A theory of endogenous time preference, and discounted utility anomalies

Svetlana Boyarchenko [*] [1]

Sergei Levendorskiǐ[†]

[*] Department of Economics, The University of Texas at Austin, 1 University Station C3100, Austin, TX 78712-0301, U.S.A. (e-mail: sboyarch@eco.utexas.edu)
[†] Department of Economics, The University of Texas at Austin, 1 University Station C3100, Austin, TX 78712-0301, U.S.A. (e-mail: leven@eco.utexas.edu)

Key words: Time preference, discounted utility anomalies, decision-making under uncertainty, optimal stopping

JEL Classification: D81, D91, C61, G31

Abstract. We explain essentially all known discounted utility anomalies as artefacts of the optimizing behavior of an individual with a time-separable utility function, who perceives a good as a source of a stochastic consumption stream, and believes that she can wait for an optimal moment to buy or sell the good. For this individual, the fair price of the corresponding utility stream is interpreted as an integral of a deterministic utility stream multiplied by certain non-exponential factors which we interpret as *endogenous discount factors*; the factors are different for gains and losses, and depend on the utility function and underlying uncertainty. We provide analytic expressions and numerical examples for discount factors assuming simple utility functions and gaussian uncertainty.

1 Introduction

In the first three decades of the twentieth century, "time preference" was analyzed mainly qualitatively, as interaction among different factors which may influence intertemporal decisions. In 1933, Paul Samuelson invented the discounted utility theory (DU theory), which compressed the influence of many

---

[1] Corresponding author.
The first author is grateful to Efe A. Ok who brought the problem of the DU anomalies to her attention. We are thankful to Max Stinchcombe, Dale Stahl and Takashi Hayashi for discussion.

feasible factors into one number: the discount rate. In continuous time models, an individual with the time-separable utility u calculates the value of consumption of a stream c_t over time interval $(0, T)$ according to the formula

$$U = \int_0^T e^{-rt} u(c_t) dt, \tag{1.1}$$

where $r > 0$ is the discount rate. In discrete time models, the counterpart of equation (1.1) is

$$U = \sum_{t=0}^{T} \delta^t u(c_t) dt, \tag{1.2}$$

where $\delta \in (0, 1)$. Due to the analytical simplicity (and probably, similarity to the compound interest formula), the exponential discounted utility model was almost instantly adopted as a standard tool in intertemporal models, although Samuelson suggested the DU model as a convenient tool only, and explicitly disawoved an idea that individuals really optimize an integral of the form (1.1). Almost 30 years later, Koopmans (1960) constructed an axiomatic theory of time preference which lead to the exponential discount factor in Samuelson model. As a result, a general feeling emerged that the DU model was justified. However, later, in many empirical studies, it was shown that the real behavior of individuals did not agree with the exponential discounting model. We will consider the following anomalies of discounted utility model (*DU anomalies*):

1. *hyperbolic discounting*, which means that the instantaneous discount rate for gains *decreases* with time (in the DU model, it is constant);
2. the *sign effect* (gains are discounted more than losses);
3. the *delay-speedup* asymmetry: if the change of the delivery time of an outcome is perceived as an acceleration from a reference point, then the imputed discount rate is larger than if the change is perceived as a delay;
4. the *negative discounting for losses*: the individual prefers to expedite a payment;
5. the *magnitude effect* (small outcomes are discounted more than large ones);
6. *preference for improving sequences*.

All these anomalies (and some others) have been well-documented. For the discussion of DU anomalies and references to the literature on each type of DU anomalies, see the excellent review Frederick et al. (2002).

To account for these anomalies, several alternative (types of) models have been developed. In the (β, δ)- model of quasi-hyperbolic discounting introduced first by Phelps and Pollak (1968), equation (1.2) is replaced by

$$U = u(c_0) + \sum_{t=1}^{T} \beta \delta^t u(c_t) dt, \tag{1.3}$$

where $\beta, \delta \in (0, 1)$. Equation (1.3) is analytically simple, and captures many qualitative features of hyperbolic discounting. Thus, as in Samuelson (1933), the discount factors are postulated. Another strand of literature initiated by Koopmans (1960) deals with the axiomatic systems for time preferences, which are consistent with DU anomalies - see Ok and Masatlioglu (2003) and the bibliography therein. Fudenberg and Levine (2004) suggested a "dual-self" model as a unified explanation for several empirical regularities. Habit formation models, reference point models and a number of other models incorporate non-standard features into the utility function. Still other alternative models depart from the DU model even further (once again, we refer the reader to the review Frederick et al. (2002) for more details and extensive bibliography).

In this paper, we neither postulate the non-standard dependence of the discount factor on time as in the quasi-hyperbolic discounted utility models nor deduce it from some superficial axioms for time preference. Instead, we derive general explicit formulas for the discount factors for gains and losses from simple plausible general assumptions about the behavior of individuals who are not exposed much to the practice of modern financial markets, and demonstrate that these discount factors exhibit the DU anomalies. Our starting point is that the individual perceives the future – hence the utility of consumption – as uncertain. The uncertainty may be caused by both changes in the anticipated consumption level and utility function per se: certainly, the satisfaction from possession of a certain widget may change (and typically, changes) in a not completely predictable fashion. The importance of these factors is well-understood, and, as Frederick et al. (2002), p. 384, notice, once these and other confound factors are accounted for, there may be no place left for "pure time preference". The main contribution of our paper is the observation that additional simple plausible behavioral assumptions suffice to deduce the DU anomalies as results of the optimizing behavior of the individual in an uncertain environment. We also show that if the uncertainty is taken into account but the optimizing behavior postulated in our model is not, then it is difficult to reproduce the DU anomalies observed in empirical studies. We provide numerical examples for simple utility functions and the Brownian motion model for shocks, which demonstrate that our model gives relatively simple analytic expressions for discount factors, and reproduces almost all known discounted utility anomalies. In addition, we show that for some specifications of the utility function, the (β, δ)-model of the hyperbolic discounting for gains arises naturally as a discrete time approximation; for losses, a similar (β, δ)-model arises but with $\beta > 1$. For other specifications of the utility function, different simple continuous time approximations seem to be natural; they can be regarded as continuous time analogs of the (β, δ)-model.

To demonstrate that under the behavioral assumptions which we make, the optimizing behavior of the individual in an uncertain environment are sufficient for a consistent explanation of the DU anomalies (certainly, the other factors considered in the literature contribute to the DU anomalies as well), we make relatively standard assumptions about properties of the utility function and underlying uncertainty. In fact, the only departure from the standard utility theory which we have to make is the first basic assumption of the prospect theory. We assume that the individual assesses the utility of gains and disutility from losses using the present consumption level as the reference point. In other words, when making decisions concerning consumption c_t in the future, the individual considers $u(c_t - c_0)$, and not $u(c_t)$ as in the standard utility theory. However, in contrast to the prospect theory, we do not need to assume the *loss aversion* of u. To be more specific, we do not assume that the instantaneous utility function for gains, $u(c) = u_G$, is smaller than the (dis)utility function, $u_L(c) = -u(-c)$ for losses in order to explain the delay-speedup asymmetry and sign effects (cf. Loewenstein (1988) and Loewenstein and Prelec (1992) who used the loss aversion assumption). We also do not impose the restriction that the utility function over losses is convex, as in the prospect theory. In fact, we demonstrate all the effects for gains (u_G is assumed concave, as usual) but for losses, the magnitude effect is of the wrong sign if the utility function over losses is convex. A concave utility function over losses allows us to reproduce the magnitude effect of the correct sign.

The rest of the paper is organized as follows. In Section 2, we describe a type of rational behavior which naturally leads to DU anomalies. In Section 3, we formalize the model, and derive general analytic formulas for the bid and ask prices of perpetual streams. These formulas are obtained under standard assumptions about the optimizing behavior of individuals with the time-separable utility under uncertainty; in fact, they are new interpretations of general results for perpetual American options obtained in Boyarchenko (2004) and Boyarchenko and Levendorskiĭ (2002), (2004a-c). In Section 4, we derive formulas for the discount factors for gains and losses, and explain DU anomalies in the framework of our model. In Section 5, we compare our model to other possible *uncertainty-based* models of endogenous time preference, and show that the latter either do not demonstrate some of the effects, or demonstrate too large effects, or effects of the wrong sign. In Section 6, we discuss implications of our model which can be tested in empirical studies of time preferences. Section 7 concludes. Technical details are presented in the appendix.

2 Two types of rational economic behavior

2.1 Exponential discounted utility model as time preference for traders in the complete market

Although Samuelson (1933) never claimed that the discounted utility model should be a realistic model of the human behavior, some 40 years later there appeared a general theory, where the behavior of economic agents can be naturally interpreted as the use of the same exponential discount factor for all traded assets. This is the theory of efficient financial markets developed in the beginning of seventieth in a series of papers by Merton (see Merton (1973) and the collection of papers (1990)) and Black and Scholes (1973), and explained in terms "arbitrage-free" and "complete" in Harrison and Kreps (1979) and Harrison and

Pliska (1981)[2]. In the simplest variant of the theory, it is assumed that the agents in the market can borrow and lend money at the riskless constant rate $r > 0$. Then the absence of arbitrage implies that there exists a probability measure $\mathbb{Q}$ such that the value of a stochastic stream c_t is calculated as follows:

$$V = \int_0^T e^{-rt} E^{\mathbb{Q}} [c_t] dt; \tag{2.1}$$

in the discrete time case, the integral is replaced by summation. A measure $\mathbb{Q}$ is called a risk-neutral measure; if the market is complete, a risk-neutral measure is unique. This means that all agents in the market value all streams using the same formula (2.1). Formula (2.1) can be interpreted as the *expected discounted utility* of a risk-neutral agent. The only difference between the deterministic discounted utility model (1.1) and its stochastic analog (2.1) is the expectation operator under the integral sign in the latter formula.

If we assume that under $\mathbb{Q}$, the stream c_t is a martingale, that is, the expected value of c_t equals c_0, then we can write (2.1) as the discounted value of the deterministic stream $\tilde{c}_t = c_0$:

$$V = \int_0^T e^{-rt} \tilde{c}_t dt. \tag{2.2}$$

Thus, we recover Samuelson's discounted utility formula (1.1) for a risk-neutral individual. We may say that (1.1) is the formula for all traders in an arbitrage-free complete financial markets: should any of them price some assets not in the accordance with (2.1), then possibilities of arbitrage will arise, which will be exploited by arbitrageurs. Of course, real financial markets are not exactly arbitrage free and complete; however, they are sufficiently close to the idealized efficient market, and one cannot expect that deviations of real markets from the ideal one suffice to explain the well-documented DU anomalies. As the matter of fact, we strongly suspect that the DU anomalies will not be observed – or will be much less prominent – should the researchers conduct their experiments with a group of professional traders. The natural question is why ordinary people do not act as traders. Surely, the latter know better how to optimize, and maybe, the former should have understood that and learned from the latter?

