How the Future Shapes Consumption with Time-Inconsistent Preferences*

James Feigenbaum[†]

Sepideh Raei[‡]

December 5, 2022

Abstract

Time-inconsistent preferences, which are modeled by relative discount functions, are a common explanation for the empirical finding that lifecycle profiles of household consumption are typically hump-shaped rather than monotonic. More precisely, a time inconsistent preference which is present-biased can generate a hump-shaped consumption profile over the lifecycle. We develop a general framework for understanding "present bias" in consumption through a "future weighting factor" that perturbs the discount factor of utility at future periods away from exponential discounting. Using our framework we derive necessary and sufficient conditions on the future weighting factors for the consumption profile to be locally concave. We find that these conditions, which are necessary for the consumption profile to be hump-shaped, are stronger than just assuming a present bias. Furthermore, we obtain necessary and sufficient conditions under which the consumption profile determined in the first period of life Pareto dominates the realized consumption profile. Pareto dominance of this initial path must arise when the log consumption profile is strictly concave and the future weighting factor at the longest delay is not too large.

JEL: D60, D90

Keywords: present bias, future bias, time-inconsistent preferences, consumption hump, commitment mechanisms

---

*We would like to thank Frank Caliendo and Scott Findley and participants at the SABE 2022 conference for their input.

[†]Utah State University, John Huntsman School of Business, Utah, United States; james.feigenbaum@usu.edu.

[‡]Utah State University, John Huntsman School of Business, Utah, United States; sepideh.raei@usu.edu.

1 Introduction

The canonical life-cycle model predicts that consumption will grow smoothly for patient individuals and decay smoothly for impatient individuals. However, from an empirical standpoint, one of the most striking aspects of people's choices of consumption over the lifecycle is that this profile is generally hump-shaped. As was first documented by **?**, average consumption increases while consumers are young, peaks when they reach middle age, and decreases afterwards.[1]

A sizeable literature is devoted to developing theoretical frameworks that modify the Lifecycle/Permanent-Income Hypothesis of Friedman and Modigliani (**?**, **?**) to address the inconsistency between model predictions and empirical evidence regarding the consumption profile, which has been referred to as the "lifecycle consumption puzzle".[2] One strand of literature that develops a set of solutions to this inconsistency by adding elements that are directly observable such as family-size effects (**?**, **?**, **?**), consumption-leisure trade-offs (**?**, **?**), wage income uncertainty and the precautionary saving motive (**?**, **?**, **?**, **?**, **?**), mortality risk (**?**, **?**), and consumer durables (**?**).

Another set of mechanisms that can explain the hump in the consumption profile relax the assumptions on preferences of the standard rational paradigm, epitomized by **?**. One of the most popular of these is to allow for time-inconsistent preferences by generalizing the discount function from an exponential function. **?** was the first to explore such deviations from Samuelson's model. **?** later proposed the hyperbolic function as a specific alternative to the exponential function, and David Laibson's dissertation (**?**) offered hyperbolic discounting as a solution to the consumption hump puzzle. Today, this strand of the literature generally attributes such consumption humps to the concept of "present bias".[3]

This paper develops a general framework for understanding the concept of "present bias" in consumption through a "weighting factor" that perturbs the discount factor of utility at future periods away from exponential discounting. It is essentially a generalization that nests the well-known hyperbolic and quasi-hyperbolic preferences, and which permits more general statements about necessary conditions for present bias and a concave log consumption profile (a known property of present bias). Using this representation, we establish conditions

[1]See **?**, **?**, **?**, **?**, **?**, and **?**.

[2]See **?** and **?** for more recent overviews.

[3]See **?**, **?**, **?** and **?** **?**, **?**, **?**, **?**. There are also papers that approach this puzzle by combining behavioral and more traditional factors, such as **?** who explain the hump-shaped wages with rule-of-thumb consumers in the economy.

under which the consumption profile is consistent with the empirical evidence. In a nutshell, whether a hump-shaped consumption profile can occur will depend on precisely how the discount function deviates from an exponential discount function. We find that present bias is a necessary but not sufficient condition for the lifecycle consumption profile to be hump-shaped. Further, we use this general setup to investigate the conditions under which all of the different selves prefer the commitment path of consumption (the plan established in period zero) to the realized path (the actual consumption decisions). That is to say the commitment path Pareto dominates the realized path.

Present bias, or as **?** called it present focus,[4] is a form of time-inconsistency in which individuals are more impatient in trade-offs between the present and the immediate future as compared to trade-offs between equivalent intervals of time in the more distant future. Individuals acting under this bias, who might have been inclined to postpone a future payoff when the options for when to take it were all in the future, are more inclined to take it at the first opportunity as the opportunity gets closer to the present.[5] Such a change is referred to as a preference reversal. Note that with exponential discounting, the concept of time can be assumed to be either absolute time, calendar time, or even waiting time, i.e. the time to consumption. As **?** showed, the equivalence of these three temporal interpretations is a consequence of the exponential function not exhibiting preference reversals. In contrast, for a nonexponential discount function that exhibits present bias, we must interpret the "time" that parameterizes the function as the delay or waiting time until we experience the consumption from the present moment.

In this paper, we propose a general representation of a discount function in the form of $D_t = D_1^t(1+\varepsilon_t)$ for $t = 1, \dots, T$, where ε_t is the **extra weight** (compared to the exponential discounting case) that we put on the discount factor t periods in the future, and $T+1$ is the life span. We call ε_t the *future weighting factor*. All forms of discount function, including the nonexponential ones, can be written as a specific case of this general function by finding the corresponding ε_t. An advantage of this novel approach is the opportunity it provides to understand the driving force behind the consumption hump. Using this framework, we find that the shape of the consumption profile at a given age depends on the dynamics of the future weighting factors at all delays within the remaining time horizon of the household.

[4]**?** use the term present focus, rather than the more common term present bias, because they believe the word bias implies a prejudgment that the behavior is a mistake, which is not true in their view.

[5]Present bias, which is viewed as a form of misoptimization that accounts for a range of behavioral "mistakes," e.g. undersaving for retirement, has yielded a large literature that emphasizes the potential for policies like forced pensions or retirement saving subsidies to protect against or correct such mistakes (for a survey on present bias see **?**)

Working in a lifecycle model where households naive about their time-inconsistent preferences repeatedly optimize a logarithmic utility function, we derive conditions on the future weighting factor such that the log consumption profile is locally concave since a hump-shaped consumption profile will have to be locally concave at the peak of the hump. In order for the log consumption profile to be concave at a given point in the lifecycle, the future weighting factor between that point and the end of the life span must be greater on average than the weighting factors at shorter delays, as only these will be relevant going forward. This translates to the discount function decaying at a slower rate than an exponential function over the remaining life span. It is, however, a stronger condition than a present bias, so present bias alone is not strong enough to guarantee a hump-shaped consumption profile.

In formal terms, suppose that intertemporal preferences from the perspective of period t can be represented by $U_t = \sum_{s=t}^{T} D_{s-t} u_s$, where u_s is the instantaneous utility experienced in period s and D_x reflects the discounting associated with a delay of $x \in \{0, 1, 2, ...\}$. A common example in which the concept of present bias is readily discernible is the β-δ or "quasihyperbolic" functional form

$$D_x = \begin{cases} 1 & \text{if } \; x = 0 \\ \beta\delta^x & \text{if } \; x > 0. \end{cases}$$

If $\beta = 1$, this reduces to an exponential discounting function, in which case the optimal plan at $t = 0$ will remain the optimal plan throughout the lifecycle. For $\beta \in (0, 1)$, the utility from consumption at all periods after the present are discounted by the factor β, and the difference $1 - \beta$ is a measure of present bias. The optimal plan at $t = 0$ will differ from the optimal plan later in life as the household will continually seek to advance consumption relative to what she originally planned. Conversely, if $\beta > 1$, the utility from consumption at all future periods would be magnified by the factor β, and $\beta - 1$ can be characterized as a measure of "future bias".

While the quasihyperbolic case only covers a measure-zero subset of the space of all possible discounting functions, because of their simplicity quasihyperbolic discount functions are often used as a proxy for other nonexponential discount functions. Indeed, the terminology of quasihyperbolic derives from this usage as an approximation to hyperbolic discount functions. If $\beta < 1$, the quasihyperbolic discount function will share with hyperbolic discount functions the property that the lifecycle profile of log consumption is concave. These discount functions also share another property to be further explained below: the household at most ages would prefer the consumption profile it would get if it could commit to its

$t = 0$ plan to what it gets in equilibrium after accounting for its changing intertemporal preferences. On the other hand, a future-biased quasihyperbolic function with $\beta > 1$ will yield log consumption profiles that are convex, and the household would usually prefer the consumption profile it actually gets to what it would get if it could commit to its initial plan.

However, as we demonstrate in this paper, the language of "present" and "future" bias are not reliable predictors of these properties. A convex log consumption profile does not always arise in association with discount functions that one would naturally think of as future-biased. For example, a pure myopic discount function is a discount function that vanishes for delays beyond some horizon. Households with such a discount function do not care about consumption in the future beyond that horizon. Nevertheless, myopia yields properties consistent with a future-biased quasihyperbolic discount function rather than properties consistent with a present-biased quasihyperbolic discount function.[6]

Our approach for exploring the driving force behind the shape of the consumption profile begins with an exponential discount function, for which we know there is no time-inconsistency and the log consumption profile will be linear. Measuring the deviation from an exponential discount factor in terms of future weighting factors provides a straightforward way of understanding the origin of a present bias, which comes from having all ε_t be positive and strictly increasing for $t > 1$, or a future bias, which comes from having all ε_t be negative and strictly decreasing for $t > 1$.[7] A present bias is a necessary, albeit not sufficient, condition to have a concave log consumption profile. Likewise, a future bias will be a necessary condition to have a convex log consumption profile.

Positive future weights mean that the discount function will be higher than an exponential discount function as the delay time increases and are necessary for a discount function to exhibit present bias everywhere. Negative future weights mean the opposite and likewise are normally associated with future bias. In the case of a myopic discounting function the ε_t will all be -1 for sufficiently high t, and this will exhibit both present and future bias at different delays. The upshot is that a myopic discount function will tend to behave more like a future-biased quasihyperbolic discount function since they both put <u>less</u> weight on future consumption relative to an exponential discount function.

Another issue related to present and future bias that has been the focus of a relatively recent literature pertains to welfare analysis. Since an individual with time-inconsistent preferences, whether present- or future-biased, will choose a consumption profile that depends

[6] A myopic discount function actually exhibits both present and future bias, depending on the time horizon.

[7] For the case of a future bias we also need the additional requirement that $\varepsilon_t > -1$.

on the time of the choosing, it is not obvious which of these consumption profiles or the preferences at what period of life should be the reference point for welfare comparison. A common solution to this problem in the literature is to use the preferences of the initial self to evaluate welfare (see for example **?**, **?**, **?**, **?**, **?**, **?**, **?** among many others). In fact, **?** show that commitment to the time-zero consumption plan can improve the objective function for all selves if the number of selves exceeds a certain threshold which turned out to be quite small in their setting.

Adding to this strand of literature, the other contribution of this paper is to specify the conditions on future weighting factors under which the commitment path will Pareto dominate the realized path in discrete time.[8] There is a similarity between the conditions for a concave log consumption profile and the conditions for Pareto dominance of the commitment path. We will exploit this similarity to strengthen the conditions for a concave log consumption profile to obtain sufficient conditions for Pareto dominance of the commitment path that are relatively easy to verify.

It is worth mentioning that in this paper we model the household's choices in discrete time. A companion paper, **?**, addresses the same issues in continuous time. We obtain analogous results in the two papers, but, aside from the obvious advantages that the majority of economists are most comfortable working in discrete time and that economic data accumulates intermittently rather than continuously, another advantage of working in discrete time is that the discount function can be fully specified in terms of a finite number of future weighting factors. In the case of a four-period model, for which there are two future weighting factors, we can graph where in the parameter space the commitment path will Pareto dominate the realized path.

This paper is organized in the following way. Section 2 describes the model environment including the general format for the discount function. Section 3 develops the condition on the discount function for a concave or convex log consumption profile. Section 4 explores the condition on the discount function under which commitment to the initial plan would almost Pareto dominate the realized plan and investigate the relationship between the concavity condition and Pareto Dominance condition, and finally, section 5 concludes.

[8] We only compare the preferences of the households' various selves regarding the commitment path and the realized path. We do not make any claims regarding Pareto efficiency as in **?**, i.e. we do not compare how the various selves value these two paths relative to other feasible consumption paths.

2 Model environment

We focus on a finite-horizon life-cycle model in which households live for $T+1$ periods. The household earns income $y_t \geq 0$ at age t for $t = 0, ..., T$, which can be consumed c_t or saved as k_{t+1} at a fixed gross interest rate $R \geq 0$. [9]

2.1 Household optimization problem

At time t, a household with existing saving k_t maximizes

$$U_t = \sum_{s=t}^{T} D_{s-t} \ln c_{s|t}$$

subject to

$$c_{s|t} + k_{s+1|t} = y_s + Rk_{s|t}, \quad s = t, \ldots, T,$$

where $D_t \geq 0$ is the discount function, and $c_{s|t}$ and $k_{s+1|t}$ are consumption and saving at period s as planned in period t.[10] We will normalize $D_0 = 1$ and will also assume that $D_1 > 0$. The latter condition ensures that it will never be optimal for the household to consume all of its remaining wealth in the present period, which would leave the future selves with utility of $-\infty$. Note that the household will solve this problem with $k_{t|t} = k_t$ and $k_{T+1|t} = 0$. To simplify notation, we will assume the household begins with $k_0 = 0$.[11]

Let us define

$$h_t = \sum_{s=t}^{T} \frac{y_s}{R^{s-t}}, \tag{1}$$

[9] It is worth mentioning that similar to **?**, in this paper we use the "choice-based" methodology which compares the solutions of dynamic programs with different decision dates. It is the methodology used originally by Strotz (1956), and now standard in behavioral macroeconomics, since the pioneering work of **?**, **?**, **?**. It is "choice based" because it not only uses a utility function that represents the preference relation, but also the budgetary constraints that the decision maker faces.

[10] The results are not qualitatively different for other CRRA utility functions, but they are more complicated so we only consider the logarithmic case. In solving the model we will proceed as though the household is naive about its time-inconsistency and does not know it will revise its plans as its preferences change. We could alternatively assume that the household is sophisticated about its time-inconsistency. However, with logarithmic period utility, the realized path (and the commitment path in Section 4) will be the same under both assumptions, so there is no loss of generality between naivete and sophistication in the results documented here. For more discussion see **?**.

[11] Our results easily generalize if the household is endowed with savings or debt at birth.

which represents the present value of the income stream from period t onward. Note that

$$h_t = y_t + \sum_{s=t+1}^{T} \frac{y_s}{R^{s-t}} = y_t + \frac{h_{t+1}}{R} \tag{2}$$

for $t < T$. We can combine the period budget constraints from t to T into a lifetime budget constraint as of t:

$$\sum_{s=t}^{T} \frac{c_{s|t} + k_{s+1|t}}{R^{s-t}} = \sum_{s=t}^{T} \frac{y_s + Rk_{s|t}}{R^{s-t}}.$$

Using (1) and (2), this simplifies to

$$\sum_{s=t}^{T} \frac{c_{s|t}}{R^{s-t}} = h_t + Rk_t \tag{3}$$

The Lagrangian of the household problem at t can then be written as

$$L_t = \sum_{s=t}^{T} \left[D_{s-t} \ln c_{s|t} - \frac{\lambda_t c_{s|t}}{R^{s-t}} \right] + \lambda_t [h_t + Rk_t]. \tag{4}$$

Therefore, the first order condition (FOC) with respect to consumption will be

$$\frac{\partial L_t}{\partial c_{s|t}} = \frac{D_{s-t}}{c_{s|t}} - \frac{\lambda_t}{R^{s-t}} = 0. \tag{5}$$

The initial consumption plan $c_{s|0}$ that is determined at $t = 0$, the first period of life, will be referred to hereafter as the **commitment path**. Note, however, that unless the discount function is exponential the household will only follow the initial plan at $t = 0$. Indeed, at each period t of life, the household will choose a new plan $c_{s|t}$, but only the choice of consumption at t, $c_t = c_{t|t}$, will adhere to this plan. As the household progresses from period to period, its preferences will unexpectedly change since we are assuming that the household is naive about the change in its future preferences. When it gets to $t + 1$, it will then have saving $k_{t+1} = k_{t+1|t}$, but it will solve (4) anew, updated to $t + 1$. The resulting consumption profile c_t, determined at each period t, will be referred to as the **realized path**.

