Heterogeneous Impatience and Dynamic Inconsistency

Chiaki Hara*
Institute of Economic Research, Kyoto University

May 31, 2012

Abstract

It has been shown in the literature that if the individual consumers have constant but unequal time discount rates, then the representative consumer has discount rates that is a strictly decreasing function of time, just as is the case of hyperbolic discounting. No contribution, however, has so far established a general relationship between the degree of heterogeneity of individual consumers' subjective discount rates on the one hand and the degree of dynamic inconsistency of the representative consumer's discount rate function on the other. In this paper, we show that the more convex the cumulant generating function of the approximately wealth-weighted distribution of individual consumers' subjective discount rates is, the more dynamically inconsistent the representative consumer is in the sense of Prelec (2004), and vice versa. Applications to the term structure of interest rates are explored and the special case where the distributions of subjective discount rates are taken from an exponential family is investigated.

JEL Classification Codes: D51, D53, D61, D81, D91, E43, G12.

Keywords: Discount factor, discount rate, representative consumer, term structure of interest rates, exponential family.

1 Introduction

Asset transactions are often motivated by the heterogeneity of consumers. More risk averse consumers try to unload the risks they are faced with and less risk averse ones will take them over with premiums. Optimistic consumers invest more in risky assets, while pessimistic consumers invest mostly into riskless bonds. More patient consumers save more to enjoy higher consumptions in the future, and impatient consumers may well borrow to enjoy higher consumption immediately. The raison d'être of asset markets is precisely to cater for diverse needs for asset transactions by heterogeneous consumers.

Heterogeneity in consumers' characteristics have implications not only on risk and intertemporal allocations but also on asset pricing. The impact on asset pricing can probably be best

*The paper is part of the academic projects on Economic Analysis of Intergenerational Issues: Searching for Further Development, funded by the Grant-in-Aid for Specially Promoted Research (grant number 22000001), and on Frontier of Game theory: Theory and Practice, funded by the Grant-in-Aid (S) (grant number 20223001), both from Japan's Ministry of Education, Culture, Sports, Science and Technology. My email address is `hara.chiaki.7x@kyoto-u.ac.jp`.

understood by constructing the representative consumer. The representative consumer is a fictitious consumer whose marginal utility process, evaluated along the aggregate consumption process, serves as a state price deflator. For example, the representative consumer of individual consumers having utility functions of constant and unequal relative risk aversion has a utility function of strictly decreasing relative risk aversion (Franke, Stapleton, and Subrahmanyam (1999) and Hara, Huang and Kuzmics (2007)); and the representative consumer of individual consumers having constant but unequal subjective discount rates has discount rates that are a strictly decreasing function of time (Weitzman (2001) and Gollier and Zeckhauser (2005)). The consequences of these are that the derivative assets with convex payoff functions is underestimated if the coefficients of constant relative risk aversion are erroneously assumed to be equal (Franke, Stapleton, and Subrahmanyam (1999) and Hara, Huang, and Kuzmics (2007)); and that the term structure of interest rates is more downward sloping in the case of heterogeneous subjective discount rates than in the case of homogeneous subjective discount rates.

Among these results, the case of heterogeneous subjective discount rates is of particular interest. The reason is that in that case, the representative consumer's discount rates are decreasing, while in the literature of the representative-consumer models of asset pricing, most notably that of Mehra and Prescott (1985), the discount rates are almost universally assumed to be constant, presumably because it has been often assumed for individual consumers. Recall that assuming constant discount rates is equivalent to assuming dynamically consistency. Although the postulate of dynamic consistency is highly relevant (for both analytical and empirical reasons) to individual consumers and their optimization problems, it is not so for the representative consumer, because he is a fictitious entity and does not solve any optimization problem.[1] Hence the decreasing discount rates for the representative agent, which emerges from heterogeneous subjective discount rates, deserve fuller analysis.

While Weitzman (2001) and Gollier and Zeckhauser (205) showed that the heterogeneity in individual consumers' subjective discount rates gives rise to a dynamically inconsistent representative consumer, they did not clarify how the degree of heterogeneity is related to the degree of dynamic inconsistency. In particular, they do not tell us whether a more heterogeneous economy give rise to a more dynamically inconsistent representative consumer. If any theoretical result allowed us to compare a heterogeneous economy only with a homogeneous economy, then that would not be of much use, because, after all, the real-world asset markets involves heterogeneous consumers, and we would have no reliable theoretical result to compare one heterogeneous economy with another. Without such a result, we would be left unsure whether a homogeneous economy could be so singular that no generalization should be drawn from the homogeneous-economy case to the heterogeneous economy case.

The purpose of this paper is to give a precise formulation to show that the more heterogeneous the subjective discount rates of the individual consumers are, the more dynamically inconsistent the representative consumer is. To do so, we need to give a notion of the "more dynamically inconsistent than" relation and the "more heterogeneous than" relation, between two

[1] The situation is quite different for committee problems, analyzed by Jackson and Yariv (2011), for example, where the committee itself has to arrive at a decision.

heterogeneous economies. For the former, we use the notion by Prelec (2004), which was introduced for a single decision maker having a utility function over sequences of "dated outcomes". For the latter, we introduce, following Jouini and Napp (2007, Lemma 4.1), the notion of an approximate wealth-weighted distribution of subjective discount rates of individual consumers. As it is a probability measure on the strictly positive part $\boldsymbol{R}_{++}$ of the real line, we can define its cumulant-generating function on the non-positive part $-\boldsymbol{R}_{+}$ of the real line. Then our main result (Theorem 2) shows that the more convex the cumulant-generating function of the wealth-weighted distribution of discount rate is, the more dynamically inconsistent the representative consumer is, and the converse also holds. To understand the convexity assumption, recall that the first and second derivatives of the cumulant-generating function at $s \in -\boldsymbol{R}_{+}$ coincide with the mean and variance of the probability measure of which the density function with respect to the wealth-weighted distribution of discount rates is proportional to the exponential function $q \mapsto \exp(sq)$. Hence its curvature, the ratio of the second derivative to the first derivative, at s is the ratio of the variance to the mean of the probability measure with density function $q \mapsto \exp(sq)$, which can be regarded as the degree of the heterogeneity of the distribution of subjective discount rates. As the convexity of the cumulant-generating function is measured by its curvatures, our main result, indeed, formalizes the notion that the more heterogeneous the subjective discount rates are, the more dynamically inconsistent the representative consumer is.

A weaker notion of the more-dynamically-inconsistent relation is, as it turns out, the single-crossing property of the discount rate functions. To be more specific, if the representative agent is more dynamically inconsistent in one economy than in another economy, then the graph of the representative consumer's discount rates, as a function of time, in the first economy crosses that in the second from the above. This weaker notion is sufficient to derive some interesting implications on asset pricing, and can be derived from a weaker condition on cumulant-generating functions. Indeed, Theorems 4 and 5 show that the cumulant-generating functions of two economies satisfy the single-crossing property if and only if the forward-rate curves and the yield curves satisfy the single-crossing property as well.

If the wealth-weighted distributions of individual consumers' subjective discount rates belong to some parametric family, the task of determining whether one such distribution has a more convex cumulant-generating function than another is relatively easy. In Section 5, we define a quasi log-linear family of distributions, and give a sufficient condition for the cumulant-generating function of one distribution in such a family to be more convex than that of the other, and also for the two cumulant-generating functions to have the single-crossing property. A natural exponential family is a quasi log-linear family. The family of Gamma distributions, as well as those of binomial, negative binomial, and Poisson distributions, is also a quasi log-linear family. The results of Section 5, therefore, can be applied to many types of distributions of subjective discount rates.

The rest of this paper is organized as follows. The setup and preliminary results are presented in Section 2. The main result (Theorem 2) is stated and proved in Section 3. In Section 4, this and other results are applied to the term structure of interest rates are provided. The analysis of quasi log-linear families are explored in Section 5. The results are summarized and a future research topic is suggested in Section 6.

2 Setup and Preliminary Results

2.1 Representative Consumer

The economy is subject to uncertainty, which is represented by a probability space $(\Omega, \mathscr{F}, P)$. The time span is $\boldsymbol{R}_+ = [0, \infty)$, which is of continuous time and infinite length, although it could be $[0, T]$ with $0 < T < \infty$, which is of finite length. The gradual information revelation is represented by a filtration $(\mathscr{F}_t)_{t \in \boldsymbol{R}_+}$. There is only one type of good on each time and state.[2]

We allow the number of consumers present in the economy to be finite or infinite. Formally, we let $(A, \mathscr{A}, \nu)$ be a finite measure space of (names of) consumers. If A is a finite set, $\mathscr{A}$ is the power set of A, and ν is the counting measure on A, then the consumption sector consists of finitely many consumers. If, on the other hand, A is the unit interval $[0, 1]$, $\mathscr{A}$ is the Borel σ-field $\mathscr{B}([0, 1])$, and ν is (the restriction of) the Lebesgue measure on $\mathscr{B}([0, 1])$, then the consumption sector consists of infinitely many consumers, each of whom is negligible in size relative to the total population of the economy. For each $B \in \mathscr{A}$, $\nu(B)$ is the proportion o consumers in B relative to the entire consumption sector.

