Cowles Foundation

for Research in Economics
at Yale University

Cowles Foundation Discussion Paper No. 1678

THE EVOLUTION OF TIME PREFERENCE
WITH AGGREGATE UNCERTAINTY

Arthur Robson and Larry Samuelson

October 2008

THE EVOLUTION OF TIME PREFERENCE
WITH AGGREGATE UNCERTAINTY

By

Arthur Robson and Larry Samuelson

October 2008

COWLES FOUNDATION DISCUSSION PAPER NO. 1678

THE EVOLUTION OF TIME PREFERENCE
WITH AGGREGATE UNCERTAINTY

Running head: The Evolution of Time Preference

Arthur Robson   Larry Samuelson*

November 4, 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 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.

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?

Irving Fisher's (1930, pp. 84–85) pioneering study of intertemporal trade-offs called attention to one reason future rewards are discounted—an intervening death may prevent us from realizing such rewards. The possibility of death has played a recurring role in discussions of discounting (e.g., Menahem E. Yaari (1965)). Ingemar Hansson and Charles Stuart (1990) and Alan R. Rogers (1994) 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 Arthur J. Robson and Balazs Szentes (2007)). Putting these ideas together leads to models in which people discount at the sum of the population growth rate and the mortality rate.

One difficulty with this argument is that the numbers don't obviously match. Studies of contemporary rates of time preference have produced estimates as high as twelve to twenty percent per year (Emily C. Lawrance (1991)). Steffan Andersen, Glenn W. Harrison, Morton I. Lau and E. Elisabet Rutström (2008, Table III), arguing that estimated discount rates fall when correcting for the confounding effects of risk aversion, find (still surprisingly high) discount rates of about ten percent . In contrast, Michael Gurven and Hillard Kaplan (2007, pp. 330-331) use data from contemporary hunter-gatherers to 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 must have been approximately zero, suggesting discount rates of a few percent. A second issue 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 (1990) and Rogers (1994), we view peoples' preferences as having been

---

[1]Recent policy discussions, especially those regarding global warming, have focussed attention on the first question (e.g., William Nordhaus (2007)), while recent work in behavioral economics has directed attention to the latter (Shane Frederick, George Loewenstein and Ted O'Donoghue (2002)).

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 sum of expected utility. This result involves 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 or abundance of 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] Furthermore, 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 mortality rate corresponding to mean survival—so that aggregate risk may lie behind the apparent 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]

---

[2]See Robson (1996) for an analogous result for populations without an age structure.

[3]Section 2.4 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 periods of sustained growth with rare, rapid and evolutionarily-neutral population collapses.

[4]Partha Dasgupta and Eric Maskin (2005) and Peter D. Sozou (1998) also present evolutionary foun-

Section 1 introduces the mechanics of age-structured populations for the simpler case of an environment with only idiosyncratic uncertainty. Section 2 examines aggregate uncertainty. Section 3 discusses some of the features that are left out of our analysis. Proofs not contained in the body of the paper are collected in Section 4.

1 Idiosyncratic Uncertainty

It is helpful to first consider the more straightforward case of idiosyncratic uncertainty, drawing on Brian Charlesworth (1994) and Alasdair I. Houston and John M. McNamara (1999), and following Robson and Larry Samuelson (2007).

1.1 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.

The assumption that survival rates are constant across ages looks restrictive. However, because we place no restrictions on the pattern of fertility, the constant-survival-rate assumption is innocuous. In particular, all of the evolutionarily relevant information is

---

dations for presently-biased preferences, including in Dasgupta and Maskin's case the possibility of preference reversals. We discuss these papers in Section 3.

[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.

contained in the agent's expected number of offspring at each age, where this expectation includes the probability that an intervening death may fix the realized number of offspring at zero. Given an arbitrary specification of age-dependent survival rates and expected offspring conditional on survival, we can find a formally equivalent description in which survival rates are constant across ages and fertilities are adjusted accordingly to preserve the expected number of offspring at each age, allowing us to apply the techniques described below. Section 3 further discusses the implications of this equivalence in a more general setting. 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 zero. None of the subsequent analysis would be affected if there were an environmental carrying capacity that would eventually cap the size of the population, as long as our S is then interpreted as the endogenously determined zero-population-growth steady-state survival rate.

We are ultimately interested in people's preferences over the wide variety of things they consume, rather than simply reproduction. We view our study of intertemporal preferences over reproduction as a necessary first step in studying preferences over consumption. Reproduction is the currency of evolution, with the various features of our preferences having survived the evolutionary screen because of their salutary effects on reproduction. We thus cannot understand the evolutionary implications of other intertemporal trade-offs without understanding trade-offs over reproduction. To make the link to preferences over consumption, we would view the fertility x_τ as being a function of the consumption of food, shelter, status, and a host of other economic goods, with intertemporal preferences over these goods induced by their implications for x_τ. 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 are shaped by the implications of the resulting decisions for reproduction.[6]

---

[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

The intertemporal trade-offs examined in our model explicitly concern the timing of reproduction within an agent's lifetime. Intertemporal trade-offs often involve intergenerational allocations.[7] Our model can be applied to examine such transfers: We can view the reproductive profiles $(x_1, \ldots, x_T)$ appearing in our analysis as the product of both consumption and intergenerational transfer decisions, so that our results would provide insight into preferences over transfers as well as consumption once the appropriate links between consumption and reproduction are made.

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.

1.2   Evolution

An agent in this environment is characterized by its reproduction profile $\{x_\tau, \tau \in \{1, 2, \ldots, T\}\}$. This profile is heritable—each agent's reproductive profile matches that of their parent. Notice that we have abstracted away from a number of realistic considerations. Reproduction is asexual in this model, there are no errors or distortions in the process of genetic transmission, all the agents apparently do is live and reproduce, there is no explicit trade-off between the quantity and quality of offspring, and so on. This allows us to focus on the basic determinants of time preferences.

We now ask which reproductive profiles will be selected by evolution. In particular, suppose a population initially contains a variety of reproductive profiles. Some agents may have offspring in many different periods, some in only a single period. Some may

example) as well as the way in which the x_τ combined to yield population growth. Since the first effect is relatively familiar, we concentrate here on the derivation of the growth rate criterion from the x_τ. Extending the analysis from reproduction to consumption is relatively straightforward if reproduction at age τ is a function of consumption at age τ (only), and becomes more complicated as we move away from this simple case (cf. Robson, Szentes and Emil Iantchev (2005)).

[7]Work on intergenerational transfers includes Rogers (1994).

have offspring early but have only a few, others may wait longer to reproduce but then have more offspring. Among these reproductive profiles, some will tend to generate more ultimate descendants than others, and if we examine the population after natural selection has had ample time to work, it will be composed almost entirely of agents bearing this descendant-maximizing reproductive profile. Subsequent mutations introducing reproductive profiles leading to fewer ultimate descendants will die out relatively quickly.

In making this idea precise, we follow the standard approach in assuming the number of agents following each reproductive profile 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 all agents of age τ (with reproductive profile $\{x_\tau, \tau \in \{1, 2, \ldots, T\}\}$) can be taken to be precisely x_τ and the proportion of survivors can be taken to be precisely S.[8]

The population of agents characterized by reproductive profile $\{x_\tau, \tau \in \{1, 2, \ldots, T\}\}$

---

[8]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 a draw from a survival lottery that yields survival with probability S and death with probability $1 - S$. The law of large numbers then ensures that average and expected numbers of total offspring, as well as average 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. Nabil Ibraheem Al-Najjar (1995)). 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.

thus evolves according to

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

(1)

(2)
$$\equiv N'(t)X,$$

where $'$ denotes transpose. Each row of the matrix X describes the fate of one of the age cohorts in the population. The second row, for example, tells us that each two-period-old agent has x_2 offspring, which survive with probability S to become period-$t+1$ one-period-olds, and each current two-period-old survives with probability S to become a three-period-old. The transition matrix X is the *Leslie matrix* (P. H. Leslie (1945,1948)).

The number of agents at time t characterized by the reproductive profile giving rise to the Leslie matrix X is given by

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

We can form one such equation for each possible reproductive profile (though we refrain from introducing such notation). Which reproductive profile will give rise to the most descendants at some point in the future? This is equivalent to asking which reproductive profile will give rise to the "largest" X^t for large t. In answering this question, we assume that the Leslie matrix X is primitive, in that there exists some $k > 0$ for which X^k is strictly positive.[9] This allows us to bring standard results in matrix theory to bear in examining X^t.

In particular, the Perron-Frobenius theorem (E. Seneta (1981, Theorem 1.1)) implies that the Leslie matrix has a "dominant" eigenvalue ϕ that is real, positive, of multiplicity

---

[9]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.

one, strictly exceeds the modulus of all other eigenvalues, and satisfies (Seneta (1981, 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 gives the "reproductive value" of an individual of each age, or the relative contribution that each such individual makes to the long run population.

This result can be more easily interpreted after premultiplying the first equation by the vector u', postmultiplying by v and then taking logs so that

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

The expression $u'X^t v$ is referred to as the total reproductive value of the population and serves as a convenient measure of the period-t population. This result then tells us that the population growth rate is given by the log of the dominant eigenvalue of the Leslie matrix. Those reproductive profiles whose Leslie matrixes have higher dominant eigenvalues will leave more ultimate descendants than others, and eventually the population will be composed virtually entirely of the reproductive profile that maximizes this eigenvalue. The effects of natural selection are thus easily characterized—given any set of alternatives, evolution will select the reproductive profile (and only that one) maximizing the dominant eigenvalue of the corresponding Leslie matrix.[10]

---

[10]The ultimate fate of the population depends on the magnitude of this dominant eigenvalue ϕ. The population grows without bound it $\phi > 1$, shrinks if $\phi < 1$, and converges to a constant state if $\phi = 1$. A more realistic model would allow the death rate to vary as does the population, increasing as the population grows in response to increasingly scarce resources and bringing the population to a steady state. We can capture this possibility in a simple way by reinterpreting the death rate S as the steady-state death rate.

1.3 Discounted expected utility

We must now turn this characterization of the evolutionary process into a statement about intertemporal preferences. What does the maximization of an eigenvalue have to do with the trade-off between x_τ and $x_{\tau'}$? The eigenvalue ϕ solves the characteristic equation[11]

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

where

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

Evolution would thus endow an agent with preferences, (or more precisely, would endow an agent with behavior consistent with preferences) whose indifference curves are described by the right side of (4), *evaluated at the optimal growth rate.* That is, a reproduction profile $(x_1, \ldots, x_T)$ giving a higher value for the right side would lead to a higher growth rate, and the fact that it is not observed indicates that it must be infeasible. A profile giving a smaller value of the right side of (4) is inferior, leading to a smaller growth rate that would doom its adherents to dwindle away as a proportion of the population.