While the FOC (5) for $t = 0$ governs the whole commitment path for consumption $c_{s|0}$ from $s = 0, \ldots, T$, along the realized path only the FOC with $s = t$ will actually matter.

This simplifies to

$$\frac{D_{t-t}}{c_{t|t}} - \frac{\lambda_t}{R^{t-t}} = 0, \tag{6}$$

so we have

$$\lambda_t = \frac{1}{c_t}$$

since $c_t = c_{t|t}$ and $D_0 = 1$. The future plan $c_{s|t}$ at t is only relevant to the extent that it determines the Lagrange multiplier λ_t. Generalizing (5), we obtain

$$c_{s|t} = \frac{D_{s-t}R^{s-t}}{\lambda_t} = D_{s-t}R^{s-t}c_t.$$

Inserting these into the lifetime budget constraint (3), we get

$$\sum_{s=t}^{T} \frac{D_{s-t}R^{s-t}c_t}{R^{s-t}} = h_t + Rk_t,$$

which reduces to

$$c_t = \frac{h_t + Rk_t}{\sum_{s=t}^{T} D_{s-t}}. \tag{7}$$

Hence, on the realized path, the budget constraint on period t can be written as

$$k_{t+1} = k_{t+1|t} = y_t + Rk_t - c_t = y_t + Rk_t - \frac{h_t + Rk_t}{\sum_{s=t}^{T} D_{s-t}}. \tag{8}$$

We can use this to calculate an effective Euler equation along the realized path. Combining (2) and (8), we get,

$$\begin{aligned}
h_{t+1} + Rk_{t+1} &= R\left(\frac{h_{t+1}}{R} + y_t + Rk_t - \frac{h_t + Rk_t}{\sum_{s=t}^{T} D_{s-t}}\right) \\
&= R\left(h_t + Rk_t - \frac{h_t + Rk_t}{\sum_{s=t}^{T} D_{s-t}}\right) \\
&= R\left(\frac{\sum_{s'=t+1}^{T} D_{s'-t}}{\sum_{s=t}^{T} D_{s-t}}\right)(h_t + Rk_t).
\end{aligned}$$

Updating (7) to $t+1$, consumption at $t+1$ is

$$c_{t+1} = R\left(\frac{\sum_{s'=t+1}^{T} D_{s'-t}}{\sum_{s=t}^{T} D_{s-t}}\right)\frac{h_t + Rk_t}{\sum_{s=t+1}^{T} D_{s-t-1}}$$

Applying (7) again in its original form, the effective Euler equation realized by the household for a general discounting function D_t with log utility is

$$c_{t+1} = R\frac{\sum_{s'=t+1}^{T} D_{s'-t}}{\sum_{s=t+1}^{T} D_{s-t-1}}c_t. \tag{9}$$

As mentioned above, since $D_1 > 0$, c_{t+1} will be strictly positive.

In the special case of an exponential discount function $D_t = \delta^t$, the ratio

$$\mathcal{D}_t = \frac{\sum_{s'=t+1}^{T} D_{s'-t}}{\sum_{s=t+1}^{T} D_{s-t-1}}$$

simplifies to the constant δ, and we get back the familiar Euler equation $c_{t+1} = \delta R c_t$. More generally, though, for a nonexponential discount function, the inverse ratio $\mathcal{D}_t^{-1}$ measures the gross rate of change in the sum of the discount functions relevant at periods $t+1$ to T as the household moves from t to $t+1$. That is to say the change from the sum $D_1 + \cdots + D_{T-t}$ applicable at t to the sum $1 + \cdots + D_{T-t-1}$ applicable at $t+1$. The richer consumption dynamics that can be obtained in equilibrium with nonexponential discounting functions stems entirely from the deviation of the $\mathcal{D}_t$ from a constant, which will depend on how the discount function D_t deviates from an exponential function.

2.2 Future Weighting Discount Function

Given a discount function $D_t \geq 0$ for $t = 0, ..., T$, we define the "future weighting factor" ε_t via

$$D_t = D_1^t(1+\varepsilon_t), \tag{10}$$

where D_1 is the discount factor for one period ahead. This future weighting factor basically captures the extra (or diminished, if negative) weight that we put on the discounting t periods in the future. Since we normalize $D_0 = 1$, by definition we will have $\varepsilon_0 = \varepsilon_1 = 0$. Note that this general form of discounting function can accommodate the standard geometric discounter, for which $D_t = \delta^t$, quasi-hyperbolic agents, for whom $D_t = \beta\delta^t$ where $\beta < 1$ (**?**), future-biased agents for whom $D_t = \beta\delta^t$ with $\beta > 1$; and the immediate successor agents, for whom $D_1 = \delta$ and $D_2 = D_3 = \cdots = 0$ (see, **?**, **?** and **?**). For example, we can represent a quasihyperbolic discount function by setting $\varepsilon_t = \beta^{1-t} - 1$;

$$D_t = \beta\delta^t$$

for $t > 0$ with $D_0 = 1$. Since

$$D_1 = \beta\delta,$$

ε_t can be calculated as

$$\frac{D_t}{D_1^t} = \frac{\beta\delta^t}{\beta^t\delta^t} = \beta^{1-t} = 1 + \varepsilon_t.$$

Hence

$$\varepsilon_t = \beta^{1-t} - 1. \tag{11}$$

Likewise, for a myopic discounting function that vanishes for $t \geq t^*$, we have $\varepsilon_t = -1$ for $t \geq t^*$.

Note that if $\varepsilon_t = 0$ for all t, the discount function will be exponential. Thus we can think of the future weighting factor, ε_t, as the parameter that measures the discount function's deviation from an exponential at the delay t. If $\varepsilon_t > 0$, the discount factor of t periods in the future will be higher than an exponential discount factor. This means utility from consumption t periods in the future will be weighted more heavily than it would be under an exponential discount function. Conversely, with $\varepsilon_t < 0$ the weight on utility consumption t periods in the future will be lower relative to an exponential discount function.[12]

To have a better understanding of the role of ε_t in determining consumption behavior, figure 2 compares the consumption profile under the commitment path and the realized path for a ten period model, $T = 10$. We consider two cases to demonstrate the role of an individual ε_t. First, we have a discount function for which ε_t is zero for all t except $t = 2$. Second, we have a discount function for which ε_t is zero for all t except $t = 8$.

In both plots, the blue dashed line shows the commitment path and the red solid line shows the realized path. In figure 1a, we see a spike in period two along the commitment path simply because $\varepsilon_2 > 0$ means that the household initially puts a higher weight on the utility from consuming two periods ahead compared to all other future periods. Hence, the spike at $t = 2$. Likewise, looking at figure 1b in which $\varepsilon_8 > 0$, the spike in the commitment

[12]To be very precise, as we have defined the future weighting factor, we are talking about a departure from exponential discounting at the rate used between period 0 and 1. In discrete time, it is natural to think of the deviation of D_t from D_1^t, and this will yield some helpful simplifications.

Figure 1: consumption profile, commitment path and realized path

eps_2_r.pdf

a . only $\varepsilon_2 > 0$

eps_8_r.pdf

b . only $\varepsilon_8 > 0$

Note: on both graphs the horizontal axis is time and vertical access is the consumption level at each period.

path is at $t = 8$.

The effect of ε_t on the realized path is much more subtle than for the commitment path. With $\varepsilon_2 > 0$, shown in figure 1a, the household continually plans to have high consumption two periods ahead, as happens at $t = 2$ on the commitment path. However, with each new period, she reoptimizes and pushes forward when she intends to have high consumption. This trend continues until the household arrives at period nine of her lifetime, at which point there no longer is a period two periods ahead. Consequently, the realized consumption path is quite smooth, as it would be with exponential discounting, for $t < 9$. She does not realize this intended high consumption two periods ahead until she can no longer defer this consumption. From this point, all future periods are discounted with the same rate. Consumption jumps up in these last two periods as she finally consumes the saving she accumulated to finance the planned extra consumption two periods ahead.

The same intuition applies to figure 1b in which $\varepsilon_8 > 0$. There, the future period with a higher discounting factor disappears after the second period. That is the reason why the realized consumption plan for $t \geq 3$ shifts upward. The high ε_8 disappears from her calculus once there no longer is a period eight periods ahead within her remaining time horizon.

Consequently, she behaves like an exponential discounter thereafter, smoothing out over all the periods with $t \geq 3$ the extra consumption that she had previously intended, at $t = 2$, to save entirely for the last period.

The consumption-hump literature has traditionally characterized the effect of the discount function on the shape of the (log) consumption profile in terms of present bias. By examining present and future bias in terms of future weighting factors, we can also get some new insight into the origin of these concepts. A discount function exhibits present bias at $t > 0$ if it gives rise to the following type of preference reversal. Suppose for some allocation $\{c_t\}_{t=0}^T$, there exists $\xi_t > 0$ and $\xi_{t+1} \in (0, c_{t+1})$ such that the household would prefer at time 0 the original allocation over a forward-shifted allocation with c_t increased by ξ_t and c_{t+1} decreased by ξ_{t+1}. However, when the household gets to time t, it in stead prefers the forward-shifted allocation over the original allocation. Thus the household would prefer not to shift consumption forward when the possibility of doing so is in the future, but it would opt to make that shift in the present. This is usually interpreted as the household putting an extra preference on consumption in the immediate present. Future bias at $t > 0$ is defined similarly except the preference reversal goes the other way. The household would prefer the forward-shifted allocation over the original allocation when t is in the future, and prefers the original allocation when it reaches time t. We say a discount function is present-biased (future-biased) if it exhibits present (future) bias at all $t > 0$.

Assuming $D_s > 0$ for all s, we can express the condition for preference reversals in terms of the perceived marginal rate of substitution between consumption at t and consumption at $t+1$ as of time $s \leq t$:

$$m_s(t) = \frac{D_{t+1-s} u'(c_{t+1})}{D_{t-s} u'(c_t)}.$$

The household will prefer the forward-shifted allocation at time 0 and the original allocation at t if

$$D_t u'(c_t)\xi_t - D_{t+1} u'(c_{t+1})\xi_{t+1} < 0 < u'(c_t)\xi_t - D_1 u'(c_{t+1})\xi_{t+1},$$

which we can rearrange as

$$m_0(t) = \frac{D_1 u'(c_{t+1})}{u'(c_t)} < \frac{\xi_t}{\xi_{t+1}} < \frac{D_{t+1} u'(c_{t+1})}{D_t u'(c_t)} = m_t(t).$$

The household will have a present bias at t if $m_0(t) < m_t(t)$ since we can then find ξ_t and

ξ_{t+1} such that $\frac{\xi_t}{\xi_{t+1}} \in (m_0(t), m_t(t))$. Since $\varepsilon_1 = 0$ by definition, this reduces to the condition

$$1 < \frac{1+\varepsilon_{t+1}}{1+\varepsilon_t},$$

or equivalently

$$\varepsilon_t < \varepsilon_{t+1}.$$

It will be helpful in the following to define the future weighting growth factor

$$\phi_t = \frac{1+\varepsilon_{t+1}}{1+\varepsilon_t} \tag{12}$$

at t, assuming $\varepsilon_t > -1$, since many of our results depend on such ratios. Note that we have the theorem that $\phi_t \gtreqless 1$ if and only if $\varepsilon_{t+1} \gtreqless \varepsilon_t$.

Thus a present-biased discount function will have strictly increasing and positive (for $t > 2$) future weighting factors. In the ϕ_t notation, a present-biased discount function will have $\phi_t > 1$ for $t > 2$. Conversely, a strictly positive and future-biased discount function will have strictly decreasing and negative (for $t > 2$) future weighting factors.[13] To put this in more graphical terms, a present-biased discount function will lie above the exponential function defined by the discounting between time delay 0 and time delay 1, and the divergence between the curves must increase with the time delay. A future-biased discount function will lie below the same exponential function, and the divergence between the curves must also increase (while avoiding zero as we discuss below).

Note that a myopic discount function that is zero for t greater than equal to some $t^* > 1$ does not fit nicely into the categories of a present- or future-biased discount function because it does not satisfy the caveat that the D_t are all positive, which is necessary for the marginal rate of substitution between c_t and c_{t+1} to be defined. There will be a future bias at $t^* - 1$ since at time zero the household would prefer not to consume anything at t^*, but its (t^*-1)-utility is only defined if $c_{t^*} > 0$. On the other hand, there will be a weak present bias at $t \geq t^*$ since at time zero the household will be indifferent between how it allocates consumption between t and $t+1$. However, at time t the household will prefer to have more consumption at t.

[13] A related property of discount functions is increasing patience (**?**). Since Prelec defines this concept in continuous time, we refer the reader to our companion paper in continuous time, **?**, for an understanding of how it translates into a property of the future weighting factors.

3 Curvature of the log consumption profile

Empirically, lifecycle profiles of household consumption are hump-shaped, and time-inconsistency is often invoked as an explanation for this phenomenon. As we discussed in the previous section, ε_t is the parameter that controls the discounting weight of future periods. In this section, we explore how the value of the future weighting factor, ε_t, determines the curvature of the log consumption profile of the household. More precisely, we establish a necessary condition on ε_t under which the log consumption profile would be locally concave (convex) at age $T - t$. This in turn is a necessary condition for the consumption profile to have a local maximum at age $T - t$. [14]

As a first step, we will rewrite the Euler equation in terms of the future weighting discount function. Replacing the general form of discounting function D_t in the household's Euler equation (9) with the form involving the future weighting discounting function (10) gives us

$$c_{t+1} = D_1 R \frac{\sum_{s'=t+1}^{T} D_1^{s'}(1 + \varepsilon_{s'-t})}{\sum_{s=t+1}^{T} D_1^{s}(1 + \varepsilon_{s-t-1})} c_t. \tag{13}$$

In this still exact form, it is more apparent that the Euler equation reduces to the usual $c_{t+1} = D_1 R c_t$ when we have an exponential discounting function and $\varepsilon_2 = \varepsilon_3 = \cdots = \varepsilon_T = 0$. Alternatively, by setting $z = s - t$, we can rewrite this exact Euler equation (13) as

$$\frac{c_{t+1}}{c_t} = D_1 R \frac{\sum_{z'=1}^{T-t} D_1^{z'}(1 + \varepsilon_{z'})}{\sum_{z=1}^{T-t} D_1^{z}(1 + \varepsilon_{z-1})}. \tag{14}$$

As we will often do in the following, it is helpful to consider how this equation simplifies in the limit of small future weighting factors. Since the zeroth-order terms that do not involve the ε_z are the same in the numerator and the denominator of (14), we can rearrange the equation to obtain

$$\frac{c_{t+1}}{c_t} = D_1 R \frac{1 + \frac{\sum_{z'=1}^{T-t} D_1^{z'} \varepsilon_{z'}}{\sum_{s'=1}^{T-t} D_1^{s'}}}{1 + \frac{\sum_{z=1}^{T-t} D_1^{z} \varepsilon_{z-1}}{\sum_{s=1}^{T-t} D_1^{s}}}.$$

[14] If the consumption profile has a local maximum at $t^a st$, it will, of course, also be necessary to have $\frac{c_{t^a st}}{c_{t^a st-1}} > 1 > \frac{c_{t^a st+1}}{c_{t^a st}}$. However, the main hurdle is constructing a model where the growth rate of consumption changes. Adjusting the model so we quantitatively get growth rates both above and below 1 is a matter of calibration. In a partial-equilibrium environment where R is a free parameter, this is trivial. In a general-equilibrium environment, it is more challenging but still less of an issue than getting a concave profile in the first place.

We define the difference operator Δ such that for a time series x_t we have

$$\Delta x_t = x_{t+1} - x_t. \tag{15}$$

To first order in the ε, the Euler equation then approximates to

$$\frac{c_{t+1}}{c_t} = D_1 R \left[1 + \frac{\sum_{z=0}^{T-t-1} D_1^z \Delta\varepsilon_z}{\sum_{s=0}^{T-t-1} D_1^s} \right] + O(\varepsilon^2), \tag{16}$$

where $O(g(x))$ represents an unspecified function smaller than $Mg(x)$ for some $M > 0$ in the limit as $x \to 0$.

