THE AMOUNT AND SPEED OF DISCOUNTING[1]

Dean T. Jamison[2] and Julian Jamison[3]

November 2007

[1]Earlier versions of this paper were presented at the Annual Meeting of the European Economic Association, Stockholm, August 2003, and at the Meeting of the International Health Economics Association in Barcelona, July 2005. The authors received valuable comments – from David Bradford, Ramanan Laxminarayan, Joseph Lipscomb, Philip Musgrove, Richard Newell, William Pizer, Paul Portney, Milton Weinstein and Martin Weitzman – on an earlier draft at a seminar at Resources for the Future. Kenneth Arrow provided valuable written comments. Johannes Hörner, Andrew Metrick and Jia Wang also provided helpful comments. The Fogarty International Center of the U.S. National Institutes of Health (through Order No. 263-MK-610514) provided partial support for preparation of this paper. Conclusions and opinions expressed are those of the authors and not of the U.S. government.

[2]Kennedy School of Government, Harvard University; and Global Health Sciences, UC San Francisco

[3]Brain and Creativity Institute, University of Southern California

Abstract

This paper introduces the concepts of amount and speed of a discounting procedure. Exponential discounting sequesters both concepts into a single parameter that needs to be disaggregated in order to characterize nonconstant rate procedures. The inverse of the present value of a unit stream of benefits provides a natural measure of the *amount* a procedure discounts the future. We propose geometrical and time horizon based measures of how rapidly a discounting procedure acquires its ultimate present value, and we prove these to be the same. This provides an unambiguous measure of the speed of discounting, a measure whose values lie between 0 (slow) and 2 (fast). Exponential discounting has a speed of 1. A commonly proposed approach to aggregating individual discounting procedures into a social one averages the individual discount functions. We point to serious shortcoming with this approach and propose an alternative that, for logarithmic utility, is market based and for which the amount and time horizon of the social procedure are the averages of the amounts and time horizons of the individual procedures. We further show that the social procedure will in general be slower than the average of the speeds of the individual procedures. We then characterize three families of discounting procedures in terms of their discount functions, their discount rate functions, their amounts, their speeds and their time horizons. A one parameter hyperbolic discounting procedure, $d(t) = (1 + rt)^{-2}$, has amount r and speed 0, and we argue that this zero-speed hyperbolic is well suited for social project evaluation.

JEL codes: D51, D90, H43 and Q51

Keywords: discounting, social discount rate, project evaluation, exchange economies, hyperbolic discounting

The Amount and Speed of Discounting

Discounting at a constant rate has the virtues of familiarity, analytic tractability, time-consistency of preferences, and a well-understood axiomatic foundation. That said, perceived shortcomings of discounting at a constant rate have led economists in diverse fields increasingly to suggest procedures in which the discount rate varies over time. Examples occur in contexts involving the full range of time horizons.

Giving noticeable weight to the far future with exponential discounting comes at the cost of entailing virtually no discounting in the short- to medium-term. Hence the literature on economic evaluation of long time-horizon environmental or health projects increasingly contains proposals for use of discount rates that decline with time, or slow discounting, in order to give a 'reasonable' weight to far-future outcomes (e.g., Cline, 1999; Weitzman, 2001; Newell and Pizer, 2003). Actual practice for the most part continues to use constant rates.[1]

A recent strand of work in economics imports findings from experimental psychology into the understanding of economic behavior (Loewenstein, 1992; Shane, Loewenstein and O'Donoghue, 2002). Psychologists – and the behavioral economists using their findings and methods – focus much of their interest on behavior over relatively short periods, behavior that often seems consistent with discount functions that decline slowly then rapidly, such as the quasi-hyperbolic (e.g. Laibson, 1997). This generates interest in fast discounting in a sense that we will make clear.

The time period of interest for financial markets – typically up to thirty years – lies between that of the behavioral and environmental economists. The rising yield curves that frequently characterize bond markets imply fast discounting, but the empirically observed variation in yield curves includes discounting that is slow as well as fast. Financial economics is thus a third strand of analysis where non-constant rate discounting is relevant and, indeed, is routine.

Our purpose in this paper is to provide a framework for the increasingly diverse literature using non-constant rate discounting by exploring the implications of distinguishing the *speed* of discounting from the total *amount* by which the future is discounted. Once these concepts are separated, the existing heterogeneous collection of proposals about how to discount can be simply

[1] Current guidelines for economic evaluation of health projects (Lipscomb, Weinstein and Torrance, 1996) recommended a constant rate of 3% per year. Stern (2006), in an influential assessment of the economics of climate changes, used 1.4% per year.

characterized and alternative proposals for aggregation of multiple individual procedures into a social one can be better evaluated. Important work continues on the nature of the factors (e.g. time preference, the return to capital, the marginal utility of consumption and the completeness of financial markets) that influence discounting or the term structure of interest rates. There is a natural division of labor between these topics and the task of characterizing the basic properties of discounting procedures. This paper addresses the latter.

Section 1 provides examples of slow and fast discounting and introduces our approach by providing definitions of the amount, the speed and the time horizon of discounting procedures. It then proves several basic propositions concerning these measures. Section 2 deals with aggregating multiple individual discounting procedures into a socially representative one. It points to problems with averaging of discount functions as an aggregation procedure and suggests an alternative with the attractive property that the means of the amounts and of the time horizons of the individual procedures equal the aggregate procedure's amount and time horizon. Our proposed aggregation procedure is also shown, under plausible assumptions, to result in discount rates identical to market-determined ones, which emphasize the preferences of impatient consumers in early years and patient consumers in later years. Section 3 presents a number of specific discount functions and characterizes their amounts, speeds and time horizons. It identifies attractive one-parameter fast and slow discount functions and argues that a particular slow procedure, which we label the zero-speed hyperbolic (ZSH) function, provides a natural alternative to the exponential for social discounting with long time horizons. Appendices provide proofs and amplifying material.

1 Measuring the Amount and Speed of Discounting

Two of the ways of defining a discounting procedure are by a discount function, $d(t)$ – where $d(0) = 1,\ d(t) \geq 0$, and $d(t)$ is nonincreasing for all t – or by a present value function, $pv(t)$, that gives the present value of a unit stream of benefits accumulated to time t :

$$pv(t) = \int_0^t d(x)dx. \tag{1}$$

We define the 'present value of $d(t)$' to be $pv(\infty)$.

In this section we first provide examples that illustrate how, for a given $pv(\infty)$, differing discounting procedures accumumulate present value at different rates. We use this to motivate two conceptually different definitions of speed of discounting,.a geometrical measure and a measure based on time horizon. We next prove several basic propositions concerning speed: our alternative definitions of speed are equivalent; our measure of speed lies between 0 and 2; procedures with declining (increasing) discount rates have speeds less than (greater than) 1; and exponential discounting has a speed of 1.

1.1 Examples and Definitions

Consider the following four discount functions:[2]

(2) exponential: $d(t) = e^{-rt}$, where $r = 0.02$

(3) hyperbolic: $d(t) = \left[1 + (\sigma^2/\mu)t\right]^{-\mu^2/\sigma^2}$, where $\mu = .04$ and $\sigma = .029$

(4) quasi-hyperbolic: $d(t) = 1$ for $t = 0$ and $d(t) = b(1+r)^{-(t-1)}$ for $1 \le t < \infty$ and t an integer, where $b = 0.6$ and $r = .0121$

(5) fast Weibull: $d(t) = e^{-rt^{1/s}}$, where $r = .000314$ and $s = 0.5$.

Equation (2) is the discount function with a constant discount rate of 0.02 and, hence, a present value [$= pv(\infty)$] of $.02^{-1} = 50$. Equation (3) represents a parameterization of the hyperbolic family that Weitzman (2001) used as an aggregation of the exponential discount functions resulting from a survey of economists, but with one parameter modified slightly to reduce present value from the 54.8 his parameters imply to 50. Equation (4) is the 'quasi-hyperbolic' function used by Laibson (1997), again with parameters modified to reduce present value from his implied 60.4 to 50. Read (2001) has suggested the formulation in equation (5), which is an exponential with the value for time transformed – in this case by squaring it – before being exponentiated. [Any discount function has an associated probability density function (pdf), and for equation (5) the

[2] Continuous time formulations simplify most derivations and, with only occasional exception, such as in equation (4), we use continuous time.

pdf is the Weibull distribution, hence our nomenclature.] We have again chosen parameters so that $pv(\infty) = 50$ for the Weibull discount function.

Although each of these discount functions has a present value of 50 they differ in how rapidly they acquire that present value. Figure 1 illustrates how each present value function rises with time to its asymptote of 50. Note several points: both the hyperbolic and quasi-hyperbolic functions rise more slowly than the exponential in the sense that, for both of them, $pv(t)$ is strictly less than it is for the exponential for all $t > 0$. They can thus, in this sense, be viewed as slower than the exponential with the same present value. Second, the Weibull is, with the indicated parameters, faster than the exponential. (The Weibull procedure can be either fast or slow.) Third, the differences among the procedures translate into major differences in the weight given the far future: while the Weibull has acquired essentially all of its present value within 150 years, and the exponential within 250 years, the hyperbolic has over 6% of its total present value still to be acquired after 500 years. Finally, since the present value functions for the hyperbolic and the quasi-hyperbolic cross, neither can be considered strictly slower than the other. Crossing of Lorenz curves provides a close analogy. Just as the Gini coefficient provides one way to complete the inequality ordering on income distributions generated by Lorenz curves, so, too, will an area based measure allow completion of the ordering of the speed of discount functions.

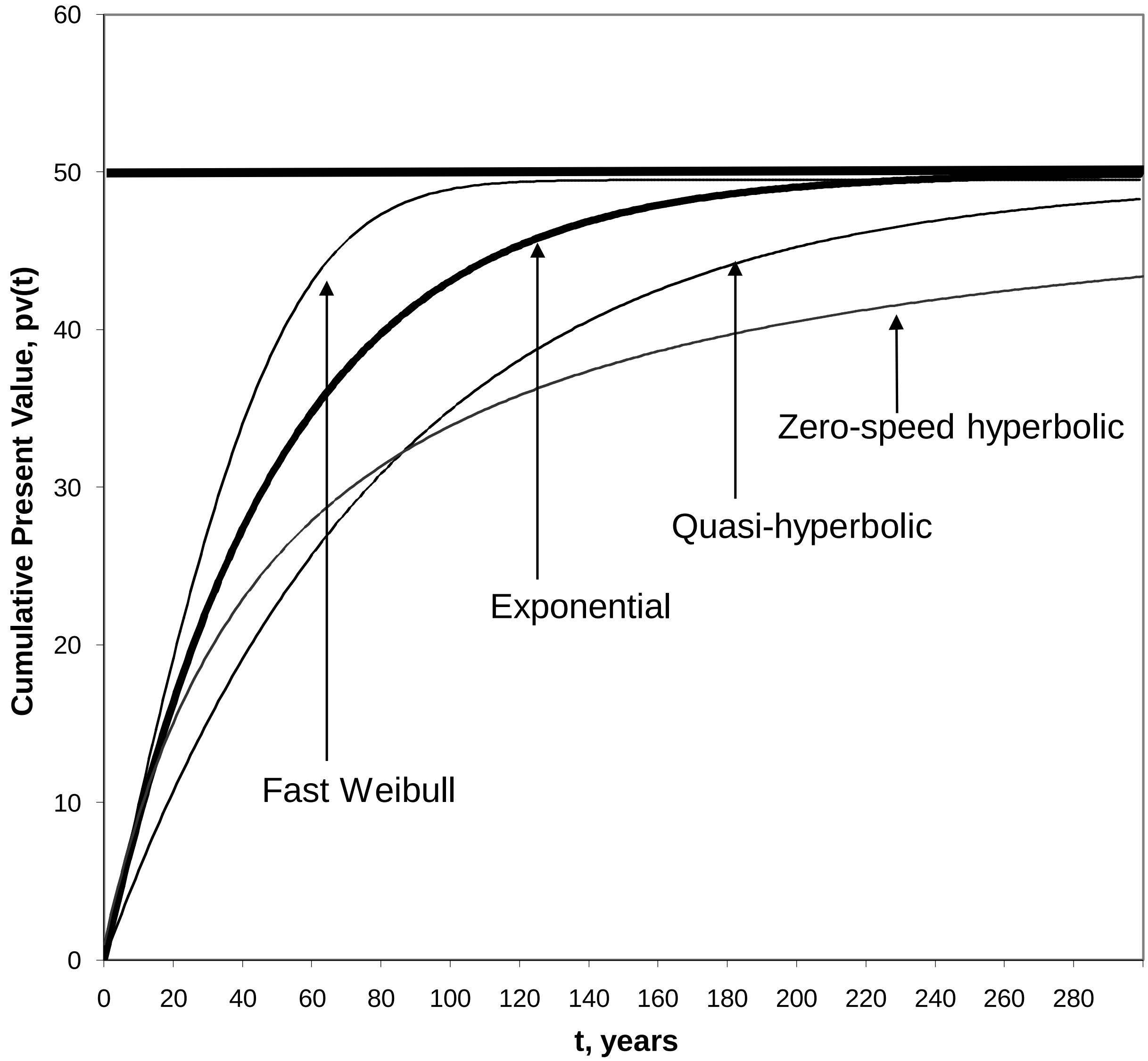
The more that one discounts the future, the less a unit stream of benefits will be worth now. It is thus natural to define the *amount* of discounting for a procedure, $\alpha(D)$, to be the inverse of its present value:

$$\alpha(D) = \left[\int_0^\infty d_D(t) dt \right]^{-1}. \tag{6}$$

If D is an an exponential procedure with constant discount rate r, then $\alpha(D)$ is, of course, simply r. This makes sense insofar as we think of higher discount rates as discounting away the future by a greater amount.

Figure 2 provides a geometrical motivation for the first definition of speed that we introduce. The greater the area between the $pv(\infty)$ of a discounting procedure and its $pv(t)$ function the slower, intuitively, it appears to be. We have denoted this area for the exponential in Figure 2 as **A**, and the following expression gives its value:

Figure 1:
Four Discounting Procedures with pv(∞) = 50



$$\mathbf{A} = \int_0^\infty [r^{-1} - \int_0^t e^{-rx} dx] \, dt, \tag{7}$$

where the inner integral is, of course, the expression for $pv(t)$ for an exponential, and r^{-1} is $pv(\infty)$ for the exponential (illustrated by the horizontal line at 50). Evaluating the integrals in equation (6) gives:

$$\mathbf{A} = r^{-2}. \tag{8}$$

Likewise $\mathbf{B}$ is the area between the quasi-hyperbolic and the line r^{-1}. The ratio $\mathbf{A}/\mathbf{B}$ provides a geometrical measure of the speed of this quasi-hyperbolic relative to that of an exponential with the same present value.

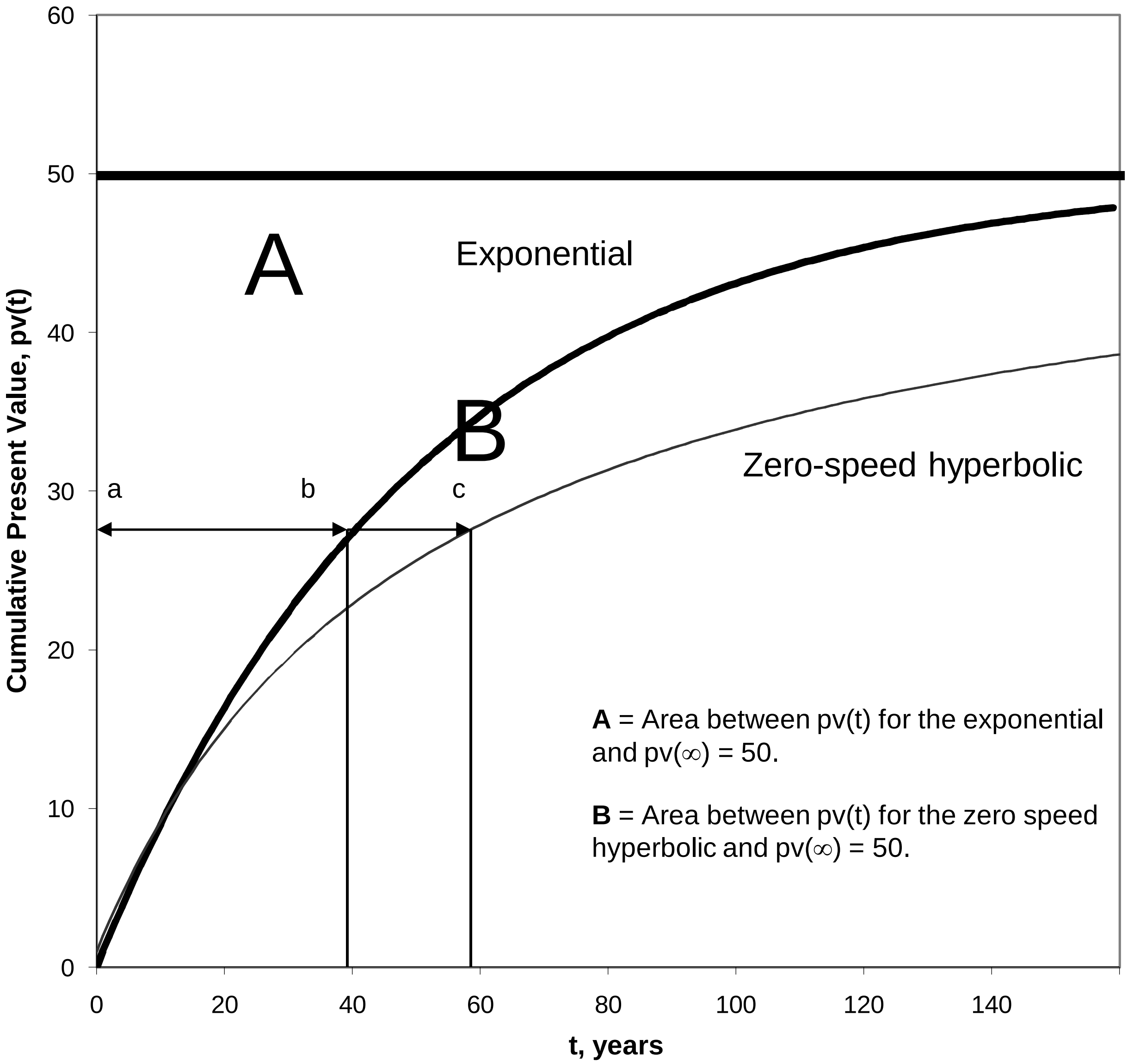
More precisely, our geometrical measure of the *speed* of a discounting procedure[3], D, is $\rho_1(D)$ and the preceding discussion suggests the following:

$$\begin{aligned}
&(9) & \mathbf{A}(D) &= \text{area between } pv_D(\infty) \text{ and the exponential with the same} \\
& & & \quad \text{present value as } D\text{, i.e. the exponential with } r = pv_D(\infty)^{-1}; \\
&(10) & \mathbf{B}(D) &= \text{area between } pv_D(\infty) \text{ and } pv_D(t); \\
& & &= \int_0^\infty [pv_D(\infty) - pv_D(t)] \, dt, \text{if this integral converges; and} \\
&(11) & \rho_1(D) &= \mathbf{A}(D)/\mathbf{B}(D).
\end{aligned}$$

Another characteristic of potential interest for a discounting procedure concerns how long a time the procedure takes to build up present value. We refer to the time required for accumulation of present value as the *time horizon* of a procedure and introduce two distinct definitions. The line segments (a,b) and (a,c) in Figure 2 show how many years it takes the exponential and quasi-hyperbolic procedures to accumulate half of their present values of 50. The times are, respectively, 35 and 58 years. We define 'median time to accumulation of present value' for a procedure to be the time required for it to accumulate 50% of its present value. We label this $\tau(D)$, which is given

[3] A discounting *procedure* can be defined in terms of its associated discount function, $d_D(t)$, its present value function, $pv_D(t)$, its discount rate function, $r_D(t)$, its yield curve or an associated probability density function. Appendix 1 states the relations among alternative definitions.

