=2}^{T}(\tau-1)x_\tau \Phi^{T-\tau}} = \left(\frac{x_1}{\Phi} + \frac{x_2}{\Phi^2} + \frac{x_3}{\Phi^3}\ldots + \frac{x_{T-1}}{\Phi^{T-1}} + \frac{x_T}{\Phi^T}\right)v_1 \tag{20}$$

$$\begin{aligned} v_2 &= \left(\frac{x_2}{\Phi} + \frac{x_3}{\Phi^2} + \ldots + \frac{x_{T-1}}{\Phi^{T-2}} + \frac{x_T}{\Phi^{T-1}}\right)v_1 \\ &\vdots \\ v_{T-1} &= \left(\frac{x_{T-1}}{\Phi} + \frac{x_T}{\Phi^2}\right)v_1 \\ v_T &= \left(\frac{x_T}{\Phi}\right)v_1. \end{aligned} \tag{21}$$

and hence depends both on ϕ and $(x_1, \ldots, x_T)$.[22] We can calculate:

$$\begin{aligned}\frac{d\Gamma}{dx_1} &= \frac{D\Gamma}{D\phi}\Phi^{T-1}\frac{d\phi}{dx_T} - \epsilon^2\frac{u_T^2 v_1^2}{2\phi^2}E\left(\frac{S}{\tilde{S}}\right)^2\left(2\sum_{j=1}^{T}\Phi^{2T-1-j}x_j\mathbb{C}_{1j}\right) + O(\epsilon^3) \\ &\vdots \\ \frac{d\Gamma}{dx_\tau} &= \frac{D\Gamma}{D\phi}\Phi^{T-\tau}\frac{d\phi}{dx_T} - \epsilon^2\left(\frac{d\Gamma}{dv_1}\frac{1}{\epsilon^2}\right)\frac{(\tau-1)\Phi^{T-\tau}\sum_{\tau=1}^{T}\Phi^\tau}{[\Phi^T + \sum_{\tau=2}^{T}(\tau-1)x_\tau\Phi^{T-\tau}]^2} \\ &\quad - \epsilon^2\frac{u_T^2 v_1^2}{2\phi^2}E\left(\frac{S}{\tilde{S}}\right)^2\left(2\sum_{j=1}^{T}\Phi^{2T-\tau-j}x_j\mathbb{C}_{\tau j}\right) + O(\epsilon^3)\end{aligned}$$

---

[22] The expressions for $v_2, \ldots, v_T$ in terms of v_1 follow from the fact that v is a right eigenvector of the Leslie matrix. The expression for v_1 then follows from the normalization $v'u = 1$.

$$\vdots$$

$$
\begin{aligned}
\frac{d\Gamma}{dx_T} &= \frac{D\Gamma}{D\phi}\frac{d\phi}{dx_T} - \epsilon^2\left(\frac{d\Gamma}{dv_1}\frac{1}{\epsilon^2}\right)\frac{(T-1)\sum_{\tau=1}^{T}\Phi^\tau}{[\Phi^T+\sum_{\tau=2}^{T}(\tau-1)x_\tau\Phi^{T-\tau}]^2} \\
&\quad - \epsilon^2\frac{u_T^2 v_1^2}{2\phi^2}E\left(\frac{S}{\tilde{S}}\right)^2\left(2\sum_{j=1}^{T}\Phi^{2T-T-j}x_j\mathbb{C}_{Tj}\right) + O(\epsilon^3),
\end{aligned}
$$

where we write this so as to take advantage of the fact that $\left(\frac{d\Gamma}{dv_1}\frac{1}{\epsilon^2}\right)$ is of order *zero*. Notice that there is some asymmetry between $D\Gamma/dx_1$ and the remaining terms, arising out of the fact that v_1 does not directly depend on x_1. Let us now suppose $x_1 = x_2 = \ldots = x_T \equiv x$, and let

$$
\begin{aligned}
\alpha &= \frac{D\Gamma}{D\phi}\frac{d\phi}{dx_T} > 0. \\
\beta &= -\left(\frac{d\Gamma}{dv_1}\frac{1}{\epsilon^2}\right)\frac{\sum_{\tau=1}^{T}\Phi^\tau}{[\Phi^T+\sum_{\tau=2}^{T}(\tau-1)x_\tau\Phi^{T-\tau}]^2} > 0 \\
\gamma &= x\frac{u_T^2 v_1^2}{2\phi^2}E\left(\frac{S}{\tilde{S}}\right)^2 > 0.
\end{aligned}
$$