Eq. (16) shows that deviations of the Euler equation from the canonical Euler equation $c_{t+1} = D_1 R c_t$ for an exponential discounting function arise because of changes in the future weighting as the delay changes by one period. The effect of a change in future weighting s periods in the future is discounted by D_1^s, so a change in the future weighting at short delays will have a bigger effect than a change at long delays.

We will now focus on the log consumption profile, which will be concave if $\log(\frac{c_{t+1}}{c_t})$ decreases with t. We can take logs of both sides of equation (14) and difference it to obtain

$$\Delta \ln c_t = \ln(D_1 R) + \ln \left(\frac{\sum_{s'=1}^{T-t} D_1^{s'}(1+\varepsilon_{s'})}{\sum_{s=1}^{T-t} D_1^s(1+\varepsilon_{s-1})} \right). \tag{17}$$

Similarly, we can define the second-order difference

$$\Delta^2 \ln c_t = \ln \left(\frac{\sum_{z'=1}^{T-t-1} D_1^{z'}(1+\varepsilon_{z'})}{\sum_{z=1}^{T-t-1} D_1^z(1+\varepsilon_{z-1})} \right) - \ln \left(\frac{\sum_{s'=1}^{T-t} D_1^{s'}(1+\varepsilon_{s'})}{\sum_{s=1}^{T-t} D_1^s(1+\varepsilon_{s-1})} \right),$$

which simplifies to

$$\Delta^2 \ln c_t = \ln \left(\frac{\sum_{z'=1}^{T-t-1} D_1^{z'}(1+\varepsilon_{z'})}{\sum_{s'=1}^{T-t} D_1^{s'}(1+\varepsilon_{s'})} \frac{\sum_{s=1}^{T-t} D_1^s(1+\varepsilon_{s-1})}{\sum_{z=1}^{T-t-1} D_1^z(1+\varepsilon_{z-1})} \right). \tag{18}$$

The log consumption profile will be concave iff $\Delta^2 \ln c_t \leq 0$ for $t = 0, ..., T-2$. If $\Delta^2 \ln c_t < 0$ for all $t = 0, ..., T-2$, then the log consumption profile will be strictly concave. The reverse inequalities will yield convex and strictly convex profiles.[15]

[15]Unlike in continuous time, for the log consumption profile to be strictly concave (convex) at $t+1$ we must have $\Delta^2 \ln c_t$ be negative (positive). If the second difference vanishes, the profile must be locally linear.

Notice that the $\ln D_1 R$ in (17) vanishes from (18). Absent the future weighting factors in (18), the argument of the logarithm is clearly one, so all of the surviving terms on the right-hand side are of first or higher order in the ε_t, corroborating again that the log consumption profile with an exponential discounting function is exactly linear. Any deviation from linearity is driven by the future weighting factors.

Consequently, if the log consumption is concave at $t+1$, we must have

$$\frac{\sum_{z'=1}^{T-t-1} D_1^{z'}(1+\varepsilon_{z'})}{\sum_{s'=1}^{T-t} D_1^{s'}(1+\varepsilon_{s'})} \frac{\sum_{s=1}^{T-t} D_1^{s}(1+\varepsilon_{s-1})}{\sum_{z=1}^{T-t-1} D_1^{z}(1+\varepsilon_{z-1})} \leq 1, \tag{19}$$

which can be simplified to[16]

$$\varepsilon_{T-t} \geq \frac{\sum_{s'=0}^{T-t-2} D_1^{s'}(1+\varepsilon_{s'+1})}{\sum_{s=0}^{T-t-2} D_1^{s}(1+\varepsilon_{s})}(1+\varepsilon_{T-t-1}) - 1. \tag{20}$$

Note that all of the future weighting factors on the right-hand side are at delays shorter than $T-t$. Thus the exact condition for concavity at $t+1$ is a lower bound on ε_{T-t} that depends on future weighting factors at shorter delays.

If $\varepsilon_{T-t-1} = -1$, so $D_{T-t-1} = 0$, there are two possibilities in terms of the shape of the log consumption profile at $t+1$. These depend on ε_{T-t}. If $\varepsilon_{T-t} = -1$ too, then $\Delta \ln c_t = \Delta \ln c_{t+1}$, and the log consumption profile will be linear (and thus both weakly concave and weakly convex) in the vicinity of $t+1$. If, on the other hand, $\varepsilon_{T-t} > -1$, $\Delta \ln c_t > \Delta \ln c_{t+1}$, and the log consumption profile will be strictly concave in the vicinity of $t+1$.

Let us define

$$\overline{\phi}_t = \frac{\sum_{s=0}^{t} D_s \phi_s}{\sum_{s'=0}^{t} D_{s'}}. \tag{21}$$

Then we can conveniently express the following result.

Proposition 1. *If $\varepsilon_s > -1$ for all $s = 0, \ldots, T-t-1$, the log consumption profile will be strictly concave locally at $t+1$ iff*

$$\phi_{T-t-1} > \overline{\phi}_{T-t-2}. \tag{22}$$

The profile will be strictly convex locally if the inequality is reversed.

This follows from (20) using (12). Note that $\phi_t \lesseqgtr \overline{\phi}_{t-1}$ iff $\phi_t \lesseqgtr \overline{\phi}_t$. So an equivalent condition for strict concavity at t is that $\phi_{T-t} > \overline{\phi}_{T-t}$.

[16] See appendix A for details on this calculation.

So the concavity condition at $t+1$ is that ϕ_{T-t-1} is bigger than a weighted average of the ϕ_s for $s = 0, ..., T-t-2$, where the weights are the D_s. That is to say, the log consumption profile will be concave when there are s periods remaining if and only if the future weighting growth factor at s is bigger than a weighted average of the future weighting growth factor at shorter delays.

As we showed in the previous section, a present-biased discount function will have $\phi_t > 1$ for all $t > 0$. Given the assumption of $\varepsilon_0 = \varepsilon_1 = 0$, we have $\phi_1 = 1$. Therefore, local concavity imposes a stronger condition on the shape of consumption profile compared to present bias.

Since a weighted average of a heterogeneous set must be less than the maximum in the set and greater than the minimum in the set, it follows immediately from Proposition 1 that if the ϕ_t are strictly increasing (decreasing) then the log consumption profile will be strictly concave (convex). Moreover, if the ϕ_t are increasing with $\phi_1 > 1$, the log consumption profile will be strictly concave. Likewise, if the ϕ_t are decreasing with $\phi_1 < 1$, the log consumption profile will be strictly convex.

Proposition 2. *For the entire log consumption profile to be strictly concave (convex), the $\Delta\varepsilon_s$ from $s = 1, \ldots, T-1$ must all be positive (negative) and the ϕ_s must all be greater (less) than the weighted average of previous ϕ_s where the weight is the discount factors. Consequently, present bias is a necessary and <u>not</u> sufficient condition for the log consumption profile to be strictly concave.*

This proposition can be proved by induction. Suppose the $\phi_i > 1$ for $s = 1, \ldots, s-1$. Then (22) implies $\phi_s > 1$, and $\varepsilon_{s+1} = \varepsilon_s + \Delta\varepsilon_s > \varepsilon_s > 0$. Note also that each successive iteration of (22) is the necessary condition for the log consumption profile to be concave one period earlier. Thus the condition that

$$\varepsilon_T > \frac{1}{D_1} \frac{\sum_{s'=1}^{T-1} D_1^{s'}(1+\varepsilon_{s'})}{\sum_{s=0}^{T-2} D_1^{s}(1+\varepsilon_{s})}(1+\varepsilon_{T-1}) - 1.$$

is the condition that the log consumption profile is strictly concave between $t = 0$ and $t = 2$. Iterating forward in time, each log consumption growth ratio will depend on one more difference $\Delta\varepsilon_s$ than the ensuing log consumption growth ratio, so $\Delta\varepsilon_s > 0$, or equivalently $\varepsilon_{s+1} > \varepsilon_s$ will be necessary to have the log consumption growth ratio decrease with time. One way to think about this result is that what often gets referred to as present bias is really a case of young households putting extra weight on consumption in the distant future. However, consumption in these future periods gradually matters less to the household as the future gets closer to the present. Therefore, the ε_t must grow with t because that implies the extra weight associated with a specific age gets smaller as we approach that age and the delay time gets shorter.

4 Pareto dominance of the commitment path

In previous sections, we described the household problem with a discounting function that depends on the time to consumption from the present rather than the absolute time when the consumption occurs. Such a household has time-inconsistent preferences and therefore, as **?** noted, the marginal rate of substitution between consumption at different times depends on when the household is evaluating the utility from these consumptions. Consequently, the household at different ages will value consumption plans differently. This multiplicity of selves can substantially complicate welfare analysis.

A common solution to tackle this complication in the literature is to use the preferences of the initial self to evaluate welfare. See, for example, (**??**), **?**, and (**???**). This approach does have its criticisms however. **?** states that there is "no normative foundation" for equating welfare with time-zero preferences.

A more recent literature explores conditions that can be imposed on the discount function under which committing to the initial plan of the time-zero self improves the welfare of all selves over the life cycle as compared to what they would actually obtain over the lifecycle, providing a justification for singling out the preferences of the time-zero self. **?** show that with quasihyperbolic discounting commitment to the time-zero consumption plan can improve the objective function for all selves if the number of selves exceeds a certain threshold which turned out to be quite small in their setting.

In this section, we use our setup to explore conditions on a general discount function in discrete time under which committing to the initial plan will Pareto dominate the realized plan. In a nutshell, exploiting results from the previous section, we find that when the log consumption profile is strictly concave, the difference in welfare between the commitment path and the realized path for each self decomposes into two terms of unambiguous but opposite sign. We can control the magnitude of the contrary term by imposing an upper bound on the future weighting factor at the longest delay. Consequently, we can obtain sufficient conditions for Pareto dominance that are applicable when the log consumption profile is strictly concave, which is easy to assess visually and, with log utility, depends only on the structure of the discount function. Note, however, that these conditions are difficult to satisfy for small T and impossible to satisfy for $T = 2$.

Our results in this section suggest that one does need to be careful when performing welfare analysis with time-inconsistent preferences. Absent other information, it will not be obvious that the consumption path chosen by the initial self will necessarily be the best choice to serve as the benchmark for the purpose of welfare exercises. On the contrary, for most of the parameter space of possible discount functions, neither the commitment path nor the realized path will Pareto dominate the other.

As a starting point, we derive an expression for the difference between realized utility and commitment utility. The realized utility as of time τ is simply the realized value of the household's objective function at time τ, which we have already dealt with in previous sections:

$$U^*_\tau = \sum_{t=\tau}^{T} D_{t-\tau} \ln(c_t).$$

In contrast, the commitment utility at time τ is

$$U^c_\tau = \sum_{t=\tau}^{T} D_{t-\tau} \ln(c_{t|0}), \tag{23}$$

which is what you obtain if you insert the original consumption path as of time 0 into the objective function at time τ. What concerns us most is ΔU_τ which is the difference in utility between the realized plan and the original plan at time τ:

$$\Delta U_\tau = U^*_\tau - U^c_\tau = \sum_{t=\tau}^{T} D_{t-\tau} \ln\left(\frac{c_t}{c_{t|0}}\right). \tag{24}$$

If $\Delta U_\tau > 0$, then following the realized consumption plan provides the household at age τ with a higher utility compared to the initial plan. Conversely, if $\Delta U_\tau < 0$, then committing to the initial plan is optimal for the household at age τ. By definition, the commitment path must maximize lifetime utility at $t = 0$, so we must have $\Delta U_0 \leq 0$. We will say that the commitment path Pareto dominates the realized path if, for all $\tau = 0, \ldots, T$, $\Delta U_\tau \leq 0$ and if, for some $s \in 0, \ldots, T$, $\Delta U_s < 0$. This Pareto dominance provides a compelling justification for helping the household commit to the initial path without having the policy maker impose her norms about which selves matter more on the household.

In what follows, we will also use the term "almost Pareto dominates", which we define as follows. If $\Delta U_\tau \leq 0$ for all $\tau \in 2, \ldots, T$ then the commitment path will almost Pareto dominates the realized path. Likewise, if $\Delta U_\tau \geq 0$ for all $\tau \in 2, \ldots, T$ then the realized path almost Pareto dominate the commitment path[17]. The reason for introducing the term "almost" is because what happens for the second self, i.e. $t = 1$ is somewhat different from what happens for the later selves as we will demonstrate later.

First let us explore ΔU_τ for the exceptional case of a discount function that is not strictly positive, including the case of a myopic discount function. Let $t_* = \min\{t \in \{2, ..., T\} : D_t = 0\}$. Then referring to the commitment consumption rule, $c_{t_*} = 0$. This does not cause anything pathological for U_0 since $D_{t_*} \ln(c_{t_*}) = 0$. However, $D_{t_*-1} \ln(c_{t_*}) = -\infty$, so $U^c_\tau = -\infty$ for $\tau = 1, \ldots, t_*$. In contrast, since we have assumed $D_1 > 0$, the realized path of consumption will be positive for all t, and U^*_τ will be finite for all τ. Thus the commitment path cannot Pareto dominate the realized path. In the myopic case where $D_t = 0$ for all $t \geq t_*$, the realized path will almost Pareto dominate the commitment path.

For the remainder of this section, we will assume the discount function is strictly positive, so the $\varepsilon_t > -1$ for all t, and the ϕ_t are all defined.

We will begin by simplifying the expression for ΔU_τ. Note that for both paths, we have

$$c_t = c_0 \prod_{s=0}^{t-1} \frac{c_{s+1}}{c_s}, \tag{25}$$

so we can rewrite (24) as

$$\Delta U_\tau = \sum_{t=\tau}^{T} \sum_{s=0}^{t-1} D_{t-\tau} \left[\ln\left(\frac{c_{s+1}}{c_s}\right) - \ln\left(\frac{c_{s+1|0}}{c_{s|0}}\right) \right]. \tag{26}$$

This is convenient because we have previously specified the evolution of the realized path in terms of the effective Euler equation (14). The initial plan $c_{t|0}$, i.e. the consumption at period t as determined at period 0, can be obtained from (6):

$$c_{t|0} = D_t R^t c_0 = D_1^t (1 + \varepsilon_t) R^t c_0. \tag{27}$$

Thus consumption growth from t to $t + 1$ along the commitment path simplifies to

$$\frac{c_{t+1|0}}{c_{t|0}} = D_1 R \frac{1 + \varepsilon_{t+1}}{1 + \varepsilon_t}. \tag{28}$$

[17] Note that the realized path can *almost* Pareto dominate the commitment path, but it cannot Pareto dominate the commitment path.

Combining these expressions, 14 and (28), for consumption growth along the two paths, the difference in utility at age τ becomes

$$\Delta U_\tau = \sum_{t=\tau}^{T} \sum_{s=0}^{t-1} D_{t-\tau} \left[\ln \left(\frac{\sum_{z=1}^{T-s} D_1^z (1+\varepsilon_z)}{\sum_{z'=1}^{T-s} D_1^{z'} (1+\varepsilon_{z'-1})} \right) - \ln \left(\frac{1+\varepsilon_{s+1}}{1+\varepsilon_s} \right) \right], \tag{29}$$

which we can rewrite as

$$\Delta U_\tau = \sum_{t=\tau}^{T} \sum_{s=0}^{t-1} D_{t-\tau} \left[\ln \left(\frac{\sum_{z=0}^{T-s-1} D_1^z \phi_z}{\sum_{z'=0}^{T-s-1} D_{z'}} \right) - \ln \phi_s \right].$$

Using (21), this simplifies to

$$\Delta U_\tau = \sum_{t=\tau}^{T} \sum_{s=0}^{t-1} D_{t-\tau} \ln \left(\frac{\overline{\phi}_{T-s-1}}{\phi_s} \right). \tag{30}$$

The remainder of this section will ultimately explore to what extent the concavity condition that we developed in the previous section can help us to determine the sign of (30). Recall that the condition for strict concavity at t is that $\phi_{T-t} > \overline{\phi}_{T-t}$. However, if the log consumption profile is everywhere strictly concave that does not generally imply that $\Delta U_\tau < 0$ since the subscripts of ϕ and $\overline{\phi}$ in (30) are different. The difference in the subscripts arises from how the future weighting factors affect the two paths. The commitment path is obtained by iterating (28), so $c_{t|0}$ depends on $\phi_1, \ldots, \phi_{t-1}$. In contrast, the realized consumption at t depends on the future weighting factors $\varepsilon_2, \ldots, \varepsilon_{T-t}$ that still affect the household's problem at age t. Their combined effect is conveyed by $\overline{\phi}_{T-t}$ instead of $\overline{\phi}_t$.