This description of preferences is self-referential, since alternatives are ranked according to a criterion expressed in terms of the optimal growth rate, which is itself determined by the optimal alternative. This self-reference is not necessary, serving only to provide a convenient and familiar description of the preferences for which evolution selects. The evolutionary criterion is clear: a reproductive profile $(x_1, \ldots, x_T)$ is better than $(\hat{x}_1, \ldots, \hat{x}_T)$ if and only if the Leslie matrix associated with the former has a larger dominant eigenvalue. This gives us a complete and unambiguous ranking of reproductive profiles, one that can be checked without reference to the optimal choice. However, there is no explicit

---

[11]This is a rearrangement of

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

functional form capturing this ranking, whereas the growth rate associated with the best profile gives us an explicit and simple way (in the form of (4)) of describing preferences.

We can move closer to our objective of talking about intertemporal trade-offs by extracting marginal rates of substitution from (4) of the form:

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

(5)

Intuitively, an agent should be willing to forego current offspring only if the return is Φ as many offspring next period. Marginal rates of substitution between $x_{\tau+1}$ and x_τ are independent of τ and independent of the magnitudes of $x_{\tau+1}$ and x_τ.

Equivalently and perhaps more informatively, we can capture the preferences represented by (4) in a utility function of the form:

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

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 $(-\ln S)$.[12] 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 these losses.

The finding that evolution will select for a discount rate equal to the sum of the growth and death rates echoes a long-standing view to which we alluded in the introduction, namely that discounting is at least partly motivated by the possibility of an intervening death. Perhaps surprisingly, however, the resulting discount rate is independent of the death rate (for a fixed fertility profile). An increase in the death rate would simply prompt a compensating decrease in the population growth rate, leaving their sum, and hence the discount rate, unchanged. We see this in (5), giving marginal rates of substitution equal to Φ, which (4) reveals to be determined by the fertility rates $(x_1,\ldots,x_T)$ only.

---

[12]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 $\delta = -\ln S$.

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)). How might we reconcile the model with the wide variety of choices we actually see people making? Suppose that newborn agents are independently (across time and agents) assigned a feasible set $X^T \subset \Re_+^T$ of possible reproductive profiles from which they must choose.[13] Some agents may find themselves in more favorable circumstances than others, with correspondingly more favorable prospects for reproduction. A reproductive strategy would now be not simply a profile $(x_1, \ldots, x_T)$, but rather a vector of such profiles, one for each of the possible feasible sets, allowing evolution to tailor the agent's reproductive profile to the appropriate feasible set. An agent would be characterized by such a vector of reproduction profiles, with this entire vector being heritable across generations. Each such vector gives rise to an "average" growth rate Φ, where this average is taken over the collection of feasible sets with which agents might be faced. Evolution will select for the vector of reproductive profiles that gives the highest average growth rate. We would then observe a potentially vast variety of behavior, but a straightforward calculation shows that all of the resulting choices would be consistent with preferences that are again described by (4) (applied to the appropriate feasible set in each case), with the common value of Φ now being the average population growth rate corresponding to the optimal vector of reproductive profiles. We could again infer that marginal rates of substitution are given by (5) and the utility function by (6).

2   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

---

[13]Robson, Szentes and Iantchev (2005)) develop a similar approach.

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.

2.1   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 (1996), 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 dominant eigenvalue ϕ of the trivial Leslie matrix $[Sx]$).

Now suppose that instead of a fraction S of the agents surviving from age 0 to age 1, 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. We have constructed this example to be particularly simple and to give a particularly striking result, but it is a quite general result that aggregate uncertainty gives a lower growth rate (in an unstructured population) than does the equivalent idiosyncratic uncertainty.

2.2   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 $\tau \in \{0, \ldots, T-1\}$ survives until the next period, with mean S. Hence, we think of each agent of age $\tau \in \{1, \ldots, T-1\}$ as first receiving a common realization $\tilde{S}_\tau$ with support contained 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 independently and identically distributed over time.

The *mean Leslie matrix* is familiar and is given by

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

and we continue to let ϕ denote the dominant 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

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

Section 3 briefly explores how the the assumption of of a common idiosyncratic death rate S across ages in (7) can be relaxed, even when there are aggregate shocks to survival rates.

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 an approximation of the long-run growth rate of total reproductive value, evaluated with the population proportions u and reproductive values v from the mean Leslie matrix.[14] This is now a product of random matrices. Not only can we not apply the

_____

[14]There is no difficulty using the eigenvectors u and v from the mean Leslie matrix in this approximation

14

Perron-Frobenius theorem, but it is no longer obvious that the limit exists. Fortunately, we have the following remarkable result (first established by H. Furstenberg and H. Kesten (1960, Theorem 2) and extended in David Tanny (1981, Theorem 7.1)):[15]

Proposition 1 *Since* $-\infty < E \ln u' \tilde{X} v < \infty$, *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.

Natural selection then favors reproductive profiles that maximize the growth rate $\ln \lambda$. Once we leave the case of only idiosyncratic uncertainty, it is no longer obvious that we can restrict attention to pure strategies. Indeed, it is well known that in populations without an age structure, but with aggregate uncertainty, mixing may be strictly better from an evolutionary point of view than any pure strategy (e.g., Theodore C. Bergstrom (1997) , W. S. Cooper and R. H. Kaplan (2004), Houston and McNamara (1999, Section 10.4)).[16] Similar forces can obviously arise in a population with an age structure. However, mixing confers no evolutionary advantage, even in the presence of aggregate uncertainty, when the set of pure strategies is convex and the evolutionary criterion depends only on the

---

of the growth rate. Proposition 1 below holds for any norm $||\tilde{X}(1) \dots \tilde{X}(t)||$. We retain our assumption that the mean Leslie matrix X is primitive. Together with the restriction that the support of $\tilde{S}$ is contained in $(0, 1]$, this ensures that asymptotically, all elements of $\tilde{X}(1) \dots \tilde{X}(t)$ grow at the same finite rate.

[15]Taking each $\tilde{S}_\tau = 1$ gives us an upper bound an $u' \tilde{X} v$, ensuring that we satisfy $E \ln u' \tilde{X} v < \infty$. Our assumption that $\tilde{S}_\tau$ has support contained in $(0, 1]$ ensures that $E \ln u' \tilde{X} v > -\infty$, so the more general sufficient condition in Tanny (1981) is satisfied.

[16]Consider, for example, agents who can amass either a small or large cache of food for the winter. Building a large cache carries a higher risk of death at the hands of predators. Winters are typically mild, with very rare harsh winters. A small cache ensures survival during a mild winter but leads to death in a harsh winter, while a large cache ensures survival in either case. The pure strategy of always choosing a small cache leads to extinction at the hands of the first harsh winter, while always collecting a large cache leads to inefficiently high mortality at the hands of predators. The optimal strategy is to mix, with most agents choosing a small cache that typically ensures survival at minimal risk, but with a few choosing large caches to ensure someone survives a harsh winter. Similar examples can be constructed when strategies are drawn from a continuum.

(idiosyncratic) number of expected offspring produced at each age, as in our case. That is, the realized Leslie matrix $\tilde{X}(t)$ in each period t depends only on the realized survival rates and the reproductive profile $(x_1, \ldots, x_T)$. As a result, a population whose members attached idiosyncratic probability p to reproductive profile $(x_1, \ldots, x_T)$ and probability $1 - p$ to profile $(x'_1, \ldots, x'_T)$ would be indistinguishable from a population whose members all chose the reproductive profile $(px_1 + (1 - p)x'_1, \ldots, px_T + (1 - p)x'_T)$. It thus suffices to consider pure strategies.

Aggregate uncertainty builds risk aversion into the evolutionary selection criterion. We see this in the example of Section 2.1, where it would be worth paying virtually any price to avoid the possibility of zero offspring.[17] Returning to our discussion in Section 1.1, might intergenerational transfers now be useful as a way of mitigating risk? Transfers that cannot be conditioned on the aggregate uncertainty add nothing new to the model. In this case, transfers are again simply tools that might be used in implementing a reproductive profile $(x_1, \ldots, x_T)$, and the implications of such transfers are captured by our analysis of reproductive profiles.

2.3 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.[18] Section 4.1 proves:

Proposition 2 *Aggregate uncertainty concerning survival reduces the population growth*

---

[17]This risk is effectively pooled across agents when uncertainty is idiosyncratic, leaving a risk neutral selection criterion.

[18]This result depends on the assumption that the idiosyncratic uncertainty is independent across periods. For example, an 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

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

rate:[19]

$$\lambda \leq \phi.$$

"Generically," the growth rate is strictly lower under aggregate uncertainty, but for exceptional circumstances such as all of the possible realized Leslie matrices having the same dominant eigenvalue and associated left eigenvector. The following example illustrates this latter possibility.

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

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

In each period, the realized Leslie matrix is independently drawn to be either X' or X''. The mean matrix X has dominant 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 dominant 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'', has the same value, x^t. As a result, the growth rate without aggregate uncertainty (i.e., with $X(t') = X$ for all t') matches that with aggregate uncertainty.

2.4 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*

(9) $$-\frac{dx_{\tau+1}}{dx_\tau} = \Phi = \frac{\phi}{S},$$

---

[19]See Philip A. Curry (2001), J. H. Gillespie (1973), and Houston and McNamara (1999, Chapter 10) (as well as Robson (1996)) for similar results for the case of $T = 1$.

and hence for discounting at the sum $\ln \phi - \ln S$ of the growth rate and death rate of the mean Leslie matrix.

As before, ϕ and S are the dominant eigenvalue and survival probability associated with the mean Leslie matrix (7). Comparing with (5), we thus see that aggregate uncertainty in death rates 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 death rate associated with mean survival.

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

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

(10)

Since the fertilities $(x_1, \ldots, x_T)$ appear only in $\ln \phi$, the arguments of Section 1.3 ensure that evolution will select for marginal rates of substitution given by (9).[20] ■

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. Indeed, this discount rate could be obtained from a Leslie matrix with no mortality at all, an observation used below. 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 death rate associated with mean survival.

---

[20]Notice that, since $E\tilde{S} = S$ and hence $E \ln \tilde{S} < \ln S$, we have $\ln \lambda < \ln \phi$, in accordance with Proposition 2.