We assume that the consumers have time-additive expected utility functions over consumption processes, which exhibit constant and equal relative risk aversion, and constant but possibly unequal discount rates. Formally, let $\gamma > 0$ and $u : \boldsymbol{R}_{++} \to \boldsymbol{R}$ satisfy $u'(x) = x^{-\gamma}$ for every $x \in \boldsymbol{R}_{++}$. Let $\rho : A \to \boldsymbol{R}_{++}$ be measurable, where $\boldsymbol{R}_{++}$ is endowed with the Borel σ-field $\mathscr{B}(\boldsymbol{R}_{++})$. Then the utility function U_a of consumer a over consumption processes is defined by

$$U_a(c^a) = E\left(\int_0^\infty \exp(-\rho(a)t) u(c_t^a)\, \mathrm{d}t\right),$$

where $c^a = (c_t^a)_{t \in \boldsymbol{R}_+}$.[3,4]

To find a Pareto efficient allocation of a given aggregate consumption process $c = (c_t)_{t \in \boldsymbol{R}_+}$ and its supporting (decentralizing) state-price deflator, it is sufficient to let $\lambda : A \to \boldsymbol{R}_{++}$ be a

[2] Just as the analysis of Gollier and Zeckhauser (2005), the subsequent analysis would still be valid even in the absence of uncertainty. Although I could simplify our model by restricting our attention to a deterministic economy, I have chosen to incorporate uncertainty to make our model immediately applicable to asset pricing theory.

[3] This and other integrals in the subsequent analysis need not be well defined without no additional assumptions on c^a and other stochastic processes. But the subsequent argument depends only on the first-order conditions of (utility or social welfare) maximization problems, which must necessarily hold whenever there is a solution to the problem under consideration. We shall therefore be implicit about these additional assumptions.

[4] Although the assumption of constant and equal relative risk aversion is quite stringent, there is a good reason to restrict our attention to this case. In fact, if consumers had unequal coefficients of constant relative risk aversion, then the representative consumer's utility function would not be a product of a function of time t and a function of current consumptions c_t, and, hence, his discount factor function would not be well defined.

measurable function and consider the constrained maximization problem

$$\begin{aligned} \max_{(c^a)_{a\in A}} \quad & \int_A \lambda(a) U_a\left(c^a\right) \mathrm{d}\nu(a) \\ \text{subject to} \quad & \int_A c^a \, \mathrm{d}\nu(a) = c. \end{aligned} \tag{1}$$

The objective function of this constrained maximization problem is additively separable across time and states:

$$\begin{aligned} & \int_A \lambda(a) U_a\left(c^a\right) \mathrm{d}\nu(a) \\ = & \int_A \lambda(a) E\left(\int_0^\infty \exp(-\rho(a)t) u(c_t^a) \, \mathrm{d}t\right) \mathrm{d}\nu(a) \\ = & E\left(\int_0^\infty \int_A \lambda(a) \exp(-\rho(a)t) u(c_t^a) \, \mathrm{d}\nu(a) \, \mathrm{d}t\right) \\ = & \int_{\Omega\times\boldsymbol{R}_+} \left(\int_A \lambda(a) \exp(-\rho(a)t) u(c_t^a(\omega)) \, \mathrm{d}\nu(a)\right) \mathrm{d}\left(P\otimes\eta\right)(\omega,t), \end{aligned}$$

where η is the Lebesgue measure. The constraint, $\int_A c^a \, \mathrm{d}\nu(a) = c$, is, of course, separable across time and states. To find the solution to the constrained maximization problem, therefore, it suffices to denote by $\mathscr{I}$ the set of all integrable functions defined on A and taking values in $\boldsymbol{R}_{++}$ and solve

$$\begin{aligned} \max_{\iota\in\mathscr{I}} \quad & \int_A \lambda(a) \exp(-\rho(a)t) u(\iota(a)) \, \mathrm{d}\nu(a) \\ \text{subject to} \quad & \int_A \iota(a) \, \mathrm{d}\nu(a) = x. \end{aligned} \tag{2}$$

for every $(x,t) \in \boldsymbol{R}_{++} \times \boldsymbol{R}_+$.

If there is a solution to (2), it must be essentially unique with respect to ν because u is strictly concave. Assuming its existence, we denote the unique solution by $k(x,t) \in \mathscr{I}$ and write $k_a(x,t)$ instead of $k(x,t)(a)$. Then, for every (x,t), by the first-order condition for the solution to (2),

$$f_a(x,t) = \frac{\left(\lambda(a)\exp(-\rho(a)t)\right)^{1/\gamma}}{h(t)} x, \tag{3}$$

for almost every $a \in A$, where

$$h(t) = \int_A \left(\lambda(a)\exp(-\rho(a)t)\right)^{1/\gamma} \mathrm{d}\nu(a). \tag{4}$$

This shows that the mutual fund theorem holds at each time t across states, but not across time, because the consumption share $(h(t))^{-1}\left(\lambda(a)\exp(-\rho(a)t)\right)^{1/\gamma}$ depends on t unless ρ is an essentially constant function of a, that is, all consumers' discount rates are equal.

The representative consumer's felicity function $v : \boldsymbol{R}_{++} \times \boldsymbol{R}_+ \to \boldsymbol{R}$ is defined as the value function of (2):

$$v(x,t) = \int_A \lambda(a) \exp(-\rho(a)t) u(f_a(x,t)) \, \mathrm{d}\nu(a). \tag{5}$$

His utility function over the aggregate consumption processes is defined through time additivity by

$$U(c) = E\left(\int_0^\infty v(c_t, t)\,\mathrm{d}t\right). \tag{6}$$

It follows from (3) that

$$v(x,t) = d(t)u(x)$$

for every $t \in \boldsymbol{R}_+$, where

$$d(t) = (h(t))^\gamma = \left(\int_A (\lambda(a)\exp(-\rho(a)t))^{1/\gamma}\,\mathrm{d}\nu(a)\right)^\gamma. \tag{7}$$

Note that the representative consumer too has constant relative risk aversion equal to γ. The function $d : \boldsymbol{R}_+ \to \boldsymbol{R}$ is the representative consumer's *discount factor function*. In order for v (and, thus, U) to be well defined, it is necessary and sufficient that $h(0) < \infty$, because $\exp(-\rho(a)t) \leq 1$ for every $a \in A$ and every $t \in \boldsymbol{R}_+$. This is equivalent to saying that the function $a \mapsto (\lambda(a))^{1/\gamma}$ is integrable with respect to ν. By multiplying a constant to λ if necessary, we can assume that $\int_A (\lambda(a))^{1/\gamma}\,\mathrm{d}\nu(a) = 1$. Then $d(0) = 1$ and d is strictly decreasing. By the dominated convergence theorem, d is continuous and satisfies $d(t) \to 0$ as $t \to \infty$.

Define the representative consumer's *discount rate function* $r : \boldsymbol{R}_+ \to \boldsymbol{R}_{++}$ by $r(t) = -d'(t)/d(t)$, then

$$\frac{d(t_2)}{d(t_1)} = \exp\left(-\int_{t_1}^{t_2} r(t)\,\mathrm{d}t\right)$$

whenever $0 \leq t_1 < t_2$. Thus r represents the representative consumer's continuously compounded instantaneous subjective discount rate as a function of time. Unlike the case of individual consumers, this is not constant but varies with t unless all individual consumers have the same discount rate.

Although we analyze the Pareto efficient allocations, if the asset markets are complete, then our analysis is applicable to the equilibrium allocations and asset prices. This is because the first welfare theorem holds in complete markets, so that the equilibrium allocations are Pareto efficient and the equilibrium asset prices are given by the marginal utility process. Since u is concave, the second welfare theorem also holds, so that every Pareto efficient allocation is an equilibrium allocation for some distribution of initial endowments. Hence an analysis of Pareto efficient allocations is nothing but an analysis of equilibrium allocations.

2.2 Comparison of discount factor functions

Prelec (2004) introduced an *at-least-as-decreasingly-impatient-as* relation between two utility functions over *timed consumptions*, that is, functions of the form $d(t)u(x)$ defined over consumption levels x consumed at time t, where $d : \boldsymbol{R}_+ \to (0,1]$ is strictly decreasing and satisfies $d(0) = 1$, and $u : \boldsymbol{R}_+ \to \boldsymbol{R}_+$ is strictly increasing and satisfies $u(0) = 0$. The following definition is a variant of the relation, is concerned with discount factor functions, rather than utility functions over timed consumptions, and which is equivalent to Prelec's definition if u is continuous.