Nevertheless, for $\tau = T$, i.e. the terminal period,

$$\begin{aligned} \Delta U_T &= \sum_{s=0}^{T-1} D_0 \left[\ln \overline{\phi}_{T-s-1} - \ln \phi_s \right] \\ &= \sum_{s=0}^{T-1} \ln \left(\frac{\overline{\phi}_s}{\phi_s} \right). \end{aligned}$$

Thus strict concavity of the log consumption profile does imply that $\Delta U_T < 0$. Likewise, strict convexity implies that $\Delta U_T > 0$.

For $\tau < T$, strict concavity of the log consumption profile is not sufficient to unambiguously sign the whole sum in (30), but it is sufficient to unambiguously sign the individual terms. We will exploit that result to establish the stronger conditions necessary to ensure that ΔU_τ for $\tau < T$ are also negative. By Proposition 2, strict concavity implies present bias, so the ϕ_t and the $\overline{\phi}_t$, which are a weighted average of the ϕ_t, must be greater than or equal to 1. This means the sign of each of the ΔU_τ for $\tau = 1, \ldots, T-1$ is the result of conflicting forces. As we will see, which sign will prevail is determined largely by ε_T, the future weighting factor at the longest delay. We will first consider how the ΔU_τ depend on ε_T in the general case. Later we will specialize to the case when the log consumption profile is strictly concave.

For $\tau \geq 1$, ΔU_τ only depends on ε_T through its dependence on ϕ_{T-1} and $\overline{\phi}_{T-1}$. We can rewrite (30) as

$$\Delta U_\tau = \sum_{t=\tau}^{T} \sum_{i=0}^{t-1} D_{t-\tau} \ln \overline{\phi}_{T-t+i} - \sum_{t=\tau}^{T} \sum_{i=0}^{t-1} D_{t-\tau} \ln \phi_i. \tag{31}$$

Note that the $D_{t-\tau}$ that appear in this equation will never depend on ε_T for $\tau > 0$, and we do not need to consider ΔU_0 since it must, by definition, be nonpositive.

To differentiate (31) with respect to ε_T, it will be helpful to compute the partial derivatives of the ϕ_t and $\overline{\phi}_t$. From (12), the former is

$$\frac{\partial \phi_t}{\partial \varepsilon_T} = \frac{\delta_{t,T-1}}{1 + \varepsilon_{T-1}}, \tag{32}$$

where δ_{ij} is the Kronecher delta, equaling 1 when i and j are the same and 0 otherwise. Likewise, the latter is, by (21),

$$\frac{\partial \overline{\phi}_t}{\partial \varepsilon_T} = \frac{1}{1 + \varepsilon_{T-1}} \frac{D_{T-1}\delta_{t,T-1}}{\sum_{s=0}^{T-1} D_s}. \tag{33}$$

The corresponding derivatives of $\ln \phi_t$ and $\ln \overline{\phi}_t$ are

$$\frac{\partial \ln \phi_t}{\partial \varepsilon_T} = \frac{\delta_{t,T-1}}{1 + \varepsilon_T} \tag{34}$$

and

$$\frac{\partial \ln \overline{\phi}_t}{\partial \varepsilon_T} = \frac{1}{1 + \varepsilon_{T-1}} \frac{D_{T-1}\delta_{t,T-1}}{\sum_{s=0}^{T-1} D_s \phi_s}. \tag{35}$$

All four of these partial derivatives are nonnegative. Since both terms in (31) include contributions from ϕ_{T-1} and $\overline{\phi}_{T-1}$ that are strictly positive, this means that the first term, which accrues from the realized utility, is unambiguously positive while the second term, which accrues from the subtraction of the commitment utility, is unambiguously negative.

This property that $\frac{\partial \phi_t}{\partial \varepsilon_s}$ and $\frac{\partial \overline{\phi_t}}{\partial \varepsilon_s}$ are nonnegative for all $t = 1, \ldots, T-1$ is unique to $s = T-1$. This elucidates the reason that ε_T is of special significance of all the future weighting factors.

An increase in ε_T will generate a spike in consumption at the end of life on the commitment path that will add to the commitment utility of all the household's selves. As we discussed above, for the singular case of $\tau = T$, this spike will unambiguously decrease ΔU_T since only c_T matters for the welfare of the final self.[18] The initial self will save more to finance this spike in terminal consumption. However, the later selves will end up diverting some of this additional saving to consumption at other ages. Thus an increase in ε_T will increase $c_{T|0}$ more than c_T, resulting in a net decrease of ΔU_T, but this is accomplished by decreasing the $c_{t|0}$ for $t < T$. This latter effect can make the ΔU_τ positive for $\tau < T$, rendering ΔU_τ nonmonotonic.

Let us now focus on the case of $\tau < T$. Partially differentiating (31) with respect to ε_T,

$$
\begin{aligned}
\frac{\partial \Delta U_\tau}{\partial \varepsilon_T} &= \frac{1}{1+\varepsilon_{T-1}} \sum_{t=\tau}^{T} \sum_{i=0}^{t-1} D_{t-\tau} \frac{D_{T-1}\delta_{T-t+i,T-1}}{\sum_{s=0}^{T-1} D_s \phi_s} - \frac{1}{1+\varepsilon_T} \sum_{t=\tau}^{T} \sum_{i=0}^{t-1} D_{t-\tau}\delta_{i,T-1} \\
&= \frac{1}{1+\varepsilon_{T-1}} \sum_{t=\tau}^{T} D_{t-\tau} \frac{D_{T-1}}{\sum_{s=0}^{T-1} D_s \phi_s} - \frac{D_{T-\tau}}{1+\varepsilon_T} \\
&= \sum_{t=\tau}^{T} D_{t-\tau} \frac{D_1^{T-1}}{\sum_{i=0}^{T-1} D_1^i (1+\varepsilon_{i+1})} - \frac{D_{T-\tau}}{1+\varepsilon_T} \\
&= D_1^{T-\tau} \left[\frac{\sum_{t=\tau}^{T} D_1^{t-1}(1+\varepsilon_{t-\tau})}{\sum_{i=0}^{T-1} D_1^i (1+\varepsilon_{i+1})} - \frac{1+\varepsilon_{T-\tau}}{1+\varepsilon_T} \right] \\
&= D_1^{T-\tau} \frac{(1+\varepsilon_T)\sum_{t=\tau}^{T} D_1^{t-1}(1+\varepsilon_{t-\tau}) - (1+\varepsilon_{T-\tau})\sum_{i=0}^{T-1} D_1^i (1+\varepsilon_{i+1})}{(1+\varepsilon_T)\sum_{i=0}^{T-1} D_1^i (1+\varepsilon_{i+1})}
\end{aligned}
$$

and finally, after cancelling like terms in the numerator, we obtain

$$
\frac{\partial \Delta U_\tau}{\partial \varepsilon_T} = D_1^{T-\tau} \frac{(1+\varepsilon_T)\sum_{t=\tau}^{T-1} D_1^{t-1}(1+\varepsilon_{t-\tau}) - (1+\varepsilon_{T-\tau})\sum_{i=0}^{T-2} D_1^i (1+\varepsilon_{i+1})}{(1+\varepsilon_T)\sum_{i=0}^{T-1} D_1^i (1+\varepsilon_{i+1})}. \tag{36}
$$

Notice that ε_T only appears once, in the first term of the numerator. That is to say, the numerator is linear in ε_T. Moreover, given our assumptions that $D_1 > 0$ and the $\varepsilon_t > -1$, both the whole denominator and the coefficient of ε_T in the numerator are positive.[19] This

[18]This accounts for why strict concavity alone can guarantee that $\Delta U_T < 0$ since strict concavity imposes a strictly positive upper bound on ε_T that depends on the other future weighting factors, which must also be positive. This is a tighter bound on ε_T than what we need to get a negative ΔU_T.

[19]For $\tau = T$, the numerator does not depend on ε_T and with those assumptions is strictly negative. Thus

yields our next result.

Proposition 3. *For $\tau = 1, \ldots, T-1$, the difference ΔU_τ of lifetime welfare between the realized path and the commitment path is a U-shaped function of the future weighting factor ε_T at the longest delay. The global minimum of ΔU_τ as a function of ε_T is*

$$\underline{\varepsilon}_T^\tau \equiv D_1^{1-\tau} \frac{\sum_{i=0}^{T-2} D_i \phi_i}{\sum_{t=0}^{T-\tau-1} D_t} (1 + \varepsilon_{T-\tau}) - 1. \tag{37}$$

Thus ΔU_τ is a strictly decreasing function of ε_T for $\varepsilon_T < \underline{\varepsilon}_T^\tau$ and ΔU_τ is strictly increasing for $\varepsilon_T > \underline{\varepsilon}_T^\tau$.

This follows immediately from (36). The location of the global minimum is obtained by solving for the ε_T that sets the numerator of (36) to zero. See Appendix B for details.

Since our objective is to characterize where the ΔU_τ are all negative, we focus on where the U_τ bottom out for $\tau = 1, \ldots, T-1$. A necessary condition for the commitment path to Pareto dominate the realized path is that the future weighting factors must be chosen so that ε_T is sufficiently close to $\underline{\varepsilon}_T^\tau$ for all $\tau \in 1, \ldots, T-1$.[20]

We can rewrite (37) as

$$\frac{1 + \underline{\varepsilon}_T^\tau}{1 + \varepsilon_{T-\tau}} = D_1^{1-\tau} \frac{\sum_{s=0}^{T-2} D_s}{\sum_{s'=0}^{T-\tau-1} D_{s'}} \overline{\phi}_{T-2}, \tag{38}$$

so $1 + \underline{\varepsilon}_T^\tau$ is proportional to the weighted average $\overline{\phi}_{T-2}$ of the $\phi_0, \ldots, \phi_{T-1}$. If $\tau = 1$, the right-hand side is exactly $\overline{\phi}_{T-2}$. For $\tau = 2, \ldots, T-1$, the right-hand side will be strictly greater than $\overline{\phi}_{T-2}$ since the discount factors are all positive.

The different behavior of $\underline{\varepsilon}_T^\tau$ for $\tau = 1$ and $\tau > 1$ has some interesting consequences in the special case where all the future weighting factors ε_t for $t < T$ vanish. In that case $\overline{\phi}_{T-2} = 1$ and $\varepsilon_{T-\tau} = 0$, so we have $\underline{\varepsilon}_T^1 = 0 < \underline{\varepsilon}_T^\tau$ for $\tau > 1$. In this limit, $\Delta U_\tau = 0$ for all τ in the limit as $\varepsilon_T = 0$ since in that limit the discount function is exactly exponential and there is no time-inconsistency. Since the global minimum of ΔU_1 is exactly zero, this implies that $\Delta U_1 \geq 0$ with equality only when $\varepsilon_T = 0$. Thus the second self will always prefer the realized path to the commitment path if $\varepsilon_2 = \cdots = \varepsilon_{T-1} = 0$. In contrast, a later self at $\tau > 1$ will definitely prefer the commitment path to the realized path if $\varepsilon_T \in (0, \underline{\varepsilon}_T^\tau]$ and even for some

ΔU_T is strictly decreasing in ε_T.

[20]Additional conditions will be necessary to guarantee that ΔU_τ is in fact negative in the neighborhood of this minimum.

$\varepsilon_T > \underline{\varepsilon}_T^\tau$ since $\Delta U_\tau < 0$ when $\varepsilon_T = \underline{\varepsilon}_T^\tau$. Nevertheless, the commitment path cannot Pareto dominate the realized path if $\varepsilon_2 = \cdots = \varepsilon_{T-1} = 0$.

The intuition behind what we explained above is that the second self will always prefer the realized path to the commitment path in this case because there is time-inconsistency between the preferences of the initial self and the later selves, but there is no inconsistency between the preferences of the later selves. The initial self puts a different weight on c_T than an exponential discounter would, but the later selves are all exponential discounters. Obviously, the second self must prefer the path that is optimal for her over the path that is optimal for the initial self or there is no time-inconsistency. Since the third and later selves have no reason to further alter the consumption path planned by the second self, the optimal plan for the second self will be the realized path. Thus the second self must prefer the realized path to the initial path.

Whether the third and later selves will also prefer the realized path to the initial path depends on how the two paths deviate. If we decrease ε_T from zero while holding the other future weights fixed at zero, the initial self will put less and less weight on the terminal consumption, going to zero as $\varepsilon_T \to -1$. Since the other selves will not put less weight on c_T, the initial path will become more and more objectionable as compared to the realized path as ε_T gets more negative. This is true for the second self as well as for the later selves.

For the third and later selves, excepting the terminal self, what happens to ΔU_T as ε_T increases from zero correctly informs our intuition for small ε_T. The initial self puts more weight on $c_{T|0}$, and the welfare of all these selves depends on c_T so that makes the initial path more preferable as compared to the realized path. However, the increase in $c_{T|0}$ does not come without a cost. The household's total wealth remains unchanged, so the initial self can only increase $c_{T|0}$ at the expense of reducing the initial allocation of consumption to earlier ages. That does not matter for the terminal self, but it will matter for $\tau = 2, \ldots, T-1$. This is a second-order effect, so it is dominated by the effect of a high $c_{T|0}$ for small ε_T, but for large enough ε_T these selves will look upon the realized path more favorably. Indeed, for these τ we can show that ΔU_τ will eventually turn positive as we keep increasing ε_T. This leads us to the following proposition:

Proposition 4. *For $\tau = 1, \ldots, T-1$, for any given choice of $\varepsilon_2, \ldots, \varepsilon_{T-1}$, we have*

$$\lim_{\varepsilon_T \to \infty} \Delta U_\tau = \lim_{\varepsilon_T \to -1} \Delta U_\tau = \infty. \tag{39}$$

See Appendix C for the proof. Note that with $\varepsilon_{T-1} \to -1$, we have $\phi_{T-1} = 0$ and

$\overline{\phi}_{T-1} > 0$, while all of the other ϕ_t and $\overline{\phi}_t$ remain positive. Therefore, the second equality in (39) holds since all of the ΔU_τ for $\tau \geq 1$ depend on $\ln \phi_{T-1}$ with a positive coefficient in (31).

As a consequence of Propositions 3 and 4, we conclude that to have $\Delta U_\tau < 0$ will require ε_T to lie within a finite interval (A^τ, B^τ) that depends on the other future weighting factors and includes $\underline{\varepsilon}_T^\tau$. We do not have a simple characterization of the exact bounds on this interval beyond what can be obtained by numerical solution of the equation $\Delta U_\tau = 0$. Nevertheless, Proposition 1 implies that the log consumption profile will be strictly concave if

$$\varepsilon_T > F = (1 + \varepsilon_{T-1})\overline{\phi}_{T-1} - 1. \tag{40}$$

With a strictly concave log consumption profile, we can sign the individual terms of (31) and thus obtain an upper bound G_τ on ε_T such that $\Delta U_\tau < 0$. In other words, $(F, G_\tau) \subseteq (A^\tau, B^\tau)$. Thus F will be an upper bound on the lower bound A^τ, and G_τ will be a lower bound on the upper bound B^τ.

However, we must emphasize that the interval (A^τ, B^τ) could be empty. This can be easily demonstrated for the case of $T = 3$, i.e. when the household lives for four periods.[21] Since there are only two future weighting factors in this case, it is feasible to graph the numerical solution of where all the ΔU_τ are negative. So, before we characterize bounds on the bounds of the intervals, let us first see what we can learn when we can diagram the whole parameter space with a two-dimensional graph.

In figure 2, we show for two calibrations of D_1 graphs of a neighborhood of the origin in which white pixels show pairs $(\varepsilon_3.\varepsilon_2)$ for which the commitment path Pareto dominates the realized path whereas the black pixels correspond to pairs where at least one self prefers the realized path. In both cases we imagine a period is twenty years so a total life span is 80 years. In 2a, $D_1 = 0.44$, or 0.96 in annual terms. In 2b, it is 1.42, or 1.02 in annual terms.