2.2 Producers and consumers for the long run

> We do not understand many things not because our notions are weak but because these things do not belong to the circle of our notions.[3] *Thoughts and aphorisms from the fruits of meditation of Kozma Prutkov, Aphorism # 66*

[2]Recall that a financial market is arbitrage-free, if it is impossible to construct a portfolio of assets traded in the market offering "something for nothing." For a rigorous definition see, for example, Duffie (2001).

[3]Transl. from Russian by the authors

Among the host of possible natural objections, we want to stress two important properties of the optimizing behavior of agents in an arbitrage-free complete market. First, according to the theory, an agent must correctly infer the risk-neutral measure chosen by the market. Since the markets are, in fact, incomplete, although close to arbitrage-free ones, an agent must be a specialist in guessing a risk-neutral measure chosen by the market; we note that this is one of the most important types of activities real traders are engaged in. In other words, they must be specialists in guessing the winner of a beauty contest - which implies that they must be prone to the herd behavior[4]. We believe that a human who is not too much exposed to the theory and practice of financial markets can be expected to rely more on her personal opinion and perception of the value of a good. Second, if an agent has derivative securities in her portfolio, such as options, she *must trade continuously* (otherwise arbitrage opportunity will arise). In other words, she *must perceive assets* in the market as something *of no lasting personal value*, and be ready to buy or sell anything at any moment. Certainly, ordinary people in real life do not perceive many consumption goods as something that can be easily disposed of. The opposite of an asset which an agent holds for an infinitesimally small time period is a perpetual consumption stream, such as the one from a fruit tree which a villager planted hundreds or thousand years ago or a small field which he cultivated after cutting a patch of a forest, and never thought about selling it. Similarly, if he had to cut the tree or leave this small field, he regarded it as a loss of a perpetual stream of fruits from the tree or wheat that grew on the field. We conjecture that the perception of utility streams from consumption of durable good, imprinted in the human mind, remains fairly close to the perception of perpetual streams. The assumption that the individual thinks about consumption as something distributed over time is important. In Subsection 5.3, we construct a variant of the model with a non-durable good which has to be consumed immediately, and demonstrate that although this *instantaneous consumption model* exhibits the sign effect, asymmetry effect and hyperbolic effect, the sizes of these effects are unreasonably large. In addition, large losses are discounted more than small ones, both for convex and concave utility functions.

Even more crucial is the next assumption: when contemplating a purchase of a durable good, the individual assumes - consciously or subconsciously - that she can buy it at any moment (and will enjoy it forever) at a price which the individual considers as fair. Similarly, when contemplating losses, the individual presumes that she is free to choose the moment to give up a stream of utility from consumption of a good, and the loss will be permanent. In this case, the individual gets a compensation, which is fair from her point of view. Our two assumptions are extreme but they make the model below especially tractable. More realistic variant should impose some bounds on the periods of consumption

[4]Notice that even very sophisticated and clever people as George Soros may fail to feel accurately the degree of deviation of the behavior of the herd of professional traders from really rational behavior and lose a lot of money during the last years of the bubble, betting that quite soon the bubble will burst. One can say that Mr. Soros failed to infer properly the beliefs of the market which had remained enthusiastic for much longer time period than it should have had, in George Soros' reasonable opinion.

and deliberation or assume that the lengths of these time periods are random. However, the qualitative conclusions will not change after this refinement of the model, and if both time periods are large then the quantitative conclusions will change insignificantly.

3 Decision-making under uncertainty: gains vs. losses, and the natural bid-ask spread

In this section, we show that if an economic agent has a time-separable utility function with standard properties, perceives a perpetual stream of consumption as a monotone function of a stochastic process with i.i.d. increments, and believes that she can freely choose the moment when to buy or sell the right for the stream then the fair price (from the point of view of the agent) of the corresponding utility stream can be naturally interpreted as an integral of a deterministic utility stream with non-exponential discounted factors; the factors are different for gains and losses, and depend on the utility function and underlying uncertainty. For other models of the DU anomalies, which produce utility-dependent discount factors, see Section 7 in Frederick et al. (2002).

3.1 The fair exercise price as the expected present value of a utility stream vs. the real options approach

As the first step, assume that the individual has a simple utility function $u(c) = c^\gamma$ over gains, where $\gamma \in (0,1)$, and that the utility stream $\{c_t\}$ follows a geometric Brownian motion: $c_t = e^{X_t}$, where $X_t = X_0 + bt + \sigma W_t$, and W_t is the standard Brownian motion. Suppose that the individual perceives the stream as constant, on average: $E[c_t] = c_0$. Since $E[e^{X_t}] = e^{(b+\sigma^2/2)t+X_0}$, it must be that $b = -\sigma^2/2$. Assume further that the utility is time separable, and $r \ge 0$ is the exogenous discount factor (if $r = 0$, then there is no exogenous time preference whatsoever). Then the utility of consumption of the individual over time period $[0,T]$ is

$$\begin{aligned} U &= \int_0^T e^{-rt} E[u(c_t)]dt \\ &= \int_0^T e^{-rt} E[e^{\gamma X_t}]dt \\ &= \int_0^T e^{-t(r+q^c)} e^{\gamma X_0} dt \\ &= \int_0^T e^{-(r+q^c)t} u(c(X_0))dt, \end{aligned} \tag{3.1}$$

where $q^c = -\gamma b - \gamma^2\sigma^2/2 = \gamma(1-\gamma)\sigma^2/2 > 0$ is the endogenous correction to the exogenous discount rate. If the latter is 0, the former *is the endogenous discount rate*. A possible interpretation of formula (3.1) is Samuelson's exponential discounted utility formula (1.1). If the individual can express her utility in dollars, she should be willing to pay up to U dollars for the consumption stream $\{c_t\}$, and ask at least U dollars as the compensation for the loss of the same stream.

We see that the naive use of (3.1) as a fair price, at which the individual should be willing to buy or sell the right for the stream, does not lead to the hyperbolic discounting. The reason is that this naive exercise rule is incorrect for the individual who is under no obligation to buy or sell the right for the consumption stream $\{c_t\}$ instantly, and can wait for a more favorable realization of the underlying uncertainty. In other words, there is the *option value for waiting*. This fact is well-known in the theory of financial and real options, and more generally, in the optimal stopping theory. If the individual can wait forever, the right to buy or sell a good may be regarded as a perpetual option. The pricing formula in this case is especially tractable. Also analytically tractable but more involved is the case of an option which can be exercised any moment up to the maturity date T, provided T is an exponential random variable independent of the underlying stochastic process. In this paper, we confine ourselves to the simple case of perpetual options, that is, $T = +\infty$.

3.2 The model

We make the following assumptions.

A1. *The consumption level at time $t > 0$, as perceived by the individual at time $t = 0$, is $c_t = c(X_t)$, where c is an increasing function, and $\{X_t\}$ is a process with i.i.d. increments.*

Thus, in the discrete time model, X_t is a random walk, and in the continuous time model, X_t is a Brownian motion or more generally, Lévy process. If $c(X_t) = e^{X_t}$, and X_t is a Brownian motion, we obtain the geometric Brownian motion model. One may argue that a geometric mean-reverting process would be preferable. Notice, however, that with an appropriate choice of a functional dependence of c_t on X_t, processes with mean-reverting features can be obtained (see Boyarchenko and Levendorskiǐ (2004a-c)).

A2G. *The utility over gains is time-separable, and the instantaneous utility $u_G(c) = u(c)$ for gains is monotone and concave.*

A2L. *The utility over losses is time-separable, and the instantaneous utility $u(c) = -u_L(-c)$ for losses is monotone.*

We will analyze how the DU anomalies change if we replace the concavity of the instantaneous utility over losses by convexity (equivalently, the convexity of u_L by concavity), as in the prospect theory.

A3. *The individual discounts the future at the constant rate $r > 0$.* To be more specific, her *ex post* evaluation of utility from consumption of a perpetual stream $\{c_t\}$ is

$$U(u_G \circ c; x) = \int_0^\infty e^{-rt} E[u_G(c(X_t))]dt, \tag{3.2}$$

where x is the level of the stochastic factor X_t at time 0: $X_0 = x$. Similarly, while evaluating the *ex post* disutility from the loss of consumption of a perpetual stream $\{c_t\}$ she uses (3.2) with u_L instead of u_G. It will become clear later on in the paper that *ex post* and *ex ante* evaluations of utility streams are different.

We cannot avoid the use of a small background discount factor (killing rate) in the model with perpetual consumption streams, and perpetual options to acquire or sell the stream. (In a more involved version of the model with exponentially distributed lengths of consumption and deliberation periods, the background factor is unnecessary). However, as we will see below, arbitrary small $r > 0$ leads to the same qualitative conclusions (unless the concavity of the utility function over losses is significant; in this case, the background discount factor should be significant as well in order to avoid the feeling of an infinite loss). Moreover, we will see that in the case of gains, under natural assumptions, it will be possible to pass to the limit $r \to 0$, and get rid of the exogenous component of the discount rate altogether; for losses, this is also possible but only under more stringent assumptions about the utility function and consumption flow. In any case, some exogenous killing rate seems to be natural [5].

In the literature on DU anomalies, it is implicit that

(i) utility of gains and disutility of losses have dollar equivalents;

(ii) the individual agrees to pay K dollars for a stream $\{c_t\}$ iff $K \le U(u_G \circ c; x)$;

(iii) the individual agrees to lose a stream $\{c_t\}$ for a compensation of K dollars iff $K \ge U(u_L \circ c; x)$.

In numerous experimental studies (typically, under an assumption that the individual perceives a consumption stream as deterministic), it is shown that the rules (ii) and (iii) do not work; in view of the real option approach, this was to be expected.

Our aim is to show that the majority of the DU anomalies can be naturally explained if we make the following crucial assumption:

A4. *While contemplating the decision to pay K dollars for a consumption stream, or accept K dollars for the loss of a consumption stream, the individual regards this possibility of exchange as a perpetual option which can be exercised any moment, and she presumes that she has a menu of options with different strikes to choose from.*

Thus, the individual contemplating gains regards herself as a holder of a perpetual American call option with the payoff $U(u_G \circ c; x) - K$. In the standard option theory, the strike K is given, and the individual chooses a moment to exercise when the spot level x becomes sufficiently high. In the context of experiments dealing with DU anomalies, the individual has to make a decision now. In effect, she is told to choose the *fair strike price*, or the highest strike price, which makes a call option optimal to exercise at the current level of the

[5] As Voland put it: "Of course man is mortal, but that's only half the problem. The trouble is that mortality sometimes comes to him so suddenly! And he cannot even say what he will be doing this evening." M. Bulgakov, *The Master and Margarita*, transl. by Michael Glenny, 1967, Hamper and Row, New York

stochastic factor. Similarly, the individual contemplating losses regards herself as a holder of a perpetual American put option with the payoff $K - U(u_L \circ c; x)$, and she exercises the option now if she perceives the strike price K as fair.