Figure 2:
The Time Horizon and Relative Speed of Discounting Procedures



Note: This figure displays pv(t) for the same exponential and quasi-hyperbolic discounting procedures shown in Figure 1.

by the following expression:

$$\tau(D) = t^* \text{ such that } \frac{\int_0^{t^*} d_D(t)\ dt}{pv(D)} = 0.5. \tag{12}$$

A plausible second definition of the speed of a procedure would relate how long an exponential (with the same α as the procedure) takes to reach half its present value to how long the procedure does. (In this example the speed would be 35/58.) A variant of this approach proves more tractable.

Just as there is a median time to accumulation of present value so, too, can one define a mean time, which can be thought of as how far from time zero, on average, the present value is being accumulated from.[4] We label the mean time to accumulation as $\theta(D)$, given by:

$$\theta(D) = \frac{\int_0^{\infty} t d_D(t)\ dt}{pv(D)}. \tag{13}$$

Our second definition of the speed of a discounting procedure relates the time horizon of the procedure, $\theta(D)$, to the time horizon of an exponential procedure that discounts the future by the same amount as D , i.e. by $\alpha(D)$. The time horizon for an exponential with rate r is simply r^{-1}, so the time horizon for the exponential that has the same amount of discounting as D will be $\alpha(D)^{-1}$. Our time horizon based definition of speed, $p_2(D)$, is the ratio of the horizon for the equivalent exponential to the time horizon for the procedure being considered:

$$\rho_2(D) = \alpha(D)^{-1} \ / \ \theta(D). \tag{14}$$

This reflects the intuition that if D has a longer time horizon than the exponential with the same present value it can be viewed as slower.

Before proceeding further it is important to deal briefly with questions of convergence. Initially proposed variants of hyperbolic discounting took the form:[5]

$$d_h(t) = 1/(1 + at). \tag{15}$$

[4] A physical analogy may be helpful. Think of d(t) as distributing mass along $[0, \infty)$. A given total mass can be spread far along the axis or concentrated toward the beginning. θ is the point such that if all the mass could be concentrated at that point it would exactly balance d(t).

[5] See, for examples, Harvey's (1986, 1994) early important work on nonconstant rate discounting and work in the psychology literature of Herrnstein (1981). Loewenstein and Prelec (1992) discuss and provide an axiomatic foundation for a 'generalized hyperbolic' that places an exponent, k, on the denominator of equation (15). Many parameterizations are possible [e.g. equation (3)] and, in this paper, we refer to the generalized hyperbolic family simply as hyperbolic. The hyperbolic will have a finite present value if and only if $k > 1$.

Like exponential discounting with a zero discount rate, $d_h(t)$ has infinite present value. To put this slightly differently, an improved outcome in each time period by a finite amount x, however large, up to time t^*, also arbitrarily large, would be more than counterbalanced in present value terms, using the discount function $d_h(t)$, by a decrement y, however small, to all outcomes after t^*. It is precisely this property of making outcome changes over any finite time horizon irrelevant compared to tiny but sustained changes in the extremely distant future – even without infinite time horizons – that leads to objections to exponential discounting with a rate of 0.

The point of this example is simply to illustrate the importance of paying attention to the issue of convergence when selecting a discounting procedure for evaluation of long time horizon investments.[6] Even with a finite but long time horizon present values can be quite sensitive to the discount rate in outer years – hence the importance of explicit consideration of the total amount, $\alpha(D)$, by which a procedure discounts the future.

Implications for $r(t)$ – the discount rate function – are that it must go to zero very slowly, or not at all, in order to ensure convergence.[7] If

$$\lim_{t\to\infty} r(t) = r^*, \ \ r^* > 0, \tag{16}$$

then in the limit $d(t+1) \leq \frac{d(t)}{1+r^*}$, which guarantees convergence. Analogous results hold for continuous time. All exponential discount functions, then, or ones that are ultimately exponential – in the sense that equation (16) holds – will yield finite present values. On the other hand, if for all t greater than some t^*, $r(t) = 0$ [but $d(t) > 0$ still] then the present value will be infinite (as is the case for the discounting procedure proposed in Weitzman, 2001, Table 2).

Having a convergent discounting procedure, however, does not necessarily entail convergence of the integrals defining our concepts of the speed of discounting [equation (10) and equation (13)]. Equation (13) illustrates the question of convergence more clearly since its integrand, $td(t)$, is the product of a function going to zero and a function going to ∞. Stronger conditions on $d(t)$

[6] The issue of convergence assumes less importance in modelling phenomena over shorter periods as in the psychological literature or some areas of science and engineering. Sokolnikoff and Redheffer (1958, p. 176) observe, for example, that "... divergent Fourier series often arise in practice, for example in the theory of Brownian motion, in problems of filtering and noise, or in analyzing the ground return to a radar system. Even when divergent the Fourier series represents the main features of $f(x)$ ...".

[7] To take an example: if $r(t) = r_0 e^{-kt}$ then r will decline from an initial level of r_o to 0 too rapidly for the present value of the procedure to be finite.

must, obviously, obtain for this integral to converge than for the integral of $d(t)$ to converge.

We define a discounting procedure to be *strongly convergent* if the integrals both of its related discount function, $d(t)$, and of $td(t)$ converge. We define a procedure to be *weakly convergent* if $d(t)$ converges but $td(t)$ fails to converge. Weakly convergent procedures have a speed of 0 and the potentially desirable property of giving infinite weight to the future – in the sense that the average time to accumulation of present value is infinite – while still having a finite present value. Weakly convergent $d(t)$s can thus provide an alternative to zero rate discounting for those who – like Ramsey (1928) – hold a preference for giving genuine weight to outcomes in the extremely distant future. Among the discounting procedures discussed in this paper only one, the hyperbolic, includes weakly convergent (or zero speed) procedures within the family. In section 3.2 we pay particular attention to the one-parameter procedure that we label the zero-speed hyperbolic (ZSH).

1.2 Main results concerning the speed of discounting

Concepts introduced in the preceding subsection are not independent of one another, and our initial proposition identifies and proves the central relation between our two definitions of speed.

Proposition 1: For all D, $\rho_1(D) = \rho_2(D)$.

Proof: We have already seen that the area $\mathbf{A}(D)$ is α^{-2}, so starting from the definition: $\rho_1(D) = \left[\alpha^2 \int_0^\infty \left(\alpha^{-1} - pv_D(t)\right) dt\right]^{-1} = \left[\alpha^2 \int_0^\infty \left(\int_0^\infty d(x)dx - \int_0^t d(x)dx\right) dt\right]^{-1} = \left[\alpha^2 \int_0^\infty \int_t^\infty d(x)dxdt\right]^{-1}$. Changing the order of integration, this is $\left[\alpha^2 \int_0^\infty \int_0^x d(x)dtdx\right]^{-1} = \alpha^{-1}\left[\alpha \int_0^\infty d(x)\left(\int_0^x dt\right) dx\right]^{-1} =$ $\alpha^{-1}\left[\alpha \int_0^\infty xd(x)dx\right]^{-1} = \alpha^{-1}\left[\theta(D)\right]^{-1} = \rho_2(D)$.

Because of the equivalence of ρ_1 and ρ_2, from this point on we unambiguously use ρ to denote speed, and the following corollary follows immediately from Proposition 1 and equation (14):

Corollary: $\alpha(D)\rho(D)\theta(D) = 1$.

We now turn to the range of values possible for the speed of discounting:

Proposition 2: For all D, $0 \leq \rho(D) \leq 2$.

Proof: Fixing α, the mean time θ is minimized by concentrating as much weight as possible as early as possible (i.e. onto small values of t in the integrand). Given that $d(t)$ must start at 1 and be weakly decreasing, this is accomplished by setting $d(t) = 1$ for $t \leq 1/\alpha$ and $d(t) = 0$ for $t > 1/\alpha$. Such a discounting procedure has a $\theta = 1/(2\alpha)$, while the equivalent exponential has

$\theta = 1/\alpha$. Hence the maximum possible ρ (corresponding to the minimum θ) is 2. Meanwhile, the maximum θ is infinite (even with a finite present value and thus $\alpha > 0$), as occurs for any weakly convergent procedure. This proves that the minimum ρ is 0.

Finally, we wish to relate the speed of a discounting procedure to its discount rate function, $r(t)$. In particular, intuition suggests that a decreasing discount rate yields a procedure that is in some sense slow. It is obvious that such a procedure is 'slower' in the sense of having a longer time horizon, θ, than the constant-rate procedure that starts at the same discount rate and stays there, but on the other hand these two will not have the same total present value. Our third proposition states that, even relative to an exponential *with the same amount of discounting*, any decreasing-rate procedure is slow.

Proposition 3: If the discount rate $r(t)$ is weakly decreasing (resp. increasing), then the corresponding discounting procedure is slow (resp. fast), i.e. $\rho \leq 1$ (resp. $\rho \geq 1$). Furthermore, this result is tight in the sense that if $r(t)$ is weakly decreasing everywhere and strictly decreasing somewhere, then the inequality is strict.

Proof: Let $d(t)$ be the discount function with decreasing rate $r(t)$, and let $\mu = pv(\infty) = \int_0^\infty d(t)dt$. We can think of d as the survivor function for a failure density[8], in which case $r(t)$ decreasing is exactly the definition of a decreasing failure rate (and μ is the mean time to failure). Then for any strictly increasing function f on $[0, \infty)$, $\int_0^\infty f(t)d(t)dt \geq \int_0^\infty f(t)e^{-t/\mu}dt$ by Theorem 4.8 (p. 32) of Barlow and Proschan (1965), with equality only if $d(t) = e^{-t/\mu}$ identically. So let $f(t) = t$: $\int_0^\infty td(t)dt \geq \int_0^\infty te^{-t/\mu}dt = \mu^2 = \left[\int_0^\infty d(t)dt\right]^2$, implying that

$$\rho = \frac{\left[\int_0^\infty d(t)dt\right]^2}{\int_0^\infty td(t)dt} \leq 1,$$

as desired. Equality implies that d must be exponential with amount $\alpha = 1/\mu$. Likewise, $r(t)$ increasing corresponds to an increasing failure rate, and all inequalities are reversed.

We can alternately interpret this conclusion as saying that the exponential is the fastest

[8] Failure analysis arises in the study of systems reliability. The primitive in these analyses is the probability of no failures before time t; this is the role played by $d(t)$ in our setting. In reliability studies this is referred to as the survivor or reliability function, and it is the inverse *cdf* of the failure density (i.e. the probability of failure at any given time). The failure rate as a function of time is $-d'/d$, which is thus exactly our discount rate $r(t)$.

discounting procedure within the family of those with weakly decreasing discount rates. As we have seen, the exponential discounting procedure $d(t) = e^{-rt}$ also has a total present value equal to its mean time horizon: $pv_d(\infty) = \theta = 1/r$. Proposition 3 allows us to answer the question of whether or not it is the unique discounting procedure with this property.

Corollary: If $d(t)$ is a discount function with monotone discount rate $r(t)$ and satisfying $pv_d(\infty) = \theta(D)$, then d is exponential: $d(t) = e^{-t/\theta}$.

Proof: From the definitions, $pv_d(\infty) = \theta$ means $\int_0^\infty d(t)dt = \frac{\int_0^\infty td(t)dt}{\int_0^\infty d(t)dt}$, i.e. $\left[\int_0^\infty d(t)dt\right]^2 = \int_0^\infty td(t)dt$. But this again corresponds to equality in the proof of Proposition 3, so d must identically equal the equivalent exponential.

2 Social Discounting

Suppose that we start with a population of individuals each of whom uses some discounting procedure. The individual procedures may differ both in their parameter values and in their actual functional forms. We wish to aggregate these procedures to achieve a social discounting procedure that appropriately reflects the preferences of all members of society. In this section we first discuss existing approaches to generating a social discounting procedure from individual ones and propose an alternative procedure that avoids the shortcomings of those in the literature. We then show our preferred aggregation procedure to generate discount rates that are equilibrium rates in a reasonably defined intertemporal exchange economy. We next prove results that relate the amount, speed and time horizon of the social discount function to the amounts, speeds and time horizons of the individual ones. We conclude by providing specific examples for when all the individual discounting procedures are exponential.

2.1 Alternatives for generating a social discounting procedure from individual ones

One obvious aggregation option is to average discount *rate* functions across individuals (ADR). For example, if person A uses a standard constant-rate procedure with $d(t) = e^{-rt}$ (i.e. a rate of r), and person B uses $d(t) = e^{-st}$, then this method would yield a social discounting procedure characterized by a constant rate equal to $\frac{1}{2}(r + s)$. One advantage of ADR is that even if only one of the individuals uses nonconvergent discounting (e.g. if $s = 0$), the aggregate social procedure will be convergent (a notion of robustness). As Weitzman (2001) has emphasized, averaging rates has a down side: when society decides how to trade off between two given time points in the future, it counts everyone's opinion on that question equally, even those who do not care much about the future. Thus while a procedure that averages rates is able to successfully aggregate *amounts* of discounting (in the sense that the amount of the aggregate procedure is the average of the amounts of the individual procedures), it does less well on the shape over time that discounting should take.

Another natural option is simply to average the discount functions of individuals (ADF). In the example with two exponentially discounting individuals, this would lead to a social discount function of $\frac{1}{2}(e^{-rt} + e^{-st})$. This aggregate procedure has a discount rate that starts at $\frac{1}{2}(r + s)$ for $t = 0$, and then declines over time to the minimum of r and s. To redress the inadequacy of aggregation by averaging rates, Weitzman (2001) advocates using ADF and interprets this process as valuing a dollar at time t according to any particular discounting procedure weighted by the "probability of correctness" of that procedure. In our example, this probability is 0.5 for each of the two individual procedures in the domain and is 0 for all others.

One immediate concern about the ADF process is that if even one individual's discounting procedure is nonconvergent, then the social procedure will also be, no matter how large the society (assuming it is finite). Indeed, even if no individual in the population has an infinite present value (nonconvergent) procedure it turns out to be quite possible for the aggregate procedure generated by ADF to have infinite present value. This shortcoming alone makes ADF unviable as a general aggregation procedure.

The convergence problem could be avoided by arbitrarily ruling out nonconvergent procedures in the first place, but it speaks to a second important concern. Basically, under ADF, individuals are weighted equally (in terms of either discount functions or discount rates) at the beginning of time, after which those who discount less are increasingly favored. ADF yields social discounting procedures whose discount rate functions are both declining over time (as patient individuals are increasingly favored) *and* that are low overall (since they start at the social average and go down). At no point in time does ADF aggregation differentially reflect the preferences of *impatient* individuals even though, as time goes on, the preferences of patient indivuals become increasingly consequential. The shape of the ADF social discount function faithfully reflects a weighted average of the individual shapes (weighted by their own sense of relative time preference), but it skews the amount of discounting toward those who discount less.