As we can see in figure 2, Pareto dominance of the commitment path holds over a larger subset of the parameter space when D_1 is 1.02 as opposed to 0.96. In both calibrations the subset where the commitment path Pareto dominates the realized path lies entirely within the first quadrant, where both ε_2 and ε_3 are positive. However when D_1 is 1.02, the subset radiates from the origin. When D_1 is 0.96, Pareto dominance only occurs when ε_2 and ε_3 are both large and positive.

As one might expect given the previous discussion of what happens when $\varepsilon_2 = \varepsilon_{T-1} = 0$,

[21] When $T = 2$, ε_2 is the only future weighting factor, so our previous argument that the second self must prefer the realized path over the commitment path applies.

Figure 2: Pixel plot of the combinations of ε_2 and ε_3 for which the commitment path Pareto dominate the realized path

D1-96.pdf	D1-102.pdf
a . for $D_1 = 0.96$	b . for $D_1 = 1.02$

Note: on both graphs, the bright area shows the region that Pareto condition holds. We have $\varepsilon_2 \in [-1, 10]$ on the y axis and $\varepsilon_3 \in [-1, 10]$ on the x axis.

the condition that $\Delta U_1 \leq 0$ is usually the tightest constraint on Pareto dominance. For $\tau > 1$, there must be discount functions in a neighborhood of the origin, i.e. where the discount function is exponential, such that $\Delta U_\tau < 0$, but that need not be true for $\tau = 1$ as demonstrated by figure 2a.

Proposition 3 established that ΔU_1 is minimized at $\varepsilon_T = 0$ when $\varepsilon_2 = \cdots = \varepsilon_{T-1} = 0$, so $\frac{\partial \Delta U_1}{\partial \varepsilon_T} = 0$ at the origin. In fact, all of the first partial derivatives of ΔU_1 vanish at the origin. We state this observation in the next proposition:

Proposition 5. *The gradient* $\nabla \Delta U_1 = 0$ *when* $\varepsilon_2 = \cdots = \varepsilon_T = 0$.

The proof is in Appendix D. This is a consequence of the fact that the commitment and realized plans are the same at $t = 0$ so $c_0 = c_{0|0}$. We have

$$\begin{aligned} D_1 \Delta U_1 &= D_1 \sum_{t=1}^{T} D_{t-1} \ln\left(\frac{c_t}{c_{t|0}}\right) \\ &= \ln\left(\frac{c_0}{c_{0|0}}\right) + \sum_{t=1}^{T} D_t \ln\left(\frac{c_t}{c_{t|0}}\right) + \sum_{t=1}^{T} (D_1 D_{t-1} - D_t) \ln\left(\frac{c_t}{c_{t|0}}\right) \\ &= \Delta U_0 + \sum_{t=1}^{T} (D_1 D_{t-1} - D_t) \ln\left(\frac{c_t}{c_{t|0}}\right). \end{aligned}$$

When the future weighting factors all vanish, both factors, $(D_1 D_{t-1} - D_t)$ and $\ln\left(\frac{c_t}{c_{t|0}}\right)$, of the last term vanish at the origin so partial derivatives of this last term also vanish at the origin. Consequently, the gradient of ΔU_1 is proportional to the gradient of ΔU_0. Since the initial self must prefer the commitment path, ΔU_0 must be maximized at the origin. Therefore, its gradient must vanish, and the gradient of ΔU_1 must also vanish. This intuition does not extend to later τ because $\ln\left(\frac{c_\tau}{c_{\tau|0}}\right)$ for $\tau \geq 1$ only vanishes at the origin, so it has a nonzero gradient. Thus we have that the story about whether the $\tau = 1$ self prefers the initial path or the realized path is different from the story for the later selves.

However, while the gradient of ΔU_1 must vanish at the origin, ΔU_1 differs from ΔU_0 in that the origin is not a global maximum of ΔU_1. On the contrary, we have already demonstrated that ΔU_1 is minimized with respect to ε_T at the origin. But the origin is only a global minimum of ΔU_1 for the trivial case of $T = 2$ when there is only the one future weighting factor, ε_2. In Appendix E we calculate the Hessian of ΔU_1 at the origin for $T = 3$. The diagonal elements are both positive, so the second self will prefer the realized path both if $\varepsilon_2 (\neq 0)$ is small in magnitude while $\varepsilon_3 = 0$ and if $\varepsilon_3 \neq 0$ while $\varepsilon_2 = 0$. Nevertheless, the determinant of the Hessian is $D_1^3(1 - D_1^3)$. Thus if $D_1 > 1$, ΔU_1 has a saddlepoint at the origin. This explains why the subset of the parameter space in figure 2b where the commitment path Pareto dominates the realized path radiates from the origin. For that choice of D_1, which is bigger than one, if we increase both ε_2 and ε_3 from zero in a right proportion, the second self will prefer the commitment path to the realized path. For $D_1 < 1$, if we depart from the origin in any direction, the second self will for some positive distance prefer the realized path over the commitment path.

Proposition 3 tells us where in the parameter space we will have $\frac{\partial \Delta U_1}{\partial \varepsilon_T} = 0$ and $\frac{\partial^2 \Delta U_1}{\partial \varepsilon_T^2} > 0$, and it is certainly reasonable to search the neighborhood of such points for candidate discount functions that have the commitment path Pareto dominate the realized path. Nevertheless it is not sufficient for ΔU_1 to be minimized with respect to ε_T to guarantee that $\Delta U_1 < 0$.

As we stated earlier, propositions 3 and (39) do imply that for a given choice of the other future weighting factors, if $\Delta U_\tau < 0$ at $\underline{\varepsilon}_T^\tau$ there will be an interval $[A^\tau, B^\tau]$ such that $\Delta U_\tau \leq 0$ for $\varepsilon_T \in [A^\tau, B^\tau]$, which includes $\underline{\varepsilon}_T^\tau$. We will then have that the commitment path Pareto dominates the realized path iff

$$\varepsilon_T \in \cap_{\tau=2}^{T-1} (A^\tau, B^\tau). \tag{41}$$

We can use our results from Section 3 and (31) to construct upper and lower bounds on ε_T that demarcate a proper subset of this band. That is to say, holding $\varepsilon_2, \ldots, \varepsilon_{T-1}$ fixed, if ε_T is within these bounds then the commitment path will Pareto dominate the realized path. These bounds will not, however, enclose all ε_T such that with these $\varepsilon_2, \ldots, \varepsilon_{T-1}$ the commitment path will Pareto dominate the realized path.

Proposition 6. *If a discount function yields a strictly concave log consumption profile (which implies lower bounds on all of the ε_t), then if, for $s = 1, \ldots, T-2$,*

$$\varepsilon_T < G_s = \min\{G_s^1, \ldots, G_s^{T-1}\} \tag{42}$$

the commitment path will Pareto dominate the realized path, where

$$G_s^\tau = \exp\left(\frac{\sum_{i=0}^{s} \left[P_i^\tau \left(\ln \frac{\phi_i}{\bar{\phi}_i} + \ln(1+\varepsilon_i) + \sum_{j=i+1}^{s} \ln \phi_j \right) + Q_i^\tau \ln \phi_i \right]}{\sum_{i'=0}^{T-1} P_{i'}^\tau} \right) - 1, \tag{43}$$

$$P_i^\tau = \sum_{t=\max\{\tau, T-i\}}^{T} D_{t-\tau}, \tag{44}$$

and

$$Q_i^\tau = \sum_{t=\max\{\tau, i+1\}}^{T} D_{t-\tau}. \tag{45}$$

See Appendix F for the proof.

Note that we actually define a sequence of upper bounds in (43), where $G_1^\tau \leq \cdots \leq G_{T-2}^\tau$. The right-hand side of (43) does not depend on ε_T, but as s is increased it includes more of the ϕ_t and $\frac{\phi_t}{\bar{\phi}_t}$. Thus as we increase s, G_s will be a looser upper bound on the lower limit A^T of (41), but it will be more complicated to calculate.

5 Concluding remarks

Present and future bias are defined as a form of time-inconsistency in which individuals' behavior regarding trade-offs in consumption at the beginning and end of the same time interval vary between the near future and the far future. The common approach for modeling this bias is with a relative discounting function, i.e. a form of discounting function which is a function of the time to consumption from the decision-making present. As a consequence, the optimal plan changes as an individual advances through the life span. A functional form that is widely used in the literature as a proxy for non-exponential discounting functions is the quasihyberbolic functional form, which is used to discuss the shape of the consumption profile and the preferences of different selves.

In this paper we proposed a general representation of relative discounting functions that allows us to focus on how the discounting function deviates from an exponential discounting function, which does not exhibit time-inconsistency. We term the perturbation away from the exponential case a *future weighting factor* ε_t. This specific format of the discounting function provides a simple way to depict a future bias by having all ε_t be negative and decreasing for $t > 1$, and a present bias by having all ε_t be positive and increasing for $t > 1$.

We find that for the log consumption profile to be locally concave, which is necessary at the peak of a hump-shaped consumption profile, a future weighting growth factor must be bigger than the weighted average of future weighting growth factor at shorter delays, where the weights are the discount factor. This means that a present bias is a necessary but not sufficient condition for the entire log consumption profile to be strictly concave.

Also, using the proposed future weighting functional form, we explored the conditions on the future weighting factors under which the consumption profile that is determined in the first period of life will Pareto dominate the realized consumption profiles chosen at each period. We find a sufficient condition for this Pareto dominance is that the realized log consumption profile be strictly concave and the future weighting factor at the longest delay not be too large. This result is especially useful because Pareto dominance of the initial path is often used to motivate how one performs welfare analysis in these models with time-inconsistent preferences, where choosing a reference consumption plan for the analysis is a point of controversy in the literature. The results of our study suggest that one has to be cautious when analyzing welfare with time-inconsistent preferences. The consumption path chosen by one's initial self is not necessarily the best choice to serve as the benchmark for welfare purposes without additional information. As a matter of fact, neither the commitment path nor the realized path will dominate each other for most of the parameter space of possible discount functions.

Appendices

A Simplifying the Concavity Condition

The log consumption profile is concave at $t+1$ iff we have

$$\frac{\sum_{z'=1}^{T-t-1} D_1^{z'}(1+\varepsilon_{z'})}{\sum_{s'=1}^{T-t} D_1^{s'}(1+\varepsilon_{s'})}\frac{\sum_{s=1}^{T-t} D_1^{s}(1+\varepsilon_{s-1})}{\sum_{z=1}^{T-t-1} D_1^{z}(1+\varepsilon_{z-1})} \leq 1.$$

We can rearrange this inequality as follows.

$$\frac{\sum_{z'=1}^{T-t-1} D_1^{z'}(1+\varepsilon_{z'})}{\sum_{s'=1}^{T-t} D_1^{s'}(1+\varepsilon_{s'})} \leq \frac{\sum_{z=1}^{T-t-1} D_1^{z}(1+\varepsilon_{z-1})}{\sum_{s=1}^{T-t} D_1^{s}(1+\varepsilon_{s-1})}$$

$$1-\frac{D_1^{T-t}(1+\varepsilon_{T-t})}{\sum_{s'=1}^{T-t} D_1^{s'}(1+\varepsilon_{s'})} \leq 1-\frac{D_1^{T-t}(1+\varepsilon_{T-t-1})}{\sum_{s=1}^{T-t} D_1^{s}(1+\varepsilon_{s-1})}$$

$$\frac{1+\varepsilon_{T-t-1}}{\sum_{s=1}^{T-t} D_1^{s}(1+\varepsilon_{s-1})} \leq \frac{1+\varepsilon_{T-t}}{\sum_{s'=1}^{T-t} D_1^{s'}(1+\varepsilon_{s'})}$$

We wish to isolate ε_{T-t}, which appears in both the numerator and the denominator of the right-hand side.

$$\frac{1+\varepsilon_{T-t-1}}{\sum_{s=0}^{T-t-1} D_1^{s+1}(1+\varepsilon_{s})} \leq \frac{1+\varepsilon_{T-t}}{\sum_{s'=1}^{T-t-1} D_1^{s'}(1+\varepsilon_{s'})+D_1^{T-t}(1+\varepsilon_{T-t})}$$

$$\frac{\sum_{s'=1}^{T-t-1} D_1^{s'}(1+\varepsilon_{s'})+D_1^{T-t}(1+\varepsilon_{T-t})}{\sum_{s=0}^{T-t-1} D_1^{s+1}(1+\varepsilon_{s})}(1+\varepsilon_{T-t-1}) \leq 1+\varepsilon_{T-t}$$

$$\frac{\sum_{s'=1}^{T-t-1} D_1^{s'}(1+\varepsilon_{s'})}{\sum_{s=0}^{T-t-1} D_1^{s+1}(1+\varepsilon_{s})}(1+\varepsilon_{T-t-1}) \leq (1+\varepsilon_{T-t})\left[1-\frac{D_1^{T-t}(1+\varepsilon_{T-t-1})}{\sum_{s=0}^{T-t-1} D_1^{s+1}(1+\varepsilon_{s})}\right]$$

$$\frac{\sum_{s'=1}^{T-t-1} D_1^{s'}(1+\varepsilon_{s'})}{\sum_{s=0}^{T-t-1} D_1^{s+1}(1+\varepsilon_{s})}(1+\varepsilon_{T-t-1}) \leq (1+\varepsilon_{T-t})\left[\frac{\sum_{z'=0}^{T-t-2} D_1^{z'+1}(1+\varepsilon_{z'})}{\sum_{z=0}^{T-t-1} D_1^{z+1}(1+\varepsilon_{z})}\right]$$

Thus we obtain the condition

$$\frac{\sum_{z=1}^{T-t-1} D_1^{z}(1+\varepsilon_{z})}{\sum_{z'=0}^{T-t-2} D_1^{z'+1}(1+\varepsilon_{z'})}(1+\varepsilon_{T-t-1}) \leq 1+\varepsilon_{T-t} \tag{46}$$

for concavity at $t+1$.