3.3 Bad news and good news principles, or why it may be optimal to keep old golf clubs in a closet

The optimal exercise rules for perpetual American call and put options with perpetual streams as payoffs were obtained in Boyarchenko (2004) for the case $u(c(X_t)) = e^{X_t}$, where X_t is a Lévy process; the generalization for $u(c(X_t)) = e^{\gamma X_t}$, where $\gamma > 0$, is straightforward. In Boyarchenko and Levendorskiĭ (2004a-c), the results were extended to arbitrary increasing u in continuous time, discrete time – continuous state space and discrete time – discrete state space models. Notice that it was Bernanke (1983) who spelt out the "bad news principle" for the first time (in a different set-up).

Bad news principle. Given the strike price K, exercise the perpetual American call option on a stream $u(c(X_t))$ the first time the expected present value of a stream $u \circ c$ calculated for the infimum process $\underline{X}_t = \inf_{0 \le s \le t} X_s$ in place of X_t, equals or exceeds K:

$$U^-(u \circ c; x) := E^x \left[\int_0^\infty e^{-rt} u(c(\underline{X}_t)) dt \right] \ge K \tag{3.3}$$

(for a Brownian motion, the rule involves "=" instead of $\ge$). Equivalently, given x, the price K is fair iff (3.3) holds with equality.

Good news principle. Given the strike price K, exercise the perpetual American put option on a stream $u(c(X_t))$ the first time the expected present value of a stream $u \circ c$ calculated for the supremum process $\bar{X}_t = \sup_{0 \le s \le t} X_s$ in place of X_t, equals or drops below K:

$$U^+(u \circ c; x) := E^x \left[\int_0^\infty e^{-rt} u(c(\bar{X}_t)) dt \right] \le K \tag{3.4}$$

(for a Brownian motion, the rule involves "=" instead of $\le$). Equivalently, given x, the price K is fair iff (3.4) holds with equality.

The bad and good news principles give the following formulas for the fair *ex ante* value of a stream of gains

$$G(u_G \circ c; x) = E^x \left[\int_0^\infty e^{-rt} u_G(c(\underline{X}_t)) dt \right], \tag{3.5}$$

and losses

$$L(u_L \circ c; x) = E^x \left[\int_0^\infty e^{-rt} u_L(c(\bar{X}_t)) dt \right]. \tag{3.6}$$

Before proceeding further, we give the following interpretation of *ex ante* pricing formulas (3.5)–(3.6). When the individual plans to purchase the good, she is pessimistic in her anticipation of changes in the relative quality of a durable good or her tastes, and disregards all possible *temporary* upward movements. On the other hand, when assessing the disutility of losses, the individual is too optimistic, and mainly takes into account possible increases of her perception of the quality of the good (maybe, I will use my old golf clubs the next year although I do not want to do it now). Notice that we do not make these plausible behavioral assumptions; we *deduce them* from assumptions A1–A4.

3.4 Natural bid-ask spread

We assume that functions $u = u(c)$ and $c = c(x)$ are monotone, and

$$u_L(c) := -u(-c) \geq u_G(c) := u(c), \ \forall \ c > 0, \tag{3.7}$$

which is a natural assumption for a concave u. It may also hold for utility functions that are concave over gains and convex over losses, as in the prospect theory. Then the fair value of gains is less than the fair value of losses, and we observe a natural *bid-ask spread* for any reasonable model for shocks, even in the case of the same u for utility of gains and disutility of losses. Indeed, the trajectories of the supremum process are not lower than the ones of the infimum process, and if the process is not trivial, they are higher: $\underline{X}_t < \bar{X}_t$ a.s., and therefore, even for $u_G = u_L$, the RHS in (3.5) is less than the RHS in (3.6).

4 Discounted utility anomalies

4.1 Discount factors for gains and losses

In the majority of empirical studies of DU anomalies, it is assumed that gains and losses are deterministic; in the framework of our paper, this means that $c(X_t)$ is deterministic and constant. In an uncertain environment, the closest analog is a martingale:

$$E^x[c(X_t)] = c(x), \quad \forall \ t \geq 0. \tag{4.1}$$

Assume (4.1), and write (3.5)–(3.6) as

$$G(u_G \circ c; x) = \int_0^\infty e^{-rt} D^c_G(x,t) u_G(c(x)) dt, \tag{4.2}$$

and

$$L(u_L \circ c; x) = \int_0^\infty e^{-rt} D^c_L(x,t) u_L(c(x)) dt, \tag{4.3}$$

where

$$D^c_G(x,t) = \frac{E^x[u_G(c(\underline{X}_t))]}{u_G(c(x))} \quad \text{and} \quad D^c_L(x,t) = \frac{E^x[u_L(c(\bar{X}_t))]}{u_L(c(x))} \tag{4.4}$$

are correction factors to the background exponential discount factor e^{-rt}. We can regard

$$D_G(x,t) := e^{-rt} D_G^c(x,t) \quad \text{and} \quad D_L(x,t) := e^{-rt} D_L^c(x,t) \tag{4.5}$$

as the (total) discount factors for gains and losses. Generally, both depend not only on time as in the exponential discounted utility model and hyperbolic model but on the utility function and current level of the stochastic factor $x = X_0$ (equivalently, on the current consumption level $c_0 = c(X_0)$) as well. Notice that although we derived the formulas for the discount factors for evaluation of perpetual streams of losses and gains, we will use them to evaluate streams over finite time intervals. More consistent approach would be to derive the formulas for the discount factors for each situation but this would have led to more involved formulas with essentially the same qualitative properties. Another motivation is: certainly, it is hard to imagine that the individual can evaluate all possible "real life options" accurately; we believe that it is more realistic to assume that the human brain has certain general matrices for time preference which she uses in all situations.

4.2 Analysis of DU anomalies

We claim that the discount rates $D_G(x,t)$ and $D_L(x,t)$ enjoy properties 1-6 (DU anomalies) listed in Introduction. The properties 2 and 3 do not require any specific assumption about the process X_t and functions u and c apart from monotonicity and (3.7); the others will be demonstrated for u and c satisfying certain natural assumptions on u and c, and for a simplest model of a process with i.i.d. increments: a Brownian motion.

4.2.1 Gains are more discounted than losses.

Under our standing assumption (3.7), we obtain from (3.5) and (3.6) the inequality

$$D_G(x,t) < D_L(x,t), \quad \forall\ x \text{ and } t > 0. \tag{4.6}$$

4.2.2 The "Delay-Speedup" Asymmetry

If the individual is asked whether she is willing to delay the delivery of a widget (say, a CD–player) and receive it at date T instead of the present date, she evaluates the expected value of utility which she will not receive during the period $[0,T]$, that is, she evaluates the disutility of losses (using the discount factor for losses). Hence, the fair value which she asks as a compensation is

$$K_{\text{del}} = \int_0^T D_L(x,t) u_L(c(x)) dt. \tag{4.7}$$

Similarly, if she expects the delivery at time $T > 0$ then the instant delivery provides an additional utility stream over the period $[0,T]$, which she discounts using the discount factor $D_G(x,t)$. The fair price which she is willing to pay for the instant delivery is

$$K_{\text{exp}} = \int_0^T D_G(x,t) u_G(c(x)) dt. \tag{4.8}$$

Inequalities (4.6) and (3.7) taken together imply that the RHS in (4.8) is less than the RHS in(4.7): $K_{\text{exp}} < K_{\text{del}}$.

In some studies, a similar effect is demonstrated for losses: respondents demanded more to expedite payment (due at time $t > 0$) than they would pay to delay it. We can deduce this effect from inequalities (4.6) and (3.7) as follows. If the individual agrees to expedite a payment, she will not lose the consumption chunk c_t, which she evaluates as gains: $u_G(c(X_t))$. She discounts this utility using the discount factor $D_G(x,t)$. In return, she suffers the disutility $u_L(c(x))$, and the balance is $-u_L(c(x)) + D_G(x,t)u_G(c(x))$. Hence the individual should demand the compensation

$$K_{\text{exp}} = u_L(c(x)) - D_G(x,t)u_G(c(x)). \tag{4.9}$$

On the other hand, if the individual can delay the payment, she enjoys the utility $u_G(c(x))$ now but suffers the disutility $u_L(c(X_t))$ later, at time t. The latter is discounted using the factor $D_L(x,t)$, and the balance (the fair price which the individual should be willing to pay for the delay) is

$$K_{\text{del}} = u_G(c(x)) - D_L(x,t)u_L(c(x)). \tag{4.10}$$

Since $D_L(x,t) > D_G(x,t)$, and $u_L \geq u_G$, the difference

$$K_{\text{exp}} - K_{\text{del}} = u_L(c(x))(1 + D_L(x,t)) - u_G(c(x))(1 + D_G(x,t))$$

is positive, which agrees with empirical studies.

4.2.3 Hyperbolic discounting: A model example for gains

Consider the instantaneous utility function for gains of a simple form $u_G(c) = c^\gamma$, where $\gamma \in (0,1)$, and assume that $c(X_t) = e^{X_t}$, where $X_t = X_0 + bt + \sigma W_t$ is the Brownian motion with drift. The shape of the curve $u(c)$ in the neighborhood of 0 is non-standard but for this oversimplified functional form explicit analytical results are available. In Subsubsection 4.2.5, we will deduce analytical formulas for a general u but these formulas will involve integration. The results presented below agree with the numerical examples for the utility function $u_G(c) = (1 + c)^\gamma - 1$, and if the current consumption level $c(x) = e^x$ is not small, then even quantitative differences are small.