We seek an aggregating process that avoids the problems associated with the simple averaging of either discount rates or functions. That is, we feel an aggregation process should satisfy two criteria:

(i) the aggregate procedure should discount the future by an amount that is the average of the individual amounts – as averaging the discount rates does[9]; and

(ii) the aggregate procedure's discount rates in the future should place greater weight on individuals who value the future more highly – as averaging the discount functions does. (This implies, combined with criterion 1, that near term discount rates will be weighted toward those of individuals who place greater weight on the present.)[10]

Both criteria can be met by averaging the *normalized discount functions* for each individual (the normalized function is $\alpha d(t)$ and has a total present value of 1). We then divide by the value at 0 of this average function in order to un-normalize and recover a valid discount function (with $d(0) = 1$, as is necessary). We label this the average normalized discount function (ANDF) aggregation process. The ANDF process results in a shape equal to the average shape, and we will prove that it has an amount exactly equal to the average amount, $\overline{\alpha}$ (and a mean time horizon equal to the average mean time, $\overline{\theta}$). In our running example, the normalized individual discount functions are re^{-rt} and se^{-st}, respectively, so the ANDF social procedure is

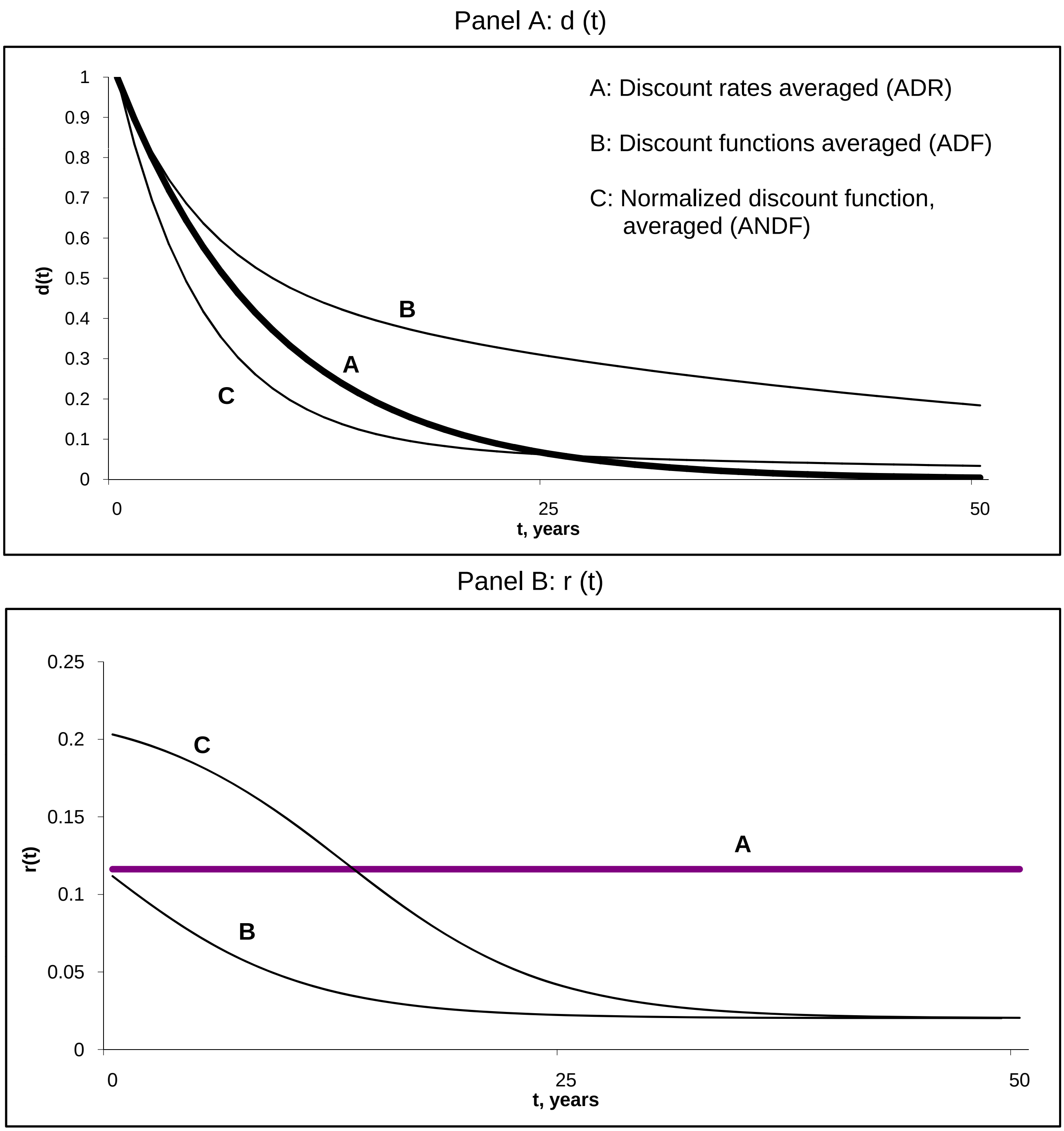
$$d_{ANDF}(t) = \frac{re^{-rt} + se^{-st}}{r+s}. \tag{17}$$

Here $\alpha = (r+s)/2$ and $\theta = (r^{-1} + s^{-1})/2$. Figure 3 illustrates the $d(t)$ and $r(t)$ resulting from each of the three ways of aggregating two exponentials, in this case with $r = 0.02$ and $s = 0.20$. For the ANDF process applied to these two exponentials, $\alpha = 0.11$, $\theta = 27.5$, and $\rho = 0.33$ (so it

[9] Note that this criterion refers to aggregation of individual procedures into a social one. It has less relevance for generating an expected value when there is underlying uncertainty in interest rates, e.g. the situation considered in Newell and Pizer (2003).

[10] Gollier and Zeckhauser (2003) develop a market-oriented aggregation procedure for exponentials that results in high discount rates (relative to the average) in early periods and low discount rates in later periods. ANDF has this feature and we conjecture that any market-derived aggregation procedure will also.

Figure 3:
Three Procedures for Aggregating Individual Discount Functions



Notes: This figure shows the functions d(t) and r(t) that result from different ways of aggregating two discount functions: $d_1(t) = e^{-.2t}$ and $d_2(t) = e^{-.02t}$.

is slow). Note in particular in Panel B of Figure 3 that the discount rate function for averaging the discount functions (ADF) is below the average discount rate for all $t > 0$ whereas ANDF falls from above the average to 0.02.

One interpretation of the ANDF is that it gives each individual a total weight of 1 to spread across the future in a way that reflects his or her preferred time path of consumption. ANDF averages those shapes and then un-normalizes, which retrieves the average amount of discounting. This separation of amount and shape ensures that there is no problem with convergence and that more patient members of society are favored at later times relative to earlier times, but not overall. Additionally, and importantly, it turns out that in a canonical intertemporal exchange economy, the market equilibrium discount rates are precisely those generated by ANDF.

2.2 Market discount rates in an intertemporal exchange economy

Consider a deterministic economy with a single non-storable good and multiple consumers with varying discount functions. Each agent receives an endowment of one unit of the good in each period.[11] There is a market interest rate r^t between periods $t-1$ and t, so that each agent can write binding contracts at time 0 regarding how much they wish to borrow or lend in each period. We will focus on a finite number of agents who take the market rate as given, but this can be formalized by replacing each agent with a unit-mass of infinitesimal agents with the same preferences. Agents have logarithmic felicity functions, so that agent i maximizes

$$\sum_{t=0}^{T} d_i^t \ln\left(1 + b_i^t - (1+r^t)b_i^{t-1}\right),$$

where d_i^t is the value of i's discount function at time t, and b_i^t (which may be negative) is the amount that i borrows in period t. We force $b_i^{-1} = b_i^T = 0$, though in general one could also consider $T = \infty$. The market interest rate function $r^1, r^2, ..., r^T$ is defined as the unique function that makes markets clear: for all t, $\sum_i b_i^t = 0$.

We can also apply the definition of ANDF in this setting. In particular, each agent has an

[11] We could add a constant growth rate g to all variables, but it wouldn't change the underlying interactions. Note also that we have switched to discrete time in order to simplify the exposition and allow comparison to the existing literature.

amount of discounting α_i given by $\left(\sum_{t=0}^{T} d_i^t\right)^{-1}$, so the social discount function in period t is the weighted average across all agents, scaled by the total weights:

$$d_{ANDF}^t = \frac{\sum_i \alpha_i d_i^t}{\sum_i \alpha_i}.$$

Our next proposition shows that in the simplest economy which allows for nonconstant individual discount rates, the market interest rate is given by the ANDF aggregation procedure.

Proposition 4: In a market as described above, with two agents and three periods (i.e. $T = 2$), the market interest rate is identical to the social interest rate according to ANDF aggregation. Namely, $r^t = \left(d_{ANDF}^{t-1}/d_{ANDF}^t\right) - 1$.

Hence if all agents use exponential discounting, but with heterogeneity in rates, the market interest rate will change over time. Specifically, it will decrease over time (corresponding to slow social discounting) starting from a rate corresponding to the least patient member of society and approaching the rate corresponding to the most patient member of the economy. If agents in the economy have fast individual discount functions (as is suggested by much of the psychology literature), it is possible that the market rate will in fact be constant over time.

Proof: We assume that agent 1 is a net lender in both periods (for instance we need that $d_1^1 > d_2^1$)[12], and define $b^t = b_2^t = -b_1^t$ to be the amount lent, for $t = 0, 1$. Thus agent 1 solves the following optimization problem:

$$\max_{b^0, b^1} \left\langle \ln(1 - b^0) + d_1^1 \ln(1 - b^1 + (1 + r^1)b^0) + d_1^2 \ln(1 + (1 + r^2)b^1) \right\rangle.$$

Then the first-order conditions are

$$\frac{1}{1 - b^0} = \frac{d_1^1(1 + r^1)}{1 - b^1 + (1 + r^1)b^0}$$

and

$$\frac{d_1^1}{1 - b^1 + (1 + r^1)b^0} = \frac{d_1^2(1 + r^2)}{1 + (1 + r^2)b^1}.$$

These can be solved for agent 1's optimal choice of b^0 and b^1 in terms of r^1 and r^2. Agent 2's optimization problem yields analogous first-order conditions:

$$\frac{1}{1 + b^0} = \frac{d_2^1(1 + r^1)}{1 + b^1 - (1 + r^1)b^0}$$

[12] The other cases are identical, mutatis mutandis.

and

$$\frac{d_2^1}{1+b^1-(1+r^1)b^0} = \frac{d_2^2(1+r^2)}{1-(1+r^2)b^1}.$$

These can also be solved for b^0 and b^1 in terms of r^1 and r^2, but of course the two solutions must be the same (market clearing). This gives us two equations in two unknowns, and we compute

$$r^1 = \frac{1 - d_1^1 d_2^1 + (d_1^2 + d_2^2)/2 - (d_1^1 d_2^2 + d_1^2 d_2^1)/2}{d_1^1 d_2^1 + (d_1^1 + d_2^1)/2 + (d_1^1 d_2^2 + d_1^2 d_2^1)/2}$$

and

$$r^2 = \frac{d_1^1 d_2^1 + (d_1^1 + d_2^1)/2 - d_1^2 d_2^2 - (d_1^2 + d_2^2)/2}{d_1^2 d_2^2 + (d_1^2 + d_2^2)/2 + (d_1^1 d_2^2 + d_1^2 d_2^1)/2}.$$

Turning now to the ANDF procedure, we see that $\alpha_1 = (1+d_1^1+d_1^2)^{-1}$ and $\alpha_2 = (1+d_2^1+d_2^2)^{-1}$. Thus the average of $\alpha_1 d_1^1$ and $\alpha_2 d_2^1$ is $\left(d_1^1 d_2^1 + (d_1^1 + d_2^1)/2 + (d_1^1 d_2^2 + d_1^2 d_2^1)/2\right)(1 + d_1^1 + d_1^2)^{-1}(1 + d_2^1 + d_2^2)^{-1}$. To find d^1_{ANDF}, we need to normalize this value by dividing by $(\alpha_1 + \alpha_2)/2$, in order to recover a value of 1 at time 0. Hence

$$d^1_{ANDF} = \frac{d_1^1 d_2^1 + (d_1^1 + d_2^1)/2 + (d_1^1 d_2^2 + d_1^2 d_2^1)/2}{1 + (d_1^1 + d_2^1)/2 + (d_1^2 + d_2^2)/2}$$

and so

$$r^1_{ANDF} = 1/d^1_{ANDF} - 1 = \frac{1 - d_1^1 d_2^1 + (d_1^2 + d_2^2)/2 - (d_1^1 d_2^2 + d_1^2 d_2^1)/2}{d_1^1 d_2^1 + (d_1^1 + d_2^1)/2 + (d_1^1 d_2^2 + d_1^2 d_2^1)/2}.$$

This is of course exactly we found above for the market interest rate. We can similarly average $\alpha_1 d_1^2$ and $\alpha_2 d_2^2$, normalize by the same factor as above to get d^2_{ANDF}, and ultimately compute

$$r^2_{ANDF} = d^1_{ANDF} \ / \ d^2_{ANDF} - 1 = \frac{d_1^1 d_2^1 + (d_1^1 + d_2^1)/2 - d_1^2 d_2^2 - (d_1^2 + d_2^2)/2}{d_1^2 d_2^2 + (d_1^2 + d_2^2)/2 + (d_1^1 d_2^2 + d_1^2 d_2^1)/2}.$$

This is likewise the same as the corresponding market interest rate, which is precisely what we wished to show.

It is important to note, however, that the preceding proposition depends on the choice of logarithmic utility and, in particular, does not hold for a more general CRRA utility function.

While it may be the case that no general aggregation algorithm will always generate the market's solutions, the point of the example is to show that ANDF plausibly captures the qualitative features of market solutions in a way that ADR and ADF cannot.

This section and the preceding one have motivated use of ANDF as a procedure for aggregating individual discounting procedures. The following section develops specific characteristics of ANDF aggregation.

2.3 The amount and time horizon of social discounting

Formally, assume that we have a collection of individuals with discounting procedures parameterized by $x \in X$ (possibly multivariate), with frequency distribution across parameters of $f(x)$, so that in particular $\int_X f(x)dx = 1$. Then if individual x uses a discount function $d(t;x)$ with associated amount $\alpha(x)$, we define the ANDF aggregate procedure $\overline{D}$ by its discount function as follows:[13]

$$d_{\overline{D}}(t) = \frac{\int_X \alpha(x)d(t;x)f(x)dx}{\int_X \alpha(x)f(x)dx}. \tag{18}$$

Because $\int_X \alpha(x)d(0;x)f(x)dx = \int_X \alpha(x)(1)f(x)dx = \int_X \alpha(x)f(x)dx$, $d_{\overline{D}}(0) = 1$ as required. The definition in equation (18) leads to a convergent aggregate procedure if there is a nonzero proportion of the population with convergent procedures.

To each individual discount function $d(t;x)$ there corresponds a discount rate function $r(t;x)$ satisfying $\dot{d}(t;x) = -r(t;x)d(t;x)$, where the superscript dot denotes a time derivative. For the aggregate procedure $\overline{D}$, since the $d(t;x)$ in the numerator is the only term involving time,

$$\dot{d}_{\overline{D}}(t) = \frac{-\int_X \alpha(x)r(t;x)d(t;x)f(x)dx}{\int_X \alpha(x)f(x)dx}, \tag{19}$$

and therefore

[13] We require only that $\int_X \alpha(x)f(x)dx > 0$, i.e. that at least some nonzero fraction of the population uses convergent discounting procedures. If $\int_X \alpha(x)f(x)dx = 0$, then there is no need to normalize (essentially all individuals already have the same α, namely $\alpha = 0$), so we define $d_{\overline{D}}(t)$ as simply $\int_X d(t;x)f(x)dx$.

$$(20) \qquad r_{\overline{D}}(t) = \frac{-\dot{d}_{\overline{D}}(t)}{d_{\overline{D}}(t)} = \frac{\int_X \alpha(x) r(t;x) d(t;x) f(x) dx}{\int_X \alpha(x) d(t;x) f(x) dx}.$$

We now prove three core results that relate the characteristics of the ANDF aggregation procedure to the corresponding characteristics of the individual procedures that were aggregated. Let $Y = \{x \in X |\, \alpha(x) = 0\}$, i.e. Y is the set of parameters that correspond to nonconvergent discounting procedures.

Proposition 5: If $\int_Y f(x)dx = 0$, then, for the ANDF process defined above, the amount $\alpha(\overline{D})$ and mean time horizon $\theta(\overline{D})$ of the aggregate procedure $\overline{D}$ are the average amount $\overline{\alpha}$ and the average mean time $\overline{\theta}$ respectively.[14]

Proof: Since $\alpha(x) = 0$ for all $x \in Y$, anytime the integrand involves $\alpha(x)$ we can switch the domain of integration between X and $X \backslash Y = X - Y$ as we wish. We first verify that $\int_0^\infty \int_X \alpha(x) d(t;x) f(x) dx dt = \int_0^\infty \int_{X \backslash Y} \alpha(x) d(t;x) f(x) dx dt = \int_{X \backslash Y} \alpha(x) f(x) \left[\int_0^\infty d(t;x) dt\right] dx = \int_{X \backslash Y} \alpha(x) f(x) \frac{1}{\alpha(x)} dx$ ($\alpha(x) > 0$ on $X \backslash Y$ so $\frac{1}{\alpha(x)}$ is well-behaved) $= \int_{X \backslash Y} f(x) dx = \int_X f(x) dx - \int_Y f(x) dx = 1 - 0 = 1$, which makes sense since this was an average normalized function. Now note that the denominator of the aggregate function $d_{\overline{D}}$ is constant in t, and thus, using the calculation we just made, the aggregate amount is

$$\alpha(\overline{D}) = \left[\int_0^\infty d_{\overline{D}}(t) dt\right]^{-1} = \left[\frac{1}{\int_X \alpha(x) f(x) dx}\right]^{-1} = \int_X \alpha(x) f(x) dx,$$

which is precisely the formula for the average amount $\overline{\alpha}$ of discounting across the population, as desired.