B Derivation of Eq. (37)

$$
\begin{aligned}
\underline{\varepsilon}_T^{\tau} &= \frac{\sum_{i=0}^{T-2} D_1^i (1+\varepsilon_{i+1})}{\sum_{t=\tau}^{T-1} D_1^{t-1}(1+\varepsilon_{t-\tau})}(1+\varepsilon_{T-\tau}) - 1 \\
&= \frac{\sum_{i=0}^{T-2} D_1^i (1+\varepsilon_{i+1})}{\sum_{t=0}^{T-\tau-1} D_1^{t+\tau-1}(1+\varepsilon_{t})}(1+\varepsilon_{T-\tau}) - 1 \\
&= D_1^{1-\tau}\frac{\sum_{i=0}^{T-2} D_1^i (1+\varepsilon_{i+1})}{\sum_{t=0}^{T-\tau-1} D_1^{t}(1+\varepsilon_{t})}(1+\varepsilon_{T-\tau}) - 1 \\
&= D_1^{1-\tau}\frac{\sum_{i=0}^{T-2} D_i \phi_i}{\sum_{t=0}^{T-\tau-1} D_t}(1+\varepsilon_{T-\tau}) - 1.
\end{aligned}
$$

C Derivation of limits of ΔU_τ

$$
\begin{aligned}
\lim_{\varepsilon_T\to\infty} \Delta U_\tau &= \lim_{\varepsilon_T\to\infty} \sum_{t=\tau}^{T}\sum_{i=0}^{t-1} D_{t-\tau} \ln \frac{\overline{\phi}_{T-t+i}}{\phi_i} \\
&= \sum_{t=\tau}^{T}\sum_{i=0}^{t-1} D_{t-\tau} \lim_{\varepsilon_T\to\infty} \ln \overline{\phi}_{T-t+i} - \sum_{t=\tau}^{T}\sum_{i=0}^{t-1} D_{t-\tau} \lim_{\varepsilon_T\to\infty} \ln \phi_i \\
&= \sum_{t=\tau}^{T}\sum_{i=0}^{t-2} D_{t-\tau} \lim_{\varepsilon_T\to\infty} \ln \overline{\phi}_{T-t+i} + \sum_{t=\tau}^{T} D_{t-\tau} \lim_{\varepsilon_T\to\infty} \ln \overline{\phi}_{T-t+(t-1)} - \sum_{t=\tau}^{T-1}\sum_{i=0}^{t-1} D_{t-\tau} \lim_{\varepsilon_T\to\infty} \ln \phi_i \\
&\quad - \sum_{i=0}^{T-1} D_{T-\tau} \lim_{\varepsilon_T\to\infty} \ln \phi_i \\
&= \sum_{t=\tau}^{T}\sum_{i=0}^{t-2} D_{t-\tau} \ln \overline{\phi}_{T-t+i} + \sum_{t=\tau}^{T} D_{t-\tau} \lim_{\varepsilon_T\to\infty} \ln \overline{\phi}_{T-1} - \sum_{t=\tau}^{T-1}\sum_{i=0}^{t-1} D_{t-\tau} \ln \phi_i - \sum_{i=0}^{T-2} D_{T-\tau} \ln \phi_i - D_{T-}
\end{aligned}
$$

$$
\lim_{\varepsilon_T\to\infty} \ln \phi_{T-1} = \lim_{\varepsilon_T\to\infty} \ln \varepsilon_T - \ln(1+\varepsilon_{T-1}) = \lim_{\varepsilon_T\to\infty} \ln \varepsilon_T
$$

$$\begin{aligned}
\lim_{\varepsilon_T\to\infty} \ln \overline{\phi}_{T-1} &= \lim_{\varepsilon_T\to\infty} \ln\left(\frac{\sum_{s=0}^{T-1} D_s\phi_s}{\sum_{s'=0}^{T-1} D_{s'}}\right) \\
&= \lim_{\varepsilon_T\to\infty} \ln\left(D_1^{T-1}(1+\varepsilon_T)\right) - \ln\left(\sum_{s'=0}^{T-1} D_{s'}\right) \\
&= \ln D_1^{T-1} - \ln\left(\sum_{s'=0}^{T-1} D_{s'}\right) + \lim_{\varepsilon_T\to\infty} \ln\varepsilon_T = \lim_{\varepsilon_T\to\infty} \ln\varepsilon_T
\end{aligned}$$

For $\tau < T$,

$$\begin{aligned}
\lim_{\varepsilon_T\to\infty} \Delta U_\tau &= \sum_{t=\tau}^{T}\sum_{i=0}^{t-2} D_{t-\tau}\ln\overline{\phi}_{T-t+i} - \sum_{t=\tau}^{T-1}\sum_{i=0}^{t-1} D_{t-\tau}\ln\phi_i - \sum_{i=0}^{T-2} D_{T-\tau}\ln\phi_i + \left(\sum_{t=\tau}^{T-1} D_{t-\tau}\right)\lim_{\varepsilon_T\to\infty}\ln\varepsilon_T \\
&= \left(\sum_{t=\tau}^{T-1} D_{t-\tau}\right)\lim_{\varepsilon_T\to\infty}\ln\varepsilon_T > 0.
\end{aligned}$$

D Gradient of ΔU_1 at the Origin

$$\Delta U_\tau = \sum_{t=\tau}^{T}\sum_{i=0}^{t-1} D_{t-\tau}\ln\frac{\overline{\phi}_{T-t+i}}{\phi_i}.$$

$$\Delta U_1 = \sum_{s=1}^{T}\sum_{i=0}^{s-1} D_{s-1}\ln\frac{\overline{\phi}_{T-s+i}}{\phi_i}$$

$$\frac{\partial\Delta U_1}{\partial\varepsilon_t} = \sum_{s=1}^{T}\sum_{i=0}^{s-1}\left[\frac{\partial D_{s-1}}{\partial\varepsilon_t}\ln\frac{\overline{\phi}_{T-s+i}}{\phi_i} + D_{s-1}\left(\frac{\partial\ln\overline{\phi}_{T-s+i}}{\partial\varepsilon_t} - \frac{1}{1+\varepsilon_{i+1}}\frac{\partial\varepsilon_{i+1}}{\partial\varepsilon_t} + \frac{1}{1+\varepsilon_i}\frac{\partial\varepsilon_i}{\partial\varepsilon_t}\right)\right]$$

$$\overline{\phi}_s = \frac{\sum_{i=0}^{s} D_i\phi_i}{\sum_{j=0}^{s} D_j}$$

$$\frac{\partial\ln\overline{\phi}_s}{\partial\varepsilon_t} = \frac{\sum_{i=0}^{s}\left(\frac{\partial D_i}{\partial\varepsilon_t}\phi_i + D_i\frac{\partial\phi_i}{\partial\varepsilon_t}\right)}{\sum_{i'=0}^{s} D_{i'}\phi_{i'}} - \frac{\sum_{j=0}^{s}\frac{\partial D_j}{\partial\varepsilon_t}}{\sum_{j'=0}^{s} D_{j'}}$$

$$\frac{\partial\phi_i}{\partial\varepsilon_t} = \frac{\partial}{\partial\varepsilon_t}\left(\frac{1+\varepsilon_{i+1}}{1+\varepsilon_i}\right) = \frac{\delta_{i+1,t}}{1+\varepsilon_i} - \frac{1+\varepsilon_{i+1}}{(1+\varepsilon_i)^2}\delta_{i,t}$$

$$\left.\frac{\partial\phi_i}{\partial\varepsilon_t}\right|_{\varepsilon=0} = \delta_{i+1,t} - \delta_{i,t}$$

$$\left.\frac{\partial \ln \overline{\phi}_s}{\partial \varepsilon_t}\right|_{\varepsilon=0} = \frac{\sum_{i=0}^{s}\left(\frac{\partial D_i}{\partial \varepsilon_t} + D_1^i(\delta_{i+1,t} - \delta_{i,t})\right)}{\sum_{i'=0}^{s} D_1^{i'}} - \frac{\sum_{j=0}^{s}\frac{\partial D_j}{\partial \varepsilon_t}}{\sum_{j'=0}^{s} D_1^{j'}} = \frac{\sum_{i=0}^{s} D_1^i(\delta_{i+1,t} - \delta_{i,t})}{\sum_{i'=0}^{s} D_1^{i'}}$$

$$\begin{aligned}
\left.\frac{\partial \Delta U_1}{\partial \varepsilon_t}\right|_{\varepsilon=0} &= \sum_{s=1}^{T}\sum_{i=0}^{s-1} D_1^{s-1}\left(\frac{\partial \ln \overline{\phi}_{T-s+i}}{\partial \varepsilon_t} - \delta_{i+1,t} + \delta_{i,t}\right) \\
&= \sum_{s=1}^{T}\sum_{i=0}^{s-1} D_1^{s-1}\left(\frac{\sum_{j=0}^{T-s+i} D_1^j(\delta_{j+1,t} - \delta_{j,t})}{\sum_{s'=0}^{T-s+i} D_1^{s'}} - \delta_{i+1,t} + \delta_{i,t}\right)
\end{aligned}$$

Suppose $T = 2$.

$$\left.\frac{\partial \Delta U_1}{\partial \varepsilon_2}\right|_{\varepsilon=0} = \sum_{s=1}^{2}\sum_{i=0}^{s-1} D_1^{s-1}\left(\frac{\sum_{j=0}^{2-s+i} D_1^j(\delta_{j+1,2} - \delta_{j,2})}{\sum_{s'=0}^{2-s+i} D_1^{s'}} - \delta_{i+1,2} + \delta_{i,2}\right)$$

$$\begin{aligned}
\left.\frac{\partial \Delta U_1}{\partial \varepsilon_2}\right|_{\varepsilon=0} &= \sum_{i=0}^{0} D_1^0\left(\frac{\sum_{j=0}^{1+i} D_1^j(\delta_{j+1,2} - \delta_{j,2})}{\sum_{s'=0}^{1+i} D_1^{s'}} - \delta_{i+1,2} + \delta_{i,2}\right) \\
&\quad + \sum_{i=0}^{1} D_1^1\left(\frac{\sum_{j=0}^{i} D_1^j(\delta_{j+1,2} - \delta_{j,2})}{\sum_{s'=0}^{i} D_1^{s'}} - \delta_{i+1,2} + \delta_{i,2}\right) \\
&= \frac{\sum_{j=0}^{1} D_1^j(\delta_{j+1,2} - \delta_{j,2})}{\sum_{s'=0}^{1} D_1^{s'}} \\
&\quad + D_1\frac{\sum_{j=0}^{0} D_1^j(\delta_{j+1,2} - \delta_{j,2})}{\sum_{s'=0}^{0} D_1^{s'}} + D_1\left(\frac{\sum_{j=0}^{1} D_1^j(\delta_{j+1,2} - \delta_{j,2})}{\sum_{s'=0}^{1} D_1^{s'}} - 1\right) \\
&= \frac{D_1}{1+D_1} + D_1\left(\frac{D_1}{1+D_1} - 1\right) = \frac{D_1 + D_1^2}{1+D_1} - D_1 = 0
\end{aligned}$$

$$\left.\frac{\partial \Delta U_1}{\partial \varepsilon_t}\right|_{\varepsilon=0} = \sum_{s=1}^{T}\sum_{i=0}^{s-1} D_1^{s-1}\left(\frac{\sum_{j=0}^{T-s+i} D_1^j(\delta_{j+1,t} - \delta_{j,t})}{\sum_{s'=0}^{T-s+i} D_1^{s'}} - \delta_{i+1,t} + \delta_{i,t}\right)$$

$$\begin{aligned}
S &= \{(s,i) \in \mathbf{Z}^2 : 1 \le s \le T \wedge 0 \le i \le s-1\} \\
S' &= \{(s,i) \in \mathbf{Z}^2 : 0 \le i \le T-1 \wedge i+1 \le s \le T\}
\end{aligned}$$

If $(s,i) \in S$, $1 \le s \le T \wedge 0 \le i \le s-1$. Thus $0 \le i \le s-1 \le T-1$, and $i+1 \le s \le T$, so $(s,i) \in S'$.

If $(s,i) \in S'$, $0 \le i \le T-1 \wedge i+1 \le s \le T$, $1 \le i+1 \le s \le T$ and $0 \le i \le s-1$.

$$\left.\frac{\partial \Delta U_1}{\partial \varepsilon_t}\right|_{\varepsilon=0} = \sum_{i=0}^{T-1} \sum_{s=i+1}^{T} D_1^{s-1} \left(\frac{\sum_{j=0}^{T-s+i} D_1^j (\delta_{j+1,t} - \delta_{j,t})}{\sum_{s'=0}^{T-s+i} D_1^{s'}} - \delta_{i+1,t} + \delta_{i,t} \right)$$

$$\begin{aligned} \sum_{i=0}^{T-1} \sum_{s=i+1}^{T} D_1^{s-1} (\delta_{i+1,t} - \delta_{i,t}) &= \sum_{s=t}^{T} D_1^{s-1} - (1 - \delta_{tT}) \sum_{s=t+1}^{T} D_1^{s-1} \\ &= D_1^{t-1} + \delta_{tT} \sum_{s=t+1}^{T} D_1^{s-1} = D_1^{t-1} \end{aligned}$$

$$V_1 = \sum_{s=1}^{T} \sum_{i=0}^{s-1} \sum_{j=0}^{T-s+i} \frac{D_1^{s+j-1} (\delta_{j+1,t} - \delta_{j,t})}{\sum_{s'=0}^{T-s+i} D_1^{s'}}$$

Let $z = s - i$, so $i = s - z$

$$V_1 = \sum_{s=1}^{T} \sum_{z=1}^{s} \sum_{j=0}^{T-z} \frac{D_1^{s+j-1} (\delta_{j+1,t} - \delta_{j,t})}{\sum_{s'=0}^{T-z} D_1^{s'}}$$

$$\begin{aligned} S &= \{(z,j) \in \mathbf{Z}^2 : 1 \le z \le s \wedge 0 \le j \le T - z\} \\ S' &= \{(z,j) \in \mathbf{Z}^2 : 0 \le j \le T-1 \wedge 1 \le z \le \min\{s, T-j\}\} \end{aligned}$$

If $(z,j) \in S$, $1 \le z \le s \wedge 0 \le j \wedge j \le T - z$. Thus $0 \le j \le T - z \le T - 1$. $1 \le z$, $z \le s$, and $z \le T - j$. Thus $1 \le z \le \min\{s, T-j\}$. So $(z,j) \in S'$.

If $(z,j) \in S'$, $0 \le j \le T-1 \wedge 1 \le z \le \min\{s, T-j\}$. Thus $1 \le z \le s$. Since $z \le T - j$, we have $j \le T - z$. Thus $0 \le j \le T - z$. Thus $(z, j) \in S$.

$$
\begin{aligned}
V_1 &= \sum_{s=1}^{T} \sum_{j=0}^{T-1} \sum_{z=1}^{\min\{s,T-j\}} \frac{D_1^{s+j-1}(\delta_{j+1,t} - \delta_{j,t})}{\sum_{s'=0}^{T-z} D_1^{s'}} \\
&= \sum_{j=0}^{T-1} \sum_{s=1}^{T} \sum_{z=1}^{\min\{s,T-j\}} \frac{D_1^{s+j-1}(\delta_{j+1,t} - \delta_{j,t})}{\sum_{s'=0}^{T-z} D_1^{s'}} \\
&= \sum_{s=1}^{T} \sum_{z=1}^{\min\{s,T-t+1\}} \frac{D_1^{s+t-2}}{\sum_{s'=0}^{T-z} D_1^{s'}} - (1 - \delta_{t,T}) \sum_{s=1}^{T} \sum_{z=1}^{\min\{s,T-t\}} \frac{D_1^{s+t-1}}{\sum_{s'=0}^{T-z} D_1^{s'}} \\
&= \sum_{s=1}^{T} \left[\sum_{z=1}^{\min\{s,T-t+1\}} \frac{D_1^{s+t-2}}{\sum_{s'=0}^{T-z} D_1^{s'}} - \sum_{z=1}^{\min\{s,T-t\}} \frac{D_1^{s+t-1}}{\sum_{s'=0}^{T-z} D_1^{s'}} + \delta_{t,T} \sum_{z=1}^{\min\{s,T-t\}} \frac{D_1^{s+t-1}}{\sum_{s'=0}^{T-z} D_1^{s'}} \right] \\
&= \sum_{s=1}^{T} \left[\sum_{z=1}^{\min\{s,T-t+1\}} \frac{D_1^{s+t-2}}{\sum_{s'=0}^{T-z} D_1^{s'}} - \sum_{z=1}^{\min\{s,T-t\}} \frac{D_1^{s+t-1}}{\sum_{s'=0}^{T-z} D_1^{s'}} \right]
\end{aligned}
$$

If $T = t = 2$,

$$
\begin{aligned}
V_1 &= \sum_{s=1}^{2} \left[\sum_{z=1}^{\min\{s,1\}} \frac{D_1^{s+2-2}}{\sum_{s'=0}^{2-z} D_1^{s'}} - \sum_{z=1}^{\min\{s,0\}} \frac{D_1^{s+2-1}}{\sum_{s'=0}^{2-z} D_1^{s'}} \right] \\
&= \sum_{s=1}^{2} \sum_{z=1}^{\min\{s,1\}} \frac{D_1^{s}}{\sum_{s'=0}^{2-z} D_1^{s'}} \\
&= \sum_{z=1}^{\min\{1,1\}} \frac{D_1^{1}}{\sum_{s'=0}^{2-z} D_1^{s'}} + \sum_{z=1}^{\min\{2,1\}} \frac{D_1^{2}}{\sum_{s'=0}^{2-z} D_1^{s'}} \\
&= \frac{D_1}{1 + D_1} + \frac{D_1^2}{1 + D_1} = D_1
\end{aligned}
$$

$$
\begin{aligned}
V_1 &= D_1^{t-1} \sum_{s=1}^{T} \left[\sum_{z=1}^{\min\{s,T-t+1\}} \frac{D_1^{s-1}}{\sum_{s'=0}^{T-z} D_1^{s'}} - \sum_{z=1}^{\min\{s,T-t\}} \frac{D_1^{s}}{\sum_{s'=0}^{T-z} D_1^{s'}} \right] \\
&= D_1^{t-1} \left[\sum_{s=1}^{T} \sum_{z=1}^{\min\{s,T-t+1\}} \frac{D_1^{s-1}}{\sum_{s'=0}^{T-z} D_1^{s'}} - \sum_{s=1}^{T} \sum_{z=1}^{\min\{s,T-t\}} \frac{D_1^{s}}{\sum_{s'=0}^{T-z} D_1^{s'}} \right] \\
&= D_1^{t-1} \left[\sum_{s=0}^{T-1} \sum_{z=1}^{\min\{s+1,T-t+1\}} \frac{D_1^{s}}{\sum_{s'=0}^{T-z} D_1^{s'}} - \sum_{s=1}^{T} \sum_{z=1}^{\min\{s,T-t\}} \frac{D_1^{s}}{\sum_{s'=0}^{T-z} D_1^{s'}} \right] \\
&= D_1^{t-1} \left[\sum_{z=1}^{\min\{1,T-t+1\}} \frac{D_1^{0}}{\sum_{s'=0}^{T-z} D_1^{s'}} + \sum_{s=1}^{T-1} \frac{D_1^{s}}{\sum_{s'=0}^{T-\min\{s,T-t\}-1} D_1^{s'}} - \sum_{z=1}^{\min\{T,T-t\}} \frac{D_1^{T}}{\sum_{s'=0}^{T-z} D_1^{s'}} \right] \\
&= D_1^{t-1} \left[\frac{1}{\sum_{s'=0}^{T-1} D_1^{s'}} + \sum_{s=1}^{T-1} \frac{D_1^{s}}{\sum_{s'=0}^{T-\min\{s,T-t\}-1} D_1^{s'}} - \sum_{z=1}^{T-t} \frac{D_1^{T}}{\sum_{s'=0}^{T-z} D_1^{s'}} \right]
\end{aligned}
$$