As above, we assume that $c(X_t)$ is a martingale. Since for a Brownian motion, $E^x[e^{X_t}] = e^{x+(b+\sigma^2/2)t}$, it must be that $b = -\sigma^2/2$. Set $\rho = \gamma\sigma$ and $\mu = -\sigma/2$, and represent γX_t in the form $\gamma X_t = \gamma X_0 + \rho Y_t$, where $Y_t = \mu t + W_t$. Now,

$$D_G^c(x,t) = e^{-\gamma x} E^x[e^{\gamma \underline{X}_t}] = E[e^{\rho \underline{Y}_t}],$$

and we can use equation 1.2.3 on p. 251 in Borodin and Salminen (2002):

$$E[e^{\rho \underline{Y}_t}] = \frac{\rho+\mu}{\rho+2\mu} e^{\rho(\rho+2\mu)t/2} \mathrm{Erfc}\left(\frac{(\rho+\mu)\sqrt{t}}{\sqrt{2}}\right) + \frac{\mu}{\rho+2\mu} \mathrm{Erfc}\left(-\frac{\mu\sqrt{t}}{\sqrt{2}}\right), \quad (4.11)$$

where Erfc is the complementary error function

$$\mathrm{Erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{+\infty} e^{-v^2} dv. \quad (4.12)$$

This special function is tabulated and can be found in any standard package (and it can be easily expressed in terms of the cumulative normal distribution function). Substituting $\mu = -\sigma/2$ and $\rho = \gamma\sigma$ into (4.11), we obtain the correction factor $D^c_G(x,t)$ for gains (to the background discount factor e^{-rt}) in terms of $\gamma \in (0,1)$ and σ:

$$D^c_G(x,t) = \frac{0.5-\gamma}{1-\gamma} e^{\gamma(\gamma-1)\sigma^2 t/2} \mathrm{Erfc}\left(\frac{2\gamma-1}{2}\sqrt{\frac{\sigma^2 t}{2}}\right) + \frac{0.5}{1-\gamma} \mathrm{Erfc}\left(\frac{1}{2}\sqrt{\frac{\sigma^2 t}{2}}\right). \quad (4.13)$$

Using (4.13), we can easily calculate $D^c_G(x,t)$ and the correction term $q^c_G(x,t) = -\log D^c_G(x,t)/t$ to the background discount rate r. In Fig. 1 and 2, it is clearly seen that $q^c_G(x,t)$ decreases with t, and it drops especially fast in a neighborhood of 0. Since the discount rate

$$q_G(x,t) = -\log D(x,t)/t = r + q^c_G(x,t), \quad (4.14)$$

we conclude that q_G enjoys the same properties for any value of the background discount rate. Hence, we observe the hyperbolic discounting effect. Notice that for a homogeneous utility function $u_G(c) = c^\gamma$ the discount rate is independent of x; this will not be the case for inhomogeneous utility functions. We also see that the discount rates increase with σ and γ.

Although it is difficult to conduct a thorough analytical study of properties of the correction factor $D^c_G(x,t)$, it is relatively easy to study its behavior for large t, and the results are rather interesting. In the appendix, we show that

a) if $\gamma \in (0, 0.5)$, then as $t \to +\infty$,

$$D_G(x,t) = \beta\delta^t + \cdots, \quad (4.15)$$

where $\beta := (1-2\gamma)/(1-\gamma) \in (0,1)$ and $\delta := e^{-(r+\gamma(1-\gamma)\sigma^2/2)} \in (0,1)$. Using (4.14), we obtain

$$q_G(x,t) = r + \frac{\gamma(1-\gamma)\sigma^2}{2} + \frac{-\log\beta}{t} + \cdots. \quad (4.16)$$

(In (4.15)–(4.16) and in approximate equations below, the dots stand for the terms which decay faster than the leading (shown) term(s).) Since for small t, $D_G(x,t)$ is closer to 1 than to β, we observe the behavior which is strikingly similar to the (β,δ)-model of hyperbolic discounting (see (1.3)). Notice also that for these γ, (4.15) can be regarded as a hyperbolic correction to the naive stochastic discounting rule (3.1).

b) if $\gamma = 0.5$, then

$$D_G(x,t) = \frac{4}{\sqrt{2\pi\sigma^2 t}} e^{-t(r+\sigma^2/8)} + \cdots , \tag{4.17}$$

and therefore from (4.14),

$$q_G(x,t) = r + \frac{\sigma^2}{8} + \frac{\log t}{2t} + \cdots ; \tag{4.18}$$

c) if $\gamma \in (0.5, 1)$, then

$$D_G(x,t) = \frac{8}{\sqrt{2\pi}(\sigma^2 t)^{3/2}} e^{-t(r+\sigma^2/8)} + \cdots , \tag{4.19}$$

and therefore from (4.14),

$$q_G(x,t) = r + \frac{\sigma^2}{8} + \frac{3\log t}{2t} + \cdots . \tag{4.20}$$

It is clearly seen that the discount rate decreases with t, and the leading terms of $q_G(x,t)$ for large t is independent of $\gamma \in (0.5, 1)$. In other words, *for utility functions with a small or moderate coefficient of the relative risk aversion,* $1-\gamma \in (0, 0.5)$*, the discount rate is hyperbolic and universal in the sense that it is essentially independent of the utility function (for large* t*, at least).* The final two remarks are: first, the correction term to the discount rate increases in σ^2, and second, it remains positive even in the limit $r \to 0$ (that is, when the background discount rate vanishes). Hence, we can use our stochastic endogenous discounting model for gains without resorting to the background exogenous discount factor, and the discounting model becomes purely endogenous.

4.2.4 Negative discounting: A model example for losses

Consider the disutility function for losses of the same form $u_L(c) = c^\gamma$. This time, we allow both $\gamma < 1$ and $\gamma > 1$. We have

$$D^c_L(x,t) = e^{-\gamma x} E^x[e^{\gamma \bar{X}_t}] = E[e^{\rho \bar{Y}_t}],$$

and we can use equation 1.1.3 on p. 250 in Borodin and Salminen (2002):

$$E[e^{\rho \bar{Y}_t}] = \frac{\rho+\mu}{\rho+2\mu} e^{\rho(\rho+2\mu)t/2} \mathrm{Erfc}\left(-\frac{(\rho+\mu)\sqrt{t}}{\sqrt{2}}\right) + \frac{\mu}{\rho+2\mu} \mathrm{Erfc}\left(\frac{\mu\sqrt{t}}{\sqrt{2}}\right). \tag{4.21}$$

Inserting $\rho = \gamma\sigma$ and $\mu = -\sigma/2$, we obtain

$$D^c_L(x,t) = \frac{0.5-\gamma}{1-\gamma} e^{\gamma(\gamma-1)\sigma^2 t/2} \mathrm{Erfc}\left(\frac{1-2\gamma}{2}\sqrt{\frac{\sigma^2 t}{2}}\right) + \frac{0.5}{1-\gamma} \mathrm{Erfc}\left(-\frac{1}{2}\sqrt{\frac{\sigma^2 t}{2}}\right) \tag{4.22}$$

Using (4.22), we can easily calculate $D^c_L(x,t)$ and the correction term $q^c_L(x,t) = -\log D^c_L(x,t)/t$ to the background discount rate r. Although (4.22) is similar to (4.13), the qualitative behavior of the discount factors for gains and losses are strikingly different. First of all, for small $t > 0$, the correction factor is greater than 1:

$$D^c_L(x,t) = 1 + \frac{2\gamma\sigma}{\sqrt{2\pi}}\sqrt{t} + \cdots \tag{4.23}$$

(for the proof of (4.23) and the statements below, see the appendix), therefore, for sufficiently small $t > 0$, the negative discounting will be observed always:

$$q_L(x,t) = r - \frac{\log(1 + 2\gamma\sigma\sqrt{t} + \cdots)/\sqrt{2\pi}}{t} = -\frac{2\gamma\sigma}{\sqrt{2\pi}}t^{-1/2} + \cdots < 0.$$

As $t \to +\infty$, the correction factor $D^c_L(x,t)$ behaves as follows.

If $\gamma \in (0,1)$, then

$$D^c_L(x,t) = 1/(1-\gamma) + \cdots, \tag{4.24}$$

and therefore, any positive background discount rate suffices to ensure that the discount rate for losses becomes negative for large t. In particular, for any $r > 0$, and large t, we obtain an approximate formula

$$D_L(x,t) = \beta\delta^t, \tag{4.25}$$

where $\beta = 1/(1-\gamma) > 1$, and $\delta = e^{-r} \in (0,1)$.

In the case $\gamma = 1$, $D^c_L(x,t)$ grows slower than any exponential function, and therefore, as in the case $\gamma \in (0,1)$, any positive background discount rate suffices to ensure that the discount rate for losses becomes negative for large t.

Finally, if $\gamma > 1$, then

$$D^c_L(x,t) = \frac{2\gamma - 1}{\gamma - 1}e^{t\gamma(\gamma-1)\sigma^2/2} + \cdots. \tag{4.26}$$

Hence, to avoid a feeling of the infinite loss, the background discount factor must be not too small: $r > \gamma(\gamma-1)\sigma^2/2$. Then, for large t, we obtain an approximate formula (4.25) with $\beta = (2\gamma-1)/(\gamma-1) > 1$ and $\delta = e^{-r+\gamma(\gamma-1)\sigma^2/2} \in (0,1)$.

We conclude that for modelling of the non-standard discounting for losses, the (β,δ) model (1.3) might be appropriate *but with* $\beta > 1$ *and* $\delta \in (0,1)$ instead of $\beta \in (0,1)$ and $\delta \in (0,1)$. In Fig. 3 and 4, we plot, for several values of γ and σ, correction factors and rates, and (total) discount factors and rates for losses.

4.2.5 Inhomogenous utility functions and magnitude effect

A homogeneous utility function $u(c) = c^\gamma$ is not quite appropriate because it exhibits incorrect behavior near 0; in addition, the discount factors are independent of x, hence, of the consumption level $c(x)$. Thus, the magnitude effect (smaller discount factors for larger gains/losses) cannot be reproduced.

Take an inhomogeneous u, say, $u(c) = (1+c)^\gamma - 1$. Then we can calculate $E^x[u(c(\bar{X}_t))]$ and $E^x[u(c(\underline{X}_t))]$ using the explicit formulas for the cumulative distribution functions of $\bar{Y}_t$ and $\underline{Y}_t$ (formulas 1.2.4 and 1.1.4 on p.251 and p.250 in Borodin and Salminen (2002)): for $y > 0$,

$$\mathbb{P}(\bar{Y}_t \le y) = 1 - \frac{1}{2}\mathrm{Erfc}\left(\frac{y - \mu t}{\sqrt{2t}}\right) - \frac{1}{2}\mathrm{Erfc}\left(\frac{y + \mu t}{\sqrt{2t}}\right), \tag{4.27}$$

and for $y < 0$,

$$\mathbb{P}(\underline{Y}_t \le y) = \frac{1}{2}\mathrm{Erfc}\left(-\frac{y - \mu t}{\sqrt{2t}}\right) - \frac{1}{2}\mathrm{Erfc}\left(-\frac{y + \mu t}{\sqrt{2t}}\right), \tag{4.28}$$

where $\mu = -\sigma/2$. We have

$$D_L^c(x,t) = \int_0^{+\infty} \frac{u_L(c(x+\sigma y))}{u_L(c(x))} d\mathbb{P}(\bar{Y}_t \le y), \tag{4.29}$$

and

$$D_G^c(x,t) = \int_{-\infty}^{0} \frac{u_G(c(x+\sigma y))}{u_G(c(x))} d\mathbb{P}(\underline{Y}_t \le y). \tag{4.30}$$

The Riemann-Stiltjes integrals on the RHS's are reduced to the Riemann integrals using the following formulas:

$$\frac{d\mathbb{P}(\bar{Y}_t \le y)}{dy} = \sqrt{\frac{2}{\pi t}} e^{-(y-\mu t)^2/(2t)} - \mu e^{2\mu y} \cdot \mathrm{Erfc}\left(\frac{y + \mu t}{\sqrt{2t}}\right),$$

and

$$\frac{d\mathbb{P}(\underline{Y}_t \le y)}{dy} = \sqrt{\frac{2}{\pi t}} e^{-(y-\mu t)^2/(2t)} + \mu e^{2\mu y} \cdot \mathrm{Erfc}\left(-\frac{y + \mu t}{\sqrt{2t}}\right).$$

In Fig.5, we plot correction factors and rates to the background discount factor and rate, and total discount factors and rates, for several values of the current consumption level c_0. The magnitude effect is clearly seen. In Fig.6, we plot correction factors and terms for losses, for several values of the current consumption level c_0. In the left panel, $\gamma = 0.75 < 1$, which means that the utility function over losses is convex. This non-standard property (assumed in the prospect theory) leads to the magnitude effect of the wrong sign: the discount factors increase with the consumption level. In the right panel, u_L is convex, hence the utility function is concave, and the magnitude effect of the correct sign is observed.