[14] Thus the bar above a particular characteristic denotes expectation with respect to the density f. Exactly because of the proposition, this will not prove to be confusing terminology.

For the mean time horizon, we compute θ for the aggregate procedure as

$$\begin{aligned}
\theta(\overline{D}) &= \alpha(\overline{D})\int_0^\infty t d_{\overline{D}}(t)dt \\
&= \alpha(\overline{D})\int_0^\infty t\left[\frac{\int_X \alpha(x)d(t;x)f(x)dx}{\alpha(\overline{D})}\right]dt \\
&= \int_0^\infty \int_X t\alpha(x)d(t;x)f(x)dxdt \\
&= \int_X \left[\alpha(x)\int_0^\infty td(t;x)dt\right]f(x)dx \\
&= \int_X \theta(x)f(x)dx \\
&= \overline{\theta}.
\end{aligned}$$

In general, if $\int_Y f(x)dx > 0$, $\alpha(\overline{D})$ will simply be the average over all strictly positive α in the population: formally, $\alpha(\overline{D}) = \int_{X\setminus Y} \alpha(x)f(x)dx$ (with an analogous outcome for θ). The intuitive interpretation is that any individual who chooses $\alpha = 0$ (i.e. a nonconvergent procedure), effectively suggesting an infinite present value, ends up spreading his/her normalized weight so thinly over time that it has no effect at all on the aggregate[15]. We turn next to the speed of the aggregate procedure.

Proposition 6: If the amount $\alpha(x)$ and the mean time $\theta(x)$ are negatively covariant within a population X, then the speed resulting from the ANDF process is lower than the average speed, i.e. $\rho(\overline{D}) \leq \overline{\rho}$, with equality only if all individuals have the same amount of discounting α and all individuals have the same speed ρ.

Proof: By Propositions 1 and 5, $\rho(\overline{D}) = \theta(\overline{D}E)/\theta(\overline{D}) = \left(\overline{\alpha}\overline{\theta}\right)^{-1}$, while $\overline{\rho} = \int_X \rho(x)f(x)dx = \int_X \left(\alpha(x)\theta(x)\right)^{-1} f(x)dx = \overline{(\alpha\theta)^{-1}}$, where the bar continues to denote expectation with respect to f, and $\overline{D}E$ refers to the exponential that discounts by the same amount as $\overline{D}$. But since $(\cdot)^{-1}$ (i.e. taking inverses) is a convex function on $\mathbb{R}$, Jensen's inequality implies that the inverse of the average is weakly less than the average of the inverses, i.e. $\left(\overline{\alpha\theta}\right)^{-1} \leq \overline{(\alpha\theta)^{-1}}$. Now $cov(\alpha,\theta) = \overline{\alpha\theta} - \overline{\alpha}\overline{\theta}$ by definition, and this is negative by assumption. Hence $\overline{\alpha\theta} \leq \overline{\alpha}\overline{\theta}$, so $\rho(\overline{D}) = \left(\overline{\alpha}\overline{\theta}\right)^{-1} \leq \left(\overline{\alpha\theta}\right)^{-1} \leq \overline{(\alpha\theta)^{-1}} = \overline{\rho}$. Furthermore, since $(\cdot)^{-1}$ is in fact strictly convex, the

[15] It is possible to construct an aggregate discounting procedure whose amount is *always* equal to the average amount, inclusive of x such that $\alpha(x) = 0$. In this case the shape of the aggregate function is identical to the aggregate as defined, so the larger present value means that $pv(t)$ does not converge to $pv(\infty)$, and thus of course $\theta(\overline{D}) = \infty$ and $\rho(\overline{D}) = 0$. Details are available upon request.

inequality is strict unless both $cov(\alpha, \theta) = 0$ *and* $(\alpha\theta)^{-1}$ is constant across the population. But these cannot hold simultaneously unless α and θ are themselves constant, which is equivalent to α and ρ being constant.

In general, we do expect α and θ to have negative covariance, since placing less total present value on the future often implies that that present value is reached more quickly. This tendency is confirmed by looking at the expressions in Table 1, but it need not hold for every possible choice of distributions f over the underlying parameters. Thus we expect an aggregate procedure to be slower overall than its components, but it is possible for it to be faster under certain circumstances.[16] In the special case of constant relative speed, however, we can rule out this possibility.

Corollary: If $\rho(x)$ is constant across x, then $\rho(\overline{D}) \leq \overline{\rho}$, with equality only if $\alpha(x)$ is also constant.

Proof: From the corollary to Proposition 1, $\rho = (\alpha\theta)^{-1}$, so $\rho(x)$ constant implies $\alpha(x)\theta(x) = C$ for all x. Then we can compute $cov(\alpha, \theta) = cov(\alpha, C/\alpha) = \overline{\alpha(C/\alpha)} - \overline{\alpha}\,\overline{(C/\alpha)} = C - \overline{\alpha}C\overline{\alpha^{-1}} = C\left(1 - \overline{\alpha}\,\overline{\alpha^{-1}}\right)$. But Jensen's inequality once again implies that $(\overline{\alpha})^{-1} \leq \overline{\alpha^{-1}}$, or $1 \leq \overline{\alpha}\,\overline{\alpha^{-1}}$. Hence $cov(\alpha, \theta) \leq 0$ and Proposition 6 applies. The inequality is clearly strict unless α is constant across the population.

We turn now to the relationship between the aggregate rate $r_{\overline{D}}(t)$ and the individual discount rates $r(t; x)$. In particular, one variable of interest is the limiting discount rate $r^*(x) = \lim_{t\to\infty} r(t; x)$, if it exists, which gives the asymptotic discount rate for individual x. For any potential social limiting rate $r \geq 0$, let $A(r) = \{x \in X \text{ s.t. } \alpha(x) > 0,\ r^*(x) \text{ exists, and } r^*(x) \leq r\}$; this is the set of individuals who use convergent procedures and whose limit is no larger than r. Finally, we define $r^*_{\min} = \inf \left\langle r \,\middle|\, \int_{A(r)} f(x)dx > 0 \right\rangle$. This is the lowest rate r such that at least some nonzero fraction of the population has a limiting rate no higher than r. Dybvig, Ingersoll and Ross (1996) and Weitzman (1998) showed that if the social discount function is constructed by simple averaging of a finite number of individual discount functions, then $r^*_{\min}$ is exactly the asymptotic discount

[16] Hara and Kuzmics (2002) prove a conceptually similar result in the context of risk aversion. Specifically, they show that the representative consumer exhibits strictly decreasing relative risk aversion, ranging from that of the most risk averse individual to that of the least risk averse as the aggregate consumption level increases.

rate for the aggregate.[17] Our next proposition states that the same is true, without assuming a finite number of individuals, for the ANDF process, i.e. that in the limit the social discount rate is in some sense the smallest of any across the population:

Proposition 7: The asymptotic social discount rate for the ANDF aggregation process is given by: $\lim_{t\to\infty} r_{\overline{D}}(t) = r^*_{\min}$.

Proof: See Appendix C.

2.4 Aggregation of exponential procedures

The aggregation processes defined above can be carried out no matter what the underlying discounting procedures are. If, however, we hope for simple closed-form outcomes, we will need to make some specific assumptions. For the remainder of this section, we assume that each individual i uses a constant-rate procedure given by $d(t) = e^{-r_i t}$. This not only simplifies the analysis, but also allows us to compute empirical results from the data collected by Weitzman (2001), who surveyed 2160 economists and asked for an exponential discount rate from each.

We first observe that when aggregating exponentials, all of which have $\rho(x) = 1$, the corollary to Proposition 6 applies, so the speed of the ANDF aggregate is strictly less than 1 unless all individuals use the same constant discount rate, i.e. are identical.

We begin by considering discrete aggregation of exponentials, i.e. the case when there is a finite population. Let the number of individuals be n, with respective constant rates $r_1, ..., r_n$ (possibly with multiplicity, of course). We assume $r_i > 0$ for all i (or equivalently that there are n individuals with $r_i > 0$ and the rest can be ignored; see comment after Proposition 5). In this case the ANDF aggregate procedure results in

$$\overline{\alpha} = \frac{1}{n}\sum_{i=1}^{n} r_i, \tag{21}$$

$$\overline{\theta} = \frac{1}{n}\sum_{i=1}^{n} r_i^{-1}, \text{ and} \tag{22}$$

[17] Analogous results concerning mortality have been in the literature for some time. Just as the lowest discount rate becomes dominant in later years so, too, in a population that is an aggregate of distinct subpopulations, the cohort mortality rate in later years will approach that of the subpopulation with the lowest mortality rate. See Vaupel, Manton and Stallard (1979) for an early treatment of this subject.

(23)
$$\rho(\overline{D}) = \frac{n^2}{\left(\sum_{i=1}^n r_i\right)\left(\sum_{i=1}^n r_i^{-1}\right)}.$$

These results follow directly from Proposition 5, given that individual i's amount α_i is r_i and i's mean time horizon θ_i is r_i^{-1}. Then, using Proposition 1, $\rho(\overline{D})$ can be computed as $(\overline{\alpha}\overline{\theta})^{-1} = \left[\frac{1}{n^2}\left(\sum_{i=1}^n r_i\right)\left(\sum_{i=1}^n r_i^{-1}\right)\right]^{-1}$, as stated.

We next study continuous populations by considering a specific but very general density on the underlying distribution of discount rates, the gamma distribution, for $x > 0$:

(24)
$$f(x) = \frac{a^b}{\Gamma(b)} e^{-ax} x^{b-1},$$

with $a, b > 0$. This has mean $\mu = b/a$ and variance $\sigma^2 = b/a^2$. Such a distribution fits the Weitzman data well (although, unless $b = 1$, it has the disadvantage of putting no weight at 0, whereas almost 2.5% of the respondents chose discount rates at or below 0). Weitzman, aggregating by averaging of the $d(t)$s, showed that the gamma distribution[18] on parameters leads to:

(25)
$$d_\gamma(t) = \left[\frac{1}{1 + t/a}\right]^b.$$

Although Weitzman labels this function 'gamma discounting' the procedure is generally known as hyperbolic, and we use the standard terminology.

The ANDF aggregation process, with the same gamma distribution for individual discount rates, leads to a social discount function (see Appendix C for proof) given by

(26)
$$d_{\overline{D}}(t) = \left[\frac{1}{1 + t/a}\right]^{1+b}.$$

So it is also hyperbolic, but with an exponent that is greater by 1 than for $d_\gamma(t)$, which resolves

[18] Axtell (2003) obtains the same result as a special case of his much more general findings on ADF aggregation of exponential procedures. He obtains his results very simply with the observation that aggregation of exponentials involves taking the Laplace transform of the aggregating distribution. He also underscores the importance of "potential problems" with convergence in showing that averaging of exponential discount functions whose parameters are exponentially distributed always leads to a nonconvergent aggregate procedure.

the convergence issue. We can compute $\overline{\alpha} = b/a$ and $\rho(\overline{D}) = 1 - 1/b$. In terms of the mean μ and variance σ^2 of the density f, we get

$$\overline{\alpha} = \mu \text{ and } \rho(\overline{D}) = 1 - \frac{\sigma^2}{\mu^2} \tag{27}$$

whereas the ADF aggregate – given in the parameterization of equation (3) – has

$$\alpha_\gamma = \mu - \frac{\sigma^2}{\mu} \text{ and } \rho_\gamma = 1 - \frac{\sigma^2}{\mu^2 - \sigma^2}. \tag{28}$$

These values imply, among other things, that even after ruling out zero discount rates, the aggregate function $d_\gamma(t)$ fails to converge when $\sigma \geq \mu$, as it almost is in Weitzman's sample (where σ is roughly 3% and μ is roughly 4%). If the ANDF aggregation process (which is always convergent) is used when $\sigma \geq \mu$, then the aggregate discounting procedure $\overline{D}$ is convergent but only weakly so.

3 Specific Discounting Procedures

This section discusses a number of discounting procedures, some generalized from the literature and some new. In Appendix B we state and prove basic properties of these and other procedures. Subsection 3.1 summarizes the properties of these procedures. Subsection 3.2 then examines in more detail a specific procedure, the zero-speed hyperbolic (ZSH), which we argue to be particularly well-suited for discounting in social decision-making.

3.1 The amount and speed of selected discounting procedures

This section states the properties of a number of familiar discounting procedures as well as characterizing and stating the properties of several that are novel. Three relatively general procedures require 2 parameters and are characterized in Table 1. The 1st column of Table 1 provides a parameterization of the hyperbolic discount function (HYP) that allows amount and speed to be expressed directly in terms of its parameters r and s: $\alpha(HYP) = r$ and $\rho(HYP) = s$. The first row in the table gives the discount function, and the second row gives the discount rate function. Mean time to acquisition of present value for the hyperbolic is $(rs)^{-1}$. Note that the hyperbolic procedure is always slow and that its speed can take on the value 0.

Table 1: Properties of Hyperbolic, Gamma and Weibull Discounting Procedures

Property	Hyperbolic	Gamma	Weibull
1. discount function, $d(t)$	$[1+r(1-s)t]^{-(1+\frac{1}{1-s})}$	$\frac{\Gamma(s+1,(s+1)rt)}{\Gamma(s+1)}$	$\exp(-rt^{1/s})$
	$(r>0;\ s<1)$	$(r>0;\ s>-1)$	$(r,s>0)$
2. discount rate, $r(t)$	$r\frac{2-s}{1+r(1-s)t}$	$\frac{[(s+1)r]^{s+1}t^s e^{-(s+1)rt}}{\Gamma(s+1,(s+1)rt)}$	$\frac{r}{s}t^{(1-s)/s}$
$r(0)$	$r(2-s)$	$\begin{cases} \infty \text{ if } s<0 \\ r \text{ if } s=0 \\ 0 \text{ if } s>0 \end{cases}$	$\begin{cases} 0 \text{ if } s<1 \\ r \text{ if } s=1 \\ \infty \text{ if } s>1 \end{cases}$
$r(\infty)$	$\begin{cases} 0 \text{ if } s<1 \\ r \text{ if } s=1 \\ \infty \text{ if } s>1 \end{cases}$	$r(s+1)$	$\begin{cases} \infty \text{ if } s<1 \\ r \text{ if } s=1 \\ 0 \text{ if } s>1 \end{cases}$
3. amount, α	r	r	$r^s/\Gamma(s+1)$
4. speed, ρ	s	$2-\frac{2}{s+2}$	$\Gamma(s)\Gamma(s+1)/\Gamma(2s)$
	$(\rho<1)$	$(0<\rho<2)$	$(0<\rho<2)$
5. median time, τ	$r^{-1}(2^{1-s}-1)/(1-s)$	*no closed form*	*no closed form*
6. mean time, θ	$(rs)^{-1}$	$\frac{s+2}{2(s+1)r}$	$r^{-s}\Gamma(2s)/\Gamma(s)$

Each probability density function, $f(t)$, defined on $[0, \infty)$ forms the basis for a discounting function through the formula:

$$d(t) = \int_t^\infty f(x)dx, \tag{29}$$

as discussed in Appendix A. One such pdf, the gamma, forms the basis for procedures that result in a strictly positive asymptotic value of r. Starting with a gamma, the result is for $d(t)$ is

$$d(t) = \frac{\Gamma(s+1, (s+1)rt)}{\Gamma(s+1)} \text{ with } s > -1 \text{ and } r > 0, \tag{30}$$

a somewhat complex formulation involving the incomplete gamma function, $\Gamma(\cdot, \cdot)$. See derivations in Appendix B, section 3. The virtues of this procedure are that it allows both fast and slow speeds and that amount and speed are simply given: $\alpha = r$ and $\rho = 2 - \frac{2}{s+2}$. For the slow or declining discount rate gamma the limiting value of $r(t)$ is $r(s+1)$.

Finally, we wish to include procedures that can be either fast or slow but that can have a limiting discount rate of 0 as time goes to infinity. A procedure that falls into this class was introduced by Read (2001):

$$d(t) = e^{-rt^{1/s}}. \tag{31}$$

This discount function's associated pdf is Weibull and we denote this as Weibull discounting. Here the parameter s either expands or contracts time relative to constant rate exponential discounting.[19] Characteristics of the Weibull discounting procedure appear in Table 1; for derivations see Appendix B, section 2. If $s = 2$ (which corresponds to contracted time), then $\alpha = r^2/2$, $\rho = 1/3 < 1$, and $\theta = 6/r^2$; this is an example of slow discounting (and more generally $\rho < 1$ exactly when $s > 1$). We label this the 'slow Weibull' discounting procedure. Having $s = 1/2$ entails squaring t in the formula for $d(t)$ effectively expanding time. We label this the 'fast Weibull'. For the fast Weibull, which appeared as an example in equation 5, $\alpha = \sqrt{4r/\pi}$, $\rho = \pi/2$, and $\theta = 1/\sqrt{r\pi}$.

[19] One can, in principle, transform time with a broad range of functions $g(t)$ to get $d(t) = e^{-rg(t)}$. For example, Roelofsma (1996) discusses Weber's Law from psychology, for which $g(t) = \ln t$. In this case, interestingly, the resulting discounting procedure can be shown to be a member of the hyperbolic family.