Definition 1 A discount factor function d_1 is *at-least-as-decreasingly-impatient-as* another discount factor function d_2 if for every $(t_0, t_1, t_2, s) \in \boldsymbol{R}_+^3 \times \boldsymbol{R}$,

$$\frac{d_2(t_0 + t_1)}{d_2(t_0)} \geq \frac{d_2(t_0 + t_1 + t_2 + s)}{d_2(t_0 + t_2)} \tag{8}$$

whenever

$$\frac{d_1(t_0 + t_1)}{d_1(t_0)} = \frac{d_1(t_0 + t_1 + t_2 + s)}{d_1(t_0 + t_2)}. \tag{9}$$

To understand this definition, first compare the two ratios, $d_1(t_0+t_1)/d_1(t_0)$ and $d_1(t_0+t_1+t_2)/d_1(t_0+t_2)$. The former is the discount factor that d_1 applies to the time interval $[t_0, t_0+t_1]$, and the latter is the discount factor that d_1 applies to the time interval $[t_0 + t_2, t_0 + t_1 + t_2]$. Since they are both applied to time intervals of length t_1, they would be equal if d_1 exhibited exponential discounting, that is, if there were a $\bar{\rho} > 0$ such that $d_1(t) = \exp(-\bar{\rho}t)$ for every $t \in \boldsymbol{R}_+$.[5] However, they can be different, and, more specifically, the former is smaller than the latter if the corresponding discount rate function is decreasing over time, just as in the case of hyperbolic discounting. To compensate the difference between the two ratios, we add an interval of length s (which is positive if the discounting rate function is decreasing over time, but negative if it is increasing) to the terminal time $t_0 + t_1 + t_2$ of the interval, so that the discount rate that d_1 applies to $[t_0 + t_2, t_0 + t_1 + t_2 + s]$ is equal to the discount rate that d_1 applies to $[t_0, t_0 + t_1]$, as shown in (9). The length s can, therefore, be considered as a measure of decreasing impatience of d_1. Then (8) states that s may too large for d_2, so that the discount factor that d_2 applies to $[t_0 + t_2, t_0 + t_1 + t_2 + s]$ may be smaller than the discount factor that d_2 applies to the time interval $[t_0, t_0 + t_1]$. In this sense, the impatience of d_1 decreases at least as rapidly as that of d_2 as the time interval under consideration is shifted into a more distant future. This is exactly the idea that Definition 1 embodies.

Prelec (2004, Proposition 1) proved the following equivalence on the at-least-as-decreasingly-impatient-as relation.

Theorem 1 (Prelec (2004)) *Let d_1 and d_2 be thrice differentiable discount factor functions and r_1 and r_2 be the corresponding discount rate functions. Then the following two conditions are equivalent.*

1. *d_1 is at least as decreasingly impatient as d_2.*
2. *$-r_1'(t)/r_1(t) \geq -r_2'(t)/r_2(t)$ for every $t \in \boldsymbol{R}_+$.*

Thanks to this theorem, to determine the ranking of the degree of decreasing impatience between two discount factor functions, it is sufficient to compare the rates of decrease of the corresponding discount rate functions. In the subsequent analysis, we identify how the rate of decrease of the representative consumer's discount rate function is related to the degree of heterogeneity of individual consumers' discount rates.

[5] Then both of the two ratios are equal to $\exp(-\bar{\rho}t_1)$.

3 Analysis

3.1 Discount rates in an economy

As we have mentioned earlier, we can assume that

$$\int_A (\lambda(a))^{1/\gamma} \, \mathrm{d}\nu(a) = 1.$$

Define a probability measure μ on $(\boldsymbol{R}_{++}, \mathscr{B}(\boldsymbol{R}_{++}))$ by letting

$$\mu(B) = \int_{\rho^{-1}(B)} (\lambda(a))^{1/\gamma} \, \mathrm{d}\nu(a) \tag{10}$$

for every $B \in \mathscr{B}(\boldsymbol{R}_{++})$. If all the $\rho(a)$'s were equal, then (3) implies that at the solution to the maximization problem (2), consumer a would consume fraction $(\lambda(a))^{1/\gamma}$ of the aggregate consumption at any time in any state. Thus the fraction $(\lambda(a))^{1/\gamma}$ can be considered as presenting the wealth share of consumer a in the case of homogeneous impatience. In the case of heterogeneous impatience, however, the consumption share is not constant across time, and the fraction $(\lambda(a))^{1/\gamma}$ does not coincide with the wealth share, which is determined by evaluating the solution of (2) by the Lagrange multipliers associated with its constraints.[6] Yet, according to Jouini and Napp (2007, Lemma 4.1), the fraction $(\lambda(a))^{1/\gamma}$ approximates the wealth share, and, as such, for each $B \in \mathscr{B}(\boldsymbol{R}_{++})$, $\mu(B)$ can be thought as approximating the proportion of the wealth owned by those who have discount rates in B. We can therefore think of μ as approximating the distribution of discount rates in terms of wealth shares.

By (7),

$$d(t) = \left(\int_{\boldsymbol{R}_{++}} \exp\left(-\frac{qt}{\gamma}\right) \mathrm{d}\mu(q) \right)^{\gamma}.$$

Let K be the cumulant-generating function of μ, that is,

$$K(s) = \ln\left(\int_{\boldsymbol{R}_{++}} \exp(sq) \, \mathrm{d}\mu(q) \right) \tag{11}$$

for every $s \in \boldsymbol{R}$ for which the integral on the right-hand side is finite. Note that if $s \leq 0$, then the integral is finite, because, then, $\exp(sq) \leq 1$ for every $q \in \boldsymbol{R}_{++}$.

It is well known if the first two moments exist (are finite), then K is twice differentiable, with $K'(0)$ equal to the mean of μ and $K''(0)$ equal to the variance of μ. Denote by $\mu(s)$ the probability measure on $\boldsymbol{R}_{++}$ such that

$$\frac{\mathrm{d}\mu(s)}{\mathrm{d}\mu}(q) = \exp(sq - K(s))$$

for every $q \in \boldsymbol{R}_{++}$. Then, by Morris (1982, Section 2), for every s, $K'(s)$ and $K''(s)$ are equal to the mean and variance of $\mu(s)$. Thus, $K'(s) > 0$ for every s, and, unless μ is concentrated

[6] Since the solution is a Pareto-efficient allocation, the second welfare theorem implies that it is an equilibrium allocation under some price system. The Lagrange multipliers constitute such a price system.

on a single point, $K''(s) > 0$ for every s. Hence $K''(s)/K'(s)$ is the ratio of the variance to the mean, and can be considered as the measure of dispersion of the probability measure $\mu(s)$.

Another interpretation of the ratio $K''(s)/K'(s)$ can be given in terms of the *variance function*, introduced by Morris (1982, Section 2). Assume that μ is not concentrated on a single point. Then $K''(s) > 0$ for every s, and hence K' is a one-to-one function from the set of s for which $K(s)$ is well defined (finite) to $\boldsymbol{R}_+$. Denote by $(K')^{-1}$ its inverse function, of which the domain is the range of K'. Then define the variance function $V = K'' \circ (K')^{-1}$. This is the function that maps the mean to the variance of distributions in the family $\{\mu(s) \mid K(s) \text{ is well defined}\}$. If $m = K'(s)$, then

$$\frac{K''(s)}{K'(s)} = \frac{V(m)}{m}. \tag{12}$$

Thus, $K''(s)/K'(s)$ can be interpreted as the slope of the chord connecting the point corresponding to $\mu(s)$ on the graph of V and the origin of $\boldsymbol{R}^2$.

The following proposition gives the fundamental relation between the representative consumer's discount rate function and the cumulant generating function of the approximate wealth-weighted distribution of individual consumers' discount rates.

Proposition 1 *For every $t \in \boldsymbol{R}_+$,*

$$r(t) = K'\left(-\frac{t}{\gamma}\right), \tag{13}$$

$$r'(t) = -\frac{1}{\gamma}K''\left(-\frac{t}{\gamma}\right), \tag{14}$$

$$-\frac{r'(t)}{r(t)} = \frac{1}{\gamma}\frac{K''(-t/\gamma)}{K'(-t/\gamma)}. \tag{15}$$

Proof of Proposition 1 By (7),

$$\ln d(t) = \gamma K\left(-\frac{t}{\gamma}\right).$$

By differentiating both sides with respect to t, we obtain (13). The other two equalities can be obtained from (13). ///

3.2 Comparison of discount rates in two economies

Now consider two economies, $n = 1, 2$, each with the space $(A_n, \mathscr{A}_n, \nu_n)$ of (names of) consumers and a weighting function $\lambda_n : A_n \to \boldsymbol{R}_{++}$ satisfying $\int_{A_n} (\lambda_n(a))^{1/\gamma}\, \mathrm{d}\nu_n(a) = 1$. Define, for each n, the discount rate function $r_n : \boldsymbol{R}_+ \to \boldsymbol{R}_{++}$, the probability measure μ_n on $\boldsymbol{R}_{++}$, and the cumulant-generating function K_n analogously to (10) and (11), using $(A_n, \mathscr{A}_n, \nu_n)$ and $\lambda_n : A_n \to \boldsymbol{R}_{++}$.

The following theorem is immediately obtained from (15) in Proposition 1.