4.2.6 Preference for improving sequences

For simplicity, we start with the deterministic discrete time model. Assume that the individual regards each drop in the consumption level as a loss. Then she calculates the utility from consumption of a decreasing consumption sequence $c_0 > c_1 > c_2 > \cdots > c_n$ (consumed at $t = t_0, t_1, \ldots, t_n$) as follows:

$$\begin{aligned} U = D_G(t_0)u_G(c_0) &+ \{D_G(t_1)u_G(c_0) - D_L(t_1)u_L(c_0 - c_1)\} \\ &+ \{D_G(t_2)u_G(c_1) - D_L(t_2)u_L(c_1 - c_2)\} \\ &+ \cdots \\ &+ \{D_G(t_n)u_G(c_{n-1}) - D_L(t_n)u_L(c_{n-1} - c_n)\}, \end{aligned}$$

which is

$$U = D_G(t_0)u_G(c_0) + \sum_{j=1}^{n} D_G(t_j)u_G(c_{j-1}) - \sum_{j=1}^{n} D_L(t_j)u(c_{j-1} - c_j). \quad (4.31)$$

Assume that u_L is differentiable at 0. Passing to the limit $\max_j(t_{j+1} - t_j) \to 0$, we obtain

$$U = \int_0^T D_G(t)u_G(c_t)dt - (u_L)'(0) \int_0^T D_L(t)d(-c_t). \quad (4.32)$$

A natural stochastic analog of (4.32) is possible only if the supremum process for $c(X_t)$ is of bounded variation

$$U(x) = \int_0^T e^{-rt}E^x[u_G(c(\underline{X}_t))]dt - (u_L)'(0) \int_0^T e^{-rt}E^x[d(-c(\bar{X}_t))]. \quad (4.33)$$

Otherwise, the negative term on the RHS becomes infinite. If the supremum process for $c(X_t)$ is of unbounded variation as in the geometric Brownian motion model for consumption, then we cannot pass to the continuous time limit, and the smaller the time step between the bursts of consumption, the larger (in absolute value) the disutility from perceived losses is. This observation alone explains the preference for improving sequences, and we will not produce numerical examples (which require more complicated processes or discrete time modelling).

5 Comparison with other uncertainty-based models

To demonstrate that all components of our approach are important for the explanation of DU anomalies, we consider two simpler standard ways to account for the uncertainty of the future, and then a variant of our model with instantaneous consumption.

5.1 Discrete time model, i.i.d. draws

A rather popular way to model the uncertain consumption in the future is to consider the discrete time model, with the consumption in each time period modeled as an i.i.d. draw from a given distribution. Assume that time periods are of length $\Delta > 0$, and the individual's discount factor per period is $\delta > 0$. As above, the expected value of consumption c_t at time $t = j \cdot \Delta$, $j = 0, 1, \ldots$, equals the current consumption level: $c_0 = E[c_t]$. If u is concave, then $\beta := E[u(c_t)]/u(c_0) < 1$ by Jensen's inequality, and we recover the (β, δ)-model, which is popular as an approximation to the hyperbolic discounting for gains. If we use a convex utility function over losses, as in the prospect theory, we obtain the same model, and there will be no negative discounting observed. In addition, if $u_L = u_G$, then there will be neither sign effect nor asymmetry effect, whereas our model produces both effects even in this case. The negative discounting for losses will be observed for *small positive* t if the utility over losses is concave, and either uncertainty is sufficiently large or the exogenous background discount rate is sufficiently small. Indeed, for a concave disutility function u_L, we obtain $\beta := E[u(c_t)]/u(c_0) > 1$.

Notice, however, that the δ is independent of the uncertainty, and therefore, to account for fairly large values of observed discounting rate for gains, it is necessary to presume that the exogenous background discount rate is rather large. In our model, it is possible to obtain a sizable discount rate for gains even for arbitrary small levels of the exogenous discount rate. Further, the magnitude effect is observed but its size is several times smaller than in our model for the same instantaneous utility function and uncertainty modeled as $c_1 = e^{X_\Delta}$, where $\Delta < 0.5$ and X_t is the Brownian motion with the drift used in our model. Notice that $\Delta < 0.5$ means a reasonable time period less than half a year.

5.2 Naive endogenous discounting revisited

In Subsection 3.1, we showed that for the simplest instantaneous utility $u(c) = c^\gamma$ and the geometric Brownian motion model for consumption stream $\{c_t\}$, the constant discount rate result, and we remain in the realm of the exponential discounting. One might have hoped that the use of more natural utility/disutility functions as in Subsubsection 4.2.5 will produce the hyperbolic effect. Unfortunately, this is not the case, and the resulting curves are essentially straight and dull. Unlike the case of i.i.d. draws, some corrections to the discount rate can be obtained but they are much smaller than in our model where the optimizing behavior of the individual is taken into account. The size of the magnitude effect is also smaller, and the sign and asymmetry effects cannot be reproduced for $u_L = u_G$, because for the same instantaneous utility function and process, the discount factors for gains and losses are the same.

5.3 A version of the model with optimizing agents, for instantaneous payoffs

Assume that the individual contemplates consumption of a non-storable good with the delivery (hence, consumption) date $t > 0$, and the uncertain consumption value $c_t = e^{X_t}$. Assume further that $u_G(c) = c^\gamma$, where $\gamma > 0$. Then $u(c(X_t)) = e^{\gamma X_t}$, and we can represent the expected discounted utility of the instantaneous consumption as $G(x) := e^{-rt}E^x[e^{\gamma X_t}] = D_0(t)e^{\gamma x}$, where $D_0(t) = e^{-(r+\gamma(1-\gamma)\sigma^2/2)t}$. For a call option with the strike K and the instantaneous payoff $G(x)$, it is well-known that the optimal exercise price is the solution to the equation

$$K = D_0(t)\frac{\beta^+ - \gamma}{\beta^+}e^{\gamma x}, \tag{5.1}$$

where $\beta^+ > 1$ is the positive root of the quadratic equation

$$\frac{\sigma^2}{2}(\beta^2 - \beta) - r = 0. \tag{5.2}$$

Equivalently, at the current level $X_0 = x$, the individual should be willing to pay K for the delivery at time t. We obtain the following formula for the discounting factor for gains:

$$D_G(t) = \frac{\beta^+ - \gamma}{\beta^+} e^{-(r+\gamma(1-\gamma)\sigma^2/2)t}, \tag{5.3}$$

which is *exactly* the formula in the quasi-hyperbolic model (in the discrete time model, we obtain a similar formula: see Boyarchenko and Levendprskiǐ (2004a)). Indeed, the first factor $(\beta^+ - \gamma)/\beta^+ \in (0, 1)$, since $0 < \gamma < 1 < \beta^+$.

A similar argument leads to the following formula for the discounting factor for the losses:

$$D_L(t) = \frac{\beta^- - \gamma}{\beta^-} e^{-(r+\gamma(1-\gamma)\sigma^2/2)t}, \tag{5.4}$$

where $\beta^- < 0$ is the positive root of the quadratic equation (5.2). This time, the first factor $(\beta^- - \gamma)/\beta^- > 1$, since $\beta^- < 0 < \gamma$.

We see that as in the model with consumption streams, the use of an over-simplified instantaneous utility function does not allow one to reproduce the magnitude effect. Consider more general utility functions. As Boyarchenko and Levendorskiǐ (2004b) notice, we can reduce the optimal stopping problem with an instantaneous payoff $G(x) := e^{-rt}E^x[u(c(X_t))]$ to a problem with the payoff stream $g(x) := (r - L_X)u(c(x))$, where $L_X = \frac{\sigma^2}{2}\partial^2 + b\partial = \frac{\sigma^2}{2}(\partial^2 - \partial^2)$ is the infinitesimal generator of the Brownian motion X_t. If g is monotone (straightforward calculations show that this is the case if $u(c) = (1+c)^\gamma - 1$, for instance), then general results of Boyarchenko and Levendorskiǐ (2004b) apply, and the resulting formulas for the discount factors for gains and losses are

$$D_G(x,t) = \frac{e^{-rt}}{\sqrt{2\pi\sigma^2 t}} \int_{-\infty}^{+\infty} e^{-(y-\sigma^2 t/2)^2/(2\sigma^2 t)} \left(g(x+y) - \frac{1}{\beta^+} g'(x+y) \right) dy, \tag{5.5}$$

and

$$D_L(x,t) = \frac{e^{-rt}}{\sqrt{2\pi\sigma^2 t}} \int_{-\infty}^{+\infty} e^{-(y-\sigma^2 t/2)^2/(2\sigma^2 t)} \left(g(x+y) - \frac{1}{\beta^-} g'(x+y) \right) dy. \tag{5.6}$$

Details are available on request. Numerical examples show that for small and moderate levels of the volatility σ and the background discount factor r, the discount factor curves are essentially straight, and agree well with the (β, δ)-model (with $\beta \in (0, 1)$ for gains, and $\beta > 1$ for losses). The discount curves become really curved only for very large values of σ and r. In all cases, the discount factors for gains are too large, and for losses, they are too large in absolute values (the negative discounting effect is very large). In addition, both for convex and concave u_L, the magnitude effect is of the wrong sign (large losses are discounted more than small ones). See Fig. 8–11.