Then each of these terms is of order ϵ^0. Let

$$K_\tau = 2\sum_{j=1}^{T}\Phi^{2T-\tau-j}\mathbb{C}_{\tau j}.$$

We then have

$$
\begin{aligned}
\frac{d\Gamma}{dx_1} &= \alpha\Phi^{T-1} - \epsilon^2\gamma K_1 + O(\epsilon^3) \\
\frac{d\Gamma}{dx_2} &= \alpha\Phi^{T-2} + \epsilon^2\beta\Phi^{T-2} - \epsilon^2\gamma K_2 + O(\epsilon^3) \\
&\vdots \\
\frac{d\Gamma}{dx_\tau} &= \alpha\Phi^{T-\tau} + \epsilon^2\beta(\tau-1)\Phi^{T-\tau} - \epsilon^2\gamma K_\tau + O(\epsilon^3) \\
&\vdots \\
\frac{d\Gamma}{dx_T} &= \alpha + \epsilon^2\beta(T-1) - \epsilon^2\gamma K_T + O(\epsilon^3).
\end{aligned}
$$

and hence, for $\tau \in 2, \ldots, T-1$,

$$\begin{aligned} -\frac{dx_2}{dx_1} &= \frac{\frac{d\Gamma}{dx_1}}{\frac{d\Gamma}{dx_2}} = \frac{\alpha\Phi^{T-1} - \epsilon^2\gamma K_1 + O(\epsilon^3)}{\alpha\Phi^{T-2} + \epsilon^2\beta\Phi^{T-2} - \epsilon^2\gamma K_2 + O(\epsilon^3)} \qquad (22) \\ -\frac{dx_{\tau+1}}{dx_\tau} &= \frac{\frac{d\Gamma}{dx_\tau}}{\frac{d\Gamma}{dx_{\tau+1}}} = \frac{\alpha\Phi^{T-\tau} + \epsilon^2\beta(\tau-1)\Phi^{T-\tau} - \epsilon^2\gamma K_\tau + O(\epsilon^3)}{\alpha\Phi^{T-\tau-1} + \epsilon^2\beta\tau\Phi^{T-\tau-1} - \epsilon^2\gamma K_{\tau+1} + O(\epsilon^3)}. \qquad (23) \end{aligned}$$

We have increasing marginal rates of substitution if, for $\tau = 2, \ldots, T-2$

$$-\frac{dx_2}{dx_1} < -\frac{dx_3}{dx_2} \qquad \text{and} \qquad -\frac{dx_{\tau+1}}{dx_\tau} < -\frac{dx_{\tau+2}}{dx_{\tau+1}}, \qquad (24)$$

which can be verified by a straightforward but tedious calculation (details available in the technical appendix).

■

5.3.2 Proof of Proposition 5

From (18), using (20)–(21), we now have (hereafter omitting the $O(\epsilon^3)$ term)

$$\begin{aligned} \Gamma &= \ln\phi + E\ln\tilde{S} - \ln S \\ &\quad -\frac{\epsilon^2 u_T^2 v_1^2}{2\phi^2} E\left(\frac{S}{\tilde{S}}\right)^2 \left[\sum_{i=1}^{T}\sum_{j=1}^{T}\Phi^{2T-i-j}x_i x_j \mathbb{C}_{ij} + \sum_{i=1}^{T-1}\sum_{j=1}^{T-1}\Phi^{2T-i-j}k_i k_j C_{ij} + \sum_{i=1}^{T}\sum_{j=1}^{T-1}\Phi^{2T-i-j}x_i k_j \mathfrak{C}_{ij}\right], \end{aligned}$$

where

$$k_i = \frac{x_{i+1}}{\Phi} + \frac{x_{i+2}}{\Phi^2} + \frac{x_{i+3}}{\Phi^3} + \ldots + \frac{x_T}{\phi^{T-i+1}}. \qquad (25)$$

We conserve on notation by letting K denote the term in square brackets and hence writing Γ as

$$\Gamma = \ln\phi + E\tilde{S} - \ln S - \frac{\epsilon^2 u_T^2 v_1^2}{2\phi^2} E\left(\frac{S}{\tilde{S}}\right)^2 K. \qquad (26)$$

The derivation of decreasing marginal rates of substitution then follows lines similar to the proof of Proposition 6, revolving around a straightforward but tedious calculation and comparisons of the derivatives of Γ, presented in the technical appendix.