Theorem 2 *The following two conditions are equivalent.*

1. *For every $t \geq 0$, $-r_1'(t)/r_1(t) > -r_2'(t)/r_2(t)$.*

2. *For every $s \leq 0$, $K_1''(s)/K_1'(s) > K_2''(s)/K_2'(s)$.*

The first condition of this theorem states that the representative consumer of the first economy is more dynamically inconsistent than the representative consumer of the second economy. Since

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{r_1(t)}{r_2(t)}\right) = \frac{r_1(t)}{r_2(t)}\left(\frac{r_1'(t)}{r_1(t)} - \frac{r_2'(t)}{r_2(t)}\right),$$

it is equivalent to the *monotone likelihood ratio condition*, in that $r_1(t)/r_2(t)$ is a strictly decreasing function of t. The second condition states that the cumulant-generating function of the distribution, measured in terms of the wealth held, of individual consumers' discount rates in the first economy is more convex than in the second economy. This condition is equivalent to saying that the variance divided by the mean of the individual consumers' discount rates is higher in the first economy than in the second whenever their distribution is transformed by a negative exponential density function. The theorem, then, asserts that the representative consumer is more dynamically inconsistent if and only if the cumulant-generating function of individual consumers' discount rates is more convex.

The next theorem deals with weaker conditions, although they are still useful to investigate the term structure of interest rates.

Theorem 3 *The following two conditions are equivalent.*

1. *For every $t \geq 0$, if $r_1(t) = r_2(t)$, then $r_1'(t) < r_2'(t)$.*

2. *For every $s \leq 0$, if $K_1'(s) = K_2'(s)$, then $K_1''(s) > K_2''(s)$.*

The condition in the first part of this theorem implies the single-crossing property, in that r_1 crosses r_2 at most once from above: if $r_1(t_0) = r_2(t_0)$, then $r_1(t) < r_2(t)$ for every $t > t_0$, and $r_1(t) > r_2(t)$ for every $t < t_0$. That is, the discount rate in the first economy is higher than in the second up to a time, after which the former is lower. The second part is the single-crossing property of the K_n', where K_1' crosses K_2' at most once from below. This condition is equivalent to saying that if the distributions of the individual consumers' discount rates are transformed by a negative exponential density function so that the means with respect to the transformed distribution are equal in the two economies, then the variance is higher in the first economy than in the second. This theorem follows directly from (13) and (14).

The following proposition gives an implication of Theorem 3 on the representative consumer's discount factor.

Proposition 2 *Suppose that for every $s \leq 0$, if $K_1'(s) = K_2'(s)$, then $K_1''(s) > K_2''(s)$. Then the function $t \mapsto d_1(t)/d_2(t)$ is either strictly decreasing everywhere, strictly increasing everywhere, or strictly decreasing up to some time, beyond which it is strictly increasing. Moreover, if there* *is a $t_1 > 0$ such that $d_1(t_1) = d_2(t_1)$, then it is strictly decreasing up to some time, beyond which it is strictly increasing.*

Proof of Proposition 2 Since $d_1(0) = d_2(0) = 1$,

$$\frac{d_1(t)}{d_2(t)} = \exp\left(\int_0^t (-r_1(\tau) + r_2(\tau))\,\mathrm{d}\tau\right). \tag{16}$$

Thus, if $r_1(t) > r_2(t)$ for every $t > 0$, then d_1/d_2 is strictly decreasing; and if $r_1(t) < r_2(t)$ for every $t > 0$, then d_1/d_2 is strictly increasing. Suppose that for every $s \leq 0$, if $K_1'(s) = K_2'(s)$, then $K_1''(s) > K_2''(s)$. Suppose that there is a $t_0 > 0$ such that $r_1(t_0) = r_2(t_0)$. Theorem 3 implies that $r_1(t) > r_2(t)$ for every $t < t_0$ and $r_1(t) < r_2(t)$ for every $t > t_0$. Since

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{d_1(t)}{d_2(t)}\right) = \frac{d_1(t)}{d_2(t)}\left(-r_1(t) + r_2(t)\right),$$

d_1/d_2 is strictly decreasing on $[0, t_0]$ and strictly increasing on $[t_0, \infty)$.

Suppose that there is a $t_1 > 0$ such that $d_1(t_1) = d_2(t_1)$. Then, by (16),

$$\int_0^{t_0} (-r_1(t) + r_2(t))\,\mathrm{d}t = 0.$$

Since the r_n are continuous, there is a $t_0 \in (0, t_1)$ such that $-r_1(t_0) + r_2(t_0) = 0$. We can then apply the argument in the previous paragraph to conclude that $d_1(t)/d_2(t)$ is strictly decreasing up to $t_0 > 0$, beyond which it is strictly increasing. ///

4 Term structure of interest rates

In this section, we explore the implication of the result in Section 3 on the term structure of interest rates. In particular, we consider the term structures of the two economies of which the cumulant-generating functions for the distribution of individual consumers' discount rates cross each other at most once, and compare the yield curves and forward rates.

4.1 Term structure of an economy

The representative consumer's marginal utility process $(d(t)u'(c_t))_{t\in\boldsymbol{R}_+}$, evaluated at the aggregate consumption process $c = (c_t)_{t\in\boldsymbol{R}_+}$, serves as a state price process. This means that the price at time t_1, relative to the current consumption, of the discount bond with maturity $t_2 > t_1$ is equal to

$$E_{t_1}\left(\frac{d(t_2)u'(c_{t_2})}{d(t_1)u'(c_{t_1})}\right) = \frac{d(t_2)}{d(t_1)}E_{t_1}\left(\frac{u'(c_{t_2})}{u'(c_{t_1})}\right) = \exp\left(-\int_{t_1}^{t_2} r(t)\,\mathrm{d}t\right)E_{t_1}\left(\frac{u'(c_{t_2})}{u'(c_{t_1})}\right). \tag{17}$$

We denote this price by $B(t_1, t_2)$. The *yield to maturity*, at time t_1, of the discount bond with maturity $t_2 > t_1$ is equal to

$$-\frac{1}{t_1 - t_2} \ln B(t_1, t_2) = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} r(t)\, \mathrm{d}t - \frac{1}{t_2 - t_1} \ln E_{t_1} \left(\frac{u'(c_{t_2})}{u'(c_{t_1})} \right). \tag{18}$$

We denote this $g(t_1, t_2)$.

Another rate that we are interested in is the *instantaneous forward rate*, determined at time t_1, for the delivery of the about-to-mature bond at time t_2 is equal to

$$-\frac{\partial}{\partial t_2} \ln B(t_1, t_2) = r(t_2) - \frac{\mathrm{d}}{\mathrm{d}t_2} \ln E_{t_1} \left(\frac{u'(c_{t_2})}{u'(c_{t_1})} \right),$$

if

$$E_{t_1} \left(\frac{u'(c_{t_2})}{u'(c_{t_1})} \right) \tag{19}$$

is a differentiable function of t_2. We denote this by $f(t_1, t_2)$.

4.2 Term structures of two economies

Let d_1 and d_2 be the discount factor functions derived from two economies, with the cumulant-generating functions K_1 and K_2. Let r_1 and r_2 be the corresponding discount rate functions. For each $= 1, 2$, let B_n, g_n, and f_n be the corresponding bond prices, yields to maturity, and instantaneous forward rates. Then

$$\frac{B_1(t_1, t_2)}{B_2(t_1, t_2)} = \exp \left(- \int_{t_1}^{t_2} (r_1(t) - r_2(t))\, \mathrm{d}t \right),$$

$$g_1(t_1, t_2) - g_2(t_1, t_2) = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} (r_1(t) - r_2(t))\, \mathrm{d}t, \tag{20}$$

$$f_1(t_1, t_2) - f_2(t_1, t_2) = r_1(t_2) - r_2(t_2). \tag{21}$$

Note that while the $B_n(t_1, \cdot)$, $g_n(t_1, \cdot)$, and $f_n(t_1, \cdot)$ are, in general, stochastic processes (because, for each fixed t_1, (19) is a stochastic process in t_2), none of the above is stochastic. Moreover, the difference in the instantaneous forward rates, (21), does not depend on the time t_1 at which it is evaluated.

Theorem 4 *Suppose that for every $s \leq 0$, if $K_1'(s) = K_2'(s)$, then $K_1''(s) > K_2''(s)$. Then for every $t_1 \geq 0$ and every $t_2 > t_1$, if $f_1(t_1, t_2) = f_2(t_1, t_2)$, then $\partial f_1(t_1, t_2)/\partial t_2 < \partial f_2(t_1, t_2)/\partial t_2$.*

This theorem states that if the cumulant-generating functions of the distributions of discount rates of two economies have the single-crossing property, then so does the instantaneous forward rates, that is, if $f_1(t_1, t_2) = f_2(t_1, t_2)$, then $f_1(t_1, t) < f_2(t_1, t)$ for every $t > t_2$ and $f_1(t_1, t) > f_2(t_1, t)$ for every $t \in [t_1, t_2)$. In words, if the two economies have an equal instantaneous forward rate applicable to some future time, then the instantaneous forward rate applicable to any earlier time is higher, and that applicable to any later time is lower, in the first economy than in the second. Theorem 4 can be proved by using Theorem 3 and (21).