6 Suggestions for empirical verification of the model of endogenous time preference

We started from simple plausible assumptions about the behavior of humans who are not exposed too much to modern financial markets, and we showed that their optimizing behavior under uncertainty naturally leads to the DU anomalies. We make a simple observation that the behavior of traders in financial markets should not exhibit these anomalies, and we conjecture that the type of behavior of traders in modern financial markets, and the type of behavior formalized in the paper, lead to two extreme forms of time preference. Naturally, there must be intermediate forms, and we surmise that the DU anomalies should be the more prominent the more the members an experimental group differs from traders either by the culture they belong to or lifestyle or upbringing or education. Roughly speaking, the less exposure to the practice and/or theory of modern financial markets of the individual, the more prominent DU anomalies she will demonstrate. For instance, we expect that an Indian from the deep forest in Amazonia and a villager in Siberia will demonstrate more of DU anomalies than a city dweller; a villager involved in trade (even of a simplest type) will demonstrate less of DU anomalies than his neighbor who works in a corn field; and the same will be true for a person with a BA degree as opposed to his sibling with only high school education. Finally, we presume that an MBA in Finance will exhibit less DU anomalies than a PhD in humanities.

These observations imply that it is interesting and important to conduct empirical studies *taking into account the differences across humans* described above. Assuming that a group subject to an experiment consists of individuals with a similar background (similar exposure to markets of a similar degree of efficiency), it is necessary to separate the group into subgroups with close characteristics along the following two dimensions: the attitude toward risk, and accuracy of perception of the uncertainty of the future. This implies that prior to an experiment aimed at the study of DU anomalies, "introductory" experiments are needed to separate subgroups with similar attitude towards risk, and ability to access the uncertainty. In these "introductory" experiments, the questions should relate to the present only. In experiments (with a particular subgroup) aimed at the study of DU anomalies proper, it is important to study the dependence of inferred discount factors on the level of uncertainty.

Once all these factors are accounted for, it will be possible to test the predictions of the model. Probably, the following essentially qualitative (not just quantitative) effects are good candidates for testing:

(i) the correct sign of the magnitude effect for losses (large losses are discounted less than small ones) is consistent with a concave utility function over losses but not with a convex one (used in the prospect theory);

(ii) the higher the uncertainty level, the larger the "sign effect" is, equivalently, the larger the ratio of the discount factor for losses to the one for gains is;

(iii) the higher the uncertainty level, the larger the discounting rate for gains is, and the more prominent the “negative discounting effect” for losses is;

(iv) the better the fit of the (β, δ)-model for gains, the smaller the γ in the formula $u(c) = c^\gamma$ or $u(c) = (1+c)^\gamma$ is (equivalently, the larger the coefficient of relative risk aversion is);

(v) the discount factors for gains should be independent of the instantaneous utility of gains, if the coefficient of the relative risk aversion is not large;

(vi) for losses, the hyperbolic effect of the opposite sign should be observed; a simplified version is: the (β, δ) model may hold with $\beta > 1$ only;

(vii) if the utility function over losses is concave then both β and δ depend on the coefficient of the relative risk aversion; if the utility function over losses is convex, δ must be independent of this coefficient.

Finally, we would like to stress once again that the DU anomalies are expected to be more prominent, when an experimental group belongs to either a non-market oriented society (as opposed to well-developed industrial one) or a profession with little (if any) exposure to trade especially in financial instruments or its education level and/or type of education is different from the education of an MBA in Finance.

7 Conclusion

We made plausible assumptions about the behavior of humans who are not exposed too much (if at all) to modern efficient financial markets. These assumptions are: the individual perceives a durable good as an (almost) perpetual utility stream and assumes that she can wait for a fair price for the stream to materialize. We added standard assumptions about the instantaneous utility function (monotonicity and concavity) over both gains and losses, and about the uncertainty of the future consumption stream. Using general optimal exercise rules for the perpetual American call option (on a utility stream) and perpetual American put option (on a disutility stream), we derived explicit formulas for the fair price of a utility stream, which the individual would be willing to pay, and the fair compensation for the loss of a utility stream which the same individual would be willing to accept. We interpreted the resulting analytical expressions as integrals of deterministic streams with discounting factors, which turn out to be different for gains and losses even if we assume that the instantaneous utility function for gains equals the instantaneous disutility function for losses. The results are obtained under the assumption that there is a background discount rate $r > 0$ (killing rate). We showed that the qualitative results for gains are independent of r, and make sense even when r vanishes. In the limit, we obtain endogenous discount rates, with no exogenous component. For losses, arbitrary small r is admissible if the utility over losses is convex as in the prospect theory (however, this assumption leads to the magnitude effect of the wrong sign); if the utility over losses is concave, then r must be not too small in order to avoid a feeling of the infinite loss.

We derived simpler approximate formulas for the discount factors for large t, and showed that depending on the utility function and the sign (whether gains or losses are considered), the (β, δ)-model of the hyperbolic discounting arises (for gains, $\beta \in (0, 1)$, and for losses, $\beta > 1$). The same models arises in a simple model with the uncertainty modelled as i.i.d. draws from a given distribution (Subsection 5.1), and a variant of our basic model with instantaneous consumption (Subsection 5.3), however, these two models have serious drawbacks.

In the continuous time, natural analogs of (β, δ) model suggested by our asymptotic formulas are: for gains,

$$D_G(t) = (1 + a\sqrt{t})^{-3}e^{-qt},$$

where $a, q > 0$, and for losses,

$$D_L(t) = \frac{1 + b\sqrt{t}}{1 + a\sqrt{t}}e^{-qt},$$

where $b > a > 0$, and $q > 0$.

We showed that the hyperbolic discounting model arises as a good approximate universal model (in the sense that the discount factors are independent of the utility function), in the following two cases: for gains, if the relative risk-aversion coefficient is moderate (in the range $(0, 0.5)$), and for losses, if this coefficient is less than 1.

We made suggestions how to design experiments to test the assumptions and implications of our model, and we showed that simpler uncertainty-based endogenous models for discount rates lead to DU anomalies of either an incorrect type, or size or sign. In particular, we showed that if the uncertainty is modelled as stationary i.i.d. distributions, then the (β, δ)-model can be obtained but δ will be identical to the exogenous discount rate r. Since the observed discount rates are too large to be explained as the natural *killing rates*, this simple model of uncertainty cannot produce reasonable *endogenous* discount rates. We conclude that in economics, it is better to use stochastic processes rather than i.i.d. draws. On the contrary, the variant of our model with *instantaneous consumption* leads to too large discounting rates for gains, and too large negative discounting for losses; also, the magnitude effect for losses is of the wrong sign for concave and convex utility functions. We surmise that when making the intertemporal decisions, the individual assumes that she would smooth consumption over time even if the payoff is due at a fixed instant in the future.

A Technical calculations

A.1 Proof of (4.15)–(4.20)

Making the change of variables $v = \sqrt{z}$ and integrating by part, it is straightforward to derive

$$\mathrm{Erfc}(x) = e^{-x^2}\left(\frac{1}{\sqrt{\pi}x} + \frac{1}{2\sqrt{\pi}x^3} + \cdots\right), \quad \text{as } x \to +\infty. \tag{A.1}$$

If $x \to -\infty$, we use

$$\mathrm{Erfc}(x) = \frac{2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-v^2} dv - \mathrm{Erfc}(-x) = 2 - \mathrm{Erfc}(-x),$$

and deduce from (A.1)

$$\mathrm{Erfc}(x) = 2 + \frac{1}{\sqrt{\pi}x} e^{-x^2} + \cdots, \quad \text{as } x \to -\infty. \tag{A.2}$$

Using (A.1)–(A.2), we derive from (4.13):

a) if $\gamma \in (0, 0.5)$, then $\gamma(1-\gamma)/2 < 1/8$, and therefore (4.15) follows from

$$\begin{aligned} D_G^c(x,t) &= \frac{0.5-\gamma}{1-\gamma} e^{\gamma(\gamma-1)\sigma^2 t/2} \cdot 2 + \frac{0.5}{1-\gamma} \frac{2}{\sqrt{\pi\sigma^2 t/2}} e^{-\sigma^2 t/8} + \cdots \\ &= \frac{1-2\gamma}{1-\gamma} e^{-t\gamma(1-\gamma)\sigma^2/2} + \cdots; \end{aligned}$$

b) if $\gamma = 0.5$, then (4.17) follows from

$$D_G^c(x,t) = \frac{2}{\sqrt{\pi\sigma^2 t/2}} e^{-\sigma^2 t/8} + \cdots;$$

c) if $\gamma \in (0.5, 1)$, then we have to keep two terms when using (A.1) because the leading terms (of the order $e^{-t\sigma^2/8} t^{-1/2}$) cancel out:

$$\begin{aligned} D_G^c(x,t) &= \frac{0.5-\gamma}{1-\gamma} e^{\gamma(\gamma-1)t/2} \frac{1}{\sqrt{\pi}(\gamma-0.5)\sqrt{\sigma^2 t/2}} e^{-(\gamma-0.5)^2\sigma^2 t/2} \\ &\quad + \frac{0.5}{1-\gamma} \frac{2}{\sqrt{\pi\sigma^2 t/2}} e^{-\sigma^2 t/8} + \frac{0.5}{1-\gamma} \frac{2}{2\sqrt{\pi}} \left(\frac{2}{\sqrt{\sigma^2 t/2}} \right)^3 e^{-\sigma^2 t/8} + \cdots \\ &= \frac{8}{\sqrt{2\pi}(\sigma^2 t)^{3/2}} e^{-t\sigma^2/8} + \cdots. \end{aligned}$$

A.2 Proof of (4.23)

We have $\mathrm{Erfc}(0) = 1$ and $\mathrm{Erfc}'(0) = -2/\sqrt{\pi}$, therefore for small $t > 0$

$$\begin{aligned} D_L^c(x,t) &= \frac{0.5-\gamma}{1-\gamma} \left(1 + \frac{\gamma(\gamma+1)}{2}\sigma^2 t + \cdots\right) \left(1 - \frac{2}{\sqrt{\pi}}(0.5-\gamma)\sqrt{\frac{\sigma^2 t}{2}} + \cdots\right) \\ &\quad + \frac{0.5}{1-\gamma} \left(1 - \frac{2}{\sqrt{\pi}} \left(-0.5\sqrt{\frac{\sigma^2 t}{2}}\right) + \cdots\right) \\ &= 1 + \frac{1}{\sqrt{\pi}(1-\gamma)} [-2(0.5-\gamma)^2 + 0.5] \sqrt{\frac{\sigma^2 t}{2}} + \cdots \\ &= 1 + 2\gamma\sqrt{\frac{\sigma^2 t}{2\pi}} + \cdots \end{aligned}$$

A.3 Proof of (4.24) and (4.26)

If $\gamma \in (0, 0.5)$, then

$$\begin{aligned} D^c_L(x,t) &= \frac{0.5-\gamma}{1-\gamma}e^{\gamma(\gamma-1)\sigma^2 t/2}\frac{1}{\sqrt{\pi}(0.5-\gamma)}\sqrt{\frac{2}{\sigma^2 t}}e^{-(0.5-\gamma)^2\sigma^2 t/2} + \frac{0.5}{1-\gamma}2 + \cdots \\ &= \frac{1}{1-\gamma} + \cdots, \end{aligned}$$

and if $\gamma > 0.5, \gamma \neq 1$, then

$$\begin{aligned} D^c_L(x,t) &= \frac{0.5-\gamma}{1-\gamma}e^{\gamma(\gamma-1)\sigma^2 t/2} \cdot 2 + \frac{0.5}{1-\gamma}2 + \cdots \\ &= \frac{2\gamma-1}{\gamma-1}e^{\gamma(\gamma-1)\sigma^2 t/2} + \frac{1}{1-\gamma} + \cdots . \end{aligned}$$

If $\gamma \in (0, 0.5)$, then the first term on the RHS decreases exponentially, and if $\gamma > 1$, it increases exponentially, which proves (4.24) and (4.26). Since $D^c_L(x,t)$ for $\gamma = 1$ grows not faster than $D^c_L(x,t)$ for any $\gamma > 1$, we conclude that in the case $\gamma = 1$, $D^c_L(x,t)$ grows slower than any exponential function.