Suppose $T = 3$ and $t = 2$.

$$
V_1 = \sum_{s=1}^{3} \left[\sum_{z=1}^{\min\{s,2\}} \frac{D_1^{s}}{\sum_{s'=0}^{3-z} D_1^{s'}} - \sum_{z=1}^{\min\{s,1\}} \frac{D_1^{s+1}}{\sum_{s'=0}^{3-z} D_1^{s'}} \right]
$$

$$
\begin{aligned}
V_1 &= \sum_{z=1}^{\min\{1,2\}} \frac{D_1^1}{\sum_{s'=0}^{3-z} D_1^{s'}} - \sum_{z=1}^{\min\{1,1\}} \frac{D_1^2}{\sum_{s'=0}^{3-z} D_1^{s'}} \\
&\quad + \sum_{z=1}^{\min\{2,2\}} \frac{D_1^2}{\sum_{s'=0}^{3-z} D_1^{s'}} - \sum_{z=1}^{\min\{2,1\}} \frac{D_1^3}{\sum_{s'=0}^{3-z} D_1^{s'}} \\
&\quad + \sum_{z=1}^{\min\{3,2\}} \frac{D_1^3}{\sum_{s'=0}^{3-z} D_1^{s'}} - \sum_{z=1}^{\min\{3,1\}} \frac{D_1^4}{\sum_{s'=0}^{3-z} D_1^{s'}} \\
&= \frac{D_1}{1+D_1+D_1^2} - \frac{D_1^2}{1+D_1+D_1^2} \\
&\quad + \frac{D_1^2}{1+D_1+D_1^2} + \frac{D_1^2}{1+D_1} - \frac{D_1^3}{1+D_1+D_1^2} \\
&\quad + \frac{D_1^3}{1+D_1+D_1^2} + \frac{D_1^3}{1+D_1} - \frac{D_1^4}{1+D_1+D_1^2} \\
&= \frac{D_1 - D_1^4}{1+D_1+D_1^2} + D_1^2 \\
&= \frac{D_1 - D_1^4 + D_1^2 + D_1^3 + D_1^4}{1+D_1+D_1^2} = \frac{D_1 + D_1^2 + D_1^3}{1+D_1+D_1^2} = D_1
\end{aligned}
$$

If $T = t = 3$,

$$
\begin{aligned}
V_1 &= \sum_{s=1}^{3} \left[\sum_{z=1}^{\min\{s,1\}} \frac{D_1^{s+1}}{\sum_{s'=0}^{3-z} D_1^{s'}} - \sum_{z=1}^{\min\{s,0\}} \frac{D_1^{s+2}}{\sum_{s'=0}^{3-z} D_1^{s'}} \right] \\
&= \frac{D_1^2}{1+D_1+D_1^2} + \frac{D_1^3}{1+D_1+D_1^2} + \frac{D_1^4}{1+D_1+D_1^2} = D_1^2
\end{aligned}
$$

$$
V_1 = D_1^{t-1} \left[\frac{1}{\sum_{s'=0}^{T-1} D_1^{s'}} + \sum_{s=1}^{T-1} \frac{D_1^s}{\sum_{s'=0}^{T-\min\{s,T-t\}-1} D_1^{s'}} - \sum_{z=1}^{T-t} \frac{D_1^T}{\sum_{s'=0}^{T-z} D_1^{s'}} \right]
$$

If $T = 3$ and $t = 2$,

$$
\begin{aligned}
V_1 &= D_1 \left[\frac{1}{\sum_{s'=0}^{2} D_1^{s'}} + \sum_{s=1}^{2} \frac{D_1^s}{\sum_{s'=0}^{3-\min\{s,1\}-1} D_1^{s'}} - \sum_{z=1}^{1} \frac{D_1^3}{\sum_{s'=0}^{3-z} D_1^{s'}} \right] \\
&= D_1 \left[\frac{1}{1+D_1+D_1^2} + \frac{D_1}{1+D_1} + \frac{D_1^2}{1+D_1} - \frac{D_1^3}{1+D_1+D_1^2} \right] \\
&= D_1 \left[\frac{1-D_1^3}{1+D_1+D_1^2} + D_1 \right] = D_1 \frac{1 - D_1^3 + D_1 + D_1^2 + D_1^3}{1+D_1+D_1^2} = D_1
\end{aligned}
$$

$$V_1 = \sum_{s=1}^{T}\sum_{z=1}^{s}\sum_{j=0}^{T-z}\frac{D_1^{s+j-1}(\delta_{j+1,t}-\delta_{j,t})}{\sum_{s'=0}^{T-z}D_1^{s'}}$$

Let $S = \{(s,z) : 1 \le s \le T \wedge 1 \le z \le s\}$ and $S' = \{(s,z) : 1 \le z \le T \wedge z \le s \le T\}$. Let $(s,z) \in S$. Then $1 \le s \le T \wedge 1 \le z \le s$, so $1 \le z \le s \le T$ and $z \le s \le T$, so $(s,z) \in S'$.

Let $(s,z) \in S'$. Then $1 \le z \le T \wedge z \le s \le T$, so $1 \le z \le s \le T$ and $1 \le z \le s$.

$$\begin{aligned}
V_1 &= \sum_{z=1}^{T}\sum_{s=z}^{T}\sum_{j=0}^{T-z}\frac{D_1^{s+j-1}(\delta_{j+1,t}-\delta_{j,t})}{\sum_{s'=0}^{T-z}D_1^{s'}}\\
&= \sum_{z=1}^{T}\sum_{s=0}^{T-z}\sum_{j=0}^{T-z}\frac{D_1^{s+z+j-1}(\delta_{j+1,t}-\delta_{j,t})}{\sum_{s'=0}^{T-z}D_1^{s'}}\\
&= \sum_{z=1}^{T}\sum_{j=0}^{T-z}\sum_{s=0}^{T-z}\frac{D_1^{s+z+j-1}(\delta_{j+1,t}-\delta_{j,t})}{\sum_{s'=0}^{T-z}D_1^{s'}}\\
&= \sum_{z=1}^{T}\sum_{j=0}^{T-z}D_1^{z+j-1}(\delta_{j+1,t}-\delta_{j,t})
\end{aligned}$$

Let $T = t = 2$.

$$\begin{aligned}
V_1 &= \sum_{z=1}^{2}\sum_{j=0}^{2-z}D_1^{z+j-1}(\delta_{j+1,2}-\delta_{j,2})\\
&= \sum_{j=0}^{1}D_1^{j}(\delta_{j+1,2}-\delta_{j,2}) + \sum_{j=0}^{0}D_1^{j+1}(\delta_{j+1,2}-\delta_{j,2}) = D_1
\end{aligned}$$

Let $T = t = 3$

$$\begin{aligned}
V_1 &= \sum_{z=1}^{3}\sum_{j=0}^{3-z}D_1^{z+j-1}(\delta_{j+1,3}-\delta_{j,3})\\
&= \sum_{j=0}^{2}D_1^{j}(\delta_{j+1,3}-\delta_{j,3})\\
&\quad + \sum_{j=0}^{1}D_1^{j+1}(\delta_{j+1,3}-\delta_{j,3})\\
&\quad + \sum_{j=0}^{0}D_1^{j+2}(\delta_{j+1,3}-\delta_{j,3})\\
&= D_1^2
\end{aligned}$$

Let $T = t = 2$

$$\begin{aligned}
V_1 &= \sum_{z=1}^{3}\sum_{j=0}^{3-z} D_1^{z+j-1}(\delta_{j+1,2} - \delta_{j,2}) \\
&= \sum_{j=0}^{2} D_1^{j}(\delta_{j+1,2} - \delta_{j,2}) \\
&\quad + \sum_{j=0}^{1} D_1^{j+1}(\delta_{j+1,2} - \delta_{j,2}) \\
&\quad + \sum_{j=0}^{0} D_1^{j+2}(\delta_{j+1,2} - \delta_{j,2}) \\
&= D_1 - D_1^2 + D_1^2 = D_1
\end{aligned}$$

$$V_1 = \sum_{z=1}^{T}\sum_{j=0}^{T-z} D_1^{z+j-1}(\delta_{j+1,t} - \delta_{j,t})$$

$$\begin{aligned}
S &= \{(z,j) : 1 \le z \le T \wedge 0 \le j \le T - z\} \\
S' &= \{(z,j) : 0 \le j \le T-1 \wedge 1 \le z \le T - j\}
\end{aligned}$$

Let $(z,j) \in S$. Then $1 \le z \le T \wedge 0 \le j \le T - z$. So $z \le T - j$, and $1 \le z \le T - j$ while $0 \le j \le T - z \le T - 1$. Thus $(z,j) \in S'$.

Let $(z,j) \in S'$. Then $0 \le j \le T - 1 \wedge 1 \le z \le T - j$. So $j \le T - z$, so $0 \le j \le T - z$. $1 \le z \le T - j \le T$. Thus $(z,j) \in S$.

$$\begin{aligned}
V_1 &= \sum_{j=0}^{T-1}\sum_{z=1}^{T-j} D_1^{z+j-1}(\delta_{j+1,t} - \delta_{j,t}) \\
&= \sum_{z=1}^{T-(t-1)} D_1^{z+t-2} - (1 - \delta_{Tt})\sum_{z=1}^{T-t} D_1^{z+t-1}
\end{aligned}$$

If $T = t$,

$$V_1 = \sum_{z=1}^{1} D_1^{z+t-2} = D_1^{T-1}$$

If $t < T$,

$$V_1 = \sum_{z=1}^{T-t+1} D_1^{z+t-2} - \sum_{z=1}^{T-t} D_1^{z+t-1}$$

Let $s = z - 1$, so $z = s + 1$.

$$\begin{aligned} V_1 &= \sum_{s=0}^{T-t} D_1^{s+1+t-2} - \sum_{z=1}^{T-t} D_1^{z+t-1} \\ &= \sum_{s=0}^{T-t} D_1^{s+t-1} - \sum_{z=1}^{T-t} D_1^{z+t-1} \\ &= D_1^{t-1} \end{aligned}$$

Thus

$$\left.\frac{\partial \Delta U_1}{\partial \varepsilon_t}\right|_{\varepsilon=0} = \sum_{i=0}^{T-1} \sum_{s=i+1}^{T} D_1^{s-1} \left(\frac{\sum_{j=0}^{T-s+i} D_1^j (\delta_{j+1,t} - \delta_{j,t})}{\sum_{s'=0}^{T-s+i} D_1^{s'}} - \delta_{i+1,t} + \delta_{i,t} \right) = D_1^{t-1} - D_1^{t-1} = 0$$

E Hessian of ΔU_1 at the Origin for $T = 3$

$$\begin{aligned} \Delta U_1 &= \sum_{s=1}^{T-1} \left[\ln \overline{\phi}_s \sum_{t=\max\{T-s-1,0\}}^{T-1} D_t - \ln \phi_s \sum_{t=\max\{0,s\}}^{T-1} D_t \right] \\ &= \sum_{s=1}^{T-1} \left[\ln \overline{\phi}_s \sum_{t=T-1-s}^{T-1} D_t - \ln \phi_s \sum_{t=s}^{T-1} D_t \right] \end{aligned}$$

Only ϕ_{T-1} and $\overline{\phi}_{T-1}$ will depend on ε_3, so

$$\Delta U_1 = \ln \overline{\phi}_{T-1} \sum_{t=0}^{T-1} D_t - D_{T-1} \ln \left(\frac{1 + \varepsilon_T}{1 + \varepsilon_{T-1}} \right)$$

$$\overline{\phi}_{T-1} = \frac{\sum_{z=0}^{T-1} D_z \phi_z}{\sum_{z=0}^{T-1} D_z}$$

$$\frac{\partial \ln \overline{\phi}_{T-1}}{\partial \varepsilon_T} = \frac{D_{T-1} \frac{1}{1+\varepsilon_{T-1}}}{\sum_{z=0}^{T-1} D_z \phi_z} = \frac{D_1^{T-1}}{\sum_{z=0}^{T-1} D_z \phi_z}$$

$$\frac{\partial \Delta U_1}{\partial \varepsilon_T} = \frac{D_1^{T-1}}{\sum_{z=0}^{T-1} D_z \phi_z} \sum_{t=0}^{T-1} D_t - \frac{D_{T-1}}{1 + \varepsilon_T}$$

If $\varepsilon_2 = \cdots = \varepsilon_{T-1} = 0$,

$$\begin{aligned}
\frac{\partial \Delta U_1}{\partial \varepsilon_T} &= D_1^{T-1} \frac{\sum_{t=0}^{T-1} D_1^t}{\sum_{z=0}^{T-2} D_1^z + D_1^{T-1}(1+\varepsilon_T)} - \frac{D_1^{T-1}}{1+\varepsilon_T} \\
&= D_1^{T-1} \frac{(1+\varepsilon_T)\sum_{t=0}^{T-1} D_1^t - \left[\sum_{z=0}^{T-2} D_1^z + D_1^{T-1}(1+\varepsilon_T)\right]}{\left(\sum_{z'=0}^{T-1} D_1^{z'} + D_1^{T-1}\varepsilon_T\right)(1+\varepsilon_T)} \\
&= D_1^{T-1} \frac{(1+\varepsilon_T)\sum_{t=0}^{T-2} D_1^t - \sum_{z=0}^{T-2} D_1^z}{\left(\sum_{z'=0}^{T-1} D_1^{z'} + D_1^{T-1}\varepsilon_T\right)(1+\varepsilon_T)} \\
&= D_1^{T-1} \frac{\sum_{t=0}^{T-2} D_1^t}{\left(\sum_{z'=0}^{T-1} D_1^{z'} + D_1^{T-1}\varepsilon_T\right)(1+\varepsilon_T)} \varepsilon_T
\end{aligned}$$

This is positive except when $\varepsilon_T = 0$.

$$\frac{\partial^2 \Delta U_1}{\partial \varepsilon_T^2} = D_1^{T-1} \sum_{t=0}^{T-2} D_1^t \frac{\left(\sum_{z=0}^{T-1} D_1^z + D_1^{T-1}\varepsilon_T\right)(1+\varepsilon_T) - \varepsilon_T\left[\sum_{z=0}^{T-1} D_1^z + D_1^{T-1}\varepsilon_T + D_1^{T-1}(1+\varepsilon_T)\right]}{\left[\left(\sum_{z'=0}^{T-1} D_1^{z'} + D_1^{T-1}\varepsilon_T\right)(1+\varepsilon_T)\right]^2}$$

$$\left.\frac{\partial^2 \Delta U_1}{\partial \varepsilon_T^2}\right|_{\varepsilon=0} = D_1^{T-1} \sum_{t=0}^{T-2} D_1^t \frac{\sum_{z=0}^{T-1} D_1^z}{\left[\sum_{z'=0}^{T-1} D_1^{z'}\right]^2} > 0$$

Thus if $\varepsilon_2 = \cdots = \varepsilon_{T-1} = 0$, $\Delta U_1 \geq 0$ with equality only if $\varepsilon_T = 0$.