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 the death rate associated with mean survival.

Gurven and Kaplan (2007, pp. 345–348) 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, and so 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 mean Leslie matrix.

To get an idea of the numbers involved, we need an idea of the upper bound on human growth rate, i.e., an idea of how fast a population would grow in the absence of any mortality. Suppose that individuals start reproducing at age 15 and stop at age 45, that the probability of giving birth in a given year is 0.15, and that here is no death risk before age 45. With the exception of the absence of death before the end of one's reproductive age, these numbers are reasonably consistent with observations of contemporary hunter-gatherers.[21] We can ignore the risk of death on the strength of the previous observation

---

[21]Kim Hill and A. Magdalena Hurtado's (1996, Chapter 8, especially Table 8.3) study of the Ache suggests a prime-age birth probability of 0.15 per year. (We cut the birth probabilities reported there in half. The 0.15 then represents the probability of a female birth, providing a valid comparison with our model of asexual reproduction.) Kendra McSweeney and Shahna Arps (2005, especially p. 14) survey indigenous populations in lowland Latin America who are recovering from prior catastrophic declines with rapid population growth. They find total fertility rates (roughly, the number of children born to a woman over the course of her child-bearing years) between 3.9 and 10.5, with a median of 7.9. A total fertility rate of 9, somewhat near the upper end of this range, coupled with a thirty-year reproductive

that an increased death risk would only prompt a compensating decrease in the growth rate, leaving the discount rate unchanged. From (4), the implied dominant eigenvalue is the solution to $1 = \sum_{\tau=15}^{45} \frac{(0.15)}{\phi^\tau}$, which yields $\phi = 1.05675$ and hence a growth and discount rate of $\ln \phi = 0.055$.

If this discount rate is the product of an evolutionary past featuring aggregate uncertainty and a zero growth rate, then we must have, from (10),

$$0 \;=\; \ln \lambda \;=\; \ln \phi - \ln S + E \ln \tilde{S} \;=\; 0.055 + E \ln \tilde{S},$$

where the second equality gives the realized growth rate as the difference between the discount rate ($\ln \phi - \ln S$) and expected log of the random survival probability. The final equality inserts our discount rate of 0.055. For simplicity, suppose that with probability $1 - p$ we have an ordinary period in which the death rate is about two percent.[22] With probability p a catastrophe with a lower survival rate of $S^\dagger$ appears. Then we have $E \ln \tilde{S} = p \ln S^\dagger + (1 - p)(-0.02) = -0.055$ (recalling that $-\ln S$ is the death rate). That is, we need catastrophes to appear with probability 0.25 if 85% of the population survives; with probability 0.1 if 70% survive; with probability 0.05 if 50% survive; or with probability 0.015 if only 10% survive.[23] To further examine the effects of aggregate uncertainty, consider this last possibility. Each agent faces a compound survival lottery featuring a 10% chance of survival with probability 0.015 and a 98% chance of survival with probability 0.985. The mean survival rate is thus 0.9668, with a corresponding continuous death rate of 0.034. If all this risk were idiosyncratic, the population would still grow at about two percent. It is the aggregate nature of the uncertainty that brings the growth rate down another two percent, to zero.[24]

---

span (see McSweeney and Arps (2005, p. 15) and especially Hill and Hurtado (1996, Table 8.3)), gives a yearly birth probability of 0.3. Halving this to account for our asexual model again gives us 0.15.

[22]Recall that Gurven and Kaplan (2007, pp. 330–3341) estimate that annual mortality rates ranged from one percent for ten-year-olds to four percent for sixty-year-olds.

[23]Though direct evidence is scarce, it seems inevitable that the ice ages would have caused sharp drops in primitive human population levels.

[24]In contrast, the catastrophic aggregate uncertainty we are discussing here makes little difference if the catastrophes are frequent and mild. Idiosyncratic survival lotteries featuring an 85% chance of survival

These calculations bring us from the discount rates of a few percent implied by a model with only idiosyncratic uncertainty to discount rates between five and six percent. This still does not bring us to the ten percent rates of Andersen, Harrison, Lau and Rutström (2008). It is significant here that estimates of the pure rate of time preference derived from actual behavior are often lower than estimates derived from experimental data, sometimes coming closer to our rough calculation of five to six percent (e.g., Robert H. Litzenberger and Cherukuri U. Rao (1971, Tables 1 and 2)).

Once again, we have a model with the rather counterfactual prediction that we should observe only a single reproductive profile. As is the case under idiosyncratic uncertainty, we can suppose that newborn agents are independently (across time and agents) assigned a feasible set $X^T \subset \Re_+^T$ of possible reproductive profiles from which they must choose. An agent is again characterized by a vector of reproduction profiles, one for each possible feasible set, with this vector being heritable across generations. Evolution will select for the vector of reproductive profiles that gives the highest average growth rate, inducing a variety of reproductive profiles each of which would be consistent with preferences that are described by (4) (applied to the appropriate feasible set in each case), with Φ now being the average population growth rate.

2.5   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 first explore a plausible case that gives rise to a present bias.

---

with probability 0.25 and a 98% chance of survival with probability 0.75 give a mean survival rate of 0.95, and hence a death and discount rate of 0.0543. That the first state of this compound lottery is aggregate has virtually no effect on discount rates.

We model survival rates as being affected by relatively small age-specific perturbations around an age-independent common shock. There are elements of preference and constraint mixed in this choice. We are more convinced that our evolutionary environment exhibited significant and correlated fluctuations in death rates than we are that these death rates exhibited any particular pattern across ages. We thus find appealing a model that incorporates both possibilities while emphasizing the former. In addition, our focus on small age-specific perturbations allows us to use a convenient approximation method as the basis for the analysis, while there are no general methods for examining our questions in the presence of large age-specific perturbations.

As before, a random variable $\tilde{S}(t)$ is drawn in each period t, identically and independently distributed over time, with support contained in $(0, 1)$ and with mean S. In the proportion $1 - \varepsilon$ of the population, each individual then receives an idiosyncratic draw giving 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, with $\tilde{S} + \hat{S}_\tau$ having support contained in $(0, 1]$.[25] 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. The random variable $\tilde{S}(t)$ is thus relevant for the entire population and is the counterpart of the common survival-rate fluctuations examined in Section 2.4. The random variables $(\hat{S}_0, \ldots, \hat{S}_{T-1})$ overlay these common shocks with age-specific survival-rate perturbations. The larger is ε, the greater is the variation across ages in the aggregate death rate. We consider the case of small ε and hence small age-specific aggregate shock. There is no restriction that the shock $\tilde{S}$ common to all ages is small, and no restriction on the idiosyncratic uncertainty.

We find that the discount rate is no longer constant over time. Given our restriction of our analysis to small values of ε, and hence small departures from common death-rate fluctuations, we can infer only that discount rates will depart slightly from constancy. However, discount rates may well exhibit more pronounced variations when age-specific

---

[25]More precisely, the random variables $\tilde{S}(t)$ and $\hat{S}_\tau(t')$ for all $t, t' = 1, 2, \ldots$ and $\tau = 1, \ldots, T$ are independent, except that the $\hat{S}_\tau(t')$ need not be independent across τ for a given t'.

survival rate fluctuations are larger.

We now write the realized Leslie matrix for period t as

(11)
$$\tilde{Z}(t) = \tilde{X}(t) + \varepsilon \tilde{H}(t),$$

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

(12)
$$\tilde{H}(t) = \begin{bmatrix} x_1\hat{S}_0(t) & \hat{S}_1(t) & 0 & \ldots & 0 \\ x_2\hat{S}_0(t) & 0 & \hat{S}_2(t) & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ x_{T-1}\hat{S}_0(t) & 0 & 0 & \ldots & \hat{S}_{T-1}(t) \\ x_T\hat{S}_0(t) & 0 & 0 & \ldots & 0 \end{bmatrix}.$$

Each of the random variables in the matrix $\tilde{H}(t)$ has a zero mean. Our analysis is based on the following approximation (cf. S Tuljapurkar (1990, Chapter 12)):

Proposition 4 *Suppose the matrices $\tilde{H}(t)$ in (11) are independent across periods and have a zero expected value. Then, almost surely,*

$$\lim_{t\to\infty} \frac{1}{t} \ln u'\tilde{Z}(1)\ldots\tilde{Z}(t)v$$

(13)
$$= \ln\phi - \ln S + E\ln\tilde{S} - \frac{\varepsilon^2}{2\phi^2}E\left(\frac{S}{\tilde{S}}\right)^2 E\{(u'\tilde{H}v)^2\} + O(\varepsilon^3).$$

Section 4.2 presents the proof.

The expression for the growth rate given by (13) contains some familiar terms. The first three terms give us the growth rate under purely common survival-rate perturbations (cf. (10)). 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. When perturbations to survival rates vary by age, the growth rate is thus that which would prevail without such variation, minus a "variance penalty."[26]

---

[26]Revisiting some previous points, it is immediate that $E\{(u'\tilde{H}v)^2\} \geq 0$, and hence that introducing

23

2.5.1 Marginal rates of substitution

When aggregate effects on survival vary across ages in a symmetric way, marginal rates of substitution decline over time:

Proposition 5 *Suppose the random variables $(\hat{S}_0, \ldots \hat{S}_{T-1})$ share common variance V and common covariance C of any pair. Then for sufficiently small ε, the marginal rate of substitution is decreasing in τ, i.e.,*

$$-\frac{dx_{\tau+1}}{dx_\tau} \geq -\frac{dx_{\tau+2}}{dx_{\tau+1}},$$

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.