References

[1] B.S. Bernanke (1983), Irreversibility, uncertainty, and cyclical investment, *Quarterly Journal of Economics*, 98, 85–106.

[2] Black, F., M. Scholes (1973), The pricing of options and corporate liabilities, *J. Political Econom.*, 81, 637-654.

[3] A.N. Borodin, P. Salminen (2002), *Handbook of Brownian Motion – Facts and Formulae*, 2d edition, Birkhäser Verlag, Basel Boston Berlin

[4] S.I. Boyarchenko (2004), Irreversible decisions and record setting news principles. *American Economic Review*, 94, 557–568.

[5] S.I. Boyarchenko and S.Z. Levendorskiĭ (2002), *Non-Gaussian Merton-Black-Scholes theory*, World Scientific, New Jercey, London, Singapore, Hong Kong

[6] S.I. Boyarchenko and S.Z. Levendorskiĭ (2004a), *Practical Guide to Real Options in Discrete Time*, (February 24, 2004). http://ssrn.com/abstract=510324

[7] S.I. Boyarchenko and S.Z. Levendorskiĭ (2004b), *American options: the EPV pricing model* (May 19, 2004) http://ssrn.com/abstract=547863; forthcoming in Annals of Finance

[8] S.I. Boyarchenko and S.Z. Levendorskiĭ (2004c), *Optimal stopping made easy* (October 26, 2004). http://ssrn.com/abstract=610621

[9] S.Frederick, G.Loewenstein, and T.Donoghue (2002), Time discounting and time preference: a critical review. *J. Econ. Literature*, 40:2, 351–401.

[10] D. Fudenberg, D.K. Levine (2004), A dual self model of impulse control, Working paper.

[11] J.M.Harrison, and D.M.Kreps (1979), Martingales and arbitrage in multi-period securities markets, *J. Econom. Theory*, 20, 381–408.

[12] J.M.Harrison, and S.R.Pliska (1981), Martingales and stochastic integrals in the theory of continuous trading, *Stochastic Processes Appl.*, 11, 215–260.

[13] D. Kahneman, and A.Tversky (1979), Prospect theory: an analysis of decision under risk, *Econometica*, 47, 263–292.

[14] T.C.Koopmans (1960), Stationary Ordinal Utility and Impatience, *Econometrica*, 28, 287–309.

[15] G.Loewenstein (1988), Frames of mind in intertemporal choice, *Management Science*, 34, 200–214.

[16] G.Loewenstein, and D.Prelec (1991), Negative time preference, *Amer. Econ. Rev.*, 81, 347–352.

[17] G.Loewenstein, and D.Prelec (1992), Anomalies in intertemporal choice: Evidence and interpretation, *Quart. J. Econ.*, 107:2, 573–597.

[18] R.C.Merton (1973), Theory of rational option pricing, *Bell J. Econom. Manag. Sci.,* 4, 141–183.

[19] R.C.Merton (1990), *Continuous-Time Finance*. Basil Blackwell, Oxford

[20] Efe A. Ok, and Y. Masatlioglu (2003), *A general theory of time preference*, Working paper

[21] E.S.Phelps, and R.Pollack (1968), On second-best national saving and game-equilibrium growth, *Rev. Econ. Stud.,* 35, 186–199.

[22] P. Samuelson (1933), A note on Measurement of Utility, *Rev. Econ. Stud.*, 4, 155-161

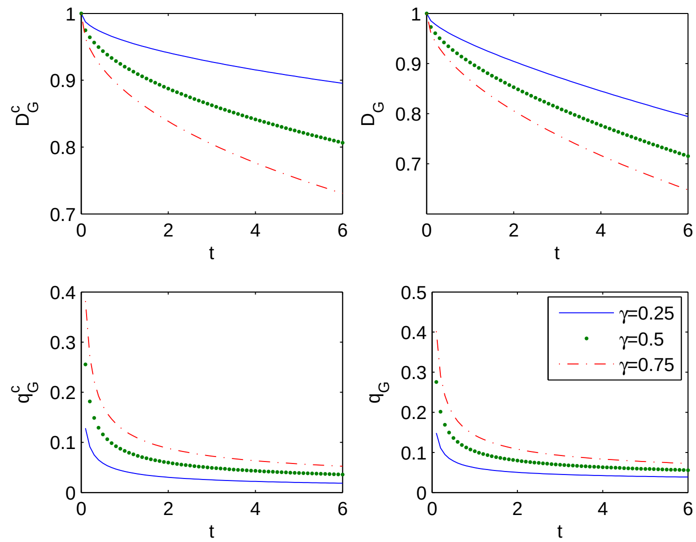

Figure 1: Discount factors and rates for gains, $u_G(c) = c^\gamma$: dependence on time and γ. Left panel: correction factors and rates. Right panel: (total) factors and rates. Parameters: $\sigma = 0.2, r = 0.02$.

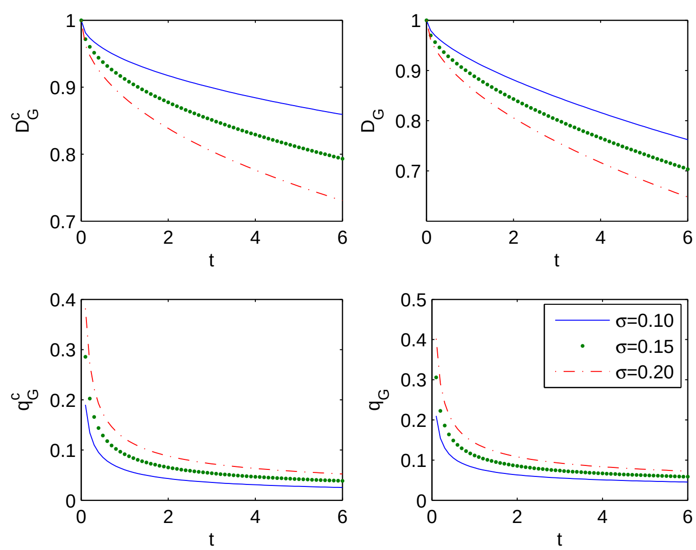

Figure 2: Discount factors and rates for gains, $u_G(c) = c^\gamma$: dependence on time and volatility σ. Left panel: correction factors and rates. Right panel: (total) factors and rates. Parameters: $\gamma = 0.75, r = 0.02$.

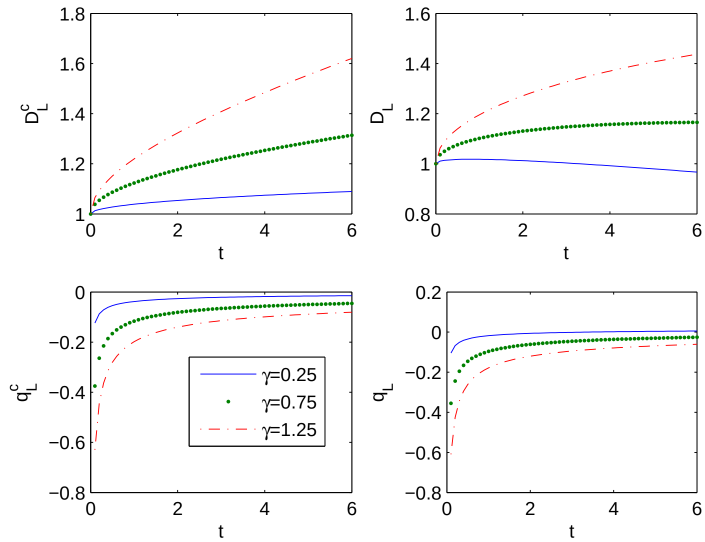

Figure 3: Discount factors and rates for losses, $u_L(c) = c^\gamma$: dependence on time and γ. Left panel: correction factors and rates. Right panel: (total) factors and rates. Parameters: $\sigma = 0.2, r = 0.02$.

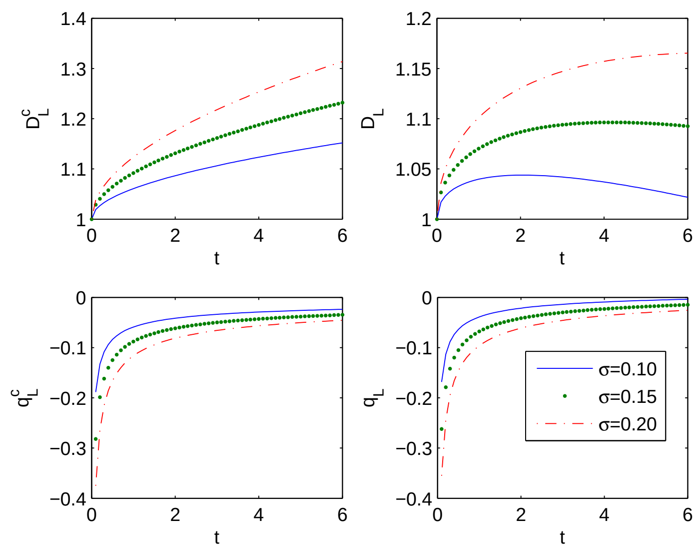

Figure 4: Discount factors and rates for losses, $u_L(c) = c^\gamma$: dependence on time and volatility σ. Left panel: correction factors and rates. Right panel: (total) factors and rates. Parameters: $\gamma = 0.75, r = 0.02$.