Not surprisingly the procedures listed in Table 1 reduce to exponentials under certain parameter values, namely the Weibull for $s = 1$ and the gamma for $s = 0$. Although we do not allow $s = 1$ explicitly in the hyperbolic procedure, if we take the limit as s approaches 1 (from below) we get an exponential with discount rate r, as expected. This can be shown directly, but it also follows from the discount rate function for the hyperbolic.

We reiterate that $s > 0$ is not required for the hyperbolic procedure. In fact, $s \leq 0$ is perfectly legitimate; this corresponds to weak convergence in our terminology. Technically, the speed as a ratio of areas will be identically 0 for any weakly convergent procedure, but it is useful to allow $s < 0$ in the definition of a hyperbolic without referring to this as a speed. That said, the case where $s = 0$ is of particular interest in that it anchors one end of a spectrum of hyperbolic procedures that is anchored at the other by the exponential $(\rho = 1)$. And, as with the exponential, the zero-speed hyperbolic is a single parameter procedure: $d(t) = (1 + rt)^{-2}$.

The preceding paragraphs have pointed to a number of special cases of the general discounting procedures characterized in Table 1. Each case results from fixing the speed of a general procedure so that the remaining single parameter in the procedure is the amount of discounting or a simple function of that amount. Table 2 presents and characterizes 5 single parameter procedures with speeds (row 4) ranging from $\rho = 0$ (zero-speed hyperbolic or ZSH) to $\rho = \pi/2$ (fast Weibull). In between are the slow Weibull ($\rho = 1/3$), the exponential ($\rho = 1$), and the gamma with $s = 1$ ($\rho = 4/3$). Among them these procedures present a menu of simple and analytically tractable procedures that span almost the full range of speeds. The fast procedures represent single parameter alternatives to the 3-parameter quasi-hyperbolics (Appendix B, Sections 6.4 and 6.5) for use in psychology and behavioral and neuroeconomics.[20] The slow procedures provide approaches to discounting for long time horizon social investments, and the next subsection further discusses the ZSH in that context.

[20] The basic observation in this literature is that individuals appear to discount very little over a 'present' period that extends a short but usually unspecified time into the future. After that the discount function drops sharply and then resumes a slow decline. The fast Weibull with speed $\pi/2$ provides a natural single parameter alternative to the 3-parameter hyperbolic. (Note that even the discrete time quasi-hyperbolic has 3 parameters – the parameters b and r in equation 4 plus the specification of the units in which time is measured.)

Table 2: Properties of Five Single Parameter Discounting Procedures

Property	Zero-speed hyperbolic ($s = 0$)	Slow Weibull ($s = 2$)	Exponential	Gamma ($s = 1$)	Fast Weibull ($s = \frac{1}{2}$)
1. discount function, $d(t)$	$(1 + rt)^{-2}$	$\exp(-rt^{1/2})$	e^{-rt}	$e^{-2rt}(1 + 2rt)$	$\exp(-rt^2)$
2. discount rate, $r(t)$	$\frac{2r}{1+rt}$	$\frac{r}{2\sqrt{t}}$	r	$\frac{4r^2t}{1+2rt}$	$2rt$
$r(0)$	$2r$	∞	r	0	0
$r(\infty)$	0	0	r	$2r$	∞
3. amount, α	r	$r^2/2$	r	r	$\sqrt{4r/\pi}$
4. speed, $\boldsymbol{\rho}$	**0**	**1/3**	**1**	**4/3**	**$\boldsymbol{\pi/2}$**
5. median time, τ	r^{-1}	–	$(\ln 2)/r$	$2r\tau = \ln(2 + 2r\tau)$	–
6. mean time, θ	∞	$6/r^2$	r^{-1}	$3/4r$	$1/\sqrt{r\pi}$

3.2 Zero-speed hyperbolic (ZSH) discounting for social choice

The preceding subsection discussed the three very general discounting procedures presented in Table 1 and provided 1-parameter special cases spanning a range of speeds in Table 2. This offers a broad menu for choice of a social discounting procedure. In selecting from this menu we would look for several things:

(i) Because the ANDF aggregation procedure yields a social procedure that is slower than the average of the speeds of the individual procedures (Proposition 6) it would in general be desirable for a social procedure to be slow ($\rho < 1$). While it is possible that the individual procedures being aggregated would be fast it is more likely that they would be exponential and in any case aggregation will result in a slowing.[21]

(ii) It would be desirable for the social discounting procedure to be a single parameter function and for that parameter to be simply related to the amount of discounting.

Table 2 presents two candidates that meet these criteria – the zero-speed hyperbolic (ZSH) with $\rho = 0$ and the slow Weibull with $\rho = 1/3$. The ZSH is a slightly simpler expression than the slow Weibull and its single parameter r is equal to the amount of discounting whereas with the slow Weibull the amount of discounting is $r^2/2$. For these reasons we propose the ZSH as a simple yet flexible procedure for social discounting.

We designate the discount and discount rate functions for the ZSH by $d_z(t)$ and $r_z(t)$, and if the amount of discounting is r and the speed 0 we have:

(32) $$d_z(t) = (1 + rt)^{-2} \text{ and}$$

(33) $$r_z(t) = 2r/(1 + rt).$$

Thus $r_z(t)$ declines from a value of $2r$ at 0 to a value of 0 at ∞, and equals r when $t = r^{-1}$, i.e. at the time by which half of the present value of the ZSH has been accumulated. (Half the present value of an equivalent exponential will have been accumulated earlier by a factor of 0.69.) Although the median time for accumulation of present value for the ZSH is r^{-1}, the mean time is infinite so the procedure puts substantial weight on the far future.

[21] For those who might prefer ADF it too will generate a slower aggregate procedure that the average of the individual procedures.

Newell and Pizer (2002) used historical volatility of discount rates in bond markets to generate an expected value for a potential social discounting function and Oxera (2002), in a study for the Office of the Deputy Prime Minister of the U.K., simplified their results into a proposed procedure for long horizon project analysis in the U.K. The proposed procedure starts with a discount rate of 3.5% per year which then declines in increments to 1% per year after 300 years. Figure 4 compares the discount and discount rate functions of the proposed U.K. procedure, which has an amount $\alpha = 0.0318$, to $d_z(t)$ and $r_z(t)$ and to the exponential with the same amount. While there are broad similarities, $r_z(t)$ starts higher (at slightly above 6%) and declines to zero rather than to 1%, which has the effect of placing substantially greater weight on the far future.

Table 3 shows the amount of present value remaining to be accumulated at various times for the proposed U.K. procedure and for ZSH and exponential procedures with amount 0.0318. After 100 years the exponential has 4.2% of its present value still to be accumulated and the proposed U.K. procedure has 7.5%. The ZSH, however, has 23.9% remaining, and even after 400 years still has as much present value to accumulate as the proposed U.K. procedure had at 100 years. The ZSH, with its ultimately 0 value of $r_z(t)$, places more weight on the far future than does a procedure with stepwise declining exponential rates that discounts by the same amount. The recently published Stern report on climate change (Stern, 2006) provides another example. It gives weight to the far future by (effectively) choosing a low discount rate of 1.4% per year. But even that low rate leaves only 6.1% of present value to be accumulated after 200 years whereas the ZSH with $r = 0.014$ has 26.3% still to be accumulated. The ZSH has 8% to accumulate after 800 years whereas, for the exponential, there is virtually nothing left by that time. In addition, short term interest rates of 2.8% with the ZSH are, while low, still more reasonable than 1.4%.

A switch to ZSH discounting, with an r of about 0.03, gives far more weight to the far future than does exponential discounting at any reasonable rate. It does this while preserving realistic discount rates over short horizons and with a tractable functional form. The ZSH results from ANDF aggregation of individual exponential discount functions whose parameters follow a gamma distribution with its standard deviation equal to its mean (equation 27). [In Weitzman's (2001) data the standard deviation was 3% around a mean of 4%, and the gamma fit fairly well except at 0%.] It would be worthwhile to conduct sensitivity analyses of long horizon policy assessments such as those of the Stern report to the use of ZSH discounting.

**Figure 4:
Comparison of Discounting Procedures with α = 0.0318 –
Exponential, Green Book and Zero-Speed Hyperbolic**

Panel A: Discount Functions -- d(t)

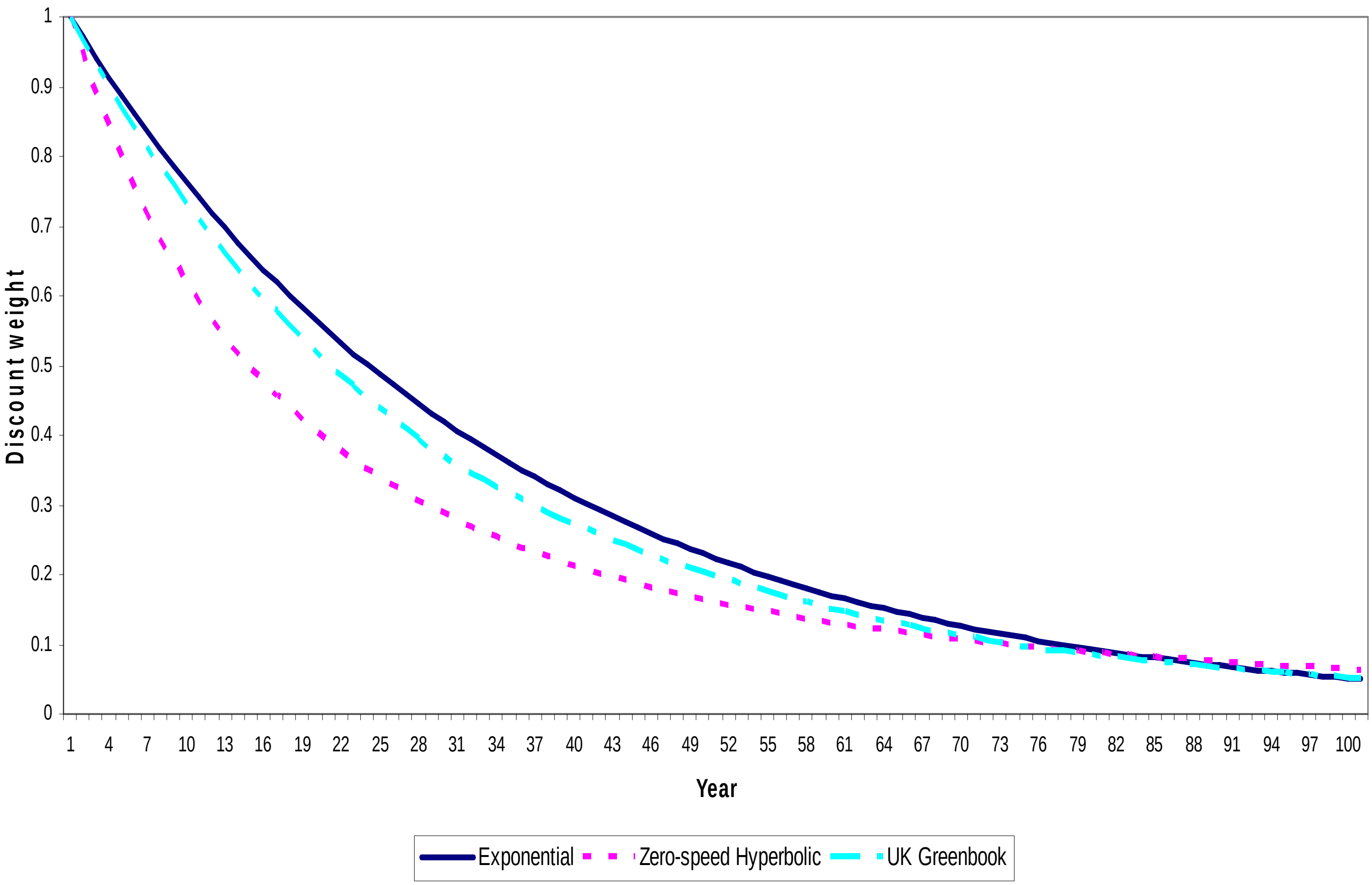


Figure 4:
Comparison of Discounting Procedures –
Exponential, Green Book and ZSH (α = 0.0318)

Panel B: Discount Rates – r(t)

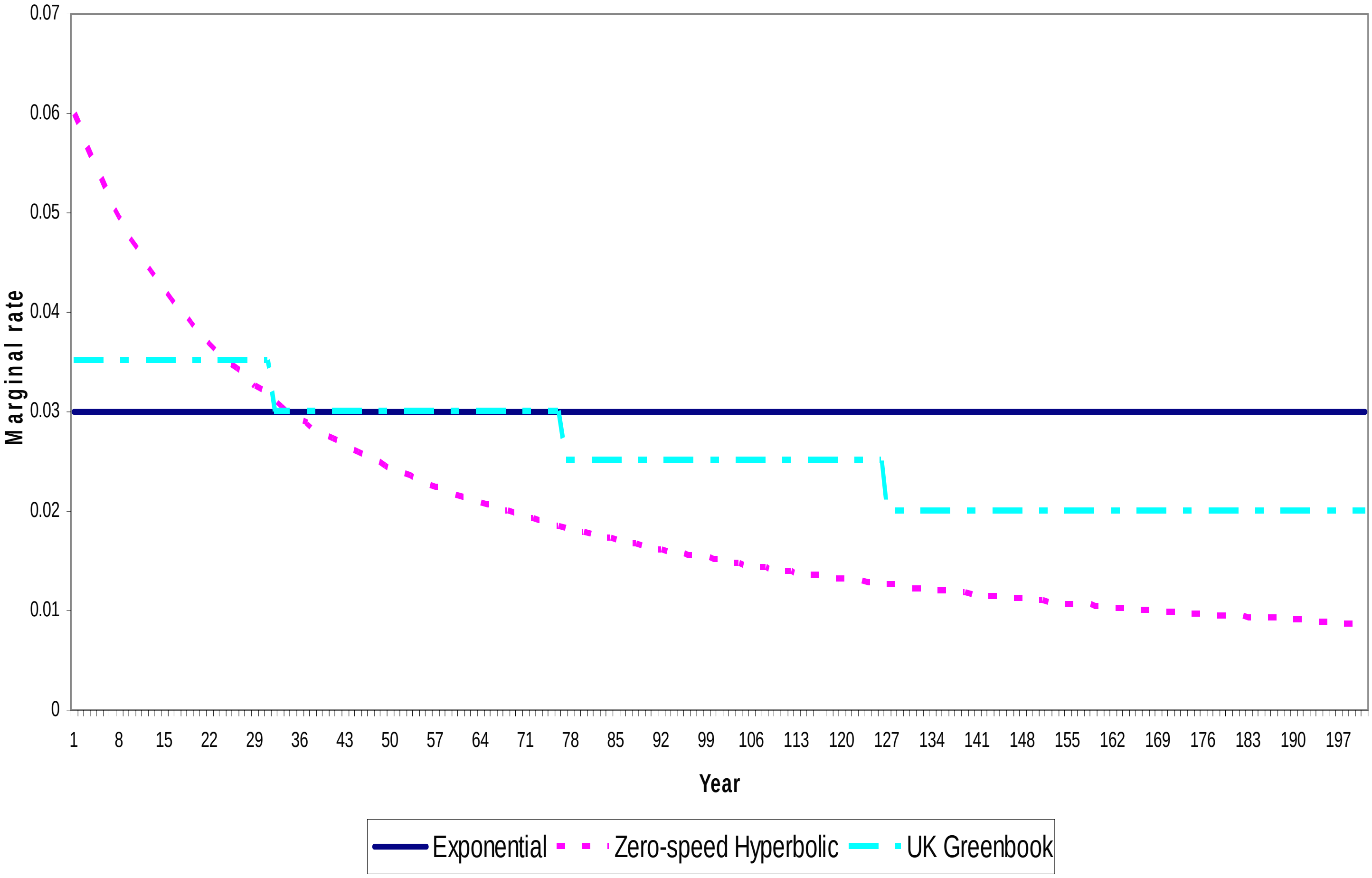


Table 3: Present Value Remaining as a Function of Time for Selected Exponential and Zero-Speed Hyperbolic (ZSH) Procedures

	Percent of present value remaining				
	$\alpha = 0.014$ (Stern Report)		$\alpha = 0.0318$		
After year	Exp	ZSH	Exp	ZSH	Proposed UK procedure
0	100	100	100	100	100
25	70.4	74.1	45.2	55.7	47.1
50	49.7	58.8	20.4	38.6	24.3
100	24.7	41.7	4.2	23.9	7.5
200	6.1	26.3	0.2	13.6	1.4
400	0.4	15.2	0	7.3	0.2
800	0	8.2	0	3.8	0

1. The amount of discounting is designated α. For exponential discounting $d(t) = e^{-\alpha t}$; for ZSH discounting $d(t) = (1 + \alpha t)^{-2}$.

2. The Stern Report on climate change (Stern, 2006) effectively used a discount rate of 1.4% per year.

3. A procedure developed for the Office of the Deputy Prime Minister in the U.K. for the U.K. 'Green Book' (Oxera, 2002) prescribes a discounting with rates fixed over time intervals, beginning at 3.5% per year and declining to 1% per year after 300 years. For this procedure it turns out that $\alpha = 0.0318$.

3.3 Conclusions

This paper proceeds from the observation that economists in a range of specialized fields use non-exponential discounting functions. Within the exponential framework, alternative discounting procedures align along the single dimension of discount rate. This simplicity, along with the intuitive appeal of the axiomatic formulation of constant rate discounting and the analytic tractability of the procedure, have ensured its dominance until recently.[22] This paper develops a framework for nonconstant rate discounting that arrays procedures along the two dimensions of amount and speed, thereby facilitating systematic comparison of procedures and enhancing their tractability.