Theorem 4 compares the instantaneous forward rates, in the two economies, determined at a fixed time t_1 but with a variable delivery time t_2. Another comparison worth exploring is the instantaneous forward rates, with a fixed time to maturity, say τ, but with a variable time t_1 at which the rates are determined. What this means, in symbols, is how

$$f_1(t_1, t_1 + \tau) - f_2(t_1, t_1 + \tau) \tag{22}$$

depends on t_1, when τ is a positive constant. In fact, by (21), (22) is equal to $r_1(t_1+\tau)-r_2(t_1+\tau)$, which, when t_1 varies, behaves in the same way as $r_1(t_2)-r_2(t_2)$ when t_2 varies. We can conclude, therefore, that the instantaneous forward rates in two economies, with a fixed time to maturity but with a variable time at which the rates are determined, also has the single-crossing property.

The next theorem shows the single-crossing property for the yields to maturity.

Theorem 5 *Suppose that for every $s \le 0$, if $K_1'(s) = K_2'(s)$, then $K_1''(s) > K_2''(s)$. Then for every $t_1 \ge 0$ and every $t_2 > t_1$, if $g_1(t_1, t_2) = g_2(t_1, t_2)$, then $\partial g_1(t_1, t_2)/\partial t_2 < \partial g_2(t_1, t_2)/\partial t_2$.*

Proof of Theorem 5 By (20) and a straightforward calculation,

$$\frac{\partial g_1}{\partial t_2}(t_1, t_2) - \frac{\partial g_1}{\partial t_2}(t_1, t_2) = \frac{1}{t_2 - t_1}\left((r_1(t_2) - r_2(t_2)) - (g_1(t_1, t_2) - g_2(t_1, t_2))\right).$$

Thus, if $g_1(t_1, t_2) = g_2(t_1, t_2)$, then

$$\frac{\partial g_1}{\partial t_2}(t_1, t_2) - \frac{\partial g_1}{\partial t_2}(t_1, t_2) = \frac{1}{t_2 - t_1}\left(r_1(t_2) - r_2(t_2)\right), \tag{23}$$

and, again by (20),

$$\int_{t_1}^{t_2} (r_1(t) - r_2(t))\, \mathrm{d}t = 0.$$

Thus, there exists a $t_0 \in (t_1, t_2)$ such that $r_1(t_0) = r_2(t_0)$. By the single-crossing property of the K_n' and Theorem 2, the r_n also have the single-crossing property. Since $t_2 > t_0$, $r_1(t_2) < r_2(t_2)$. By (23),

$$\frac{\partial g_1}{\partial t_2}(t_1, t_2) - \frac{\partial g_1}{\partial t_2}(t_1, t_2) < 0.$$

///

5 Quasi-log-linear families

If the distributions of individual consumers' discount rates are taken from a family of some special type, then the cumulant-generating function K admits a simple expression. In this section, we explore implications of the expression and present four examples.

5.1 General results

First, we give the definition of such a family.

Definition 2 A family $\mathscr{M}$ of probability measures on $(\boldsymbol{R}_{++}, \mathscr{B}(\boldsymbol{R}_{++}))$ is a *quasi-log-linear family* if there exist a σ-finite measure Λ on $\boldsymbol{R}_{++}$, a non-negative integer L, an open subset Θ of $\boldsymbol{R}$, an open subset Ξ of $\boldsymbol{R}^L$, and two mappings $g : \Xi \times \boldsymbol{R}_{++} \to \boldsymbol{R}$ and $h : \Theta \times \Xi \to \boldsymbol{R}$ such that:

1. $h(\cdot, \xi) : \Theta \to \boldsymbol{R}$ is twice differentiable for every $\xi \in \Xi$;
2. for every probability measure μ on $(\boldsymbol{R}_{++}, \mathscr{B}(\boldsymbol{R}_{++}))$, $\mu \in \mathscr{M}$ if and only if there exists a $(\theta, \xi) \in \Theta \times \Xi$ such that the function

$$q \mapsto \exp\left(\theta q + g(\xi, q) + h(\theta, \xi)\right) \tag{24}$$

from $\boldsymbol{R}_{++}$ into $\boldsymbol{R}_{++}$ is a version of the Radon-Nikodym derivative $\mathrm{d}\mu/\mathrm{d}\Lambda$; and

The name "quasi-log-linear" comes from the fact that the logarithm of the version of the Radon-Nikodym derivative has a linear term θq. If $L = 0$, then (24) is reduced to

$$q \mapsto \exp\left(\theta q + g(q) + h(\theta)\right). \tag{25}$$

This is what Morris (1982) defined as a natural exponential family. Thus a quasi-log-linear family is a multi-dimensional extension, where the linear term $\theta^1 q$ is retained. Also, if, for each $\ell = 0, 1, \ldots, L$, there exists a $g^\ell : \boldsymbol{R}_{++} \to \boldsymbol{R}$ such that

$$g(\xi, q) = g^0(q) + \sum_{\ell=1}^{L} \xi^\ell g^\ell(q)$$

for every $(\xi, q) = (\xi^1, \ldots, \xi^L, q) \in \Xi \times \boldsymbol{R}_{++}$, then (24) is a canonical expression of an exponential family. We will see in the next subsection that this extension allows us to compare more distributions and derive rankings in terms of the at-least-as-decreasingly-impatient-as relation.

Since every $\mu \in \mathscr{M}$ has a strictly positive (version of the) Radon-Nikodym derivative, it is equivalent to the reference measure Λ. Thus, the probability measures in $\mathscr{M}$ are equivalent to one another. Hence, for all $\mu_1 \in \mathscr{M}$ and $\mu_2 \in \mathscr{M}$, corresponding to $(\theta_1, \xi_1) \in \Theta \times \Xi$ and $(\theta_2, \xi_2) \in \Theta \times \Xi$, the Radon-Nikodym derivative $\mathrm{d}\mu_1/\mathrm{d}\mu_2$ exists and the function

$$q \mapsto \frac{\exp\left(\theta_1 q + g(\xi_1, q) + h(\theta_1, \xi_1)\right)}{\exp\left(\theta_2 q + g(\xi_2, q) + h(\theta_2, \xi_2)\right)} = \exp\left((\theta_1 - \theta_2)q + (g(\xi_1, q) - g(\xi_2, q)) + (h(\theta_1, \xi_1) - h(\theta_2, \xi_2))\right).$$

is a version of it. Therefore, assuming that $\Lambda \in \mathscr{M}$, $(0, 0) \in \Theta \times \Xi$, $h(0, 0) = 0$, and $g(0, q) = 0$ for every $q \in \boldsymbol{R}_{++}$ would lose no generality.

In the rest of this subsection, we let $\mathscr{M}$ be a quasi-log-linear family, with the Radon-Nikodym derivatives defined by (24). Denote the distribution in $\mathscr{M}$ of which the Radon-Nikodym derivative is (24) by $\mu(\theta, \xi)$ and its cumulant-generating function by $K(\cdot, \theta, \xi)$. When the wealth-weighted distribution of individual consumers' discount rates is $\mu(\theta, \xi)$, denote the representative consumer's discount rate by $r(\cdot, \theta, \xi)$.

Lemma 1 *For every $(\theta, \xi) \in \Theta \times \Xi$ and every $s \in \boldsymbol{R}$ with $\theta + s \in \Theta$,*

$$\frac{\partial K}{\partial s}(s, \theta, \xi) = -\frac{\partial h}{\partial \theta^1}(\theta + s, \xi), \tag{26}$$

$$\frac{\partial^2 K}{\partial s^2}(s, \theta, \xi) = -\frac{\partial^2 h}{\partial (\theta^1)^2}(\theta + s, \xi),$$

$$\frac{\dfrac{\partial^2 K}{\partial s^2}(s, \theta, \xi)}{\dfrac{\partial K}{\partial s}(s, \theta, \xi)} = \frac{\dfrac{\partial^2 h}{\partial (\theta^1)^2}(\theta + s, \xi)}{\dfrac{\partial h}{\partial \theta^1}(\theta + s, \xi)},. \tag{27}$$

When $L = 0$, $\theta \in \boldsymbol{R}$ and the results of the lemma can be more simply written as

$$\frac{\partial K}{\partial s}(s, \theta) = -h'(\theta + s), \tag{28}$$

$$\frac{\partial^2 K}{\partial s^2}(s, \theta) = -h''(\theta + s), \tag{29}$$

$$\frac{\dfrac{\partial^2 K}{\partial s^2}(s, \theta)}{\dfrac{\partial K}{\partial s}(s, \theta)} = \frac{h''(\theta + s)}{h'(\theta + s)}.$$

Proof of Lemma 1 By a straightforward calculation,

$$\exp K(s, \theta, \xi) = \int_{\boldsymbol{R}_{++}} \exp\left((\theta + s)q + g(\xi, q) + h(\theta, \xi)\right) \, \mathrm{d}\mu(q) = \frac{\exp h(\theta, \xi)}{\exp h(\theta + s, \xi)}.$$

Hence

$$K(s, \theta, \xi) = h(\theta, \xi) - h(\theta + s, \xi).$$

By differentiating both sides with respect to s, we complete the proof. ///

The following propositions compares two representative consumers' discount rate functions in terms of $h(\theta)$ in the Radon-Nikodym derivative $\mathrm{d}\mu(\theta)/\mathrm{d}\Lambda$. They follow directly from Lemma 1 and Theorems 2 and 3.