The most striking aspect of Proposition 5 is that a present bias emerges despite the complete symmetry of the aggregate age-specific survival shocks. These aggregate shocks have independent and identical distributions across periods, and within periods have identical (and possibly independent) distributions across ages, but still induce asymmetries in discounting across ages. Alternative considerations that might lie behind nonexponential discounting, such as age-dependent idiosyncratic death rates or the time structure of

---

variation in the effects of aggregate uncertainty across ages cannot increase the population growth rate. Independence across τ of the $\hat{S}_\tau(t)$ involved in the construction 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'', $\varepsilon = 1$, $E \ln \tilde{S} - \ln S = 0$ and $E\{(u'Hv)^2\} = 0$.

the technology transforming consumption into fertility, rely on asymmetries across ages for their effects. A new insight derived from the study of aggregate uncertainty is that nonexponential discounting can arise in a setting devoid of temporal asymmetries.[27]

2.5.2 Why not exponential?

What lies behind these results? Suppose there is aggregate uncertainty only in one survival rate S_τ, so that survival from age τ to $\tau + 1$ is uncertain. For much the same reason that aggregate uncertainty reduces the growth rate in a single-age population (cf. Section 2.1), this reduces the value of period-τ' births, for all $\tau' > \tau$. As a result, the discount rate between periods τ and $\tau + 1$ is increased, since it now takes more period-$\tau + 1$ births to counteract a given decrease in period-τ births. (Section 4.3 illustrates this claim.) However, marginal rates of substitution between other adjacent periods are unaffected. The marginal rate of substitution thus falls as we move beyond period τ, introducing a present bias. At the same time, the marginal rate of substitution between periods τ and $\tau + 1$ is now *higher* than the marginal rate of substitution in earlier periods, pushing discounting away from a present bias. We must in general consider aggregate uncertainty in more than one survival rate, leading to contending forces.

To strip away some of the complication, suppose that there are only three age classes ($T = 3$) and that $\tilde{S} = S \in (0, 1)$ with probability one, so there is no common component to the aggregate survival shocks. Then applying (13) and then (4), we can calculate (Section 4.3 provides details),

$$
\begin{aligned}
\ln \lambda &= \ln \phi - \frac{\varepsilon^2 u_1^2 v_1^2}{2 S^2} \left[V_0 + \frac{V_1}{\Phi^2} \left(\frac{x_2}{\Phi} + \frac{x_3}{\Phi^2} \right)^2 + \frac{V_2}{\Phi^4} \left(\frac{x_3}{\Phi} \right)^2 \right] \\
(14) \quad &= \ln \left[x_1 + \frac{x_2}{\Phi} + \frac{x_3}{\Phi} \right] + \ln S - \frac{\varepsilon^2 u_1^2 v_1^2}{2 S^2} \left[V_0 + \frac{V_1}{\Phi^2} \left(\frac{x_2}{\Phi} + \frac{x_3}{\Phi^2} \right)^2 + \frac{V_2}{\Phi^4} \left(\frac{x_3}{\Phi} \right)^2 \right],
\end{aligned}
$$

where V_τ is the variance of the aggregate shock to the period-τ survival rate. This ex-

---

[27]To be more precise, our age-specific aggregate shocks are "exchangeable" (cf. William Feller (1971, pp. 228–230)). The terminal age T effectively builds aging into our model, but this is not the source of the present bias, since the effect would still arise if we worked without such an upper bound and an infinite sequence of positive fertilities x_t.

pression immediately suggests that exponential discounting is not to be expected.[28] The important question is: Given $V_0 = V_1 = V_2 = V > 0$, how do the contending forces introduced by the aggregate shocks to the various survival rates combine to affect the marginal rates of substitution $-dx_2/dx_1$ and $-dx_3/dx_2$?

To be more precise, let us further simply by (innocuously) assuming $\Phi = 1$, giving

$$(15) \qquad \ln \lambda = \ln \phi - \frac{\varepsilon^2 u_1^2 v_1^2 V}{2S^2} \left[1 + (x_2 + x_3)^2 + (x_3)^2 \right].$$

The growth rate $\ln \lambda$ depends on the various x_τ in a number of implicit ways (e.g., through v_1). However, these implicit dependencies alone generate a constant discount rate. Departures from exponential discounting hinge on the explicit appearances of the x_τ in (14).[29] Taking the relevant derivatives and using the fact that the x_τ enter terms of order ε^2 (i.e., ignoring higher-order terms in ε), we find that if the term in brackets in (15) were linear, of the form $1 + x_2 + x_3 + x_3$, then the effect of increasing x_2 would be one half the effect of increasing x_3, which would in turn be consistent with exponential discounting (cf. Section 4.3). However, since the bracketed term is $1 + (x_2 + x_3)^2 + (x_3)^2$ (with the squares reflecting its origin as a variance), the effect of increasing $x_2 > 0$ is more than one half the effect of increasing x_3, causing the discount rate to fall as we move away from the present.

2.6   Robustness

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

---

[28]The first term in (14) gives the growth rate that would prevail without age-dependent mortality perturbations. The fertilities x_1, x_2, and x_3 appear here, with each x_τ divided by $\Phi^{\tau-1}$. As we have seen, these terms alone give us constant marginal rates of substitution (equalling Φ) and hence constant discount rates. The final term, arising out of the variance penalty, again includes the fertilities x_1 x_2 and x_3, but now divided by various powers of Φ. Once we mix these powers with the regular relationship between x_τ and $\Phi^{\tau-1}$ of the initial term, we cannot expect constant marginal rates of substitution.

[29]Section 4.3 sketches a proof of these observations.

of discounted expected utility. How robust is the result that discount rates are pushed in the direction of a present bias?

The building block for our analysis, that age-specific aggregate uncertainty only in the survival rate S_τ increases the discount rate between periods τ and $\tau + 1$, is quite general. However, the combined effects of age-specific perturbations to multiple ages are more fragile. Our present bias result rests on two assumptions, namely that perturbations to survival rates that are not common across ages are relatively small and are symmetric across ages. This strikes us as a natural setting, fueled by the belief that environmental fluctuations affecting survival rates are likely to be felt across all ages.

However, two examples illustrate how different specifications can lead to different results. Robson and Samuelson (forthcoming) present an example, with age-specific perturbations that are no by means small, in which the optimal discount rate between any pair of ages is zero, no matter what the population growth rate and death rate.[30] The present section explores another departure from our maintained assumptions that leads to a future bias, i.e., to marginal rates of substitution that increase as one moves away from the present.

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 (1994, 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. We can again write the realized Leslie matrix for period t as in (11), with $\tilde{X}(t)$ given by X from (7) for each t and with the perturbation matrix

---

[30]The discount rate is thus constant across ages in this example, but its magnitude is nonetheless surprising.

$\tilde{H}(t)$ now given by

(16)
$$\begin{bmatrix} x_1\hat{\mathbb{S}}_1(t) & 0 & 0 & \ldots & 0 \\ x_2\hat{\mathbb{S}}_2(t) & 0 & 0 & \ldots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ x_{T-1}\hat{\mathbb{S}}_{T-1}(t) & 0 & 0 & \ldots & 0 \\ x_T\hat{\mathbb{S}}_T(t) & 0 & 0 & \ldots & 0 \end{bmatrix},$$

where each of the random variables $\mathbb{S}_\tau$ in the matrix $\tilde{H}(t)$ again has a zero mean. We have:[31]

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, the marginal rate of substitution is increasing in τ, i.e.,*

$$-\frac{dx_{\tau+1}}{dx_\tau} \leq -\frac{dx_{\tau+2}}{dx_{\tau+1}},$$

strictly so if $\Phi \neq 1$ and $C < V$.

Though we find the model of age-specific aggregate survival rates of Section 2.5.1 the most natural of those we have considered, it is clear that the present bias of Proposition 5 is not universal. Notice, however, that here again we have agents who are pushed away from exponential discounting despite aggregate shocks that are symmetric across ages—the aggregate shocks of Proposition 6 have independent and identical distributions across periods, and within period have identical (and possibly independent) distributions across parental ages. Our robust finding is thus that aggregate uncertainty *per se* can push preferences away from exponential discounting, with the nature of the departure—whether present bias, future bias, or possibly something more complicated—depending upon details of the aggregate uncertainty.

---

[31]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.

28

3 Discussion

Present bias. We have found that evolutionarily-induced intertemporal preferences may exhibit a present bias. However, the preferences in our model do not generate 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 trade-off between x_9 and x_{10} that confers evolutionary advantages when made at age 1 will still confer such advantages when made 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.

We are not disappointed that the model does not generate preference reversals. We think present bias may well be a more basic phenomenon than preference reversals, and it seems more readily generated by evolutionary optimization.[32] More importantly, our analysis suggests that we can expect discount rates to vary systematically with age, addressing intertemporal choices over longer spans of time than those typically covered in preference-reversal experiments. In contrast to most models of age-dependent discounting, these variations do not reflect changes in the death rate. Hence, a present bias arising out of aggregate uncertainty could offset increasing impatience arising out of increased mortality, thus providing a possible explanation for the surprising patience of older individuals found in some studies. David M. Bishai (2004), for example, finds evidence from wage differentials that the rate of time preference declines with age.[33]

Dasgupta and Maskin (2005) and Sozou (1998) 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 re-

---

[32]Simply because it is inconsistent with exponential discounting, present bias *per se* is sometimes considered anomalous (as in Frederick, Loewenstein and O'Donoghue (2002) and Richard H. Thaler (1981)). More typically, it is assumed that a present bias must imply preference reversals.

[33]Eric Bettinger and Robert Slonim (2007) provide complementary evidence on the impatience of children. Declining impatience among children could reflect decreasing mortality, but this factor alone would imply rising impatience among adults.

alized. A source of future food may be seized by a hungry rival or access blocked by a predator.[34] 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 suggests 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.

Generalizations. Our analysis is based on an age-independent mortality rate. However, we would expect mortality to vary systematically over one's life span, especially near the beginning and end. We would then expect these variations to induce age-dependent discount-rate patterns beyond those appearing in our constant-death-rate model, tending to increase discounting among young children—who act as if there is no tomorrow—and

---

[34]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 (1998)) or by the prospect that the consumption opportunity may arrive early (in Dasgupta and Maskin (2005)). Karl Wärneryd (2007) presents an alternative model of presently-biased discounting based on intergenerational transfers, noting that with sexual reproduction one typically expects a child to carry a copy of one's genes with probability $\frac{1}{2}$ and a grandchild to do so with probability $\frac{1}{4}$, and that a tendency to select mates from somewhat interrelated groups can push this exponential sequence toward a present bias.

the elderly, reflecting then the typical human U-shaped mortality pattern.[35]

Our analysis provides the basic tools for examining age-dependent survival rates. Suppose that a life history now consists of a profile $(x_1, \ldots, x_T)$ of expected offspring and a profile $(S_0, \ldots, S_{T-1})$ of survival probabilities. Then, analogous to (4), the dominant eigenvalue of the Leslie matrix is given by

$$(17) \qquad 1 = \frac{S_0 x_1}{\phi} + \frac{S_0 S_1 x_2}{\phi^2} + \frac{S_0 S_1 S_2 x_3}{\phi^3} + \ldots + \frac{S_0 \cdots S_{T-1} x_T}{\phi^T}.$$

Notice first that taking $S'_\tau = S$ for all τ and $x'_\tau = S^{-\tau} S_0 S_1 \cdots S_{\tau-1} x_\tau$ for any $S \in (0,1)$ gives an equivalent system with an identical growth rate. As a result, we come immediately to Section 1.1's observation that any analysis in which idiosyncratic survival rates vary by age can be translated into an equivalent analysis with identical survival rates. Next, it follows from (17) that marginal rates of substitution are given by

$$-\frac{dx_{\tau+1}}{dx_\tau} = \frac{\phi}{S_\tau}.$$

This gives us the expected result that marginal rates of substitution will be higher when survival rates are lower. Turning to the case of aggregate uncertainty, we can think of the survival rates in (8) as being given by products $S_\tau \tilde{S}_\tau(t)$, allowing us to reformulate and extend Propositions 1–4 to a combination of arbitrary age-dependent idiosyncratic shocks with multiplicative aggregate shocks. There is thus considerable scope for pushing our model beyond its current focus to capture other considerations.