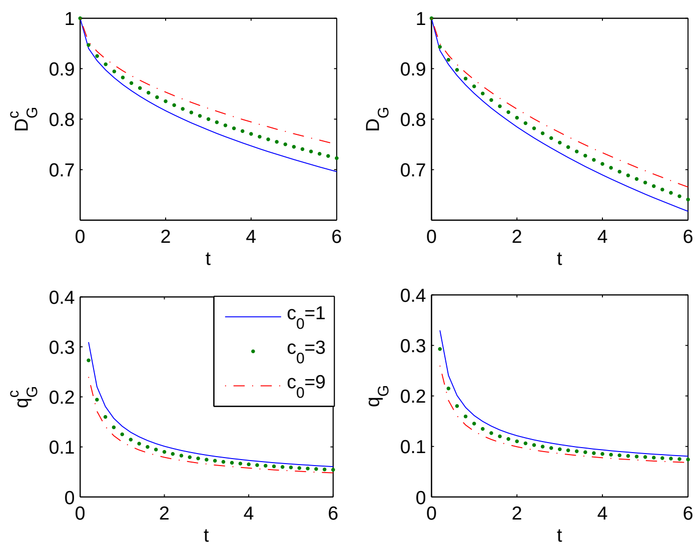

Figure 5: Magnitude effect for gains: $u_G(c) = (1+c)^\gamma - 1$. Left panel: correction factors and rates. Right panel: (total) factors and rates. Parameters: $\sigma = 0.2, \gamma = 0.5, r = 0.02$.

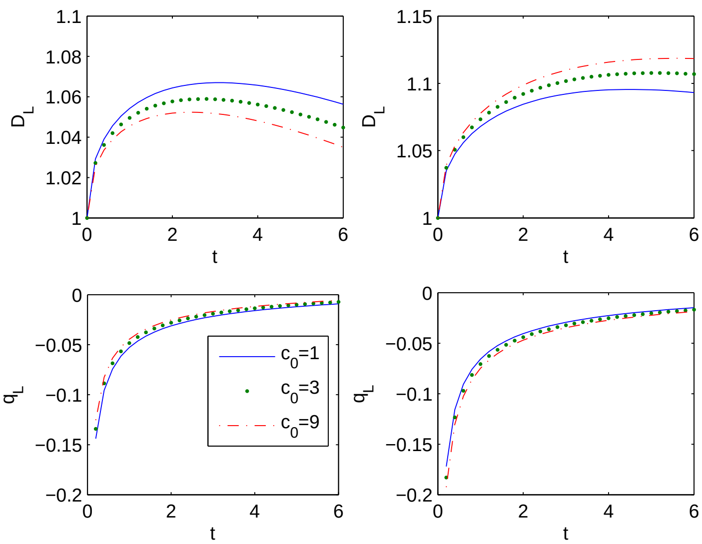

Figure 6: Magnitude effect for losses: $u_L(c) = (1+c)^\gamma - 1$. Left panel: the wrong sign of the magnitude effect (a convex utility function over losses, $\gamma = 0.75$). Right panel: the correct sign of the magnitude effect (a concave utility function over losses, $\gamma = 1.25$). Parameters: $\sigma = 0.1, r = 0.02$.

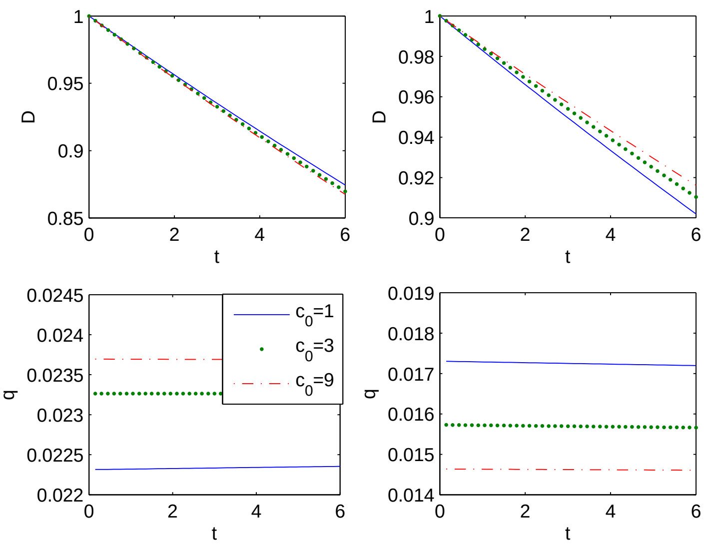

Figure 7: Naive endogenous discount factors and rates: $u(c) = (1 + c)^\gamma - 1$. Left panel: a concave utility function over gains (convex utility function over losses), $\gamma = 0.75$. Right panel: a convex utility function over gains (concave utility function over losses), $\gamma = 1.25$. Parameters: $\sigma = 0.2, r = 0.02$.

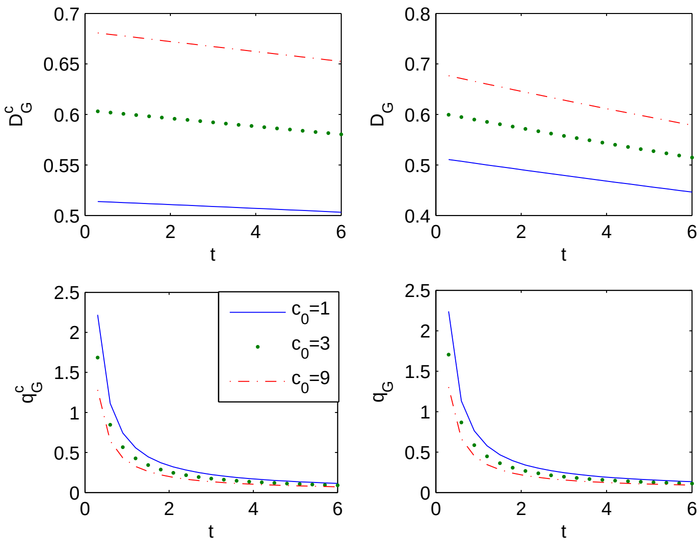

Figure 8: Instant consumption model, gains: $u_G(c) = (1+c)^\gamma - 1$. Left panel: correction factors and rates. Right panel: (total) factors and rates. Parameters: $\sigma = 0.2, \gamma = 0.25, r = 0.02$.

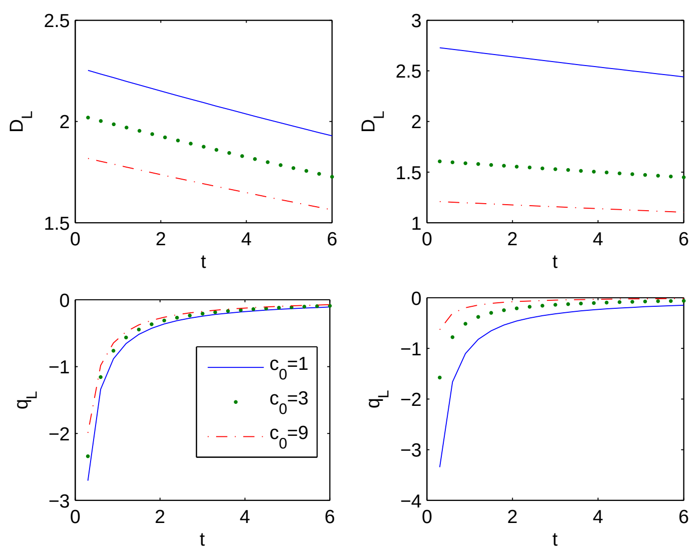

Figure 9: Instant consumption model, losses: $u_L(c) = (1+c)^\gamma - 1$. Left panel: a convex utility function over losses, $\gamma = 0.25$. Right panel: a concave utility function over losses, $\gamma = 1.25$. Parameters: $\sigma = 0.2, r = 0.02$.

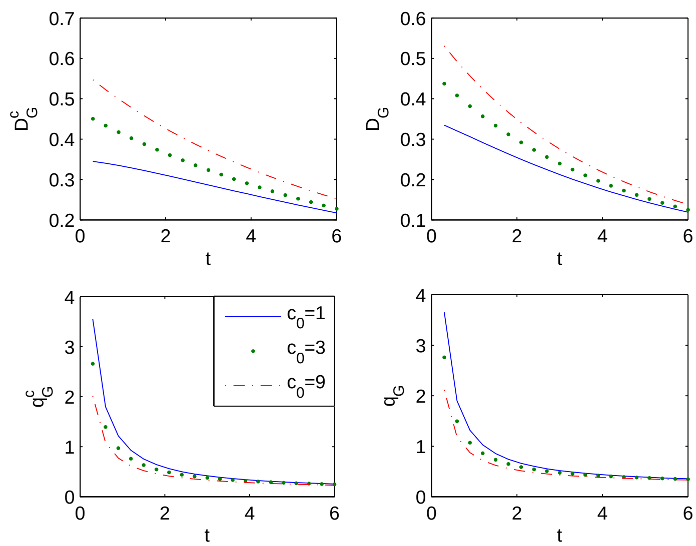

Figure 10: Instant consumption model, large $\sigma(=0.9)$ and $r(=0.1)$, gains, $u_G(c)=(1+c)^\gamma-1$. Left panel: correction factors and rates. Right panel: (total) factors and rates.

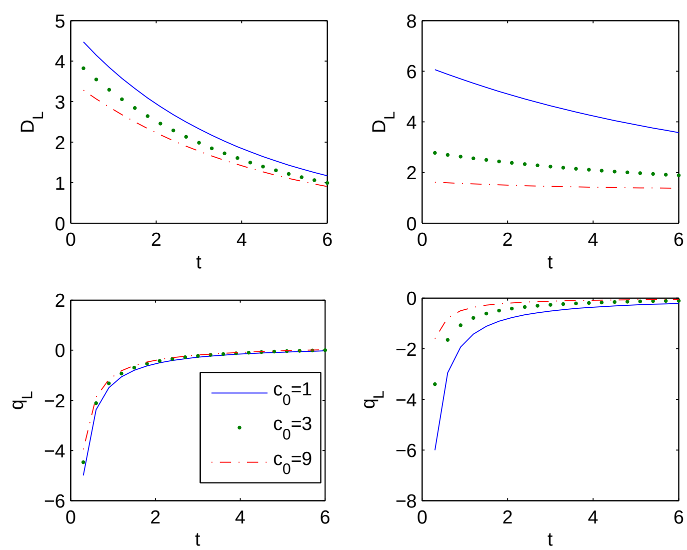

Figure 11: Instant consumption model, large $\sigma(=0.9)$ and $r(=0.1)$, losses: $u_L(c)=(1+c)^\gamma-1$. Left panel: a convex utility function over losses, $\gamma=0.25$. Right panel: a concave utility function over losses, $\gamma=1.25$.