A question that is antecedent to discounting in the context of assessing long-term investments, addressed by Arrow (1999), concerns the extent to which decisions made today will have their influence attenuated (or eliminated) by compensatory decisions of subsequent generations. This issue is very real, and arises also in the design of other sequential decision procedures, such as foreign aid programs and transfers across levels of government. It has not been our purpose in this paper to address that issue but, more simply, to improve the tools available in circumstances where discounting is being used. Similarly, although we acknowledge the existence and practical relevance of time-inconsistency issues with regard to variable rate discounting – initially addressed by Robert H. Strotz (1956) – our focus has been on a different facet of the problem.

Key results of the paper include:

(i) precise formulations for concepts of 'amount', 'speed', and 'time horizon' of discounting procedures;

(ii) proofs of key relations among the concepts of amount, speed, and time horizon;

(iii) identification of inadequacies in existing approaches to aggregating individual discounting

[22] The axiomatic foundation for constant rate discounting goes back to Koopmans (1960) who introduced a 'stationarity' axiom concerning preferences over time streams of outcomes and established plausible conditions that require discounting to be exponential. A number of authors have explored weaker axiom systems that allow representation of intertemporal preferences by a discounting procedure that is not necessarily constant rate – see Jamison (1969), Fishburn and Rubinstein (1982), Bleichrodt and Gafni (1996) and Bleichrodt and Johannesson (2001). The simple existence of an axiomatic foundation, therefore, is no argument in favor of constant- over variable-rate discounting; the question is one of assessing the descriptive, normative, and tractability consequences of adding the strong stationarity axiom.

procedures into a socially representative one and formulation of an alternative process – averaging of normalized discount functions or ANDF – that overcomes these inadequacies;

(iv) proof that under reasonable assumptions the ANDF aggregate procedure will be slower than the average of the speeds of the individual procedures;

(v) descriptions of a range of existing and new discounting procedures and provision of closed form characterizations relating amount, speed, and time horizon to their underlying parameters; and

(vi) use of a particular variant of hyperbolic discounting, which we label the zero-speed hyperbolic or ZSH, would provide an analytically tractable way of giving substantial weight to the far future in policy analyses while preserving reasonable discount rates in the short term.

Frederick, Loewenstein and O'Donoghue (2002), Chapman (2003) and Groom et al. (2005) provide valuable compilations of the recent literature on time preference and discounting. Poulos and Whittington (2000) extend the literature on discounting of lifes saved to several developing country contexts. Transforming the empirical literature into discounting procedures for policy application will require two additional steps. First, to the extent practical, data underlying the reported literature will need to be characterized in terms of estimates of the amount and speed of individual discounting. Second, the ANDF aggregation algorithm can be used (through our Proposition 6) to generate candidate social discounting procedures. We feel that the approach to discounting that we propose both undermines many of the practical objections to expanded use of nonconstant rate procedures and provides a needed framework for integrating and comparing results in the existing literature.

4 Appendix A - Relations between the discount function and other defining functions

A discounting procedure can be defined in multiple ways, each of which leads to a discount function, $d(t)$, that assigns to each future time t a coefficient by which the net outcome at that time can be converted to its present value or the time 0 'equivalent' of the time t outcome. Table 4 lists five functions that can, in addition to $d(t)$ itself, be used to define a discounting procedure and relates each of these to $d(t)$. Four of the potential defining functions – the discount rate, discount factor, yield curve, and present value – are familiar, although they often appear in different contexts. We add a fifth defining function in addition to these four, which we label the 'associated probability density function' (or *pdf*).

Table 4 brings together in one place the relation between each of the defining functions and $d(t)$. For reference the relation to $d(t)$ is given for both discrete and continuous time formulations. Row 1, for example, shows the relation of $d(t)$ to the discount rate function $r(t)$ by giving $d(t)$ as a function of $r(t)$ and vice versa. Row 2 shows the relation of discount factors – the standard representation of discounting in game theory, for instance – to $d(t)$.

Row 3 relates the yield curve from the finance literature to $d(t)$. The relation here is a useful one for interpreting results in the health-related literature on assessment of discount rates since they are frequently reported (implicitly) as yield curves.[23] Row 4 relates $d(t)$ to the cumulative present value function. Finally, Row 5 results from the observation that the inverse cumulative[24] of any *pdf* defined on $[0, \infty)$ will, in fact, be a discount function with a present value equal to the expectation of a random variable with that *pdf*. Thus if f is a *pdf* on $[0, \infty)$, then

$$d_f(t) = 1 - \int_0^t f(x)dx = \int_t^\infty f(x)dx \tag{34}$$

will be a discount function.

[23] See Malkiel (1998) or Cochrane (2001, Chapter 19) for discussions of yield curves and the term structure of interest rates. Cropper, Aydede and Portney (1994), in an important early paper on discounting of lives saved, report a 16.8% discount rate over 5 years declining to 3.4% over 100 years, a 'yield curve' formulation. Cairns and van der Pol (1997) provide further strong evidence for slow discounting in the context of saving lives.

[24] The inverse cumulative of a *pdf* is 1 minus the cumulative.

Table 4: Relations Between the Discount Function and Other Defining Functions

Definitional function	Discrete time	Continuous time
1. discount rate, $r(t)$	$d(t) = \prod_{i=1}^{t} [1 + r(i)]^{-1}$	$d(t) = e^{-\int_0^t r(x)dx}$
	$r(t) = \frac{d(t-1)}{d(t)} - 1$	$r(t) = \frac{-d'(t)}{d(t)}$
2. discount factor, $\delta(t)$	$d(t) = \prod_{i=1}^{t} \delta(i)$	$d(t) = e^{\int_0^t \ln \delta(x)dx}$
	$\delta(t) = \frac{d(t)}{d(t-1)}$	$\delta(t) = e^{d'(t)/d(t)}$
3. yield curve, $y(t)$	$d(t) = [1 + y(t)]^{-t}$	$d(t) = e^{-y(t)t}$
	$y(t) = [d(t)]^{-1/t} - 1$	$y(t) = -\frac{\ln d(t)}{t}$
4. present value, $pv(t)$	$d(t) = pv(t) - pv(t-1)$	$d(t) = pv'(t)$
	$pv(t) = \sum_{i=0}^{t} d(i)$	$pv(t) = \int_0^t d(x)dx$
5. associated *pdf*, $f(t)$	$d(t) = \sum_{i=t}^{\infty} f(i)$	$d(t) = \int_t^\infty f(x)dx$
	$f(t) = d(t) - d(t+1)$	$f(t) = -d'(t)$

Note: Our conventions here are the usual ones, namely that the empty product is equal to 1 and the unit stream begins to accrue immediately at time 0. In discrete time, $r(t)$ refers to the discount rate between time $t-1$ and time t, and similarly for $\delta(t)$. Also in discrete time, $pv(t)$ is the present value accumulated up through and inclusive of time t.

5 Appendix B - Derivation of properties of specific discounting procedures

Table 1 provided formulas for discount rates, amounts, relative speeds, median time horizons and mean time horizons for the hyperbolic, Weibull and Gamma discounting procedures. Results not proved in the text are proved in Sections 6.1, 6.2 and 6.3 below. In addition Sections 6.4 and 6.5 present derivations for the characteristics of quasi-hyperbolic discounting procedures, generalized to continuous time in two separate ways.

5.1 Hyperbolic

If $d(t) = [1 + r(1-s)t]^{-\frac{1}{1-s}-1}$ then

$$r(t) = -(-\frac{1}{1-s} - 1)r(1-s)\frac{[1 + r(1-s)t]^{-\frac{1}{1-s}-2}}{[1 + r(1-s)t]^{-\frac{1}{1-s}-1}} = \frac{r(2-s)}{1 + r(1-s)t}$$

and

$$\int_0^\infty d(t)dt = \frac{1}{r(1-s)}\frac{1}{-1/(1-s)}\left[1 + r(1-s)t\right]^{-\frac{1}{1-s}}\Big|_0^\infty = \frac{1}{r}$$

and so $\alpha = r$ as claimed. For the speed, we begin by observing that

$$\int_0^\infty td(t)dt = \int_0^\infty \frac{t}{[1 + r(1-s)t]^{1+\frac{1}{1-s}}}dt = \frac{1}{r^2(1-s)^2}\int_0^\infty \frac{x}{[1+x]^{1+\frac{1}{1-s}}}dx,$$

with the change of variable $x = r(1-s)t$. The beta function $B(y,z)$ can be written

$$B(y,z) = \int_0^\infty \frac{x^{y-1}}{[1+x]^{y+z}}dx,$$

so inserting $y = 2$ and $z = \frac{1}{1-s} - 1 = \frac{s}{1-s}$ yields

$$\int_0^\infty td(t)dt = \frac{1}{r^2(1-s)^2}B(2, \frac{s}{1-s}).$$

But the beta function can be expressed in terms of the gamma function (Gradshteyn and Ryzhik 1980, section 8.38) as

$$B(y,z) = \frac{\Gamma(y)\Gamma(z)}{\Gamma(y+z)},$$

which itself has the property that $\Gamma(y+1) = y\Gamma(y)$. Plugging in (and recalling that $\Gamma(n) = (n-1)!$ for integer n), we find that

$$B(2, \frac{s}{1-s}) = \frac{\Gamma(2)\Gamma(\frac{s}{1-s})}{\Gamma(\frac{s}{1-s}+2)} = \frac{1 \cdot \Gamma(\frac{s}{1-s})}{(\frac{s}{1-s}+1)\Gamma(\frac{s}{1-s}+1)} = \frac{\Gamma(\frac{s}{1-s})}{(\frac{1}{1-s})(\frac{s}{1-s})\Gamma(\frac{s}{r-s})} = \frac{(1-s)^2}{s}.$$

Finally,

$$\rho = \left[\alpha^2 \int_0^\infty td(t)dt\right]^{-1} = \left[r^2 \frac{1}{r^2(1-s)^2} \frac{(1-s)^2}{s}\right]^{-1} = s$$

as desired. And of course θ is then $(rs)^{-1}$. For the median time, we set $pv(\tau) = pv(\infty)/2$:

$$\begin{aligned}
\frac{1-[1+r(1-s)\tau]^{-\frac{1}{1-s}}}{r} &= \frac{1}{2r} \implies \\
1+r(1-s)\tau &= 2^{1-s} \implies \\
\tau &= \frac{2^{1-s}-1}{r(1-s)}.
\end{aligned}$$

Note that, as expected, $\lim_{s\to 1} \tau = r^{-1}\ln 2$ (use L'Hôpital's rule), which is the value for the standard exponential.

5.2 Weibull

For $d(t) = \exp(-rt^{1/s})$, we can immediately observe that $r(t) = (r/s)t^{1/s-1}$. To calculate amount and speed, we will use the following formula (Gradshteyn and Ryzhik 1980, section 3.4781):

$$\int_0^\infty x^{y-1}\exp(-rx^z)dx = \frac{1}{z}r^{-y/z}\Gamma(y/z)$$

with $z = 1/s$. First substituting $y = 1$, we find

$$\int_0^\infty d(t)dt = \frac{1}{1/s}r^{-s}\Gamma(s) = \frac{s\Gamma(s)}{r^s} = \frac{\Gamma(s+1)}{r^s}$$

using the same identity $\Gamma(y+1) = y\Gamma(y)$ as before. The inverse of this is then α, as claimed. To calculate the speed, we instead substitute $y = 2$, implying that

$$\int_0^\infty td(t)dt = \frac{1}{1/s}r^{-2s}\Gamma(2s) = \frac{s\Gamma(2s)}{r^{2s}}$$

so that

$$\rho = \left[\alpha^2 \int_0^\infty td(t)dt\right]^{-1} = \frac{s^2[\Gamma(s)]^2}{r^{2s}}\frac{r^{2s}}{s\Gamma(2s)} = s\frac{[\Gamma(s)]^2}{\Gamma(2s)} = \frac{\Gamma(s)\Gamma(s+1)}{\Gamma(2s)},$$

also as claimed. The mean time follows immediately:

$$\theta = \alpha\int_0^\infty td(t)dt = \frac{\Gamma(2s)}{r^s\Gamma(s)}.$$

Unfortunately, there is no closed form integral for $d(t)$ and thus no closed form expression for the median time horizon.

We can further investigate several properties of the Weibull discounting procedure. The binomial coefficient $\binom{n}{m}$ is by definition

$$\binom{n}{m} = \frac{n!}{m!(n-m)!} = \frac{\Gamma(n+1)}{\Gamma(m+1)\Gamma(n-m+1)},$$

and the formulation in terms of gamma functions can be used to define a continuous binomial coefficient for non-integer n, m. In that case, the relative speed is the inverse of

$$\frac{\Gamma(2s)}{\Gamma(s)\Gamma(s+1)} = \frac{\Gamma((2s-1)+1)}{\Gamma((2s-1)-s+1)\Gamma(s+1)} = \binom{2s-1}{s},$$

which is strictly increasing in s for $s > 0$ (since it is an extension of the binomial coefficient). Thus the speed is strictly decreasing in s, and of course it equals $\binom{2-1}{1} = 1$ when $s = 1$. This validates our claim that the transformed time discounting procedure is slow (i.e. $\rho < 1$) exactly when $s > 1$; and it is fast ($\rho > 1$) when $s < 1$. It also provides a simple way to calculate ρ when s is an integer. The claim is also a direct consequence of Proposition 3, once we note that the derivative of the discount rate has the same sign as $1 - s$:

$$r'(t) = (1-s)\frac{r}{s^2}t^{(1/s)-2}.$$

So when $s > 1$, the discount rate is declining over time (slow discounting), and vice-versa.

Finally, we examine the limit cases for extreme values of s. As $s \to \infty$, $d(t)$ starts to look almost constant at a value of e^{-r}, so the procedure is less and less convergent (though of course for any finite s there is no actual problem). This means that $\alpha \to 0$. Since $\Gamma(y)$ grows "factorially" in y, we find that ρ goes to 0 as well (and θ grows without bound), which does not automatically follow from the result for α. That is, the limit discounting procedure is slow even relative to the equivalent exponential (which has, of course, a zero discount rate in the limit). In the other extreme, as $s \to 0$, $d(t)$ starts to appear constant at 1 until $t = 1$, where it precipitously drops to 0. This suggests that α should be near 1, and indeed

$$\lim_{s\to 0} \alpha = \lim_{s\to 0} \frac{r^s}{\Gamma(s+1)} = \frac{r^0}{\Gamma(1)} = 1$$

since both numerator and denominator are continuous at $s = 0$. For the speed, we make use of the doubling formula for the gamma function:

$$\Gamma(2y) = \frac{2^{2y-1}}{\sqrt{\pi}}\Gamma(y)\Gamma(y+\frac{1}{2}),$$

from which we obtain

$$\rho = \frac{\Gamma(s)\Gamma(s+1)}{\Gamma(2s)} = \frac{\sqrt{\pi}}{2^{2s-1}}\frac{\Gamma(s+1)}{\Gamma(s+1/2)}$$

and thus (using the known value $\Gamma(1/2)=\sqrt{\pi}$)

$$\lim_{s\to 0}\rho=\lim_{s\to 0}\frac{\sqrt{\pi}}{2^{2s-1}}\frac{\Gamma(s+1)}{\Gamma(s+1/2)}=\frac{\sqrt{\pi}}{2^{0-1}}\frac{1}{\sqrt{\pi}}=2,$$

again by continuity at $s=0$ of all functions involved. This is the "fastest" that we can get with this type of discounting procedure. Therefore, θ approaches $1/2$, and it is clear in this case that τ approaches $1/2$ as well.

5.3 Gamma

We start with a gamma formulation $x^s e^{-(s+1)rx}$, so $s=0$ is the basic exponential case. Parameter restrictions are $s>-1$ and $r>0$. We need to first turn this into a legitimate *pdf* function: $\int_0^\infty x^s e^{-(s+1)rx}dx=\frac{\Gamma(s+1)}{[(s+1)r]^{s+1}}$, so we normalize to get

$$f(x)=\frac{[(s+1)r]^{s+1}}{\Gamma(s+1)}x^s e^{-(s+1)rx}$$

as our starting point. We turn this *pdf* into a discount function as in the paper:

$$d(t)=\int_t^\infty f(x)dx=\frac{[(s+1)r]^{s+1}}{\Gamma(s+1)}\int_t^\infty x^s e^{-(s+1)rx}dx=\frac{[(s+1)r]^{s+1}}{\Gamma(s+1)}\frac{\Gamma(s+1,(s+1)rt)}{[(s+1)r]^{s+1}}$$

(where $\Gamma(x,z)=\int_z^\infty e^{-y}y^{x-1}dy$ is by definition the incomplete gamma function), ending up with

$$d(t)=\frac{\Gamma(s+1,(s+1)rt)}{\Gamma(s+1)}.$$

The discount rate function is thus

$$r(t)=\frac{-d'(t)}{d(t)}=\frac{[(s+1)r]^{s+1}\,t^s e^{-(s+1)rt}}{\Gamma(s+1,(s+1)rt)}.$$

So $r(0)=0$ for $s>0$, $r(0)=r$ for $s=0$, and $r(0)=\infty$ for $s<0$. To find $r(\infty)$, use L'Hôpital's

Rule:

$$\lim_{t\to\infty} r(t) = \lim_{t\to\infty} \frac{[(s+1)r]^{s+1}\, t^{s-1} e^{-(s+1)rt}\, [s-(s+1)rt]}{-\,[(s+1)r]^{s+1}\, t^{s} e^{-(s+1)rt}} = \lim_{t\to\infty} \left[(s+1)r - \frac{s}{t}\right] = (s+1)r$$

for any s.