Implications. If our evolutionary model of discounting is on the right track, what sorts of behavior should we expect to see? First, we should not be surprised if discount rates exceed the sum of growth rates and death rates, with the gap being larger the more important was aggregate uncertainty in our evolutionary environment. In addition, we should not be surprised if discount rates are not constant. We have considered only small variations in survival rates across ages, giving rise to concomitantly small departures from

---

[35]At the same time, intergenerational transfers 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.

exponential discounting. Larger variations in death rates across ages might well rise to larger effects on discount rates.

Next, the role of mortality risk, long considered central in discounting, is more subtle than it first appears. Different populations that have equivalent fertility patterns and different death rates may well nonetheless exhibit identical discount factors. In our simple model, with arbitrary patterns of idiosyncratic uncertainty and uniform aggregate mortality shocks, any change in the death rate is matched by a corresponding change in the growth rate, leaving discount rates untouched. However, variations in death rates across agents within a given population, and hence within agents whose discount rates are shaped by the same population growth rate, should be directly reflected in discount rates.[36]

Perhaps most importantly, our analysis provides yet another indication that idiosyncratic and aggregate uncertainty can have quite different effects, and hence may enter our preferences quite differently. A standard finding in psychological studies of risk attitudes is that a feeling of control is important if inducing people to be comfortable with risk.[37] Risks arising out of situations in which people feel themselves unable to affect the outcome cause considerably more apprehension than risks arising out of circumstances people perceive themselves to control. Why might this be the case? The first task facing evolution in an attempt to induce different behavior in the face of idiosyncratic and aggregate risks is to give us a way of recognizing these risks. "Control" may be a convenient stand-in for an idiosyncratic risk. If so, then our seemingly irrational fear of uncontrolled risk may be a mechanism inducing an evolutionarily rational fear of aggregate risk.

---

[36]Margo Wilson and Martin Daly (1997) 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.

[37]See Paul Slovic, Baruch Fischhoff and Sarah Lichtenstein (1982) for an early contribution to this literature and Slovic (2000) for a more recent introduction.

4 Appendix

4.1 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 and satisfying $-\infty < E \ln u' \tilde{X} v < \infty$, with

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

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

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

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

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

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 $E\{\tilde{N}'(t)v\} = N'(t)v$, where v is the right eigenvector of X. We can then apply Jensen's inequality to show that this expectation is never higher under aggregate uncertainty than under the corresponding deterministic process:

$$\frac{\ln N'(t)v}{t} = \frac{\ln E\{\tilde{N}'(t)v\}}{t} \geq \frac{E\{\ln \tilde{N}'(t)v\}}{t}.$$

The inequality is strict if the distribution of $\tilde{N}(t)$ is nondegenerate.[38] The argument is completed by noting that the long-run average growth rate under the mean matrix is $\lim_{t\to\infty} \frac{\ln N'(t)v}{t}$ and under aggregate uncertainty is $\lim_{t\to\infty}\{\ln \tilde{N}'(t)v\}/t$, and that almost surely $\lim_{t\to\infty} \frac{\ln \tilde{N}'(t)v}{t} = \lim_{t\to\infty} \frac{E\ln \tilde{N}'(t)v}{t}$ (cf. Patrick Billingsley (1986, Theorem 25.12, p. 348)). ∎

4.2   Proof of Proposition 4

A key observation throughout the remaining proofs is that the growth rate

$$\lim_{t\to\infty} \frac{1}{t} \ln\left[u'\tilde{Z}(1)\ldots\tilde{Z}(t)v\right] = \lim_{t\to\infty} \frac{1}{t}E\ln\left[u'\tilde{Z}(1)\ldots\tilde{Z}(t)v\right] \equiv \Lambda(\varepsilon)$$

is jointly analytic in the matrix elements and the perturbation parameter ε.[39]

Taylor's theorem allows us to write

$$\Lambda(\varepsilon) = \Lambda(0) + \varepsilon\frac{d\Lambda(0)}{d\varepsilon} + \frac{\varepsilon^2}{2}\frac{d^2\Lambda(0)}{d\varepsilon^2} + \frac{\varepsilon^3}{6}\frac{d^3\Lambda(\varepsilon')}{d\varepsilon^3}$$

for some $\varepsilon' \in [0,\varepsilon]$. Define now the analytic function

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

so that

$$
\begin{aligned}
\Lambda(\varepsilon) &= \lim_{t\to\infty} \frac{1}{t}F(t,\varepsilon) \\
&= \lim_{t\to\infty} \frac{1}{t}F(t,0) + \varepsilon\lim_{t\to\infty} \frac{1}{t}\frac{dF(t,0)}{d\varepsilon} + \frac{\varepsilon^2}{2}\lim_{t\to\infty} \frac{1}{t}\frac{d^2F(t,0)}{d\varepsilon^2} + \frac{\varepsilon^3}{6}\frac{d^3\Lambda(\varepsilon')}{d\varepsilon^3}.
\end{aligned}
$$

---

[38]Example 1 shows that weak equality can obtain if the that different realizations of $\tilde{X}$ have different identical eigenvectors and different eigenvalues.

[39]The first inequality, which holds almost surely, follows from Billingsley (1986, Theorem 25.12, p. 348). Analyticity is shown by David Ruelle (1979, Theorem 3.1). Note that our assumptions imply there is an integer $k > 0$ such that any k-fold product of realized Leslie matrices is strictly positive. We can then represent our population process as an infinite product of randomly chosen strictly positive k-fold products of Leslie matrices. Taking C (in Ruelle's notation) to be the nonnegative orthant then ensures that Ruelle's sufficient condition is satisfied (Ruelle (1979, p. 69).

The second inequality follows from the analyticity of $F(t, \varepsilon)$ and $\Lambda(\varepsilon)$, which implies any that derivative of $\frac{1}{t} F(t, \varepsilon)$ converges to the corresponding derivative of $\Lambda(\varepsilon)$.[40]

We now note that

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

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

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

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

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

with analogous notation.

Then, using the facts that $u'X = u'\phi$, $Xv = \phi v$, $E\tilde{H}(t) = 0$, and that $\tilde{X}(i)$ and X differ in that the former involves the realized survival rate $\tilde{S}(i)$ and the latter the mean survival rate S, we have

$$\frac{1}{t} F(t, 0) = \ln \phi + E \ln \tilde{S} - \ln S$$
$$\frac{1}{t} \frac{dF(t,0)}{d\varepsilon} = 0$$
$$\frac{1}{t} \frac{d^2 F(t,0)}{d\varepsilon^2} = -\frac{E \left(u'H(1)v \right)^2}{\phi^2} E \left(\frac{\tilde{S}}{S} \right)^2$$

---

[40]See Nelson Dunford and Jacob T. Schwartz (1988, p. 228). Note that any real analytic function can be extended on a neighborhood to a complex analytic function. This result provides an independent proof that Λ is analytic.

for all t, and hence in the limit as t gets arbitrarily large.

Substituting into our Taylor expansion of $\Lambda(\varepsilon)$, we have

$$\Lambda(\varepsilon) = \ln \phi + E \ln \tilde{S} - \ln S - \frac{\varepsilon^2}{2} E \left(\frac{\tilde{S}}{S} \right)^2 \frac{E \left(u' H(1) v \right)^2}{\phi^2} + O(\varepsilon^3).$$

∎

4.3 Variances and (Non)Exponential Discounting

Set $T = 3$ and let $\tilde{S} = S \in (0, 1)$ with probability one, so there is no common component to the aggregate shocks to survival. Assume also that the age-dependent aggregate shocks are contemporaneously independent with variances V_0, V_1, and V_2. The variance component $\{(u' \tilde{H} v)\}$ in (13) is then[41]

$$
\begin{aligned}
u' \tilde{H} v &= [u_1, u_2, u_3] \begin{bmatrix} \hat{S}_0 x_1 & \hat{S}_1 & 0 \\ \hat{S}_0 x_2 & 0 & \hat{S}_2 \\ \hat{S}_0 x_3 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \\
&= v_1 \sum_{\tau=1}^{3} u_\tau x_\tau \hat{S}_0 + v_2 u_1 \hat{S}_1 + v_3 u_2 \hat{S}_2 \\
&= v_1 u_1 \Phi \hat{S}_0 + v_2 u_1 \hat{S}_1 + v_3 u_2 \hat{S}_2.
\end{aligned}
$$

Squaring and taking the expectation, using the independence of the aggregate shocks across ages, we have[42]

$$
\begin{aligned}
E(u' \tilde{H} v)^2 &= v_1^2 u_1^2 \Phi^2 V_0 + v_2^2 u_1^2 V_1 + v_3^2 u_2^2 V_2 \\
&= u_1^2 \Phi^2 \left[V_0 v_1^2 + V_1 \frac{v_2^2}{\Phi^2} + V_2 \frac{v_3^2}{\Phi^4} \right].
\end{aligned}
$$

(18)

If we convert the reproductive values v_2 and v_3 to their age-one equivalents, we find[43]

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

(19)

---

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

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

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

so that

$$E(u'\tilde{H}v)^2 = u_1^2 v_1^2 \Phi^2 \left[V_0 + \frac{V_1}{\Phi^2} \left(\frac{x_2}{\Phi} + \frac{x_3}{\Phi^2} \right)^2 + \frac{V_2}{\Phi^4} \left(\frac{x_3}{\Phi} \right)^2 \right]$$

and the long run growth rate is given by the following expression (using (13) and the degeneracy of $\tilde{S}$, and ignoring the $O(\varepsilon)$ error term)

$$\Lambda = \ln \lambda = \ln \phi - \frac{\varepsilon^2}{2} E \left(u'\tilde{H}v \right)^2 = \ln \phi - \frac{\varepsilon^2 u_1^2 v_1^2}{2S^2} \left[V_0 + \frac{V_1}{\Phi^2} \left(\frac{x_2}{\Phi} + \frac{x_3}{\Phi^2} \right)^2 + \frac{V_2}{\Phi^4} \left(\frac{x_3}{\Phi} \right)^2 \right].$$

In order to find complete expressions for the derivatives of Λ with respect to x_1 x_2 and x_3, 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, upon taking the derivatives, we find that only the explicit dependence of Λ on x_2 and x_3 introduces a distortion away from exponential discounting.