■

References

[1] Brian Charlesworth. *Evolution in Age-Structured Populations.* Cambridge University Press, Cambridge, 1994.

[2] Philip A. Curry. Decision making under uncertainty and the evolution of interdependent preferences. *Journal of Economic Theory*, 98:357–369, 2001.

[3] Partha Dasgupta and Eric Maskin. Uncertainty and hyperbolic discounting. *American Economic Review*, 95(4):1290–1299, 2005.

[4] Irving Fisher. *The Theory of Interest, as Determined by Impatience to Spend Income and Opportunity to Invest It.* MacMillan, New York, 1930.

[5] Shane Frederick, George Loewenstein, and Ted O'Donoghue. Time discounting and time preference: A critical view. *Journal of Economic Literature*, 40:351–401, 2002.

[6] H. Furstenberg and H. Kesten. Products of random matrices. *Annals of Mathematical Statistics*, 31(2):457–469, 1960.

[7] J. H. Gillespie. Polymorphism in random environments. *Theoretical Population Biology*, 4:193–195, 1973.

[8] Michael Gurven and Hillard Kaplan. Longevity among hunter-gatherers: A cross-cultural examination. *Population and Development Review*, 33(2):321–365, 2007.

[9] Ingemar Hansson and Charles Stuart. Malthusian selection of preferences. *American Economic Review*, 80:529–544, 1990.

[10] Alasdair I. Houston and John M. McNamara. *Models of Adaptive Behavior.* Cambridge University Press, Cambridge, 1999.

[11] Emily C. Lawrance. Poverty and the rate of time preference: Evidence from panel data. *Journal of Political Economy*, 99(1):54–77, 1991.

[12] P. H. Leslie. On the use of matrices in certain population mathematics. *Biometrica*, 33:183–212, 1945.

[13] P. H. Leslie. Some further remarks on the use of matrices in certain population mathematics. *Biometrica*, 35:213–245, 1945.

[14] Nabil Ibraheem Al-Najjar. Decomposition and characterization of risk with a continuum of random variables. *Econometrica*, 63(5):1195–1264, 1995.

[15] William Nordhaus. The *stern review* on the economics of climate change. *Journal of Economic Literature*, 45(3):686–702, 2007.

[16] Arthur J. Robson. A biological basis for expected and non-expected utility. *Journal of Economic Theory*, 68:397–424, 1996.

[17] Arthur J. Robson and Balazs Szentes. Evolution of time preference by natural selection: Comment. *American Economic Review*, 2007. Forthcoming.

[18] Alan R. Rogers. Evolution of time preference by natural selection. *American Economic Review*, 84:460–481, 1994.

[19] E. Seneta. *Non-Negative Matrices and Markov Chains*. Springer Verlag, New York, 1981.

[20] Peter D. Sozou. On hyperbolic discounting and uncertain hazard rates. *Proceedings of the Royal Society of London, Series B*, 265:2015–2020, 1998.

[21] David Tanny. On multitype branching processes in a random environment. *Advances in Applied Probability*, 13(3):464–497, 1981.

[22] S. Tuljapurkar. *Population Dynamics in Variable Environments*. Springer-Verlag, Berlin, 1990.

[23] Margo Wilson and Martin Daly. Life expectancy, economic inequality, homicide, and reproductive timing in chicago neighborhoods. *British Medical Journal*, 314(7089):1271, 1997.

[24] Menahem E. Yaari. Uncertain lifetime, life insurance, and the theory of the consumer. *Review of Economic Studies*, 32(1):137–158, 1965.