Proposition 3 *For all $(\theta_1, \xi_1) \in \Theta \times \Xi$ and $(\theta_2, \xi_2) \in \Theta \times \Xi$, the following two conditions are equivalent.*

1. *For every $t \geq 0$,*

$$-\frac{\dfrac{\partial r}{\partial t}(t, \theta_1, \xi_1)}{r(t, \theta_1, \xi_1)} \geq -\frac{\dfrac{\partial r}{\partial t}(t, \theta_2, \xi_2)}{r(t, \theta_2, \xi_2)}$$

2. For every $s \leq 0$,

$$\frac{\dfrac{\partial^2 h}{\partial \theta^2}(\theta_1 + s, \xi_1)}{\dfrac{\partial h}{\partial \theta}(\theta_1 + s, \xi_1)} \geq \frac{\dfrac{\partial^2 h}{\partial \theta^2}(\theta_2 + s, \xi_2)}{\dfrac{\partial h}{\partial \theta^1}(\theta_2 + s, \xi_2)}.$$

Proposition 4 *For all $(\theta_1, \xi_1) \in \Theta \times \Xi$ and $(\theta_2, \xi_2) \in \Theta \times \Xi$, the following two conditions are equivalent.*

1. For every $t \geq 0$, if $r(t, \theta_1, \xi_1) = r(t, \theta_2, \xi_2)$, then

$$-\frac{\partial r}{\partial t}(t, \theta_1, \xi_1) > -\frac{\partial r}{\partial t}(t, \theta_2, \xi_2).$$

2. For every $s \leq 0$, if

$$\frac{\partial h}{\partial \theta}(\theta_1 + s, \xi_1) = \frac{\partial h}{\partial \theta}(\theta_2 + s, \xi_2),$$

then

$$\frac{\partial^2 h}{\partial \theta^2}(\theta_1 + s, \xi_1) < \frac{\partial^2 h}{\partial \theta^2}(\theta_2 + s, \xi_2).$$

The following proposition eases the task of checking the more-convex-than relation between two cumulant-generating functions in a quasi-linear exponential family.

Proposition 5 *Suppose that the reference probability measure Λ of the quasi-linear exponential family $\mathscr{M}$ is not concentrated on any single point, and that*

$$\frac{\partial h}{\partial \theta}(\theta, \xi) \frac{\partial^3 h}{\partial \theta^3}(\theta, \xi) \gtreqless \left(\frac{\partial^2 h}{\partial \theta^2}(\theta, \xi) \right)^2 \tag{30}$$

for every $(\theta, \xi) \in \Theta \times \Xi$. Then, for all $\theta_1 \in \Theta$, $\theta_2 \in \Theta$, and $\xi \in \Xi$, if $\theta_1 > \theta_2$, then

$$-\frac{\dfrac{\partial r}{\partial t}(t, \theta_1, \xi)}{r(t, \theta_1, \xi)} \gtreqless -\frac{\dfrac{\partial r}{\partial t}(t, \theta_2, \xi)}{r(t, \theta_2, \xi)} \tag{31}$$

for every $t \geq 0$.

The conditions of this proposition can be written in a particularly simple form when $L = 0$. Indeed, (30) can written as

$$h'(\theta) h'''(\theta) \gtreqless \left(h''(\theta) \right)^2. \tag{32}$$

While Proposition 5 allows us only to compare two distributions parameterized by (θ_1, ξ_1) and (θ_2, ξ_2) with $\xi_1 = \xi_2$, if $L = 0$, then this means that it allows us to compare any two distributions in $\mathscr{M}$. For these reasons, Proposition 5 is most useful when $\mathscr{M}$ is a natural exponential family in the sense of Morris (1982).

Proof of Proposition 5 Differentiate both sides of (27) with respect to θ^1, then we obtain

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\dfrac{\partial^2 K}{\partial s^2}(s,\theta,\xi)}{\dfrac{\partial K}{\partial s}(s,\theta,\xi)}\right) = \frac{1}{\left(\dfrac{\partial h}{\partial \theta}(\theta+s,\xi)\right)^2}\left(\frac{\partial h}{\partial \theta}(\theta+s,\xi)\frac{\partial^3 h}{\partial \theta^3}(\theta+s,\xi) - \left(\frac{\partial^2 h}{\partial \theta^2}(\theta+s,\xi)\right)^2\right). \tag{33}$$

By assumption,

$$\frac{\partial h}{\partial \theta}(\theta+s,\xi)\frac{\partial^3 h}{\partial \theta^3}(\theta+s,\xi) - \left(\frac{\partial^2 h}{\partial \theta^2}(\theta+s,\xi)\right)^2 \gtrless 0.$$

Thus,

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\dfrac{\partial^2 K}{\partial s^2}(s,\theta,\xi)}{\dfrac{\partial K}{\partial s}(s,\theta,\xi)}\right) \gtrless 0.$$

Hence, if $\theta_1 > \theta_2$, then

$$\frac{\dfrac{\partial^2 K}{\partial s^2}(s,\theta_1,\xi)}{\dfrac{\partial K}{\partial s}(s,\theta_1,\xi)} \gtrless \frac{\dfrac{\partial^2 K}{\partial s^2}(s,\theta_2,\xi)}{\dfrac{\partial K}{\partial s}(s,\theta_2,\xi)}.$$

Now (31) follows from (15). ///

5.2 Examples

In this subsection, we give four examples of quasi-log-linear families from which the approximate wealth-weighted distributions of individual consumers' discount rates are taken. They consist of gamma, binomial, Poisson, and negative binomial distributions. This list does not contain, most notably, uniform and log-normal distributions: uniform distributions do not constitute a quasi-log-linear family, because no pair of them shares the same support; and log-normal distributions do not constitute a quasi-log-linear family, because no linear term θq can appear in any parameterization (θ, ξ) of the family of log-normal distributions.

Of the four families, the binomial and Poisson families are one-parameter families ($L = 0$), and, thus, natural exponential families in the sense of Morris (1982). The gamma and negative binomial families are two-parameter families ($L = 1$), and, thus, contain natural exponential families, each corresponding to a fixed value of the parameter $\xi \in \Xi$. Our examples, therefore, involves four (kinds of) natural exponential families.

We now argue that there is a good reason to investigate these four natural exponential families. Each such family consists of probability measures $\mu(\theta)$ having Radon-Nikodym derivatives (25). By (28) and (29), the mean and variance of $\mu(\theta)$ are equal to $-h'(\theta)$ and $-h''(\theta)$. Unless the $\mu(\theta)$ are concentrated on a single point, h' is a strictly decreasing function and, hence, its inverse function $(h')^{-1}$ is well defined on its domain. Define a function V on the domain by letting $V(m) = -h''\left((h')^{-1}(-m)\right)$ for every m in the domain. Then the function V, called the *variance function* by Morris (1982), maps the mean to the variance of probability measures in the natural exponential family. The distinctive feature of the four examples is that V is a quadratic function. Morris (1982, Section 4) found all six natural exponential families having a quadratic V, and the two not listed here are normal distributions and *generalized hyperbolic secant distributions*, which both have support $\boldsymbol{R}$ and, thus, cannot be used as a distribution of discount rates. Our examples, therefore, exhaust all the natural exponential families with quadratic variance functions that can be used as a distribution of discount rates.

First consider the family of gamma distributions, each with parameters α and β, that is, its density function (with respect to Lebesgue measure) is given by

$$q \mapsto \frac{\beta^{\alpha} q^{\alpha-1}}{\Gamma(\alpha)} \exp(-\beta q).$$

Gamma distributions are most commonly used as distributions of subjective time discount rates, as in Weitzman (2001), Gollier and Zeckhauser (2005), and Hara (2007). Weitzman (2001) also provided an empirical justification for the use of gamma functions: he asked more than two thousand PhD-level economists which discount factor should be used to discount the cost and benefit of mitigating climate change, and the answers turned out to follow a gamma distribution with mean 3.96% and standard deviation 2.94%.

Its cumulant generating function $K(\cdot,(\alpha,\beta))$ is given by

$$K(s,(\alpha,\beta)) = \alpha(\ln\beta - \ln(\beta - s)).$$

Thus

$$\frac{\dfrac{\partial^2 K}{\partial s^2}(s,(\alpha,\beta))}{\dfrac{\partial K}{\partial s}(s,(\alpha,\beta))} = \frac{1}{\beta - s}.$$

Hence, Theorem 2 tells us that the smaller the value of β, the more dynamically inconsistent the representative consumer. Note that the values of the α are irrelevant for the ranking of dynamic inconsistency.