Letting $\Phi = 1$ is innocuous and simplifies the notation. Now, if only $\hat{S}_2$ were nondegenerate, we would have

$$\Lambda = \ln \phi - \frac{\varepsilon^2 V_2 u_1^2 v_1^2}{2S^2} \left[(x_3)^2 \right],$$

depressing only the partial derivative with respect to x_3 and giving

$$-\frac{dx_3}{dx_2} > -\frac{dx_2}{dx_1} = \Phi = 1,$$

so that the rate of discount *increases* on this account alone. Alternatively, if only $\hat{S}_1$ were nondegenerate, we would have

$$\Lambda = \ln \phi - \frac{\varepsilon^2 V_1 u_1^2 v_1^2}{2S^2} \left[(x_2 + x_3)^2 \right],$$

depressing the partial derivatives with respect to both x_2 and x_3. This increases $-\frac{dx_2}{dx_1}$ but can be shown to leave $-\frac{dx_3}{dx_2}$ constant. That is,

$$1 = \Phi = -\frac{dx_3}{dx_2} < -\frac{dx_2}{dx_1}.$$

In the case that $V_0 = V_1 = V_2 = V$, we have

$$\Lambda = \ln \phi - \frac{\varepsilon^2 u_1^2 v_1^2 V}{2S^2} \left[1 + (x_2 + x_3)^2 + (x_3)^2 \right].$$

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

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

Using the fact that the explicit dependence of Λ on x_2 and x_3 affects only the term of order ε^2, we can calculate that we obtain 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 (19), 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 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 the same unambiguous result when $\Phi \neq 1$.

4.4   Proof of Propositions 5 and 6

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

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

The random variables $\tilde{S}(t)$, $\hat{S}_{\tau'}(t')$ and $\hat{\mathbb{S}}_{\tau''}(t'')$ are all independent, except that $\hat{S}_\tau(t)$, $\hat{S}_{\tau'}(t)$, $\hat{\mathbb{S}}_{\tau''}(t)$ and $\hat{\mathbb{S}}_{\tau'''}(t)$ need not be independent for a given t. Let $C_{\tau\tau'}$ denote the contemporaneous covariance between $\hat{S}_\tau$ and $\hat{S}_{\tau'}$, let $\mathbb{C}_{\tau\tau'}$ denote the contemporaneous covariance between $\hat{\mathbb{S}}_\tau$ and $\hat{\mathbb{S}}_{\tau'}$, and let $\mathfrak{C}_{\tau\tau'}$ denote the contemporaneous covariance between $\hat{\mathbb{S}}_\tau$ and $\hat{S}_{\tau'}$. To make the notation more compact, let

$$x_\tau\hat{\mathbb{S}}_\tau \equiv \hat{\mathbb{Z}}_\tau$$
$$\hat{S}_\tau = \hat{Z}_\tau.$$

Expanding on our previous notation, let $\Lambda(x, \varepsilon) = \lim_{t\to\infty}\frac{1}{t}\ln u'Z(1)\dots Z(t)v$. From

38

Proposition 4, we can write

$$\Lambda(x,\varepsilon) = \Lambda(x,0) + \frac{\varepsilon^2}{2}\Lambda_{\epsilon\epsilon}(x,0) + O(\varepsilon^3)$$

Since $\Lambda(x,\varepsilon)$ is analytic in (x,ε) it follows readily that the Taylor series for $d\Lambda/dx_\tau(x,\varepsilon)$ for each τ is of the form

$$\frac{d\Lambda}{dx_\tau}(x,\varepsilon) = \frac{d\Lambda}{dx_\tau}(x,0) + \frac{\varepsilon^2}{2}\frac{d^3\Lambda}{d\epsilon^2 dx_\tau}(x,0) + O(\varepsilon^3)$$

This allows us to examine marginal rates of substitution by examining derivatives, with respect to the x_τ, of the second-order Taylor expansion of $\Lambda(x,\varepsilon)$.

Expanding (13), we have

$$\begin{aligned}
\Lambda &= \ln\phi + E\ln\tilde{S} - \ln S - \frac{\varepsilon^2}{2\phi^2}E\left(\frac{S}{\tilde{S}}\right)^2 E\left\{\left(\left[\sum_{i=1}^{T}u_i\hat{\mathbb{Z}}_i, u_1\hat{Z}_1, u_2\hat{Z}_2, \ldots, u_{T-1}\hat{Z}_{T-1}\right]v\right)^2\right\} + O(\varepsilon^3) \\
&= \ln\phi + E\ln\tilde{S} - \ln S - \frac{\varepsilon^2}{2\phi^2}E\left(\frac{S}{\tilde{S}}\right)^2 E\left\{\left(v_1\sum_{i=1}^{T}u_i\hat{\mathbb{Z}}_i + \sum_{i=1}^{T-1}v_{i+1}u_i\hat{Z}_i\right)^2\right\} + O(\varepsilon^3) \\
&= \ln\phi + E\ln\tilde{S} - \ln S \\
&\quad - \frac{\varepsilon^2}{2\phi^2}E\left(\frac{S}{\tilde{S}}\right)^2\left(v_1^2\sum_{i=1}^{T}\sum_{j=1}^{T}u_iu_jx_ix_j\mathbb{C}_{ij} + \sum_{i=1}^{T-1}\sum_{j=1}^{T-1}v_{i+1}v_{j+1}u_iu_jC_{ij} + v_1\sum_{i=1}^{T}\sum_{j=1}^{T-1}x_iu_iu_jv_{j+1}\mathbb{C}_{ij}\right) \\
\text{(20)}\quad &\quad + O(\varepsilon^3).
\end{aligned}$$

4.4.1 Proof of Proposition 6

The proof of Proposition 6 is notationally less involved, and so we present this argument first. From (20), we have

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

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 (again ignoring the $O(\varepsilon^3)$ error term)

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

Then we can take the derivatives

$$
(22) \qquad \frac{d\Lambda}{dx_\tau} = \frac{D\Lambda}{D\phi}\frac{d\phi}{dx_\tau} + \frac{d\Lambda}{dv_1}\frac{dv_1}{dx_\tau} + \frac{d\Lambda}{dx_\tau}.
$$

Note that u_T depends only on ϕ, while v_1 is given by,

$$
(23) \qquad v_1 = \frac{\sum_{\tau=1}^{T}\Phi^\tau}{\Phi^T + \sum_{\tau=2}^{T}(\tau-1)x_\tau\Phi^{T-\tau}}.
$$

This expression for v_1 follows from

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

and $v'u = 1$.[44] Hence, v_1 depends both on ϕ and $(x_1, \ldots, x_T)$. We can calculate:

---

[44]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. Note that $v_1 = \left(\frac{x_1}{\Phi} + \frac{x_2}{\Phi^2} + \frac{x_3}{\Phi^3} \ldots + \frac{x_{T-1}}{\Phi^{T-1}} + \frac{x_T}{\Phi^T}\right)v_1$.

$$\frac{d\Lambda}{dx_\tau} = \frac{D\Lambda}{D\phi}\Phi^{T-\tau}\frac{d\phi}{dx_T} - \varepsilon^2\left(\frac{d\Lambda}{dv_1}\frac{1}{\varepsilon^2}\right)\frac{(\tau-1)\Phi^{T-\tau}\sum_{\tau=1}^{T}\Phi^\tau}{[\Phi^T + \sum_{\tau=2}^{T}(\tau-1)x_\tau\Phi^{T-\tau}]^2}$$

$$- \varepsilon^2\frac{u_T^2 v_1^2}{2\phi^2}E\left(\frac{S}{\tilde{\tilde{S}}}\right)^2\left(2\sum_{j=1}^{T}\Phi^{2T-\tau-j}x_j\mathbb{C}_{\tau j}\right) + O(\varepsilon^3)$$

where we note that $\left(\frac{d\Lambda}{dv_1}\frac{1}{\varepsilon^2}\right)$ is of order zero. Let us now suppose $x_1 = x_2 = \ldots = x_T \equiv x$, and let

$$\alpha = \frac{D\Lambda}{D\phi}\frac{d\phi}{dx_T} > 0.$$

$$\beta = -\left(\frac{d\Lambda}{dv_1}\frac{1}{\varepsilon^2}\right)\frac{\sum_{\tau=1}^{T}\Phi^\tau}{[\Phi^T + \sum_{\tau=2}^{T}(\tau-1)x_\tau\Phi^{T-\tau}]^2} > 0$$

$$\gamma = x\frac{u_T^2 v_1^2}{2\phi^2}E\left(\frac{S}{\tilde{\tilde{S}}}\right)^2 > 0.$$

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

$$\frac{d\Lambda}{dx_\tau} = \alpha\Phi^{T-\tau} + \varepsilon^2\beta(\tau-1)\Phi^{T-\tau} - \varepsilon^2\gamma K_\tau + O(\varepsilon^3)$$

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

(25)
$$-\frac{dx_{\tau+1}}{dx_\tau} = \frac{\frac{d\Lambda}{dx_\tau}}{\frac{d\Lambda}{dx_{\tau+1}}} = \frac{\alpha\Phi^{T-\tau} + \varepsilon^2\beta(\tau-1)\Phi^{T-\tau} - \varepsilon^2\gamma K_\tau + O(\varepsilon^3)}{\alpha\Phi^{T-\tau-1} + \varepsilon^2\beta\tau\Phi^{T-\tau-1} - \varepsilon^2\gamma K_{\tau+1} + O(\varepsilon^3)}.$$

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

(26)
$$-\frac{dx_{\tau+1}}{dx_\tau} < -\frac{dx_{\tau+2}}{dx_{\tau+1}},$$

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

41

4.4.2 Proof of Proposition 5

From (20), using (23)–(24), we now have (hereafter omitting the $O(\varepsilon^3)$ term)

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

where

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

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

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

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 Λ, presented in the technical appendix. ∎

References

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

[2] Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström. 2008. "Eliciting risk and time preferences." *Econometrica*, 76(3):583–618.