We may now calculate the amount α and relative speed ρ.

$$\begin{aligned}
\int_0^\infty d(t)dt &= \int_0^\infty \left(\frac{[(s+1)r]^{s+1}}{\Gamma(s+1)} \int_t^\infty x^s e^{-(s+1)rx} dx \right) dt \\
&= \frac{[(s+1)r]^{s+1}}{\Gamma(s+1)} \int_0^\infty \int_t^\infty x^s e^{-(s+1)rx} dx dt \\
&= \frac{[(s+1)r]^{s+1}}{\Gamma(s+1)} \int_0^\infty \int_0^x x^s e^{-(s+1)rx} dt dx \\
&= \frac{[(s+1)r]^{s+1}}{\Gamma(s+1)} \int_0^\infty x^s e^{-(s+1)rx} \int_0^x dt dx \\
&= \frac{[(s+1)r]^{s+1}}{\Gamma(s+1)} \int_0^\infty x^{s+1} e^{-(s+1)rx} dx \\
&= \frac{[(s+1)r]^{s+1}}{\Gamma(s+1)} \frac{\Gamma(s+2)}{[(s+1)r]^{s+2}} = \frac{1}{r}
\end{aligned}$$

so $\alpha = r$ quite simply. For the mean time and relative speed:

$$\begin{aligned}
\int_0^\infty t d_s(t)dt &= \int_0^\infty t \left(\frac{[(s+1)r]^{s+1}}{\Gamma(s+1)} \int_t^\infty x^s e^{-(s+1)rx} dx \right) dt \\
&= \frac{[(s+1)r]^{s+1}}{\Gamma(s+1)} \int_0^\infty \int_t^\infty t x^s e^{-(s+1)rx} dx dt \\
&= \frac{[(s+1)r]^{s+1}}{\Gamma(s+1)} \int_0^\infty \int_0^x t x^s e^{-(s+1)rx} dt dx \\
&= \frac{[(s+1)r]^{s+1}}{\Gamma(s+1)} \int_0^\infty x^s e^{-(s+1)rx} \int_0^x t dt dx \\
&= \frac{[(s+1)r]^{s+1}}{\Gamma(s+1)} \int_0^\infty (\frac{1}{2}) x^{s+2} e^{-(s+1)rx} dx \\
&= \frac{[(s+1)r]^{s+1}}{\Gamma(s+1)} \frac{\Gamma(s+3)}{2\,[(s+1)r]^{s+3}} = \frac{(s+2)}{2(s+1)r^2}.
\end{aligned}$$

But θ is α times this integral, and therefore $\theta = \frac{s+2}{2(s+1)r}$. Finally, $\rho = (\alpha\theta)^{-1}$ so

$$\rho = \left(\frac{s+2}{2(s+1)}\right)^{-1} = 2 - \frac{2}{s+2}.$$

Thus the relative speed depends only on s (not r) and is increasing in s toward a limit of 2 (the maximum possible relative speed of any discounting procedure) as s increases to infinity, with a lower limit of 0 as s goes to -1. Unfortunately, there is no closed form for the median time τ, since one cannot analytically integrate the incomplete gamma function in $d(t)$.

5.4 Split function quasi-hyperbolic

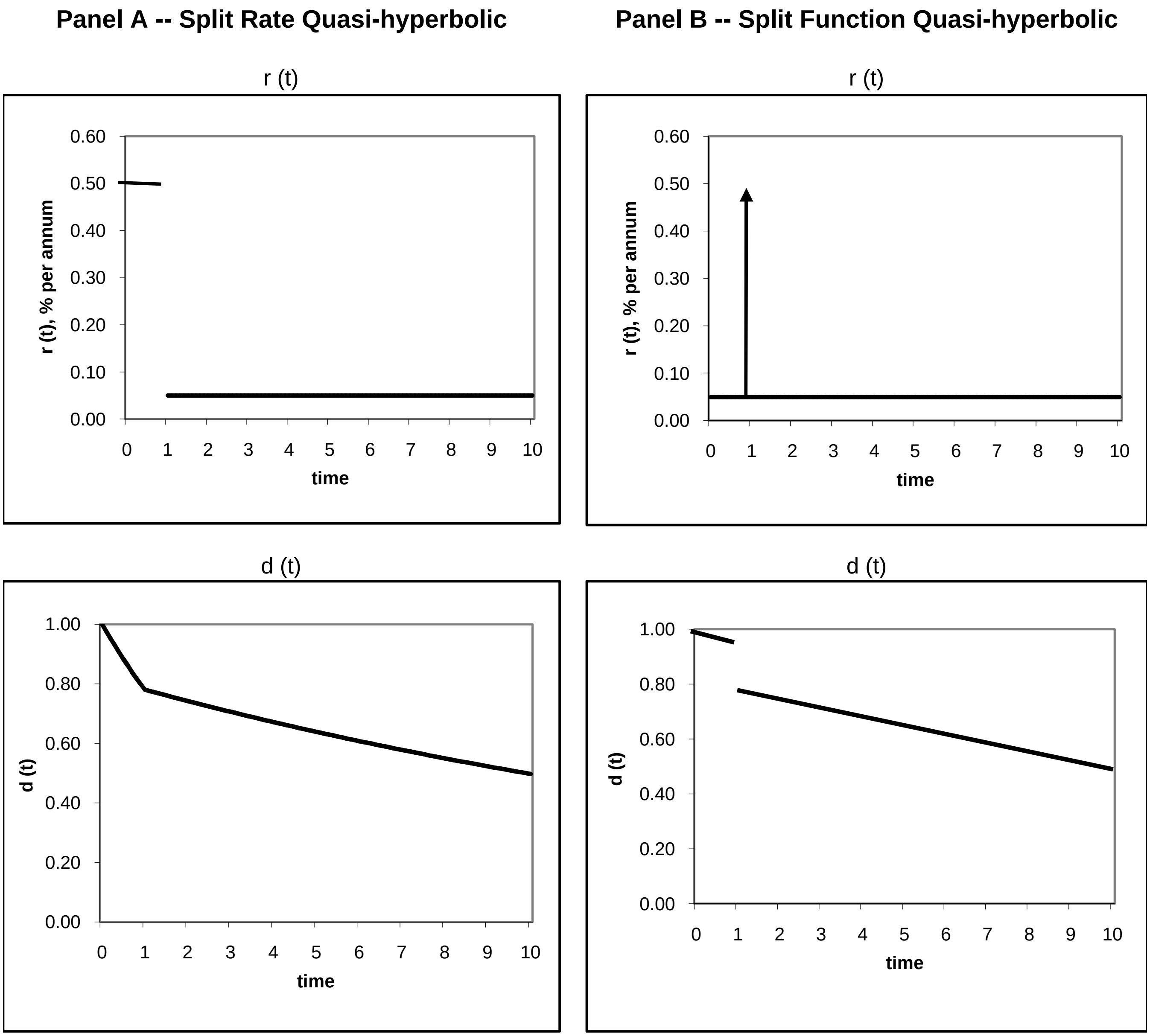
Another class of functions derives from continuous time versions of the discrete time quasi-hyperbolic discount function introduced as equation (4). This function can be generalized in a variety of ways. First, and most obviously, the point of discontinuity can be made to be at any time, not just 1 (which was in any case already a parameter because of need for the choice of time scale). Second, an option that we do not pursue is to introduce discontinuities at multiple time points. Third, the generalization to continuous time can be undertaken either by introducing a discontinuity in $r(t)$ – and, hence, in the derivative of $d(t)$ – or by introducing a discontinuity in $d(t)$ itself. We label these the 'split rate quasi-hyperbolic' and the 'split function quasi-hyperbolic'. The latter allows $r(t)$ to remain constant [except for a spike to infinity at the time of the discontinuity in $d(t)$] and has been utilized by Harris and Laibson (2000) and, in quite a different way, by Cline (1999). The most natural generalization of the discrete-time quasi-hyperbolic is the split function version, and this subsection includes a definition of this version and shows its properties. In the next subsection, 6.5, we explore properties of the split rate quasi-hyperbolic procedures. Figure 5 illustrates the split function and split rate quasi-hyperbolic procedures.

We first consider the procedure that splits the discount function rather than the discount rate:

$$d(t) = \begin{cases} e^{-rt} & \text{if } t \leq t^* \\ \lambda e^{-rt} & \text{if } t > t^* \end{cases}$$

where $r > 0$ and $\lambda < 1$. The discount rate for this procedure is constant at r except at $t = t^*$, where it is infinite. If we again denote e^{-rt^*} by β, we see that

Figure 5:
Alternative Generalizations to Continuous Time of the Quasi-hyperbolic Discounting Procedure



Note: A quasi-hyperbolic is defined in discrete time, with values of the discount function given by $(1, \beta\delta, \beta\delta^2, \beta\delta^3, \ldots)$, where δ is a discount factor and β is a constant. This figure shows two alternative continuous time formulations that generalize a quasi-hyperbolic with $\beta = 0.78$ and $\delta = 0.95$. In Panel A r(t) is discontinuous at t = 1, dropping from 0.25 to 0.05, leading to a change in the slope of d(t) at 1 where d(t) = 0.78. In Panel B r(t) = 0.05 for all t except t = 1 where it is infinite. If r(t) is an appropriately formulated Dirac delta function, then d(t) can be made to drop from 0.95 to 0.78 at t = 1. Choice of the unit of measurement for time determines the time of discontinuity which is assumed, without loss of generality, to occur at t = 1 in these examples.

$$\int_0^\infty d(t)dt = \int_0^{t^*} e^{-rt}dt + \lambda\int_{t^*}^\infty e^{-rt}dt = (1-\beta)\frac{1}{r} + \lambda\frac{1}{r}\beta = \frac{1-(1-\lambda)\beta}{r},$$

and the amount of discounting is the inverse of this:

$$\alpha = \frac{r}{1-(1-\lambda)\beta}.$$

For the speed,

$$\begin{aligned}\int_0^\infty td(t)dt &= \int_0^{t^*} te^{-rt}dt + \lambda\int_{t^*}^\infty te^{-rt}dt\\ &= \frac{1}{r}((1-\beta)\frac{1}{r} - t^*\beta) + \lambda\frac{1}{r}(t^*\beta + \frac{1}{r}\beta)\\ &= \frac{1}{r^2}[1-(1-\lambda)\beta(1+rt^*)]\end{aligned}$$

and so

$$\rho = \frac{[1-(1-\lambda)\beta]^2}{1-(1-\lambda)\beta(1+rt^*)}.$$

From here, θ follows as always. To find bounds on ρ for this procedure, we first consider the extreme case $\lambda = 0$: here $\rho = f(\beta) = (1-\beta)^2/(1-\beta+\beta\ln\beta)$ (same as in the last subsection). So once more the maximum value for ρ is 2. Note that if $\lambda > 0$, then ρ is maximized at an interior choice of β, but the $\arg\max$ goes to 1 (and the maximum value goes to 2) as λ goes to 0. On the other hand, fixing λ and letting β go to 1, ρ is approximately λ – so we can make it as small as we want (greater than 0). Thus the order of limits makes a big difference! Finally, we point out that for any fixed λ, ρ goes to 1 as β goes to 0.

For the median time, we again distinguish two cases: either τ is smaller than t^* (and τ occurs before the λ jump) or it is larger than t^* (where the expression for $pv(t)$ is different due to the λ factor). The boundary case will be if $\tau = t^*$, which occurs exactly when $pv(t^*) = pv(\infty)/2$, i.e. if $1/r - \beta/r = [1-(1-\lambda)\beta]/2r$. This is true when $\beta = 1/(1+\lambda)$, i.e. for $t^* = r^{-1}\ln(1+\lambda)$. If t^* is larger than this (so that τ is in the initial range), $pv(\tau) = 1/r - e^{-r\tau}/r$ and we need to set this equal to $pv(\infty)/2 = [1-(1-\lambda)\beta]/2r$. Solving,

$$\begin{aligned}1-e^{-r\tau} &= \frac{1-(1-\lambda)\beta}{2} \implies \\ e^{-r\tau} &= \frac{1+(1-\lambda)\beta}{2} \implies \\ \tau &= r^{-1}\ln\frac{2}{1+(1-\lambda)\beta}.\end{aligned}$$

If t^* is smaller than the cutoff value (so that $\tau > t^*$), then $pv(\tau) = \int_0^{t^*} e^{-rt}dt + \beta\int_{t^*}^{\tau} \lambda e^{-r(t-t^*)}dt = (1-\beta)/r + \lambda\left(\beta - e^{-r\tau}\right)/r$ and we solve

$$\begin{aligned}\frac{1-\beta}{r} + \frac{\lambda\left(\beta - e^{-r\tau}\right)}{r} &= \frac{1-(1-\lambda)\beta}{2r} \implies \\ \lambda\left(\beta - e^{-r\tau}\right) &= \frac{1}{2}\left[\lambda\beta - (1-\beta)\right] \implies \\ e^{-r\tau} &= \frac{\lambda\beta + (1-\beta)}{2\lambda} \implies \\ \tau &= r^{-1}\ln\frac{2\lambda}{1-(1-\lambda)\beta}.\end{aligned}$$

5.5 Split rate quasi-hyperbolic

If $d(t) = e^{-rt}$ for $t \le t^*$ and $d(t) = \beta e^{-s(t-t^*)}$ for $t > t^*$ (where $\beta = e^{-rt^*}$), then the discount rate is immediate (it is in fact the defining characteristic for this procedure), though it is undefined at $t = t^*$ where d is not differentiable[25]. For the amount,

$$\int_0^\infty d(t)dt = \int_0^{t^*} e^{-rt}dt + \beta\int_{t^*}^{\infty} e^{-s(t-t^*)}dt = (1-\beta)\frac{1}{r} + \beta\frac{1}{s},$$

from which it is clear that the inverse is indeed a [weighted] harmonic mean:

$$\alpha = \frac{rs}{\beta r + (1-\beta)s}.$$

For the speed, we calculate

[25] If $s > r$, say, we can think of $r(t)$ as having a Dirac delta function spike at t^*.

$$\begin{aligned}
\int_0^\infty td(t)dt &= \int_0^{t^*} te^{-rt}dt + \beta \int_{t^*}^\infty te^{-s(t-t^*)}dt \\
&= \left[-\frac{1}{r}te^{-rt} - \frac{1}{r^2}e^{-rt} \right] \Bigg|_0^{t^*} + \beta \int_0^\infty (t+t^*)e^{-st}dt \\
&= \frac{1}{r}(-t^*\beta - \frac{\beta}{r} + 0 + \frac{1}{r}) + \beta \int_0^\infty te^{-st}dt + \beta t^* \int_0^\infty e^{-st}dt \\
&= \frac{1}{r}((1-\beta)\frac{1}{r} - t^*\beta) + \beta\frac{1}{s^2} + \beta t^* \frac{1}{s} \\
&= \frac{1}{rs}\left[(1-\beta)\frac{s}{r} - t^*\beta s + \beta\frac{r}{s} + \beta t^* r \right]
\end{aligned}$$

so that the relative speed is the inverse of

$$\alpha^2 \int_0^\infty td(t)dt = \frac{\beta r^2 + (1-\beta)s^2 + \beta t^* rs(r-s)}{[\beta r + (1-\beta)s]^2},$$

as claimed. The formula for θ also follows from this computation. It is easy to calculate that $\rho \gtrless 1$ as $s \gtrless r$. To find bounds for ρ, we first consider $s >> r$: in this case, ρ is approximately $f(\beta) \equiv (1-\beta)^2/(1-\beta+\beta\ln\beta)$. The function f is monotonically increasing in β, with $\lim_{\beta\to 0} f(\beta) = 1$ and $\lim_{\beta\to 1} f(\beta) = 2$ (use L'Hôpital's rule twice), so ρ can get arbitrarily close to 2. On the other side, for $r >> s$, ρ is approximately $\beta/(1+st^*)$, which is obviously minimal for β near 0 (i.e. rt^* very large), and hence ρ near 0.