The above proof did not rely on the fact that the gamma distributions constitute a quasi-log-linear family. To give a proof that does use the fact, take the reference σ-finite measure Λ as the Lebesgue measure on $\boldsymbol{R}_{++}$, Let $L = 1$, $\Theta = (-\infty, 0)$, and $\Xi = (-1, \infty)$. Define $g : \Xi \times \boldsymbol{R}_{++} \to \boldsymbol{R}$ by $g(\xi, q) = \xi \ln q$, and $h : \Theta \times \Xi \to \boldsymbol{R}$ by $h(\theta, \xi) = (\xi + 1)\ln(-\theta) - \ln\Gamma(\xi + 1)$. Let $\mathscr{M}$ be the set of all probability measures on $(\boldsymbol{R}_{++}, \mathscr{B}(\boldsymbol{R}_{++}))$ for which there is a $(\theta, \xi) \in \Theta \times \Xi$ such that

$$\frac{\mathrm{d}\mu}{\mathrm{d}\Lambda}(q) = \exp\left(\theta q + g(\xi, q) + h(\theta, \xi)\right).$$

Denote the above measure by $\mu(\theta, \xi)$. Define $\varphi : \Theta \to \boldsymbol{R}^2_{++}$ by $\varphi(\theta, \xi) = (\xi + 1, -\theta)$. Then φ is one to one and onto. Moreover, for every $(\alpha, \beta) \in \boldsymbol{R}^2_{++}$, $\varphi^{-1}(\alpha, \beta) = (-\beta, \alpha - 1)$ and, hence,

$$\frac{\mathrm{d}\mu(\varphi^{-1}(\alpha,\beta))}{\mathrm{d}\Lambda}(q) = \exp\left(-\beta q + (\alpha - 1)\ln q + \alpha \ln\beta - \ln\Gamma(\alpha)\right) = \frac{\beta^{\alpha} q^{\alpha-1}}{\Gamma(\alpha)} \exp(-\beta q).$$

Hence $\mu(\theta, \xi)$ is a probability measure and $\mathscr{M}$ coincides with the set of the gamma distributions.

Moreover,

$$\frac{\partial h}{\partial \theta}(\theta,\xi) = -\frac{1+\xi}{\theta},$$
$$\frac{\partial^2 h}{\partial \theta^2}(\theta) = \frac{1+\xi}{\theta^2}.$$

Thus

$$\frac{\dfrac{\partial^2 h}{\partial \theta^2}(\theta)}{\dfrac{\partial h}{\partial \theta}(\theta)} = -\frac{1}{\theta}.$$

By Proposition 3, the representative consumer is more dynamically inconsistent the larger the value of $\theta = -\beta$, that is, the smaller the value of β.

Next, we consider the family of the binomial distributions that put, for some positive constant ε and positive integer n, positive probabilities only on $\varepsilon, 2\varepsilon, \ldots, n\varepsilon$. To be more specific, let $\alpha \in (0,1)$ and, for each $m \in \{1,2,\ldots,n\}$, the distribution $\mu(\alpha)$ gives $m\varepsilon$ a probability

$$\begin{pmatrix} n-1 \\ m-1 \end{pmatrix} \alpha^{m-1}(1-\alpha)^{n-m}. \tag{34}$$

Take Λ as the counting measure on $\{\varepsilon, 2\varepsilon, \ldots, n\varepsilon\}$. Let $L = 0$ and $\Theta = \boldsymbol{R}$. Define $g : \boldsymbol{R}_{++} \to \boldsymbol{R}$ by $g(q) = \ln(n-1)! - \ln \mathrm{B}(n - q/\varepsilon, q/\varepsilon - 1)$, where B is the beta function, and $h : \Theta \to \boldsymbol{R}$ by $h(\theta) = -\varepsilon\theta - (n-1)\ln(1 + \exp(\varepsilon\theta))$ and Let $\mathscr{M}$ be the set of all probability measures μ on $(\boldsymbol{R}_{++}, \mathscr{B}(\boldsymbol{R}_{++}))$ for which there is a $\theta \in \Theta$ such that

$$\frac{\mathrm{d}\mu}{\mathrm{d}\Lambda}(q) = \exp\left(\theta q + g(q) + h(\theta)\right).$$

Denote the above measure by $\mu(\theta)$. Define $\varphi : \Theta \to (0,1)$ by $\varphi(\theta) = \exp(\varepsilon\theta)/(1+\exp(\varepsilon\theta))$. Then φ is one to one and onto. Moreover, for every $\alpha \in \boldsymbol{R}_{++}$, $\varphi^{-1}(\alpha) = \varepsilon^{-1}(\ln\alpha - \ln(1-\alpha))$ and hence

$$\begin{aligned}
&\frac{\mathrm{d}\mu(\varphi^{-1}(\alpha))}{\mathrm{d}\Lambda}(q) \\
=&\exp\left(\varphi^{-1}(\alpha)q + g(q) + h(\varphi^{-1}(\alpha))\right) \\
=&\exp\left((\ln\alpha - \ln(1-\alpha))\left(\frac{q}{\varepsilon} - 1\right) - (n-1)\ln\left(1 + \frac{\alpha}{1-\alpha}\right) + \ln\frac{(n-1)!}{\mathrm{B}\left(n - \dfrac{q}{\varepsilon}, \dfrac{q}{\varepsilon} - 1\right)}\right) \\
=&\exp\left(\left(\frac{q}{\varepsilon} - 1\right)\ln\alpha - \left(\frac{q}{\varepsilon} - n\right)\ln(1-\alpha) + \ln\frac{(n-1)!}{\mathrm{B}\left(n - \dfrac{q}{\varepsilon}, \dfrac{q}{\varepsilon} - 1\right)}\right) \\
=&\frac{(n-1)!}{\mathrm{B}\left(n - \dfrac{q}{\varepsilon}, \dfrac{q}{\varepsilon} - 1\right)}\alpha^{q/\varepsilon - 1}(1-\alpha)^{n - q/\varepsilon},
\end{aligned}$$

and the last term is equal to

$$\begin{pmatrix} n-1 \\ m-1 \end{pmatrix} \alpha^{m-1}(1-\alpha)^{n-m}$$

if $q = m\varepsilon$. Hence, $\mu(\theta)$ is a probability measure and $\mathscr{M}$ coincides with the set of the Poisson distributions. Moreover,

$$\begin{aligned} h'(\theta) &= -\varepsilon - (n-1)\frac{\varepsilon \exp(\varepsilon\theta)}{1+\exp(\varepsilon\theta)}, \\ h''(\theta) &= -(n-1)\frac{\varepsilon^2 \exp(\varepsilon\theta)}{(1+\exp(\varepsilon\theta))^2}, \\ h'''(\theta) &= -(n-1)\frac{\varepsilon^3 \exp(\varepsilon\theta)}{(1+\exp(\varepsilon\theta))^3}\left(1-\exp(\varepsilon\theta)\right), \\ h'(\theta)h'''(\theta) - (h''(\theta))^2 &= (n-1)\frac{\varepsilon^4 \exp(\varepsilon\theta)}{(1+\exp(\varepsilon\theta))^4}\left(1+(n-2)\exp(2\varepsilon\theta)\right), \end{aligned}$$

and the last term is strictly positive whenever $n \geq 2$. Hence, by (32), the larger the "success" probability $\alpha = \varphi(\theta) = 1/(1+\exp(-\varepsilon\theta))$ of the binomial distribution, the more dynamically inconsistent the representative consumer is.

Note that the binomial distributions we used in the above example are unusual, in that each discount rate $m\varepsilon$ with $m = 1, 2, , \dots, n$ is given the probability (34), rather than

$$\begin{pmatrix} n \\ m \end{pmatrix} \alpha^{m}(1-\alpha)^{n-m}. \tag{35}$$

We used these distributions because we did not wish to give any positive probability on the zero discount rate (which would give rise to complications in the model of infinite time horizon). However, even if we gave each discount rate $m\varepsilon$ with $m = 0, 1, 2, \dots, n$ the probability (35) (which would not give rise to complications in models of finite time horizon), we would still obtain the same conclusion: the larger the success probability α, the more dynamically inconsistent the representative consumer is.

Third, we consider the family of the Poisson distributions, and show that whether a change in the parameter value of the Poisson distribution leads to a change in the degree of dynamic inconsistency depends on whether the zero discount rate is given a positive probability. We shall do so by using the variance function in Morris (1982), which maps the mean to the variance of the distributions in this family.

Recall that the Poisson distribution with parameter α gives each nonnegative integer n a probability $\exp(-\alpha)\alpha^n/n!$. Its mean and variance are both equal to α. Thus, if we denote by V the variance function of the family of Poisson distributions, then $V(m) = m$, and hence $V(m)/m = 1$ for every m. Note that V does not depend on the value of α.