[3] Bergstrom, Theodore C. 1997. "Storage for good times and bad: Of rats and men." University of California, Santa Barbara.

[4] Bettinger, Eric, and Robert Slonim. 2007. "Patience among children." *Journal of Public Economics*, 91(1–2):343–363.

[5] Billingsley, Patrick. 1986. *Probability and Measure*. John Wiley and Sons, New York.

[6] Bishai, David M. 2004. "Does time preference change with age?" *Journal of Population Economics*, 17(4):583–602.

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

[8] Cooper, William S., and Robert H. Kaplan. 2004. "Adaptive "coin-flipping": A decision theoretic examination of natural selection for random individual variation." *Journal of Theoretical Biology*, 94(1):135–151.

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

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

[11] Dunford, Nelson, and Jacob T. Schwartz. 1988. *Linear Operators Part I: General Theory*. John Wiley and Sons, New York. Wiley Classics Library Edition.

[12] Feller, William. 1971. *An Introduction to Probability Theory and Its Applications, Vol. II*. John Wiley, New York, 2nd edition.

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

[14] Frederick, Shane, George Loewenstein, and Ted O'Donoghue. 2002. "Time discounting and time preference: A critical view." *Journal of Economic Literature*, 40(2):351–401.

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

[16] Gillespie, John H. 1973. "Polymorphism in random environments." *Theoretical Population Biology*, 4(2):193–195.

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

[18] Hansson, Ingemar, and Charles Stuart. 1990. "Malthusian selection of preferences." *American Economic Review*, 80(3):529–544.

[19] Hill, Kim, and A. Magdalena Hurtado. 1996. *Ache Life History.* Aldine de Gruyter, New York.

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

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

[22] Leslie, P. H. 1945. "On the use of matrices in certain population mathematics." *Biometrica*, 33(3):183–212.

[23] Leslie, P. H. 1948. "Some further notes on the use of matrices in population mathematics." *Biometrica*, 35(1–2):213–245.

[24] Litzenberger, Robert H., and Cherukuri U. Rao. 1971. "Estimates of the marginal rate of time preference and average risk aversion of investors in electric utility shares: 1960–66." *The Bell Journal of Economics and Management Science*, 2(1):265–277.

[25] McSweeney, Kendra, and Shahna Arps. 2005. "A 'demographic turnaround': The rapid growth of indigenous populations in lowland Latin America." *Latin American Research Review*, 40(1):3–29.

[26] Nordhaus, William. 2007. "The *Stern Review* on the economics of climate change." *Journal of Economic Literature*, 45(3):686–702.

[27] Robson, Arthur J. 1996. "A biological basis for expected and non-expected utility." *Journal of Economic Theory*, 68(2):397–424.

[28] Robson, Arthur J., and Larry Samuelson. 2007. "The evolution of intertemporal preferences." *American Economic Review*, 97(2 (May)):496–500.

[29] Robson, Arthur J., and Larry Samuelson. Forthcoming. "The evolutionary foundations of preferences." In Jess Benhabib, Alberto Bisin, and Matthew Jackson, editors, *The Social Economics Handbook*. Elsevier, New York.

[30] Robson, Arthur J., and Balazs Szentes. 2008. "Evolution of time preference by natural selection: Comment." *American Economic Review*, 98(3):1178–1188.

[31] Robson, Arthur J., Balazs Szentes, and Emil Iantchev. 2005. "An evolutionary approach towards time preference." University of Chicago.

[32] Rogers, Alan R. 1994. "Evolution of time preference by natural selection." *American Economic Review*, 84(2):460–481.

[33] Ruelle, David. 1979. "Analycity properties of the characteristic exponents of random matrix products." *Advances in Mathematics*, 32(1):68–80.

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

[35] Slovic, Paul. 2000. *The Perception of Risk.* Earthscan Publications, London.

[36] Slovic, Paul, Baruch Fishhoff, and Sarah Lichtenstein. 1982. "Why study risk perception?" *Risk Analysis*, 2(2):83–93.

[37] Sozou, Peter D. 1998. "On hyperbolic discounting and uncertain hazard rates." *Proceedings of the Royal Society of London, Series B*, 265(1409):2015–2020.

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

[39] Thaler, Richard H. 1981. "Some empirical evidence on dynamic inconsistency." *Economics Letters*, 8(3):201–207.

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

[41] Wärneryd, Karl. 2007. "Sexual reproduction and time-inconsistent preferences." *Economics Letters*, 95(1):14–16.

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

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

THE EVOLUTION OF TIME PREFERENCE WITH AGGREGATE UNCERTAINTY

Technical Appendix: Details of Proofs

Not for publication

Equation numbers such as (17) refer to equations in the paper, while (A1) denotes an equation in the technical appendix.

Proof of Proposition 6

We begin the argument with equations (25)–(26) of the paper. These indicate that we have increasing impatience if, for $\tau = 2, \ldots, T - 2$,

$$-\frac{dx_{\tau+1}}{dx_\tau} < -\frac{dx_{\tau+2}}{dx_{\tau+1}}$$

or

$$
\frac{\frac{d\Lambda}{dx_\tau}}{\frac{d\Lambda}{dx_{\tau+1}}} = \frac{\alpha\Phi^{T-\tau} + \epsilon^2\beta(\tau-1)\Phi^{T-\tau} - \epsilon^2\gamma K_\tau + O(\epsilon^3)}{\alpha\Phi^{T-\tau-1} + \epsilon^2\beta\tau\Phi^{T-\tau-1} - \epsilon^2\gamma K_{\tau+1} + O(\epsilon^3)}
$$

$$
< \frac{\alpha\Phi^{T-\tau-1} + \epsilon^2\beta\tau\Phi^{T-\tau-1} - \epsilon^2\gamma K_{\tau+1} + O(\epsilon^3)}{\alpha\Phi^{T-\tau-2} + \epsilon^2\beta(\tau+1)\Phi^{T-\tau-2} - \epsilon^2\gamma K_{\tau+2} + O(\epsilon^3)} = \frac{\frac{d\Lambda}{dx_{\tau+1}}}{\frac{d\Lambda}{dx_{\tau+2}}}.
$$

For ϵ sufficiently small, this inequality is implied by[45]

$$\epsilon^2\alpha\Phi^{T-\tau}\beta(\tau+1)\Phi^{T-\tau-2} - \epsilon^2\alpha\Phi^{T-\tau}\gamma K_{\tau+2}$$

$$\epsilon^2\alpha\Phi^{T-\tau-2}\beta(\tau-1)\Phi^{T-\tau} - \epsilon^2\alpha\Phi^{T-\tau-2}\gamma K_\tau$$

$$< 2\epsilon^2\alpha\Phi^{T-\tau-1}\beta\tau\Phi^{T-\tau-1} - 2\epsilon^2\alpha\Phi^{T-\tau-1}\gamma K_{\tau+1}.$$

The terms involving β cancel one another. Then dividing by $-\epsilon^2\alpha\gamma$, it suffices that

(A3) $$\Phi^{T-\tau}K_{\tau+2} + \Phi^{T-\tau-2}K_\tau > 2\Phi^{T-\tau-1}K_{\tau+1}.$$

---

[45]Cross multiplication gives identical terms of order ϵ^0 on both sides. The next largest terms, of order ϵ^2, are collected below.

Dividing by $\Phi^{T-\tau}$ and substituting for K, this is

$$2\sum_{j=1}^{T}\Phi^{2T-(\tau+2)-j}\mathbb{C}_{\tau+2,j} + \Phi^{-2}\left(2\sum_{j=1}^{T}\Phi^{2T-\tau-j}\mathbb{C}_{\tau,j}\right)$$

$$> \Phi^{-1}\left(4\sum_{j=1}^{T}\Phi^{2T-(\tau+1)-j}\mathbb{C}_{\tau+1,j}\right).$$

Recalling our assumption that each covariance is equal to C and each variance equal to V, we see that the terms in the summation corresponding to values of j other than τ, $\tau+1$ and $\tau+2$ cancel. Dividing by $\Phi^{2T-\tau}$, it suffices that

$$2\Phi^{-\tau-2}C + 2\Phi^{-\tau-3}C + 2\Phi^{-\tau-4}V$$

$$+\quad 2\Phi^{-\tau-2}V + 2\Phi^{-\tau-3}C + 2\Phi^{-\tau-4}C$$

$$>\quad 4\Phi^{-\tau-2}C + 4\Phi^{-\tau-3}V + 4\Phi^{-\tau-4}C.$$

Dividing by $2\Phi^{-\tau}$, we can rearrange to obtain the sufficient condition

$$2\Phi^{-3}C + \Phi^{-4}V + \Phi^{-2}V \quad > \quad \Phi^{-2}C + \Phi^{-4}C + 2\Phi^{-3}V.$$

Multiplying by Φ^4, this is equivalent to

$$2\Phi C + V + \Phi^2 V > \Phi^2 C + C + 2\Phi V$$

or

$$V(\Phi-1)^2 > C(\Phi-1)^2,$$

and so the result follows. ∎

Proof of Proposition 5

We begin with equation (28) of the paper, giving

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

Then, analogously to (22) of the paper, we are interested in derivatives of the form (where $d\Lambda/d\phi$ is derived analogously to (21)

$$\frac{d\Lambda}{dx_\tau} = \frac{D\Lambda}{D\phi}\Phi^{T-\tau}\frac{d\phi}{dx_T} + \frac{d\Lambda}{dv_1}\frac{dv_1}{dx_\tau} - \frac{\epsilon^2 u_T^2 v_1^2}{2\phi^2}E\left(\frac{S}{\tilde{S}}\right)^2\frac{dK}{dx_\tau}.$$

Then following the reasoning that took us from steps (22) of the paper to (A3), we have decreasing marginal rates of substitution if[46]

$$\Phi^{T-\tau}\frac{dK}{dx_{\tau+2}} + \Phi^{T-\tau-2}\frac{dK}{dx_\tau} < 2\Phi^{T-\tau-1}\frac{dK}{dx_{\tau+1}},$$

or

(A4)
$$\Phi^2\frac{dK}{dx_{\tau+2}} + \frac{dK}{dx_\tau} < 2\Phi\frac{dK}{dx_{\tau+1}}.$$