For the median time, we need to distinguish two cases: either τ is smaller than t^* (in which case τ occurs while the discount rate is still r and we can use the appropriate expression for $pv(t)$) or it is larger than t^* (where the rate is s and the expression for $pv(t)$ is different). The boundary case will be if $\tau = t^*$, which occurs exactly when $pv(t^*) = pv(\infty)/2$, i.e. if $1/r - \beta/r = [\beta r + (1-\beta)s]/2rs$. This is true if $\beta r = (1-\beta)s$ (which is intuitively reasonable), i.e. for $\beta = e^{-rt^*} = \frac{s}{r+s}$. Thus the cutoff value for t^* is $r^{-1}\ln\frac{r+s}{s}$. If t^* is larger than this (so that τ is in the initial range), $pv(\tau) = 1/r - e^{-r\tau}/r$ and we need to set this equal to $pv(\infty)/2 = [\beta r + (1-\beta)s]/2rs$. Solving,

$$
\begin{aligned}
1-e^{-r\tau} &= \frac{1}{2}\frac{\beta r+(1-\beta)s}{s} \quad \Longrightarrow \\
e^{-r\tau} &= \frac{s-\beta(r-s)}{2s} \quad \Longrightarrow \\
\tau &= r^{-1}\ln\frac{2s}{s-\beta(r-s)}.
\end{aligned}
$$

If t^* is smaller than the cutoff value (so that $\tau > t^*$), then $pv(\tau) = \int_0^{t^*} e^{-rt}dt + \beta\int_{t^*}^{\tau} e^{-s(t-t^*)}dt = (1-\beta)/r + \beta\left(1-e^{-s(\tau-t^*)}\right)/s$ and we solve

$$
\begin{aligned}
\frac{1-\beta}{r} + \frac{\beta\left(1-e^{-s(\tau-t^*)}\right)}{s} &= \frac{1}{2}\frac{\beta r+(1-\beta)s}{rs} \quad \Longrightarrow \\
\beta r e^{-s(\tau-t^*)} &= \frac{1}{2}\left[\beta r+(1-\beta)s\right] \quad \Longrightarrow \\
e^{-s(\tau-t^*)} &= \frac{\beta r+(1-\beta)s}{2\beta r} \quad \Longrightarrow \\
\tau &= t^* + s^{-1}\ln\frac{2\beta r}{\beta r+(1-\beta)s}.
\end{aligned}
$$

6 Appendix C - Other Proofs

6.1 Proof of Proposition 7

Proposition 7 relates the limiting value of the ANDF social discount rate function to the minimum of the limiting values of the individual discount rate functions. Recall that

$$
r_{\overline{D}}(t) = \frac{\int_X \alpha(x) r(t;x) d(t;x) f(x) dx}{\int_X \alpha(x) d(t;x) f(x) dx},
$$

which we can rewrite as a weighted average

$$
r_{\overline{D}}(t) = \int_X \beta(t;x) r(t;x) dx
$$

with time-dependent weights $\beta(t;x)$ given by

$$\beta(t;x) = \frac{\alpha(x)d(t;x)f(x)}{\int_X \alpha(x)d(t;x)f(x)dx}.$$

Obviously, $\int_X \beta(t;x)dx = 1$ for any t.

For any $\varepsilon > 0$, let us partition X into A_1, $A_2(\varepsilon)$, and $A_3(\varepsilon)$ as follows: $A_1 = \{x \in X \text{ s.t. } r^*(x) < r^*_{\min}\}$; $A_2(\varepsilon) = \{x \in X \text{ s.t. } r^*_{\min} \leq r^*(x) \leq r^*_{\min} + \varepsilon\}$; and $A_3(\varepsilon) = \{x \in X \text{ s.t. } r^*_{\min} + \varepsilon < r^*(x)\}$. Then we can write $\lim_{t\to\infty} r_{\overline{D}}(t) = \lim_{t\to\infty} \int_X \beta r dx = \lim_{t\to\infty} \int_{A_1} \beta r dx + \lim_{t\to\infty} \int_{A_2(\varepsilon)} \beta r dx + \lim_{t\to\infty} \int_{A_3(\varepsilon)} \beta r dx$.

For any x with $\alpha(x) = 0$, $\beta(t;x) = 0$ identically. So ignoring any such x and noting that $\int_{A(r^*_{\min})} f(x)dx = 0$ by definition of $r^*_{\min}$, we see that $\int_{A_1} \beta(t;x)dx = 0$ as well (for all t), since $\alpha(x)$ and $d(t;x)$ are finite (and α, d, f are all weakly positive). But then $\lim_{t\to\infty} \int_{A_1} \beta(t;x)r(t;x)dx = 0$, so these values of x have no effect on $\lim_{t\to\infty} r_{\overline{D}}(t)$. This establishes that $\lim_{t\to\infty} r_{\overline{D}}(t) \geq r^*_{\min}$; it remains to show that $\lim_{t\to\infty} r_{\overline{D}}(t) \leq r^*_{\min}$.

Also by definition of $r^*_{\min}$, $\int_{A(r^*_{\min}+\varepsilon)} f(x)dx > 0$ for all $\varepsilon > 0$. But, as we just saw, $\int_{A_1} f(x)dx = 0$. Hence, for any ε, we can pick $x_2 \in A_2(\varepsilon)$ with $\alpha(x_2) > 0$ and $f(x_2) > 0$; now take any $x_3 \in A_3(\varepsilon)$. Then

$$\frac{\beta(t;x_3)}{\beta(t;x_2)} = \frac{\alpha(x_3)d(t;x_3)f(x_3)}{\alpha(x_2)d(t;x_2)f(x_2)}$$

Rewriting $d(t;x)$ as $\exp\left(-\int_0^t r(\tau;x)d\tau\right)$, we get

$$\frac{\beta(t;x_3)}{\beta(t;x_2)} = \frac{\alpha(x_3)f(x_3)}{\alpha(x_2)f(x_2)} \frac{\exp\left(-\int_0^t r(\tau;x_3)d\tau\right)}{\exp\left(-\int_0^t r(\tau;x_2)d\tau\right)} = M \exp\left(-\int_0^t \left[r(\tau;x_3) - r(\tau;x_2)\right] d\tau\right),$$

where $M = \frac{\alpha(x_3)f(x_3)}{\alpha(x_2)f(x_2)}$ is constant in t (and is finite by the choice of x_2). Therefore

$$\lim_{t\to\infty} \frac{\beta(t;x_3)}{\beta(t;x_2)} = M \lim_{t\to\infty} \exp\left(-\int_0^t \left[r(\tau;x_3) - r(\tau;x_2)\right] d\tau\right).$$

But of course $x_2 \in A_2(\varepsilon)$ and $x_3 \in A_3(\varepsilon)$ imply that $\lim_{t\to\infty} r(t;x_3) > r^*_{\min} + \varepsilon \geq \lim_{t\to\infty} r(t;x_2)$, from which it is clear that $\lim_{t\to\infty} \int_0^t \left[r(\tau;x_3) - r(\tau;x_2)\right] d\tau = \infty$ and thus

$$\lim_{t\to\infty}\frac{\beta(t;x_3)}{\beta(t;x_2)}=0.$$

This implies that, for any $x_3 \in A_3(\varepsilon)$, we must have $\lim_{t\to\infty}\beta(t;x_3) = 0$. This in turn yields $\lim_{t\to\infty}\int_{A_3(\varepsilon)}\beta(t;x)dx = 0$ (so that $\lim_{t\to\infty}\int_{A_2(\varepsilon)}\beta(t;x)dx = 1$), but it also implies the stronger conclusion that $\lim_{t\to\infty}\int_{A_3(\varepsilon)}\beta(t;x)r(t;x)dx = 0$. Therefore only $x \in A_2(\varepsilon)$ influence $\lim_{t\to\infty} r_{\overline{D}}(t)$, and in particular $\lim_{t\to\infty} r_{\overline{D}}(t) \le r^*_{\min} + \varepsilon$ for any $\varepsilon > 0$. But that means exactly that $\lim_{t\to\infty} r_{\overline{D}}(t) \le r^*_{\min}$, as needed.

6.2 Aggregation of exponentials with specific parameter distributions

We first start with an underlying gamma distribution for the parameter of individual exponential discount functions:

$$f(x)=\frac{a^b}{\Gamma(b)}e^{-ax}x^{b-1},$$

where $x \in (0,\infty)$ denotes the discount rate parameter in a standard exponential, and $a, b > 0$. Thus $\alpha(x) = x$, $d(t;x) = e^{-xt}$, and $f(x)$ is as above. Then

$$\int_0^\infty \alpha(x)d(t;x)f(x)dx = \frac{a^b}{\Gamma(b)}\int_0^\infty x^b e^{-(a+t)x}dx = \frac{a^b}{\Gamma(b)}\frac{\Gamma(b+1)}{(a+t)^{b+1}} = \frac{b}{a}\left[\frac{a}{a+t}\right]^{b+1}.$$

But

$$\overline{\alpha} = \int_0^\infty \alpha(x)f(x)dx = \frac{a^b}{\Gamma(b)}\int_0^\infty x^b e^{-ax}dx = \frac{a^b}{\Gamma(b)}\frac{\Gamma(b+1)}{a^{b+1}} = \frac{b}{a},$$

so $d_{\overline{D}}(t) = [1+t/a]^{-(1+b)}$ as claimed. The speed follows from Theorem 3 with $r = b/a$ and $s = 1 - 1/b$.

Next, in order to allow for a positive weight at 0, we consider the following as the underlying distribution on the parameter:

$$f(x)=\frac{a^2}{a+b}e^{-ax}(1+bx).$$

We have $\alpha(x) = x$ and $d(t;x) = e^{-xt}$ as before, so

$$
\begin{aligned}
\int_0^\infty \alpha(x)d(t;x)f(x)dx &= \frac{a^2}{a+b}\int_0^\infty xe^{-(a+t)x}(1+bx)dx \\
&= \frac{a^2}{a+b}\left[\frac{1}{(a+t)^2}+\frac{2b}{(a+t)^3}\right] \\
&= \frac{a^2}{a+b}\left[\frac{a+2b+t}{(a+t)^3}\right]
\end{aligned}
$$

and

$$
\overline{\alpha} = \int_0^\infty \alpha(x)f(x)dx = \frac{a^2}{a+b}\int_0^\infty xe^{-ax}(1+bx)dx = \frac{a^2}{a+b}\left[\frac{1}{a^2}+\frac{2b}{a^3}\right] = \frac{a+2b}{a(a+b)}.
$$

Thus

$$
\begin{aligned}
d_{\overline{D}}(t) &= \frac{a^3}{a+2b}\left[\frac{a+2b+t}{(a+t)^3}\right] \\
&= \left(1+\frac{t}{a+2b}\right)\left[\frac{a}{a+t}\right]^3 \\
&= \left[\frac{a}{a+2b}(1+t/a)+\left(1-\frac{a}{a+2b}\right)\right]\left[\frac{1}{1+t/a}\right]^3 \\
&= \frac{a}{a+2b}\left[\frac{1}{1+t/a}\right]^2+\frac{2b}{a+2b}\left[\frac{1}{1+t/a}\right]^3,
\end{aligned}
$$

as claimed. Finally, the first of these weighted hyperbolics is [just barely] weakly convergent, with $r = 1/a$ and $s = 0$, so the overall procedure will have $\theta = \infty$ and $\rho = 0$.

References

[1] Arrow, Kenneth J. "Discounting, Morality and Gaming," in Portney, Paul R. and John P. Weyant, eds., *Discounting and intergenerational equity*. Washington D.C.: Resources for the Future, 1999. Pp.13-22.

[2] Axtell, Robert. "A General Theory of Discounting." Unpublished Paper, The Brookings Institution, 2003.

[3] Barlow, Richard E. and Proschan, Frank. *Mathematical theory of reliability*. New York: John Wiley and Sons, 1965. (Reprinted in 1996 by the Society for Industrial and Applied Mathematics, Philadelphia.)

[4] Bleichrodt, Han and Gafni, Amiram. "Time Preference, the Discounted Utility Model and Health." *Journal of Health Economics*, 1996, *15*, pp. 49-66.

[5] Bleichrodt, Han and Johannesson, Magnus. "Time Preference and Health." *Journal of Mathematical Psychology*, 2001, *45*, pp. 265-282.

[6] Cairns, John A. and van der Pol, Marjon M. "Saving Future Lives. A Comparison of Three Discounting Models," *Health Economics*, 1997, *6*, pp. 341-350.

[7] Chapman, Gretchen B. "Time Discounting of Health Outcomes," in Loewenstein, George, Read, Daniel, and Roy F. Baumeister, eds., *Time and decision.* New York: Russell Sage Foundation, 2003. Pp. 395-418.

[8] Cline, William R. "Discounting for the Very Long Term," in Portney, Paul R. and John P.Weyant, eds., *Discounting and intergenerational equity*. Washington, D.C.: Resources for the Future, 1999. Pp. 131-140.

[9] Cochrane, John H. *Asset pricing.* Princeton and Oxford: Princeton University Press, 2001.

[10] Cropper, Maureen L. Aydede, S.L. and Portney, Paul R. "Preferences for Life Saving Programs: How the Public Discounts Time and Age." *Journal of Risk and Uncertainty*, 1994, *8*, pp. 243-265.

[11] Fishburn, Peter C. and Rubinstein, Ariel. "Time Preference." *International Economic Review*, 1982, *23*, pp. 677-694.

[12] Frederick, Shane, Loewenstein, George, and O'Donoghue, Ted. "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature*, June 2002, *XL*, pp. 351-401.

[13] Gollier, Christian. "Discounting an Uncertain Future." *Journal of Public Economics, 2002, v. 85, 149-166.*

[14] Gollier, Christian and Zeckhauser, Richard. "Aggregation of Heterogeneous Time Preferences." *Journal of Political Economy, 2005, v. 113, 878-896.*

[15] Gradshteyn, I.S. and Ryzhik, I.M. *Tables of Integrals, Series, and Products.* San Diego: Academic Press, 1980.

[16] Groom, Ben, Hepburn, Cameron, Koundiri, Phoebe and Pearce, David. "Declining Discount Rates: The Long and the Short of it", *Environmental and Resource Economics, 2005, v. 32, 445-493.*

[17] Hara, Chiaki and Kuzmics, Christoph. "Representative Consumer's Risk Aversion and Efficient Risk-Sharing Rules." Unpublished paper, October 2002.

[18] Harris, Christopher and Laibson, David. "Instantaneous Gratification." Unpublished paper, May 2001.

[19] Harvey, Charles M. "Value Functions for Infinite Period Planning." *Management Science*, 1986, *32*, pp. 1123-39.

[20] Harvey, Charles M. "The Reasonableness of Non-Constant Discounting." *Journal of Public Economics*, 1994, *53*, pp. 31-51.

[21] Herrnstein, Richard J. "Self-Control as Response Strength," in Bradshaw, C.M., E. Szabadi and C.F. Lowe, eds., *Quantification of steady-state operant behavior.* Amsterdam: Elsevier/North-Holland, 1981.

[22] Jamison, Dean T. *Conjoint measurement of time preference and utility.* Rand Report RM6029-PR. Santa Monica, California: Rand Corporation, 1969.

[23] Keyfitz, Nathan. "Demography." In Eatwell, John, Milgate, Murray, and Newman, Peter, eds., *The new Palgrave: a dictionary of economics*, vol. 1. London: Macmillan, 1998. Pp. 796-802.

[24] Koopmans, Tjalling C. "Stationary Ordinal Utility and Impatience." *Econometrica*, April 1960, *28(2)*, pp. 287-309.

[25] Laibson, David. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*, May 1997, *112(2)*, pp. 443-477.

[26] Lewgwiler, Yvan. "Heterogeneous Patience and the Term Structure of Real Interest Rates", *American Economic Review, 2005, v. 95, 890-896.*

[27] Lipscomb, Joseph, Weinstein, Milton C. and Torrance, George W. "Time Preference," in Gold, Marthe R., Russell, Louise B., Siegal, Joanna A. and Milton C. Weinstein, eds., *Cost-effectiveness in health and medicine.* New York and Oxford: Oxford University Press, 1996. Pp. 214-246.

[28] Loewenstein, George. "The Fall and Rise of Psychological Explanations in the Economics of Intertemporal Choice," in Loewenstein, George and Jon Elster, eds., *Choice over time.* New York: Russell Sage Foundation, 1992. Pp. 3-34.

[29] Loewenstein, George and Prelec, Drazen. "Anomalies in Intertemporal Choice: Evidence and an Interpretation." *Quarterly Journal of Economics*, May 1992, *107(2)*, pp. 573-597.

[30] Malkiel, Burton G. "Term Structure of Interest Rates," in Eatwell, J., Milgate, M., and P. Newman, eds., *The new palgrave dictionary of economics, Vol. 4.* London: Macmillan Reference, Ltd., 1998. Pp. 629-631.

[31] Newell, Richard and Pizer, William. "Discounting the Benefits of Climate Change Mitigation: How much Do Uncertain Rates Increase Valuations?" *Journal of Environmental Economics and Management*, 2003, v. 46, 52-71.

[32] Oxera. 2002. *A Social Time Preference Rate for Use in Long-term Discounting.* Oxford: U.K.: Oxera.

[33] Poulos, Christine and Whittington, Dale. "Time Preference for Life-Saving Programs: Evidence from Six Less Developed Countries." *Environmental Science and Technology*, 2000, *34*, pp. 1445-1455.

[34] Ramsey, Frank P. "A Mathematical Theory of Saving." *The Economic Journal*, December 1928, *XXXVIII(152)*, pp. 543-559.

[35] Read, Daniel. "Is Time-Discounting Hyperbolic or Subadditive?" *Journal of Risk and Uncertainty*, 2001, *23(1)*, pp. 5-32.

[36] Roelofsma, Peter H.M.P. "Modelling Intertemporal Choices." *Acta Psychologica*, 1996, *93*, pp. 5-22.

[37] Rudin, Walter. *Principles of mathematical analysis*, Third Edition. New York: McGraw-Hill, 1976.

[38] Sokolnikoff, I.S. and Redheffer, Raymond A. *Mathematics of physics and modern engineering.* New York: McGraw-Hill, 1958.

[39] Stern, N. 2006. *Review of the Economics of Climate Change.* London: Her Majesty's Treasury.

[40] Strotz, Robert H. "Myopia and Inconsistency in Dynamic Utility Maximization." *Review of Economic Studies*, 1956, *23(3)*, pp. 165-180.

[41] Vaupel, James W., Manton, Kenneth G. and Stallard, Eric. "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality." *Demography*, August 1979, *16(3)*, pp. 439-454.

[42] Weitzman, Martin L. "Gamma Discounting." *American Economic Review*, March 2001, *91(1)*, pp. 260-271.

[43] Weitzman, Martin L. "Why the Far-Distant Future Should Be Discounted at Its Lowest Possible Rate." *Journal of Environmental Economics and Management*, Novemeber 1998, *36(3)*, pp. 201-208.