Let $\varepsilon > 0$. Consider, first, the distribution that gives each discount rate $n\varepsilon$ with $n = 0, 1, 2, \dots$ a probability $\exp(-\alpha)\alpha^n/n!$. Denote by $K_1(\cdot, \alpha)$ its cumulant generating function parameterized by α. Then $\partial K_1(s,\alpha)/\partial s$ is equal to ε times the mean of the Poisson distribution with parameter $\alpha\exp(s)$; and $\partial^2 K_1(s,\alpha)/\partial s^2$ is equal to ε^2 times the variance of the Poisson distribution with parameter $\alpha\exp(s)$. Thus, if we denote by V_1 the variance function of the family of these distributions, then $V_1(m) = \varepsilon m$ and, hence, $V_1(m)/m = \varepsilon$ for every m. Therefore, by (12),

$$\frac{\dfrac{\partial^2 K_1}{\partial s^2}(s,\alpha)}{\dfrac{\partial K_1}{\partial s}(s,\alpha)} = \varepsilon.$$

Since this does not depend on the value of α, we can conclude that, if the zero discount rate is given a positive probability, then the representative consumers are all equally dynamically inconsistent, regardless of the values of α.

Consider, next, the distribution that gives each discount rate $n\varepsilon$ with $n = 1, 2, \dots$ a probability $\exp(-\alpha)\alpha^{n-1}/(n-1)!$. Denote by $K_2(\cdot,\alpha)$ its cumulant generating function parameterized by α. Then $\partial K_2(s,\alpha)/\partial s$ is equal to ε plus ε times the mean of the Poisson distribution with parameter $\alpha\exp(s)$; and $\partial^2 K_1(s,\alpha)/\partial s^2$ is equal to ε^2 times the variance of the Poisson distribution with parameter $\alpha\exp(s)$. Thus, if we denote by V_2 the variance function of the family of these distributions, then $V_2(m) = \varepsilon(m-\varepsilon)$ and, hence, $V_1(m)/m = \varepsilon - \varepsilon^2/m$ for every m. Therefore, by (12),

$$\frac{\dfrac{\partial^2 K_2}{\partial s^2}(s,\alpha)}{\dfrac{\partial K_2}{\partial s}(s,\alpha)} = \varepsilon - \frac{\varepsilon^2}{\dfrac{\partial K_2}{\partial s}(s,\alpha)}.$$

Since $\partial K_2(s,\alpha)/\partial s$ is a strictly increasing function of α, we can conclude that, if the zero discount rate is is given zero probability, then the larger the value of α, the more dynamically inconsistent the representative consumer is.

Finally, we consider the family of the negative binomial distributions that put, for some positive constant ε, positive probabilities only on $0, \varepsilon, 2\varepsilon, \dots,$. To be more specific, let $\alpha \in (0,1)$ and $\beta \in \boldsymbol{R}_{++}$. Then, for each non-negative integer n, the distribution $\mu(\alpha,\beta)$ gives $n\varepsilon$ a probability

$$\begin{pmatrix} n+\beta-1 \\ n \end{pmatrix} \alpha^\beta (1-\alpha)^n = \frac{\Gamma(n+\beta)}{\Gamma(\beta)n!}\alpha^\beta(1-\alpha)^n.$$

Note that that the zero discount rate is given a positive probability. This is to avoid difficult calculations, but, as mentioned in the example of Poisson distributions, it would cause no complication in models of finite time horizon.

Take Λ as the counting measure on $\{0, \varepsilon, 2\varepsilon, \dots\}$. Let $L = 1$, $\Theta = -\boldsymbol{R}_{++}$, and $\Xi = \boldsymbol{R}_{++}$. Define $g : \Xi \times \boldsymbol{R}_{++} \to \boldsymbol{R}$ by $g(\xi, q) = \ln\Gamma(q/\varepsilon + \xi) - \ln\Gamma(q/\varepsilon + 1)$, and $h : \Theta \times \Xi \to \boldsymbol{R}$ by $h(\theta,\xi) = \xi\ln(1-\exp(\varepsilon\theta)) - \ln\Gamma(\xi)$. Let $\mathscr{M}$ be the set of all probability measures μ on $(\boldsymbol{R}_{++}, \mathscr{B}(\boldsymbol{R}_{++}))$ for which there is a $(\theta,\xi) \in \Theta \times \Xi$ such that

$$\frac{\mathrm{d}\mu}{\mathrm{d}\Lambda}(q) = \exp\left(\theta q + g(\xi,q) + h(\theta,\xi)\right).$$

Denote the above measure by $\mu(\theta, \xi)$. Define $\varphi : \Theta \times \Xi \to (0,1) \times \boldsymbol{R}_{++}$ by $\varphi(\theta, \xi) = (1 - \exp(\varepsilon\theta), \xi)$. Then φ is one to one and onto. Moreover, for every $(\alpha, \beta) \in (0,1) \times \boldsymbol{R}_{++}$, $\varphi^{-1}(\alpha, \beta) = (\varepsilon^{-1} \ln(1-\alpha), \beta)$ and hence

$$
\begin{aligned}
&\frac{\mathrm{d}\mu(\varphi^{-1}(\alpha,\beta))}{\mathrm{d}\Lambda}(q) \\
&= \exp\left(\left(\frac{1}{\varepsilon}\ln(1-\alpha)\right) q + g(\beta, q) + h\left(\frac{1}{\varepsilon}\ln(1-\alpha), \beta\right)\right) \\
&= \exp\left(\frac{q}{\varepsilon}\ln(1-\alpha) + \ln\Gamma\left(\frac{q}{\varepsilon} + \beta\right) - \ln\Gamma\left(\frac{q}{\varepsilon}+1\right) + \beta\ln\alpha - \ln\Gamma(\beta)\right) \\
&= \frac{\Gamma\left(\frac{q}{\varepsilon}+\beta\right)}{\Gamma(\beta)\Gamma\left(\frac{q}{\varepsilon}+1\right)}\alpha^{\beta}(1-\alpha)^{q/\varepsilon},
\end{aligned}
$$

and the last term is equal to

$$\frac{\Gamma(n+\beta)}{\Gamma(\beta)n!}\alpha^{\beta}(1-\alpha)^{n}.$$

if $q = n\varepsilon$. Hence, $\mu(\theta)$ is a probability measure and $\mathscr{M}$ coincides with the set of the negative binomial distributions. Moreover,

$$
\begin{aligned}
\frac{\partial h}{\partial \theta}(\theta, \xi) &= \xi \frac{-\varepsilon \exp(\varepsilon\theta)}{1 - \exp(\varepsilon\theta)}, \\
\frac{\dfrac{\partial^2 h}{\partial \theta^2}(\theta,\xi)}{\dfrac{\partial h}{\partial \theta}(\theta,\xi)} &= \frac{\varepsilon}{1-\exp(\varepsilon\theta)},
\end{aligned}
$$

and the last term is an increasing function of θ. Thus, the smaller the value of $\alpha = 1 - \exp(\varepsilon\theta)$, the more dynamically inconsistent the representative consumer is. As in the case of gamma distributions, the value of the other parameter, β, is irrelevant for the degree of dynamic inconsistency.

The above two-parameter family of negative binomial distributions, parameterized by (α, β) is not an exponential family, because $\Gamma(n+\beta)$ is not multiplicatively separable between n and β. Hence, although this family is a quasi-log-linear family, it is not an exponential family.

6 Conclusion

In this paper, we have given a precise formulation to the notion that the more heterogeneous the individual consumers' subjective discount rates are, the more dynamically inconsistent the representative consumer is. The result allowed us to compare the term structures of interest rates of two heterogeneous economies. When the wealth-weighted distributions of subjective discount rates are drawn from what we termed a quasi log-linear family, the task of determining the more-heterogeneous-than relation is easy, and some well known distributions are shown to constitute quasi log-linear families.

The analysis of this paper should be extended to the case where asset markets are incomplete.

Since individual consumers have fewer instruments to transfer purchasing power across time and states, the impact of the heterogeneity of subjective discount rates on the representative consumer's discount rates will be less pronounced than in the case of complete markets. To increase the relevance of the results of this paper, it is important to determine exactly how much the impact is reduced.

References

[1] Günter Franke, Richard C. Stapleton, and Marti G. Subrahmanyam, 1999, When are options overpriced? The Black-Scholes model and alternative characterisations of the pricing kernel, European Finance Review 3, 79–102.

[2] Christian Gollier and Richard Zeckhauser, 2005, Aggregation of heterogeneous time preferences, Journal of Political Economy 113, 878–896.

[3] Chiaki Hara, James Huang, and Christoph Kuzmics, 2007, Efficient risk-sharing rules and the representative consumer's risk aversion, Journal of Economic Theory 137, 652–672.

[4] Matthew O. Jackson and Leeat Yariv, 2011, Collective dynamic choice: The necessity of time inconsistency, Available at SSRN: `http://ssrn.com/abstract=1699444` or `http://dx.doi.org/10.2139/ssrn.1699444`.

[5] Elyès Jouini and Clotilde Napp, 2007, Consensus consumer and intertemporal asset pricing with herogeneous belief, Review of Economic Studies 74, 1149–1174.

[6] Carl N. Morris, 1982, Natural exponential families with quadratic variance functions, Annals of Statistics 10, 65–80.

[7] Drazen Prelec, 2004, Decreasing impatience: A criterion for non-stationary time preference and hyperbolic discounting, Scandinavian Journal of Economics 106, 511–532.

[8] Martin L. Weitzman, 2001, Gamma discounting, American Economic Review 91, 260–271.