$$\begin{aligned}
\Delta U_1 &= (D_1 + D_1^2(1+\varepsilon_2))\left(\ln\left(1 + \frac{D_1}{1+D_1}\varepsilon_2\right) - \ln(1+\varepsilon_2)\right) \\
&\quad + (1 + D_1 + D_1^2(1+\varepsilon_2)) \ln\left(\frac{1 + D_1 + D_1^2 + D_1\varepsilon_2 + D_1^2\varepsilon_3}{1 + D_1 + D_1^2 + D_1^2\varepsilon_2}\right) - D_1^2(1+\varepsilon_2)\ln\left(\frac{1+\varepsilon_3}{1+\varepsilon_2}\right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \Delta U_1}{\partial \varepsilon_2} &= D_1^2\left(\ln\left(1 + \frac{D_1}{1+D_1}\varepsilon_2\right) - \ln(1+\varepsilon_2)\right) + (D_1 + D_1^2(1+\varepsilon_2))\left(\frac{\frac{D_1}{1+D_1}}{1 + \frac{D_1}{1+D_1}\varepsilon_2} - \frac{1}{1+\varepsilon_2}\right) \\
&\quad + D_1^2 \ln\left(\frac{1 + D_1 + D_1^2 + D_1\varepsilon_2 + D_1^2\varepsilon_3}{1 + D_1 + D_1^2 + D_1^2\varepsilon_2}\right) \\
&\quad + (1 + D_1 + D_1^2(1+\varepsilon_2))\left[\frac{D_1}{1 + D_1 + D_1^2 + D_1\varepsilon_2 + D_1^2\varepsilon_3} - \frac{D_1^2}{1 + D_1 + D_1^2 + D_1^2\varepsilon_2}\right] \\
&\quad - D_1^2 \ln\left(\frac{1+\varepsilon_3}{1+\varepsilon_2}\right) + D_1^2 \frac{1+\varepsilon_2}{1+\varepsilon_2}
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial \Delta U_1}{\partial \varepsilon_2} &= D_1^2\left(\ln\left(1+\frac{D_1}{1+D_1}\varepsilon_2\right)-\ln(1+\varepsilon_2)\right)+(D_1+D_1^2(1+\varepsilon_2))\left(\frac{D_1}{1+D_1+D_1\varepsilon_2}-\frac{1}{1+\varepsilon_2}\right)\\
&+D_1^2\ln\left(\frac{1+D_1+D_1^2+D_1\varepsilon_2+D_1^2\varepsilon_3}{1+D_1+D_1^2+D_1^2\varepsilon_2}\right)\\
&+(1+D_1+D_1^2(1+\varepsilon_2))\left[\frac{D_1}{1+D_1+D_1^2+D_1\varepsilon_2+D_1^2\varepsilon_3}-\frac{D_1^2}{1+D_1+D_1^2+D_1^2\varepsilon_2}\right]\\
&-D_1^2\ln\left(\frac{1+\varepsilon_3}{1+\varepsilon_2}\right)+D_1^2
\end{aligned}
$$

As a check,

$$
\begin{aligned}
\left.\frac{\partial \Delta U_1}{\partial \varepsilon_2}\right|_{\varepsilon_2=\varepsilon_3=0} &= (D_1+D_1^2)\left(\frac{D_1}{1+D_1}-1\right)\\
&+(1+D_1+D_1^2)\left[\frac{D_1-D_1^2}{1+D_1+D_1^2}\right]+D_1^2\\
&= -\frac{D_1+D_1^2}{1+D_1}+D_1-D_1^2+D_1^2=0
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \Delta U_1}{\partial \varepsilon_2} &= (D_1+D_1^2(1+\varepsilon_2))\left(\frac{D_1}{1+D_1+D_1\varepsilon_2}-\frac{1}{1+\varepsilon_2}\right)+D_1^2\\
&+D_1^2\ln\left(\left(1+\frac{D_1}{1+D_1}\varepsilon_2\right)\frac{1+D_1+D_1^2+D_1\varepsilon_2+D_1^2\varepsilon_3}{(1+D_1+D_1^2+D_1^2\varepsilon_2)(1+\varepsilon_3)}\right)\\
&+(1+D_1+D_1^2(1+\varepsilon_2))\left[\frac{D_1}{1+D_1+D_1^2+D_1\varepsilon_2+D_1^2\varepsilon_3}-\frac{D_1^2}{1+D_1+D_1^2+D_1^2\varepsilon_2}\right]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \Delta U_1}{\partial \varepsilon_2 \partial \varepsilon_3} &= D_1^2\left[\frac{D_1^2}{1+D_1+D_1^2+D_1\varepsilon_2+D_1^2\varepsilon_3}-\frac{1}{1+\varepsilon_3}\right]\\
&-D_1^3\frac{1+D_1+D_1^2(1+\varepsilon_2)}{(1+D_1+D_1^2+D_1\varepsilon_2+D_1^2\varepsilon_3)^2}
\end{aligned}
$$

$$
\begin{aligned}
\left.\frac{\partial^2 \Delta U_1}{\partial \varepsilon_2 \partial \varepsilon_3}\right|_{\varepsilon_2=\varepsilon_3=0} &= D_1^2\left[\frac{D_1^2}{1+D_1+D_1^2}-1\right]-D_1^3\frac{1+D_1+D_1^2}{(1+D_1+D_1^2)^2}\\
&= -D_1^2\frac{1+D_1}{1+D_1+D_1^2}-\frac{D_1^3}{1+D_1+D_1^2}
\end{aligned}
$$

$$
\left.\frac{\partial^2 \Delta U_1}{\partial \varepsilon_2 \partial \varepsilon_3}\right|_{\varepsilon_2=\varepsilon_3=0} = -\frac{D_1^2}{1+D_1+D_1^2}(1+2D_1)
$$

$$\left.\frac{\partial^2 \Delta U_1}{\partial \varepsilon_3^2}\right|_{\varepsilon_2=\varepsilon_3=0} = \frac{D_1^2}{1+D_1+D_1^2}(1+D_1)$$

According to Mathematica,

$$\left.\frac{\partial^2 \Delta U_1}{\partial \varepsilon_2^2}\right|_{\varepsilon_2=\varepsilon_3=0} = \frac{D_1(1+D_1+4D_1^2+3D_1^3)}{(1+D_1)(1+D_1+D_1^2)}$$

$$\begin{aligned}
\Delta U_1 &= \frac{D_1^2+D_1^3}{2(1+D_1+D_1^2)}\varepsilon_3^2 - \frac{D_1^2+2D_1^3}{1+D_1+D_1^2}\varepsilon_2\varepsilon_3 + \frac{D_1+D_1^2+4D_1^3+3D_1^4}{2(1+D_1)(1+D_1+D_1^2)}\varepsilon_2^2 + O(\varepsilon^3) \\
&= \frac{1}{2}\frac{1}{1+D_1+D_1^2}\begin{bmatrix} \varepsilon_2 & \varepsilon_3 \end{bmatrix}\begin{bmatrix} D_1^2(1+D_1) & -D_1^2(1+2D_1) \\ -D_1^2(1+2D_1) & \frac{D_1+D_1^2+4D_1^3+3D_1^4}{(1+D_1)} \end{bmatrix}\begin{bmatrix} \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} + O(\varepsilon^3)
\end{aligned}$$

$$\begin{aligned}
&\begin{vmatrix} D_1^2(1+D_1) & -D_1^2(1+2D_1) \\ -D_1^2(1+2D_1) & \frac{D_1+D_1^2+4D_1^3+3D_1^4}{(1+D_1)} \end{vmatrix} \\
&= D_1^3+D_1^4+4D_1^5+3D_1^6-D_1^4(1+2D_1)^2 \\
&= D_1^3+D_1^4+4D_1^5+3D_1^6-D_1^4-4D_1^5-4D_1^6 \\
&= D_1^3-D_1^6 \\
&= D_1^3(1-D_1^3)
\end{aligned}$$

If $D_1 > 1$, $\Delta U_1 < 0$ is possible. However, if $D_1 < 1$, the determinant is nonnegative. Thus in a deleted neighborhood of $(\varepsilon_2, \varepsilon_3) = (0,0)$, ΔU_1 must be strictly nonnegative.

F Sufficient Upper Bound on ε_T for Pareto Dominance of the Commitment Path

$$\Delta U_\tau = \sum_{t=\tau}^{T}\sum_{i=0}^{t-1} D_{t-\tau}\ln\frac{\overline{\phi}_{T-t+i}}{\phi_i}.$$

We can rewrite this as

$$\Delta U_\tau = \sum_{t=\tau}^{T}\sum_{j=T-t}^{T-1} D_{t-\tau}\ln\overline{\phi}_j - \sum_{t=\tau}^{T}\sum_{i=0}^{t-1} D_{t-\tau}\ln\phi_i,$$

where $j = T - t + i$. The first terms are all positive while the second terms are all negative.

$S = \{(t,i) : \tau \le t \le T \wedge 0 \le i \le t-1\}$. $S' = \{(t,i) : 0 \le i \le T-1 \wedge \max\{\tau, i+1\} \le t \le T\}$. Let $(t,i) \in S$, so $\tau \le t \le T \wedge 0 \le i \le t-1$. Then $0 \le i \le t-1 \le T-1$. We have both $\tau \le t$ and $i+1 \le t$, so $\max\{\tau, i+1\} \le t \le T$. Thus $(t,i) \in S'$.

Now let $(t,i) \in S'$, so $0 \le i \le T-1 \wedge \max\{\tau, i+1\} \le t \le T$. Then $\tau \le t \le T$. $0 \le i \le t-1$. Thus $(t,i) \in S$.

Let $S = \{(t,j) : \tau \le t \le T \wedge T-t \le j \le T-1\}$. Let $S' = \{(t,j) : 0 \le j \le T-1 \wedge \max\{\tau, T-j\} \le t \le T\}$. Let $(t,j) \in S$. Then $\tau \le t \le T \wedge T-t \le j \le T-1$. So $0 \le T-t \le j \le T-1$, and we have both $\tau \le t$ and $T-j \le t$, so $\max\{\tau, T-j\} \le t \le T$. Thus $(t,j) \in S'$. Let $(t,j) \in S'$. Then $0 \le j \le T-1 \wedge \max\{\tau, T-j\} \le t \le T$. $\tau \le t \le T$, and $T-t \le j \le T-1$. Thus $(t,j) \in S$.

$$\Delta U_\tau = \sum_{j=0}^{T-1} \sum_{t=\max\{\tau,T-j\}}^{T} D_{t-\tau} \ln \overline{\phi}_j - \sum_{i=0}^{T-1} \sum_{t=\max\{\tau,i+1\}}^{T} D_{t-\tau} \ln \phi_i.$$

The first terms are all positive and the second terms are all negative. Let us define

$$P_i^\tau = \sum_{t=\max\{\tau,T-i\}}^{T} D_{t-\tau} \tag{47}$$

and

$$Q_i^\tau = \sum_{t=\max\{\tau,i+1\}}^{T} D_{t-\tau} \tag{48}$$

Thus we have

$$\begin{aligned}
\Delta U_\tau &= \sum_{i=0}^{T-1} \left[P_i^\tau \ln \overline{\phi}_i - Q_i^\tau \ln \phi_i \right] \\
&= \sum_{i=0}^{T-1} \left[P_i^\tau \left(\ln \frac{\overline{\phi}_i}{\phi_i} + \ln \phi_i \right) - Q_i^\tau \ln \phi_i \right] \\
&= \sum_{i=0}^{T-1} \left[P_i^\tau \left(\ln \frac{\overline{\phi}_i}{\phi_i} + \sum_{j=i}^{T-1} \ln \phi_j - \sum_{j=i+1}^{T-1} \ln \phi_j \right) - Q_i^\tau \ln \phi_i \right] \\
&= \sum_{i=0}^{T-1} \left[A_i^\tau \left(\ln \frac{\overline{\phi}_i}{\phi_i} + \ln \frac{1+\varepsilon_T}{1+\varepsilon_i} - \sum_{j=i+1}^{T-1} \ln \phi_j \right) - Q_i^\tau \ln \phi_i \right]
\end{aligned}$$

$$\Delta U_\tau = \sum_{i=0}^{T-1} P_i^\tau \ln(1+\varepsilon_T) + \sum_{i=0}^{T-1}\left[P_i^\tau\left(\ln\frac{\overline{\phi}_i}{\phi_i} - \ln(1+\varepsilon_i) - \sum_{j=i+1}^{T-1}\ln\phi_j\right) - Q_i^\tau \ln\phi_i\right]$$

Suppose that $s < T-1$. Suppose that

$$\varepsilon_T \leq B_s^\tau = \exp\left(\frac{\sum_{i=0}^{s}\left[P_i^\tau\left(\ln\frac{\phi_i}{\overline{\phi}_i} + \ln(1+\varepsilon_i) + \sum_{j=i+1}^{s}\ln\phi_j\right) + Q_i^\tau\ln\phi_i\right]}{\sum_{i'=0}^{T-1} P_{i'}^\tau}\right) - 1.$$

Then we will have

$$\sum_{i=0}^{T-1} P_i^\tau \ln(1+\varepsilon_T) + \sum_{i=0}^{s}\left[P_i^\tau\left(\ln\frac{\overline{\phi}_i}{\phi_i} - \ln(1+\varepsilon_i) - \sum_{j=i+1}^{s}\ln\phi_j\right) - Q_i^\tau\ln\phi_i\right] \leq 0$$

since

$$\begin{aligned} 0 \;\;\geq\;\; & \sum_{i=0}^{s}\left[P_i^\tau\left(\ln\frac{\overline{\phi}_i}{\phi_i} - \ln(1+\varepsilon_i) - \sum_{j=s+1}^{T-1}\ln\phi_j\right) - Q_i^\tau\ln\phi_i\right] \\ & + \sum_{i=s+1}^{T-1}\left[P_i^\tau\left(\ln\frac{\overline{\phi}_i}{\phi_i} - \ln(1+\varepsilon_i) - \sum_{j=i+1}^{T-1}\ln\phi_j\right) - Q_i^\tau\ln\phi_i\right] \end{aligned}$$

we have

$$\begin{aligned} 0 \;\;\geq\;\; & \sum_{i=0}^{T-1} P_i^\tau \ln(1+\varepsilon_T) + \sum_{i=0}^{s}\left[P_i^\tau\left(\ln\frac{\overline{\phi}_i}{\phi_i} - \ln(1+\varepsilon_i) - \sum_{j=i+1}^{s}\ln\phi_j\right) - Q_i^\tau\ln\phi_i\right] \\ & + \sum_{i=0}^{s}\left[P_i^\tau\left(\ln\frac{\overline{\phi}_i}{\phi_i} - \ln(1+\varepsilon_i) - \sum_{j=s+1}^{T-1}\ln\phi_j\right) - Q_i^\tau\ln\phi_i\right] \\ & + \sum_{i=s+1}^{T-1}\left[P_i^\tau\left(\ln\frac{\overline{\phi}_i}{\phi_i} - \ln(1+\varepsilon_i) - \sum_{j=i+1}^{T-1}\ln\phi_j\right) - Q_i^\tau\ln\phi_i\right] \\ = \;\; & \sum_{i=0}^{T-1} P_i^\tau \ln(1+\varepsilon_T) + \sum_{i=0}^{s}\left[P_i^\tau\left(\ln\frac{\overline{\phi}_i}{\phi_i} - \ln(1+\varepsilon_i) - \sum_{j=i+1}^{T-1}\ln\phi_j\right) - Q_i^\tau\ln\phi_i\right] \\ & + \sum_{i=s+1}^{T-1}\left[P_i^\tau\left(\ln\frac{\overline{\phi}_i}{\phi_i} - \ln(1+\varepsilon_i) - \sum_{j=i+1}^{T-1}\ln\phi_j\right) - Q_i^\tau\ln\phi_i\right] \\ = \;\; & \sum_{i=0}^{T-1} P_i^\tau \ln(1+\varepsilon_T) + \sum_{i=0}^{T-1}\left[P_i^\tau\left(\ln\frac{\overline{\phi}_i}{\phi_i} - \ln(1+\varepsilon_i) - \sum_{j=i+1}^{T-1}\ln\phi_j\right) - Q_i^\tau\ln\phi_i\right] = \Delta U_\tau. \end{aligned}$$