To verify (A4), we must first calculate dK/dx_τ. This is (see (27) for k_i)

$$
\begin{aligned}
\frac{dK}{dx_\tau} &= \sum_{j=1}^{T}\Phi^{2T-\tau-j}x_j\mathbb{C}_{\tau j} + \sum_{i=1}^{T}\Phi^{2T-i-\tau}x_i\mathbb{C}_{i\tau} + \sum_{i=1}^{T-1}\sum_{j=1}^{T-1}\Phi^{2T-i-j}\left(\frac{dk_i}{dx_\tau}k_j + \frac{dk_j}{dx_\tau}k_i\right)C_{ij}\\
&\quad + \sum_{j=1}^{T-1}\Phi^{2T-\tau-j}k_j\mathfrak{C}_{\tau j} + \sum_{i=1}^{T}\sum_{j=1}^{T-1}\Phi^{2T-i-j}x_i\frac{dk_j}{dx_\tau}\mathfrak{C}_{ij}\\
&= \Phi^{2T}\left(2\sum_{j=1}^{T}\Phi^{-\tau-j}x_j\mathbb{C}_{\tau j} + 2\sum_{i=1}^{T-1}\sum_{j=1}^{T-1}\Phi^{-i-j}\frac{dk_i}{dx_\tau}k_jC_{ij}\right.\\
&\quad \left. + \sum_{j-1}^{T-1}\Phi^{-\tau-j}k_j\mathfrak{C}_{\tau j} + \sum_{i=1}^{T}\sum_{j=1}^{T-1}\Phi^{-i-j}x_i\frac{dk_j}{dx_\tau}\mathfrak{C}_{ij}\right)\\
&= \Phi^{2T}\left(2\sum_{j=1}^{T}\Phi^{-\tau-j}x_j\mathbb{C}_{\tau j} + 2\sum_{i=1}^{\tau-1}\sum_{j=1}^{T-1}\Phi^{-i-j}\Phi^{-(\tau-i)}k_jC_{ij}\right.\\
&\quad \left. + \sum_{j=1}^{T-1}\Phi^{-\tau-j}k_j\mathfrak{C}_{\tau j} + \sum_{i=1}^{T}\sum_{j=1}^{\tau-1}\Phi^{-i-j}x_i\Phi^{-(\tau-j)}\mathfrak{C}_{ij}\right),
\end{aligned}
$$

where the first equality collects like terms and the second uses (27) to take derivatives of k_i. Inserting in (A4) we then have increasing impatience if

---

[46] At this point, we simply write dK/dx_τ rather than taking the derivative explicitly; the corresponding derivative in moving from (20) to (A3) is $2\sum_{j=1}^{T}\Phi^{2T-\tau-j}x_j\mathbb{C}_{\tau j}$, the notation for which we subsequently simplify by letting $K_\tau \equiv 2\sum_{j=1}^{T}\Phi^{2T-\tau-j}\mathbb{C}_{\tau j}$ when deriving (A3).

$$2\Phi^2 \sum_{j=1}^{T} \Phi^{-(\tau+2)-j} x_j \mathbb{C}_{\tau+2,j} + 2\Phi^2 \sum_{i=1}^{\tau+1}\sum_{j=1}^{T-1} \Phi^{-i-j}\Phi^{-((\tau+2)-i)} k_j C_{ij}$$

$$+ \Phi^2 \sum_{j=1}^{T-1} \Phi^{-(\tau+2)-j} k_j \mathfrak{C}_{\tau+2,j} + \Phi^2 \sum_{i=1}^{T}\sum_{j=1}^{\tau+1} \Phi^{-i-j} x_i \Phi^{-((\tau+2)-j)} \mathfrak{C}_{ij}$$

$$+ 2 \sum_{j=1}^{T} \Phi^{-\tau-j} x_j \mathbb{C}_{\tau j} + 2 \sum_{i=1}^{\tau-1}\sum_{j=1}^{T-1} \Phi^{-i-j}\Phi^{-(\tau-i)} k_j C_{ij}$$

$$+ \sum_{j-1}^{T-1} \Phi^{-\tau-j} k_j \mathfrak{C}_{\tau j} + \sum_{i=1}^{T}\sum_{j=1}^{\tau-1} \Phi^{-i-j} x_i \Phi^{-(\tau-j)} \mathfrak{C}_{ij}$$

$$< \quad 4\Phi \sum_{j=1}^{T} \Phi^{-(\tau+1)-j} x_j \mathbb{C}_{\tau+1,j} + 4\Phi \sum_{i=1}^{\tau}\sum_{j=1}^{T-1} \Phi^{-i-j}\Phi^{(-(\tau+1)-i)} k_j C_{ij}$$

$$+ 2\Phi \sum_{j=1}^{T-1} \Phi^{-(\tau+1)-j} k_j \mathfrak{C}_{\tau+1,j} + 2\Phi \sum_{i=1}^{T}\sum_{j=1}^{\tau} \Phi^{-i-j} x_i \Phi^{-((\tau+1)-j)} \mathfrak{C}_{ij}.$$

It is then helpful to tackle this inequality in parts. We begin with the first and fifth terms on the left, and the first on the right. These are precisely the terms that entered the calculations in proving Proposition 6, leading to increasing impatience. In this case, given our assumption that the random variables $\hat{\mathbb{Z}}_\tau$, for $\tau = 1, \ldots, T$, are perfectly correlated, these terms cancel.

Now we work on the second and sixth terms on the left and the second on the right. We have

$$2\Phi^2 \sum_{i=1}^{\tau+1}\sum_{j=1}^{T-1} \Phi^{-i-j}\Phi^{-((\tau+2)-i)} k_j C_{ij} + 2 \sum_{i=1}^{\tau-1}\sum_{j=1}^{T-1} \Phi^{-i-j}\Phi^{-(\tau-i)} k_j C_{ij}$$

$$< \quad 4\Phi \sum_{i=1}^{\tau}\sum_{j=1}^{T-1} \Phi^{-i-j}\Phi^{-((\tau+1)-i)} k_j C_{ij}$$

if

$$2 \sum_{i=1}^{\tau+1}\sum_{j=1}^{T-1} \Phi^{-i-j}\Phi^{-(\tau-i)} k_j C_{ij} + 2 \sum_{i=1}^{\tau-1}\sum_{j=1}^{T-1} \Phi^{-i-j}\Phi^{-(\tau-i)} k_j C_{ij}$$

$$< \quad 4 \sum_{i=1}^{\tau}\sum_{j=1}^{T-1} \Phi^{-i-j}\Phi^{-(\tau-i)} k_j C_{ij}$$

if

$$2\sum_{i=\tau}^{\tau+1}\sum_{j=1}^{T-1}\Phi^{-i-j}\Phi^{-(\tau-i)}k_jC_{ij} < 4\sum_{i=\tau}^{\tau}\sum_{j=1}^{T-1}\Phi^{-i-j}\Phi^{-(\tau-i)}k_jC_{ij}$$

if

$$2\sum_{j=1}^{T-1}\Phi^{-\tau+-j}k_jC_{\tau+1,j} < 2\sum_{j=1}^{T-1}\Phi^{-\tau-j}k_jC_{\tau j}$$

if

$$\Phi^{-\tau}k_\tau C + \Phi^{-(\tau+1)}k_{\tau+1}V < \Phi^{-\tau}k_\tau V + \Phi^{-(\tau+1)}k_{\tau+1}$$

if

$$2\Phi^{-2\tau}C\left(k_\tau - \frac{k_{\tau+1}}{\Phi}\right) < 2\Phi^{-2\tau}V\left(k_\tau - \frac{k_{\tau+1}}{\Phi}\right)$$

if

(A5)
$$2\Phi^{-2\tau}x_{\tau+1}C < 2\Phi^{-2\tau}x_{\tau+1}V,$$

which follows from $C < V$ given $x_{\tau+1} > 0$.

Now we turn to the third and seventh term on the left and the third on the right. Here we have

$$\Phi^2\sum_{j=1}^{T-1}\Phi^{-(\tau+2)-j}k_j\mathfrak{C}_{\tau+2,j} + \sum_{j=1}^{T-1}\Phi^{-\tau-j}k_j\mathfrak{C}_{\tau j} = 2\Phi\sum_{j=1}^{T-1}\Phi^{-(\tau+1)-j}k_j\mathfrak{C}_{\tau+1,j}.$$

Taking out a factor $\Phi^{-\tau}$, this holds if

$$\sum_{j=1}^{T-1}\Phi^{-j}k_j\mathfrak{C}_{\tau+2,j} + \sum_{j=1}^{T-1}\Phi^{-j}k_j\mathfrak{C}_{\tau j} = 2\sum_{j=1}^{T-1}\Phi^{-j}k_j\mathfrak{C}_{\tau+1,j}.$$

However, each of the terms $\mathfrak{C}_{\tau\tau'}$ represents the common covariance C between one of the random variables $\hat{Z}_1, \ldots, \hat{Z}_{\tau-1}$ and the $\hat{\mathbb{Z}}_0$. Hence, this equality holds.

Finally, we work on the fourth and eighth terms on the left, and the fourth term on the right. Here, we have

$$\Phi^2\sum_{i=1}^{T}\sum_{j=1}^{\tau+1}\Phi^{-i-j}x_i\Phi^{-((\tau+2)-j)}\mathfrak{C}_{ij} + \sum_{i=1}^{T}\sum_{j=1}^{\tau-1}\Phi^{-i-j}x_i\Phi^{-(\tau-j)}\mathfrak{C}_{ij} = 2\Phi\sum_{i=1}^{T}\sum_{j=1}^{\tau}\Phi^{-i-j}x_i\Phi^{(-(\tau+1)-j)}\mathfrak{C}_{ij}.$$

Taking out $\Phi^{-\tau}$, this holds if

$$\sum_{i=1}^{T}\sum_{j=1}^{\tau+1}x_i\Phi^{-i}\mathfrak{C}_{ij} + \sum_{i=1}^{T}\sum_{j=1}^{\tau-1}x_i\Phi^{-i}\mathfrak{C}_{ij} = \sum_{i=1}^{T}\sum_{j=1}^{\tau}x_i\Phi^{-i}\mathfrak{C}_{ij}.$$

Once again, each of the terms $\mathfrak{C}_{\tau\tau'}$ represents the common covariance C between one of the random variables $\hat{Z}_1, \ldots, \hat{Z}_{\tau-1}$ and the $\hat{\mathbb{Z}}_0$. These terms are thus constant in j, and hence the equality can be verified by simply counting the number of terms on each side. The desired result (A4) then follows from (A5). ∎